LINEAR ESTIMATION OF REGRESSION COEFFICIENTS*

BY

ROLF K. ADENSTEDT AND BENNETT EISENBERG

Lehigh University

1. Introduction. We describe in this paper a general approach to the problem of linearly estimating a regression coefficient, with some applications. Let T be an arbitrary index set and $X(t)$, $t \in T$, a zero-mean random process with known covariance function $R(s, t) = EX(s)X(t)$. The process $Y(t) = X(t) + c\beta(t)$, with known regression terms $\beta(t)$, is observed for t in a fixed subset S of T. These observations are used to form linear unbiased estimators $\hat{c}$ of the unknown regression constant c. Of particular interest is the best linear unbiased estimator (BLUE) $\hat{c}_{\mathrm{BLU}}$, i.e. the $\hat{c}$ having minimum variance. A similar problem may be posed for several regression constants.

Regression analysis has generally been restricted to special cases, i.e. special forms of T, S, R and $\beta(t)$. The earliest results dealt with finitely many uncorrelated random variables. More recent investigations have been directed toward stationary disturbances $X(t)$. Grenander and Szegö in [3] and [5] considered efficiency of the least squares estimator for the case $\beta(t) \equiv 1$. Vitale [8] and Adenstedt [1] constructed more general asymptotically efficient estimators. Grenander [4] considered certain general regressions $\beta(t)$ in the stationary case and Rosenblatt [7] extended Grenander's results to vector-valued processes.

Most authors have approached the problem of finding $\hat{c}_{\mathrm{BLU}}$ by using properties of the covariance matrix R in the case of finite S, and treating the case of infinite S by a limiting procedure. In this paper we instead consider the problem in a Hilbert space setting, without restricting S and the $\beta(t)$. We find in Sec. 2 that linear unbiased estimators may be viewed as the elements of a hyperplane, leading to a useful representation of $\hat{c}_{\mathrm{BLU}}$ (Theorem 1). This representation is used in Sec. 3 to show that the BLUE coincides with the maximum likelihood estimator in the Gaussian case. In Sec. 4 we establish a procedure for obtaining lower bounds for Var $\hat{c}_{\mathrm{BLU}}$. Secs. 5 and 6 deal with the case of estimating the mean for a stationary process. Theorems 3 and 4 generalize results of Grenander and Szegö concerning asymptotic efficiency of the sample mean. In Sec. 7 we give examples of applications to nonstationary processes. Finally, in Sec. 8 we generalize Theorem 1 for the case of several regression variables and apply the result to some special examples.

2. Representation of the BLUE. We retain the notation of the general problem outlined above. All quantities are assumed real; generalizations to complex-valued processes are straightforward. So that linear unbiased estimators exist, we always assume that $\beta(t) \neq 0$ for at least one value of t in S.

* Received May 10, 1973.

Linear estimators are limits (in mean-square) of finite linear combinations

$$\hat{c}_n = \sum_{\nu=1}^{n} a_{n\nu} Y(t_{n\nu}) = \sum a_{n\nu} X(t_{n\nu}) + c \sum a_{n\nu}\beta(t_{n\nu}), \qquad t_{n\nu} \in S, \tag{1}$$

with which we may associate the corresponding forms

$$\phi_n = \sum_{\nu=1}^{n} a_{n\nu} X(t_{n\nu}), \qquad t_{n\nu} \in S. \tag{2}$$

A finite estimator (1) is unbiased if $\sum a_{n\nu}\beta(t_{n\nu}) = 1$. Then we have $\operatorname{Var} \hat{c}_n = E\phi_n^2$, which we write as $||\phi_n||^2$, for the corresponding form (2). While the correspondence between (1) and (2) is straightforward for finite combinations, the appropriate correspondence for limits presents some difficulties. To avoid these, we now adopt a slightly different point of view.

Denote by μ_c the probability measure on sample path space Ω for a process equal in law to $X(t) + c\beta(t)$. Then, for $\omega \in \Omega$, we regard $X(t, \omega) = Y(t, \omega) = \omega(t)$, evaluation of the sample path at t. The processes differ only in that $X(t)$ is considered as an element in $L^2(\mu_0)$, whereas $Y(t)$ is considered as an element in $L^2(\mu_c)$. In this way we restrict attention to a single set of sample paths. An estimator $\hat{c}(\omega)$ is now a function of the sample path. It is unbiased if $\int \hat{c}(\omega)\, d\mu_c = c$ for every c, and is linear if it is in the subspace $H_Y(S)$ of $L^2(\mu_c)$ spanned by the $Y(t)$ for t in S.

A singular case arises when there is a form (2) with $\phi_n = 0$ but $\sum a_{n\nu}\beta(t_{n\nu}) \neq 0$ or, more generally, when there is a sequence of elements (2) with $\int \phi_n^2\, d\mu_0 \to 0$ but $\sum a_{n\nu}\beta(t_{n\nu}) = 1$. Then $\sum a_{n\nu}[X(t_{n\nu}) + c\beta(t_{n\nu})]$ converges to c in $L^2(\mu_0)$, or equivalently $\sum a_{n\nu} Y(t_{n\nu})$ converges to c in $L^2(\mu_c)$. Thus c is precisely determined by linear combinations of observations; in other words, $\hat{c}_{BLU} = c$.

