ON THE ESTIMATION OF REGRESSION COEFFICIENTS IN THE CASE OF AN AUTOCORRELATED DISTURBANCE[1]

By Ulf Grenander

University of California, Berkeley and University of Stockholm

1. Introduction. The following stochastic model has been used in various applications. A time series x_ν is the sum of a mean value m_ν and a disturbance y_ν. The disturbance is supposed to consist of independent and identically distributed stochastic variables with mean zero. The mean value of x_ν is a linear combination

$$m_\nu = \sum_{n=1}^{s} c^{(n)} \varphi_\nu^{(n)}$$

of certain known sequences $\{\varphi_\nu^{(n)}\}$, $n = 1, 2, \cdots s$, the regression variables, but with unknown regression coefficients $c^{(n)}$. For an observed sample $x_0, x_1, \cdots x_N$, the problem of estimating the c's is usually solved by applying the method of least squares (L.S.). As is well known this procedure is optimal in the sense that the estimates obtained are best linear unbiased (B.L.U.) estimates.

The problem studied in this paper arises when the disturbance is still stationary but allowed to be autocorrelated. If the correlation matrix of y_ν is known, the B.L.U. estimates can be constructed although their form is not so simple as in the first case. It is no longer generally true that they coincide with the L.S. estimates. In the applications the correlation matrix of the disturbance is seldom known. As this is not needed for the construction of the L.S. estimates which are optimal in the case of a nonautocorrelated disturbance, it seems natural to ask if they have some optimum property for large samples.

Looking at the problem from another point of view, we ask if it can happen that the knowledge of the correlation matrix does not contain any information relative to our problem of inference for large samples.

The main result is given in Theorems 3 and 5 and their corollaries. They express the asymptotic efficiency of the L.S. estimates in terms of the spectrum of the process and the spectrum of the regression sequence to be defined below. As a consequence of these results, we show that in the case of trigonometric or polynomial regression the L.S. estimates are asymptotically efficient.

The problem of estimating a constant mean value of a stationary process, studied in [2] and [3], can be considered as a special case of our present problem. Several authors have studied related questions. We refer especially to [5].

2. The disturbance. The process x_ν is observed at the points $\nu = 0, 1, \cdots, N$. It is convenient to allow the process to take complex values.

Received 3/10/53, revised 6/2/53.

[1] This paper was prepared with the partial support of the Office of Naval Research at the University of Chicago and at the University of California, Berkeley.

We shall suppose that the disturbance y_ν has mean zero and finite moments of the second order. Introduce the covariance matrix

$$M_N = \begin{Bmatrix} \rho_{00}, & \rho_{01}, & \cdots & \rho_{0N} \\ \rho_{10}, & \rho_{11}, & \cdots & \rho_{11N} \\ \rho_{N0}, & \rho_{N1}, & \cdots & \rho_{NN} \end{Bmatrix}; \qquad \rho_{\nu\mu} = E\bar{y}_\nu y_\mu ,$$

where, as usual, $\bar{y}_\nu$ is the complex conjugate of y_ν. Supposing the disturbance to be stationary (in the wide sense) we can write

$$\rho_{\nu\mu} = \rho_{\nu-\mu} = E\bar{y}_\nu y_\mu .$$

As is well known there exists a bounded, nondecreasing function $F(\lambda)$ defined in the interval $(-\pi, \pi)$ such that

$$\rho_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF(\lambda).$$

Further the process itself has a similar representation

$$y_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dZ(\lambda),$$

where $Z(\lambda)$ is an orthogonal process and

$$E \mid Z(\lambda_2) - Z(\lambda_1) \mid^2 = \mid F(\lambda_2) - F(\lambda_1) \mid.$$

We shall consider the class of processes which have an absolutely continuous spectrum with a *continuous* and *positive* spectral density. Then

$$F(b) - F(a) = \int_a^b f(\lambda)\, d\lambda$$

where $f(\lambda)$ is the spectral density. This class of processes shall be denoted by Y.

3. The regression variables. In order not to overshadow the idea of the proof, we shall at first consider the case of only one regression variable φ_ν. We shall deal only with the case when there exists a consistent estimate of c. We shall later see that this implies that

$$\textstyle\sum_{\nu=0}^{\infty} \mid \varphi_\nu \mid^2 = \infty .$$

As we shall see in Section 5, the asymptotic efficiency of the L.S. estimates is determined by certain properties of the sequence $\varphi_0, \varphi_1, \varphi_2, \cdots$. In order to specify these properties, the most straightforward thing to do would be to assume that the sequence has a Fourier-Stieltjes representation. But then we could not deal with even such a simple case as $\varphi_\nu = A + B\nu$, $B \neq 0$. Instead we shall use a method which is an extension of generalized harmonic analysis.

Introduce the notation $\Phi(N) = \sum_{\nu=0}^{N} \mid \varphi_\nu \mid^2$. This is an unbounded nondecreasing sequence for $N = 0, 1, \cdots$.

DEFINITION 1. If for every $n \geqq 0$ the limits

$$R_n = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_\nu \bar{\varphi}_{\nu+n}$$

exist, we shall call the sequence $\{\varphi_\nu\}$ *regular* and write $\{\varphi_\nu\} \varepsilon R$.

DEFINITION 2. Let $\{\varphi_\nu\} \varepsilon R$. If for every $h \geqq 0$ the sequences $\{\varphi_{\nu+h}\}$ all generate the same set of

$$R_n = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_{\nu+h} \bar{\varphi}_{\nu+h+n},$$

we shall say that $\{\varphi_\nu\}$ is a *stationary sequence*.

DEFINITION 3. If the sequence $\Phi(N)$ satisfies

$$\lim_{N \to \infty} \frac{\Phi(N+h)}{\Phi(N)} = 1$$

for every integer h we shall say that $\Phi(N)$ is *slowly increasing*.

THEOREM 1. *In order that a regular sequence be stationary it is necessary and sufficient that $\Phi(N)$ be slowly increasing.*

PROOF: Suppose that $\{\varphi_\nu\}$ is stationary. Then

$$R_n = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_\nu \bar{\varphi}_{\nu+n} = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_{\nu+h} \bar{\varphi}_{\nu+n+h}.$$

Since $\Phi(N) \to \infty$ as $N \to \infty$, we get

$$\lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_\nu \bar{\varphi}_{\nu+n} = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N+h} \varphi_\nu \bar{\varphi}_{\nu+n}.$$

Thus

$$R_n = \lim_{N \to \infty} \frac{\Phi(N+h)}{\Phi(N)} R_n.$$

As $R_0 = 1$ the sequence $\Phi(N)$ must be slowly increasing. By reversing the order of the proof the sufficiency of the condition follows.

THEOREM 2. *A stationary sequence has a spectrum in the sense that for every n*

$$R_n = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_\nu \bar{\varphi}_{\nu+n} = \int_{-\pi}^{\pi} e^{in\lambda} \, d\varphi(\lambda), \tag{1}$$

where $\varphi(\lambda)$ is a distribution function defined in $(-\pi, \pi)$.

PROOF. For $n < 0$ some of the first terms in the sum appearing in (1) are not defined. Since $\Phi(N)$ tends to infinity with N, we can assign arbitrary values to them as the limits do not depend upon them. Defining $\varphi_\nu = 0$ for $\nu < 0$, we see that for $n < 0$

$$R_{-n} = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_\nu \bar{\varphi}_{\nu-n} = \lim_{N \to \infty} \frac{1}{\Phi(N)} \sum_{\nu=n}^{N} \varphi_\nu \bar{\varphi}_{\nu-n}$$

$$= \lim_{N \to \infty} \frac{\Phi(N-n)}{\Phi(N)} \frac{1}{\Phi(N-n)} \overline{\sum_{0}^{N-n} \varphi_\nu \bar{\varphi}_{\nu+n}}.$$

Thus, since $\Phi(N)$ is slowly increasing, we see that $R_{-n} = \bar{R}_n$, that is, the matrix $R_q = \{R_{n-m}; n, m = 0, 1, \cdots q\}$ is Hermitian.

A similar reasoning proves the second equality in

$$X'R_qX = \sum_{n,m=1}^{q} R_{n-m} x_n \bar{x}_m = \lim_{N\to\infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \left| \sum_{n=1}^{q} x_n \bar{\varphi}_{\nu+n} \right|^2 \geqq 0.$$

As all the matrices R_q are then nonnegative, the existence of a bounded nondecreasing function $\varphi(\lambda)$ satisfying (1) follows. Because $1 = R_0 = \int_{-\pi}^{\pi} d\varphi(\lambda)$ we see that $\varphi(\lambda)$ is a distribution function in $(-\pi, \pi)$.