In the non-singular case, the linear map L given by

$$L \sum_{\nu=1}^{n} a_\nu X(t_\nu) = \sum a_\nu \beta(t_\nu), \qquad t_\nu \in S,$$

is well-defined and bounded on the linear hull of the $X(t)$, $t \in S$. Then L extends to a bounded linear functional on the subspace $H_X(S)$ of $L^2(\mu_0)$ spanned by the $X(t)$ for t in S. This implies that $X(t) \to X(t) + c\beta(t)$, or equivalently $X(t) \to Y(t)$, extends to a bounded linear operator. Conversely, $Y(t) \to X(t)$ is always bounded. Thus elements of $H_X(S)$ are also in $H_Y(S)$, and conversely.

By the Riesz representation theorem, there is now a unique ψ in $H_X(S)$ such that $L\phi = \int \phi\psi\, d\mu_0$ for $\phi \in H_X(S)$. ψ is determined from

$$EX(t)\psi = \beta(t), \qquad t \in S; \qquad \psi \in H_X(S). \tag{3}$$

For $t_\nu \in S$, we have $\int \sum_1^n a_\nu Y(t_\nu)\, d\mu_c = c \sum a_\nu \beta(t_\nu) = c \int \sum a_\nu X(t_\nu)\psi\, d\mu_0$ and also $\operatorname{Var} \sum a_\nu Y(t_\nu) = ||\sum a_\nu X(t_\nu)||^2$. Taking limits of finite linear combinations, we obtain

$$\int \phi\, d\mu_c = c \int \phi\psi\, d\mu_0, \qquad \operatorname{Var} \phi = ||\phi||^2$$

for ϕ in $H_X(S)$. It follows that unbiased linear estimators $\hat{c}$ are precisely the elements of the hyperplane $\int \hat{c}\psi\, d\mu_0 = 1$ in $H_X(S)$, and that $\hat{c}_{BLU}$ is the element of minimum norm in this hyperplane. By the Cauchy–Schwartz inequality then $||\hat{c}|| \geq 1/||\psi||$, with equality if, and only if, $\hat{c} = \psi/||\psi||^2$. Thus $\hat{c}_{BLU}$ is determined as $\psi/||\psi||^2$, and has variance $1/||\psi||^2$.

We summarize the results in a theorem:

THEOREM 1. The BLUE for c on the basis of observations $Y(t) = X(t) + c\beta(t)$, $t \in S$, satisfies

$$\hat{c}_{\mathrm{BLU}}(\omega) = \psi(\omega)/\|\psi\|^2, \qquad \operatorname{Var} \hat{c}_{\mathrm{BLU}} = 1/\|\psi\|^2,$$

where ψ solves (3). If (3) has no solution then $\hat{c}_{\mathrm{BLU}} = c$ and c is perfectly estimable.

The theorem could also be derived from the relation $EX(t)\hat{c}_{\mathrm{BLU}} = \beta(t) \operatorname{Var} \hat{c}_{\mathrm{BLU}}$, $t \in S$, which is similar to the integral equation obtained by Grenander in [3]. We find the above representation particularly attractive in that the solution of (3) is linear in $\beta(t)$.

3. Relation to maximum-likelihood estimation. Theorem 1 may be used to show that $\hat{c}_{\mathrm{BLU}}$ is also the maximum-likelihood estimator when the process $X(t)$ is Gaussian. We now regard μ_c as a Gaussian measure and restrict attention to the σ-field generated by the $X(t)$ for t in S. The claim then follows from the result that the Radon–Nikodym derivative

$$d\mu_c/d\mu_0 = \exp\,(c\psi - \tfrac{1}{2}c^2\,\|\psi\|^2), \tag{4}$$

where ψ solves (3). This is equivalent to a somewhat different form given in [3]. Clearly (4) is maximum for $c = \psi/\|\psi\|^2$.

The proof of (4) is not difficult. Using a moment-generating function argument, it suffices to show that

$$\int \exp\,[\textstyle\sum a_\nu X(t_\nu)]\,d\mu_c = \displaystyle\int \exp\,[\textstyle\sum a_\nu X(t_\nu) + c\psi - \tfrac{1}{2}c^2\,\|\psi\|^2]\,d\mu_0 \tag{5}$$

for finite linear combinations $\sum_1^n a_\nu X(t_\nu)$, $t_\nu \in S$. Carrying out the integrations in (5)' we obtain, with use of (3) for the right side,

$$\begin{aligned}\exp\,[-\tfrac{1}{2}c^2\,\|\psi\|^2 &+ \tfrac{1}{2}\,\|\textstyle\sum a_\nu X(t_\nu) + c\psi\|^2]\\ &= \exp\left[c\textstyle\sum a_\nu \displaystyle\int X(t_\nu)\psi\,d\mu_0 + \tfrac{1}{2}\,\|\textstyle\sum a_\nu X(t_\nu)\|^2\right]\\ &= \exp\,[c\textstyle\sum a_\nu\beta(t_\nu) + \tfrac{1}{2}\displaystyle\sum_{\nu,\mu} a_\nu a_\mu R(t_\nu, t_\mu)].\end{aligned}$$

The last expression is seen to be the same as the left side of (5).

The derivative (4) would also be used in the Gaussian case for the signal-detection problem of testing the hypothesis that $c = 0$. The likelihood ratio test will be based on the BLUE $\psi/\|\psi\|^2$.

4. Lower bounds for the variance of the BLUE. Calculation of $\hat{c}_{\mathrm{BLU}}$, or equivalently solving (3), is in most practical situations difficult if not impossible because $R(s, t)$ is not known precisely. Therefore adequate approximations are needed. In this section we describe a procedure which, in the spirit of the Cramér–Rao inequality, gives a lower bound for $\operatorname{Var} \hat{c}_{\mathrm{BLU}}$. The procedure will be used in the examples of the following sections to obtain estimators nearly as good as the BLUE.

Denote by $H_X(T)$ the Hilbert space spanned by the $X(t)$ for t in T. Note $H_X(S)$ is a subspace of $H_X(T)$. While solution of (3) may be difficult, we may more easily be able to construct a $\tilde{\psi} \in H_X(T)$ satisfying

$$EX(t)\tilde{\psi} = \beta(t), \qquad t \in S. \tag{6}$$

The equation need not be satisfied for $t \notin S$. Clearly such a $\tilde{\psi}$ exists if, and only if, c is not perfectly estimable from observations over S. Note $\tilde{\psi}$ is not unique.

If P is the projection operator onto $H_X(S)$ and $\tilde{\psi} \in H_X(T)$ satisfies (6), then $EX(t)P\tilde{\psi} = EX(t)\tilde{\psi} = \beta(t)$ for $t \in S$, so that $\psi = P\tilde{\psi}$ is the solution of (3). Since $\|\psi\| = \|P\tilde{\psi}\| \leq \|\tilde{\psi}\|$, we are led to

THEOREM 2. If $\tilde{\psi}$ is in $H_X(T)$ and satisfies (6), then Var $\hat{c}_{BLU} \geq 1/\|\tilde{\psi}\|^2$. If no such $\tilde{\psi}$ exists, then Var $\hat{c}_{BLU} = 0$.

5. Estimating the mean of a stationary sequence. Assume that $X(t)$, $t = 0, \pm 1, \cdots$, is a zero-mean wide-sense stationary sequence with covariance function $R(s, t) = R(s - t)$. c is to be estimated from observation of $Y(t) = X(t) + c$ for $t = 1, 2, \cdots, N$. This corresponds to $\beta(t) \equiv 1$ in the general case. Of interest is the efficiency relative to the BLUE of the sample mean $\bar{Y} = N^{-1} \sum_1^N Y(t)$, which does not involve knowledge of R in calculation.

Grenander and Szegö show in [3] and [5] that

$$\text{Var } \bar{Y} \sim \text{Var } \hat{c}_{BLU} \sim f(0)/N, \qquad N \to \infty, \tag{7}$$

where $R(t) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda)\, d\lambda$ and the spectral density f is positive and continuous. This implies that

$$\lim_{N \to \infty} \text{Var } \hat{c}_{BLU}/\text{Var } \bar{Y} = 1. \tag{8}$$

The same result is derived in [1] under weaker assumptions on f; the proof is based on approximation by trigonometric polynomials. We shall establish a somewhat stronger result than (7) under still weaker assumptions, employing a different method of proof.

It is important to note that the standard definition of relative (asymptotic) efficiency as used in (8) is not necessarily the best one from a practical standpoint. Of more interest perhaps is a comparison, for two different estimators, of the minimum sample sizes needed to attain a given accuracy of estimation. On account of (7), one could still say that $\bar{Y}$ is asymptotically as good as $\hat{c}_{BLU}$ in this new sense. Thus (7) is a better result than (8). The latter formula is particularly empty in meaning when consistent estimation is not possible, for then (8) simply states that Var $\hat{c}_{BLU}$ and Var $\bar{Y}$ approach the same nonzero limit as $N \to \infty$. One could then hardly say that $\bar{Y}$ is as good as $\hat{c}_{BLU}$ asymptotically unless its variance approaches this limit as rapidly as does that of $\hat{c}_{BLU}$. In the following theorem we find that efficiency of $\bar{Y}$ relative to the BLUE, even in this wider sense, depends strongly on the behavior of the spectral distribution function at the origin.

THEOREM 3. Let $X(t)$ have spectral measure

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

on $[-\pi, \pi]$, where $dF_s(\lambda)$ is singular with respect to $d\lambda$.

(a) Then

$$\lim_{N \to \infty} \text{Var } \bar{Y} = \lim_{N \to \infty} \text{Var } \hat{c}_{BLU} = dF(0). \tag{9}$$

(b) Assume also that f is positive and continuous at $\lambda = 0$, that $F_s(\lambda) - F_s(-\lambda)$ is constant in $0 < |\lambda| < \delta$ for some $\delta > 0$, and that

$$\int_{-\pi}^{\pi} \frac{|p(\lambda)|^2 \, d\lambda}{f(\lambda)} < \infty \tag{10}$$

for some trigonometric polynomial $p(\lambda)$. Then

$$\operatorname{Var} \bar{Y} - dF(0) \sim \operatorname{Var} \hat{c}_{\mathrm{BLU}} - dF(0) \sim f(0)/N, \qquad N \to \infty. \tag{11}$$

Proof. (a) By dominated convergence we always have

$$\operatorname{Var} \bar{Y} = \int_{-\pi}^{\pi} N^{-2} \left| \sum_{t=1}^{N} e^{it\lambda} \right|^2 dF(\lambda) \to dF(0)$$

as $N \to \infty$. This establishes (9) for the case $dF(0) = 0$. If $dF(0) > 0$, consistent estimation is impossible because there is a $\tilde{\psi}$ with $EX(t)\tilde{\psi} = 1$ for all t, namely $\tilde{\psi} = dZ(0)/dF(0)$, where $X(t) = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda)$ is the spectral representation of the process. By Theorem 2, $\operatorname{Var} \hat{c}_{\mathrm{BLU}} \geq 1/\|\tilde{\psi}\|^2 = dF(0)$. But also $\operatorname{Var} \hat{c}_{\mathrm{BLU}} \leq \operatorname{Var} \bar{Y} \to dF(0)$, and (9) follows.

(b) By our assumptions, $F_s(\lambda) - F_s(-\lambda) = dF(0)$ for $0 < |\lambda| < \delta$. Using dominated convergence and properties of the Féjer kernel, we therefore obtain

$$N \operatorname{Var} \bar{Y} - N\, dF(0) = N^{-1} \int_{-\pi}^{\pi} \left| \sum_{1}^{N} e^{it\lambda} \right|^2 dF(\lambda) - N\, dF(0)$$

$$= \int_{|\lambda| \geq \delta} N^{-1} \left| \sum e^{it\lambda} \right|^2 dF_s(\lambda) + (2\pi N)^{-1} \int_{-\pi}^{\pi} \left| \sum e^{it\lambda} \right|^2 f(\lambda)\, d\lambda \to f(0)$$

as $N \to \infty$. To prove (11) it remains then to show that

$$\liminf_{N \to \infty} N[\operatorname{Var} \hat{c}_{\mathrm{BLU}} - dF(0)] \geq f(0). \tag{12}$$

Let $a_1 + \cdots + a_N = 1$. Then

$$\int_{-\pi}^{\pi} \left| \sum_{t=1}^{N} a_t e^{it\lambda} \right|^2 dF(\lambda) - dF(0) \geq (2\pi)^{-1} \int_{-\pi}^{\pi} \left| \sum a_t e^{it\lambda} \right|^2 f(\lambda)\, d\lambda.$$

$\operatorname{Var} \hat{c}_{\mathrm{BLU}} - dF(0)$ is just the minimum over all such a_t of the left side, while the minimum over a_t of the right side represents $\operatorname{Var} \hat{c}_{\mathrm{BLU}}$ when $F_s \equiv 0$. Thus it suffices to prove (12) when F_s vanishes, and we now assume this.

Without loss of generality we may write the polynomial in (10) as $p(\lambda) = \sum_{-\alpha}^{0} b_\nu e^{i\nu\lambda}$ and assume that $p(0) \neq 0$. Define

$$\tilde{\psi} = \int_{-\pi}^{\pi} g(\lambda)\, dZ(\lambda), \qquad g(\lambda) = \frac{p(\lambda)}{p(0) f(\lambda)} \sum_{t=1}^{N+\alpha} e^{it\lambda}.$$

Then, by a simple calculation,

$$EX(t)\tilde{\psi} = (2\pi)^{-1} \int_{-\pi}^{\pi} g(\lambda) e^{-it\lambda} f(\lambda)\, d\lambda = 1, \qquad t = 1, 2, \cdots, N.$$

Also,

$$\frac{\|\tilde{\psi}\|^2}{N + \alpha} = \frac{1}{2\pi(N + \alpha)} \int_{-\pi}^{\pi} \frac{|p(\lambda)|^2}{|p(0)|^2 f(\lambda)} \left| \sum_{1}^{N+\alpha} e^{it\lambda} \right|^2 d\lambda \to 1/f(0)$$

as $N \to \infty$. Using Theorem 2, we conclude that $\operatorname{Var} \hat{c}_{\mathrm{BLU}} \geq 1/\|\tilde{\psi}\|^2 \sim f(0)/N$, which implies (12) for $F_s \equiv 0$ and completes the proof.

The condition (10) says essentially that $1/f$ is integrable, but allows f to have zeros removable by $|p(\lambda)|^2$. The condition on F_s certainly allows superposition of a finite discrete spectrum on an absolutely continuous spectrum.

6. Estimating the mean of a stationary process. We consider the analogue of the previous section for continuous time. $X(t)$, $-\infty < t < \infty$, is a zero-mean wide-sense stationary process with spectral measure

$$dF(\lambda) = dF_s(\lambda) + (2\pi)^{-1}f(\lambda)\,d\lambda \tag{13}$$

on the line. Again $dF_s(\lambda)$ is singular with respect to $d\lambda$. Observations $Y(t) = X(t) + c$ for $0 \le t \le T$ are used to estimate c. The sample mean becomes $\bar{Y} = T^{-1}\int_0^T Y(t)\,dt$ (mean-square integration), and its efficiency relative to $\hat{c}_{BLU}$ is of interest.

The continuous-time analogue of (7) is proved in [3] and [5] for the case of a nondeterministic process $X(t)$ with $F_s \equiv 0$. Restrictions are placed on f and the moving average representation of the process. The proof involves relating the estimation problem to the problem of prediction. We make somewhat different assumptions and establish the direct analogue of Theorem 3.

THEOREM 4. Let $X(t)$ have spectral measure (13).

(a) Then

$$\lim_{T\to\infty} \operatorname{Var} \bar{Y} = \lim_{T\to\infty} \operatorname{Var} \hat{c}_{BLU} = dF(0). \tag{14}$$

(b) Assume in addition that f is positive and continuous at $\lambda = 0$, that $F_s(\lambda) - F_s(-\lambda)$ is constant in $0 < |\lambda| < \delta$ for some $\delta > 0$, and that

$$\int_{-\infty}^{\infty} \frac{|p(\lambda)|^2\,d\lambda}{(1+\lambda^2)^n f(\lambda)} < \infty \tag{15}$$

for some positive integer n and function p of the form

$$p(\lambda) = \sum_{\nu=1}^{m} b_\nu \exp(-it_\nu\lambda).$$

Then

$$\operatorname{Var} \bar{Y} - dF(0) \sim \operatorname{Var} \hat{c}_{BLU} - dF(0) \sim f(0)/T, \qquad T \to \infty. \tag{16}$$

Proof. (a) We proceed as in Theorem 3.

$$\operatorname{Var} \bar{Y} = \int_{-\infty}^{\infty} T^{-2} \left| \int_0^T e^{it\lambda}\,dt \right|^2 dF(\lambda) \to dF(0), \qquad T \to \infty,$$

always holds and implies (14) when $dF(0) = 0$. If $dF(0) > 0$ and $X(t) = \int_{-\infty}^{\infty} e^{it\lambda}dZ(\lambda)$ is the spectral representation, then $\tilde{\psi} = dZ(0)/dF(0)$ satisfies $EX(t)\tilde{\psi} = 1$ for $-\infty < t < \infty$. Therefore $dF(0) = 1/\|\tilde{\psi}\|^2 \le \operatorname{Var} \hat{c}_{BLU} \le \operatorname{Var} \bar{Y} \to dF(0)$ and again (14) follows.

(b) As in Theorem 3, we find that

$$T \operatorname{Var} \bar{Y} - T\,dF(0) = T^{-1} \int_{-\infty}^{\infty} \left| \int_0^T e^{it\lambda}\,dt \right|^2 dF(\lambda) - T\,dF(0)$$

$$= \int_{|\lambda|\ge\delta} T^{-1} \left| \int_0^T e^{it\lambda}\,dt \right|^2 dF_s(\lambda) + (2\pi T)^{-1} \int_{-\infty}^{\infty} \left| \int_0^T e^{it\lambda}\,dt \right|^2 f(\lambda)\,d\lambda \to f(0)$$

as $T \to \infty$. We have used dominated convergence and properties of the continuous Féjer kernel. It remains only to be shown that

$$\liminf_{T\to\infty} T[\operatorname{Var} \hat{c}_{\mathrm{BLU}} - dF(0)] \geq f(0). \tag{17}$$

By reasoning as in Theorem 3, we find that $\operatorname{Var} \hat{c}_{\mathrm{BLU}} - dF(0)$ is no less than the infimum over all n, $a_1 + \cdots + a_n = 1$ and $0 \leq t_1, \cdots, t_n \leq T$ of $(2\pi)^{-1} \int_{-\infty}^{\infty} |\sum a_\nu \exp(it_\nu\lambda)|^2 f(\lambda)\, d\lambda$, which is just $\operatorname{Var} \hat{c}_{\mathrm{BLU}}$ when $F_s \equiv 0$. Thus it suffices to prove (17) when F_s vanishes, and we assume this from here on.

In the remainder of the proof, n is the integer and $p(\lambda) = \sum_{\nu=1}^{m} b_\nu \exp(-it_\nu\lambda)$ the function in (15). Without loss of generality we may assume that the $t_\nu \geq 0$ and that $p(0) \neq 0$. We set $\alpha = \max\{t_1, \cdots, t_m\}$.

In [3], Grenander shows that for the spectral density $(1 + \lambda^2)^{-n}$

$$\hat{c}_{\mathrm{BLU}} = (2n + T)^{-1}\left\{\int_0^T Y(t)\, dt + \sum_{k=0}^{n-1} \binom{n}{k+1}[(-1)^k Y^{(k)}(0) + Y^{(k)}(T)]\right\},$$

with $\operatorname{Var} \hat{c}_{\mathrm{BLU}} = (2n + T)^{-1}$. Here $Y^{(k)}(t) = d^k Y(t)/dt^k$ in the mean-square sense. With use of Theorem 1 and the spectral representation, it follows that the function

$$g_T(\lambda) = \int_0^T e^{it\lambda}\, dt + \sum_{k=0}^{n-1} \binom{n}{k+1}(i\lambda)^k[(-1)^k + e^{iT\lambda}] \tag{18}$$

satisfies

$$(2\pi)^{-1} \int_{-\infty}^{\infty} g_T(\lambda) e^{-it\lambda}(1 + \lambda^2)^{-n}\, d\lambda = 1, \qquad 0 \leq t \leq T. \tag{19}$$

With $g_{T+\alpha}(\lambda)$ defined as in (18), let

$$\tilde{\psi} = \int_{-\infty}^{\infty} g(\lambda)\, dZ(\lambda), \qquad g(\lambda) = \frac{p(\lambda)}{p(0)f(\lambda)}\, g_{T+\alpha}(\lambda)(1 + \lambda^2)^{-n}.$$

With use of (19), a simple calculation then reveals that

$$EX(t)\tilde{\psi} = (2\pi)^{-1} \int_{-\infty}^{\infty} g(\lambda) e^{-it\lambda} f(\lambda)\, d\lambda = 1, \qquad 0 \leq t \leq T.$$

Therefore $\operatorname{Var} \hat{c}_{\mathrm{BLU}} \geq 1/\|\tilde{\psi}\|^2$. Writing

$$h(\lambda) = \frac{|p(\lambda)|^2}{|p(0)|^2 f(\lambda)} (1 + \lambda^2)^{-2n},$$

we find that

$$(2\pi T)^{-1} \int_{-\infty}^{\infty} \left|\int_0^T e^{it\lambda}\, dt\right|^2 h(\lambda)\, d\lambda \to h(0) = 1/f(0)$$

as $T \to \infty$. Also

$$(2\pi T)^{-1} \int_{-\infty}^{\infty} \left|g_T(\lambda) - \int_0^T e^{it\lambda}\, dt\right|^2 h(\lambda)\, d\lambda \to 0, \qquad T \to \infty,$$

since the integral has a bound independent of T. The triangle inequality for norms in $L^2(h, -\infty, \infty)$ therefore yields

$$\frac{||\tilde{\psi}||^2}{T+\alpha} = \frac{1}{2\pi(T+\alpha)} \int_{-\infty}^{\infty} |g_{T+\alpha}(\lambda)|^2 h(\lambda)\, d\lambda \to 1/f(0)$$

as $T \to \infty$. Thus Var $\hat{c}_{BLU} \geq 1/||\tilde{\psi}||^2 \sim f(0)/T$, and (17) is established for $F_s \equiv 0$. This concludes the proof.

The condition (15) stipulates that $f(\lambda)$ does not decrease too rapidly as $|\lambda| \to \infty$, but allows f to have zeros. One may show that (15) implies that $X(t)$ is nondeterministic.

7. Application to nonstationary processes. Our procedures are not restricted to stationary processes. As an example we consider the model

$$\begin{aligned} Y(t) = X(t) + c &= c + \eta_0, & t &= 0, \\ &= c + \eta_t + \xi_1 + \cdots + \xi_t, & t &= 1, 2, \cdots \end{aligned}$$

Here $\eta_0, \eta_1, \cdots, \xi_1, \xi_2, \cdots$ are uncorrelated random variables with zero means and $E\eta_t^2 = \lambda^2 > 0$, $E\xi_t^2 = \sigma^2 > 0$.

This type of model is used by Chernoff and Zacks [2] in connection with estimation of a mean value subject to change over time. They give an explicit but computationally difficult-to-use formula for $\hat{c}_{BLU}$ based on observations $Y(0), \cdots, Y(N)$. $\hat{c}_{BLU}$ is not consistent as $N \to \infty$. Mustafi [6] shows that

$$\lim_{N\to\infty} \operatorname{Var} \hat{c}_{BLU} = \lambda^2(1 - A), \tag{20}$$

where

$$A = 1 + \frac{b}{2} - \left[b\left(1 + \frac{b}{4}\right)\right]^{1/2}, \qquad b = \sigma^2/\lambda^2. \tag{21}$$

Her proof is based on Chernoff and Zacks' representation.

We can establish (20) very easily by considering the estimation problem using infinite data $Y(0), Y(1), \cdots$. The $X(t)$ have covariances $R(s, t) = \lambda^2[\delta_{st} + b \min(s, t)]$, where $b = \sigma^2/\lambda^2$ and δ_{st} is Kronecker's delta. Thus the sums

$$\psi_a = \lambda^{-2} \sum_{s=0}^{\infty} a^s X(s), \qquad |a| < 1,$$

converge in mean-square. A straightforward computation shows that

$$EX(t)\psi_a = \lambda^{-2} \sum_{s=0}^{\infty} a^s R(t, s) = 1 - \frac{1 - a^t}{(1-a)^2} [(1-a)^2 - ba]$$

for $t = 0, 1, \cdots$. Since A, defined in (21), satisfies $0 < A < 1$ and $(1 - A)^2 - bA = 0$, $EX(t)\psi_A = 1$ for all t. Also $||\psi_A||^2 = \lambda^{-2}(1 - A)^{-1}$, and it follows that

$$\hat{c}_{BLU} = (1 - A) \sum_{t=0}^{\infty} A^t Y(t), \qquad \operatorname{Var} \hat{c}_{BLU} = \lambda^2(1 - A)$$

for infinite data. Of course this implies (20) for the estimators based on finite data.

As a byproduct to this approach, we find that the easily-calculated unbiased estimators

$$\hat{c}_N = \frac{1 - A}{1 - A^{N+1}} \sum_{t=0}^{N} A^t Y(t)$$

are asymptotically efficient in the usual sense that Var $\hat{c}_{BLU}$/Var $\hat{c}_N \to 1$ as $N \to \infty$. We cannot say that $\hat{c}_N$ is asymptotically as good as $\hat{c}_{BLU}$ in the expanded sense described in Sec. 5 without a more detailed analysis. However, it is not difficult to show that Var $\hat{c}_N = \lambda^2(1 - A) + O(A^N)$, so that the optimum variance is at least approached exponentially fast.

We also mention a continuous-time model sometimes encountered in applications. Assume that $X(t)$, $0 \leq t \leq T$, is continuous in mean and that the regression function $\beta(t)$ is in the range of $R(s, t)$. Thus

$$\beta(t) = \int_0^T R(t, s)a(s)\, ds, \qquad 0 \leq t \leq T,$$

for some function $a(s)$. This relation immediately yields $EX(t) \int_0^T a(s)X(s)\, ds = \beta(t)$ for $0 \leq t \leq T$, so that $\psi = \int_0^T a(s)X(s)\, ds$ characterizes the BLUE based on observations $Y(t) = X(t) + c\beta(t)$. We have $||\psi||^2 = \int_0^T a(t)\beta(t)\, dt$, whence

$$\hat{c}_{\text{BLU}} = \int_0^T a(t)Y(t)\, dt \Big/ \int_0^T a(t)\beta(t)\, dt,$$

$$\text{Var}\, \hat{c}_{\text{BLU}} = 1 \Big/ \int_0^T a(t)\beta(t)\, dt.$$

The results might be compared with the commonly-used correlator estimator $\int_0^T \beta(t)Y(t)\, dt / \int_0^T [\beta(t)]^2\, dt$.

8. Case of several regression constants. We have restricted attention to a single regression constant c so as not to obscure the results. More generally, however, we might observe $Y(t) = X(t) + \sum_{j=1}^m c_j\beta_j(t)$ for $t \in S$, where the $\beta_j(t)$ are known and the c_j are to be estimated. We discuss this situation and some applications briefly.

We retain the viewpoint and notation of Sec. 2, with the exception that μ_c represents measure on path space Ω for a process equal in law to $X(t) + \sum c_j\beta_j(t)$. An estimator $\hat{c}_j$ for c_j is now unbiased if $\int \hat{c}_j\, d\mu_c = c_j$ for all $c_1, \cdots, c_m$, and is linear if it is in $H_Y(S)$. To guarantee existence of unbiased linear estimators for each c_j, we assume that the regression functions $\beta_1(t), \cdots, \beta_m(t)$ for $t \in S$ are linearly independent. The respective BLUE's are denoted by $\hat{c}_{j,\text{BLU}}$.

We may encounter singularity in that one or more of the c_j can be perfectly estimated. In the non-singular case there are bounded linear functionals $L_1, \cdots, L_m$ on $H_X(S)$ such that $L_jX(t) = \beta_j(t)$ for $t \in S$, and there are unique elements $\psi_1, \cdots, \psi_m$ in $H_X(S)$ such that $L_j\phi = \int \phi\psi_j\, d\mu_0$ for $\phi \in H_X(S)$. The ψ_j are determined from

$$EX(t)\psi_j = \beta_j(t), \quad t \in S; \quad \psi_j \in H_X(S) \quad (j = 1, \cdots, m). \tag{22}$$

Arguing in a manner as in Sec. 2, we find that

$$\int \phi\, d\mu_c = \sum c_j \int \phi\psi_j\, d\mu_0, \qquad \text{Var}\, \phi = ||\phi||^2 \tag{23}$$

for ϕ in $H_X(S)$. Therefore unbiased linear estimators $\hat{c}_j$ for c_j are precisely the elements of the sets in $H_X(S)$ defined by

$$\int \hat{c}_j\psi_k\, d\mu_0 = \delta_{jk}, \qquad k = 1, \cdots, m. \tag{24}$$

Introduce the vector notation

$$\hat{c} = \begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_m \end{bmatrix}, \quad \hat{c}_{\mathrm{BLU}} = \begin{bmatrix} \hat{c}_{1,\mathrm{BLU}} \\ \vdots \\ \hat{c}_{m,\mathrm{BLU}} \end{bmatrix}, \quad \psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_m \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix},$$

where $\hat{c}$ is linear and unbiased for c. (24) can be written $\int \hat{c}\psi^T \, d\mu_0 = I$, the $m \times m$ identity matrix. Also define the covariances matrices

$$\sigma(\hat{c}) = \int \hat{c}\hat{c}^T \, d\mu_0 , \qquad V = \int \psi\psi^T \, d\mu_0 .$$

Define the usual partial ordering on symmetric matrices by writing $A \geq B$ whenever $A - B$ is positive semidefinite. If we can find a minimum $\sigma(\hat{c})$ in this ordering, clearly the diagonal terms $||\hat{c}_i||^2$ will be minimum and $\hat{c}_{\mathrm{BLU}}$ obtained.

The covariance matrix V is non-singular, for if $\psi^T\alpha = 0$ for some constant vector α then $0 = \int \hat{c}\psi^T\alpha \, d\mu_0 = I\alpha = \alpha$. Thus V^{-1} exists and $0 \leq \int (\hat{c} - V^{-1}\psi)(\hat{c} - V^{-1}\psi)^T \, d\mu_0 = \sigma(\hat{c}) - \int \hat{c}\psi^T V^{-1} \, d\mu_0 - \int V^{-1}\psi\hat{c}^T \, d\mu_0 + V^{-1}VV^{-1} = \sigma(\hat{c}) - V^{-1}$, i.e. $\sigma(\hat{c}) \geq V^{-1}$. Equality holds if, and only if, $\hat{c} = V^{-1}\psi$. We have proved

THEOREM 5. *In the above notation the BLUE and its covariance matrix are*

$$\hat{c}_{\mathrm{BLU}}(\omega) = V^{-1}\psi(\omega), \qquad \sigma(\hat{c}_{\mathrm{BLU}}) = V^{-1},$$

where the components $\psi_1, \cdots, \psi_m$ of ψ satisfy (22) and have covariance matrix V. Moreover, $\sigma(\hat{c}) - \sigma(\hat{c}_{\mathrm{BLU}})$ is positive semidefinite for any linear unbiased estimator $\hat{c}$. If (22) has no solution for some j, then a linear combination of the c_i is perfectly estimable.

The theorem may be used to show that $\hat{c}_{\mathrm{BLU}}$ is again the maximum likelihood estimator in the Gaussian process case. The claim follows from the representation

$$d\mu_c/d\mu_0 = \exp\,(c^T\psi - \tfrac{1}{2}c^T V c)$$

of the Radon–Nikodym derivative, which may be established by the method of Sec. 3.

We may also obtain lower bounds for the covariance matrix $\sigma(\hat{c}_{\mathrm{BLU}})$ by the method of Sec. 4. Suppose $\tilde{\psi}_1, \cdots, \tilde{\psi}_m$ are in $H_X(T)$ and satisfy $EX(t)\tilde{\psi}_j = \beta_j(t)$ for $t \in S$, $j = 1, \cdots, m$. Then the solutions of (22) are $\psi_j = P\tilde{\psi}_j$, where P is the projection onto $H_X(S)$. If $\tilde{V}$ is the covariance matrix of the $\tilde{\psi}_j$ and α is any constant vector, then $\alpha^T V\alpha = ||\alpha^T\psi||^2 = ||P\alpha^T\tilde{\psi}||^2 \leq ||\alpha^T\tilde{\psi}||^2 = \alpha^T\tilde{V}\alpha$. Therefore $\tilde{V} \geq V$, and it follows that $\sigma(\hat{c}_{\mathrm{BLU}}) = V^{-1} \geq \tilde{V}^{-1}$.

Theorem 5 may be used to prove a generalization of part of the Gauss–Markov theorem for the case of correlated errors and arbitrary parameter. Suppose in the above setting we wish to estimate $\gamma = \alpha^T c = \sum \alpha_j c_j$, where α is a given non-zero vector. By (23), any unbiased linear estimator $\hat{\gamma}$ satisfies $\int \hat{\gamma}\psi_j \, d\mu_0 = \alpha_j$ for $j = 1, \cdots, m$, or $\int \hat{\gamma}\psi \, d\mu_0 = \alpha$ in vector form. Therefore $\alpha^T V^{-1}\alpha = \int \hat{\gamma}\alpha^T V^{-1}\psi \, d\mu_0$ and $(\alpha^T V^{-1}\alpha)^2 = [\int \hat{\gamma}\alpha^T V^{-1}\psi \, d\mu_0]^2 \leq ||\hat{\gamma}||^2 \, ||\alpha^T V^{-1}\psi||^2 = ||\hat{\gamma}||^2 \, \alpha^T V^{-1}\alpha$. Thus we arrive at $\mathrm{Var}\,\hat{\gamma} \geq \alpha^T V^{-1}\alpha = \mathrm{Var}\,\alpha^T\hat{c}_{\mathrm{BLU}}$, which implies that $\hat{\gamma}_{\mathrm{BLU}} = \sum \alpha_j \hat{c}_{j,\mathrm{BLU}}$.

Another application of Theorem 5 is to best-line fitting. Suppose we wish to find the best linear function $Y = c_1 + c_2 t$ to fit the data $Y(t) = X(t) + c_1 + c_2 t$, $t \in S$. This corresponds to $\beta_1(t) = 1$, $\beta_2(t) = t$. If $S = \{t_1, \cdots, t_N\}$ is finite and the $X(t_n)$ are uncorrelated with a common variance λ^2, then it is easily seen that

$$\psi_1 = \lambda^{-2} \sum_{n=1}^{N} Y(t_n), \qquad \psi_2 = \lambda^{-2} \sum t_n Y(t_n),$$

$$V = \lambda^{-2} \begin{pmatrix} N & \sum t_n \\ \sum t_n & \sum t_n^2 \end{pmatrix} .$$

It follows that

$$\begin{pmatrix} \hat{c}_{1,\mathrm{BLU}} \\ \hat{c}_{2,\mathrm{BLU}} \end{pmatrix} = \begin{pmatrix} N & \sum t_n \\ \sum t_n & \sum t_n^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum Y(t_n) \\ \sum t_n Y(t_n) \end{pmatrix}$$

with covariance matrix V^{-1}. This is just the least-squares solution.

Also of interest is line fitting when $S = [0, T]$ and $X(t)$ is a Brownian motion with a random initial value. Thus $Y(t) = \eta + W(t) + c_1 + c_2 t$, where η has mean zero and variance λ^2 and is uncorrelated with the Brownian motion $W(t)$ having variance parameter σ^2. $X(t) = \eta + W(t)$ has covariance function $\lambda^2 + \sigma^2 \min(s, t)$. We find that

$$\psi_1 = \lambda^{-2} Y(0), \qquad \psi_2 = \sigma^{-2}[Y(T) - Y(0)],$$

$$V = \begin{pmatrix} \lambda^{-2} & 0 \\ 0 & \sigma^{-2} T \end{pmatrix},$$

and therefore

$$\hat{c}_{1,\mathrm{BLU}} = Y(0), \qquad \hat{c}_{2,\mathrm{BLU}} = T^{-1}[Y(T) - Y(0)].$$

The solution differs radically from that of the previous paragraph.

References

[1] R. K. Adenstedt, *On large sample estimation for the mean of a stationary random sequence*, to appear

[2] H. Chernoff and S. Zacks, *Estimating the current mean of a normal distribution which is subjected to changes in time*, Ann. Math. Stat. **35**, 999–1018 (1964)

[3] U. Grenander, *Stochastic processes and statistical inference*, Ark. Mat. **1**, 195–277 (1950)

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

[5] U. Grenander and G. Szegö, *Toeplitz forms and their applications*, Berkeley, 1958

[6] C. Mustafi, *Inference problems about parameters which are subjected to changes over time*, Ann. Math. Stat. **39**, 840–854 (1968)

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

[8] R. A. Vitale, *An asymptotically efficient estimate in time series analysis*, Quart. Appl. Math. **30**, 421–440 (1973)