The set of points of increase of $\varphi(\lambda)$ is called the spectrum of the sequence $\{\varphi_\nu\}$ and is denoted $S(\varphi)$. Although a stationary sequence determines the spectral distribution function $\varphi(\lambda)$, it is evident that the sequence is far from determined by $\varphi(\lambda)$. This is equivalent to the fact that the covariance matrix of a stochastic process does not determine the realization of it. Hence to a given spectrum corresponds a multitude of possible sequences.

In Sections 8 and 9 we shall study the spectra of sequences that appear in certain cases as regression variables.

4. The L.S. estimate. The L.S. estimate coincides with the B.L.U. estimate calculated under the hypothesis that the disturbance is uncorrelated. It is

$$c_{LS}^* = \sum_{\nu=0}^{N} x_\nu \bar{\varphi}_\nu \Big/ \sum_{\nu=0}^{N} | \varphi_\nu |^2,$$

and its variance, still calculated under the same hypothesis, is

$$D^2[c_{LS}^*] = \frac{1}{\sum_{\nu=0}^{N} | \varphi_\nu |^2} = \frac{1}{\Phi(N)}. \tag{2}$$

In the following sections we shall study the asymptotic behavior of the L.S. estimate. As we deal with a *linear problem of inference* it is natural to define the efficiency of a linear estimate α^* as

$$e_N(\alpha^*) = D(\alpha_{\text{opt}}^*)/D(\alpha^*),$$

where α_{opt}^* denotes the B.L.U. estimate. The asymptotic efficiency $e(\alpha^*)$ is taken as the limit of $e_N(\alpha^*)$ if this limit exists.

Because of the linear nature of the problem, it is also natural to call a linear estimate consistent if and only if it converges in the mean to the true value.

It is clear that in the case of a normal process these linear definitions coincide with the usual ones. We shall, however, not make any assumptions regarding the distribution of the process beyond properties of the first and second moments.

Consider two spectral intensities $f(\lambda)$ and $g(\lambda)$ corresponding to two processes in the class Y. Introduce

$$\max f(\lambda) = f_2 \qquad \max g(\lambda) = g_2$$
$$\min f(\lambda) = f_1 \qquad \min g(\lambda) = g_1.$$

For a certain linear combination $c^* = \sum_{\nu=0}^{N} c_\nu x_\nu$ we have, using the spectral representation of the process,

$$c^* = \int_{-\pi}^{\pi} \sum_{\nu=0}^{N} c_\nu e^{i\nu\lambda} \, dZ(\lambda).$$

Hence, if $f_x(\lambda)$ is the true spectral density,

$$D^2[c^*] = \int_{-\pi}^{\pi} \left| \sum_{\nu=0}^{N} c_\nu e^{i\nu\lambda} \right|^2 f_x(\lambda) \, d\lambda.$$

Calculating the variance under the two hypotheses $f_x(\lambda) = f(\lambda)$ and $f_x(\lambda) = g(\lambda)$ we have

$$\text{(3)} \qquad 0 < \frac{f_1}{g_2} \leqq \frac{D_f^2[c^*]}{D_g^2[c^*]} = \frac{\int | \sum c_\nu e^{i\nu\lambda} |^2 f(\lambda) \, d\lambda}{\int | \sum c_\nu e^{i\nu\lambda} |^2 g(\lambda) \, d\lambda} \leqq \frac{f_2}{g_1} < \infty.$$

Suppose that $\Phi(N) \to \infty$. Taking $g(\lambda) = \pi/2$ and $f(\lambda)$ as the true spectral density, we get from (3)

$$D_f^2[c_{LS}^*] \leqq 2\pi f_2/\Phi(N) \to 0,$$

so that there exists a consistent estimate of c.

On the other hand, suppose that c^* is a consistent estimate of c. Then there exists a consistent unbiased estimate, say c_U^*. But it follows from (3) that

$$\frac{1}{\Phi(N)} = D_g^2[c_{LS}^*] \leqq D_g^2[c_U^*] \leqq \frac{1}{2\pi f_1} D_f^2[c_U^*] \to 0.$$

Thus, the existence of a consistent estimate is equivalent to $\Phi(N) \to \infty$.

5. The asymptotic efficiency of the L.S. estimate.

THEOREM 3. *Suppose that*

(a) *the regression sequence φ_ν is stationary and denote its spectral distribution function by $\varphi(\lambda)$.*

(b) *the disturbance $y_\nu \in Y$ and denote its spectral density by $f(\lambda)$.*

Then the asymptotic efficiency of the L.S. estimate is given by the expression

$$\text{(4)} \qquad e^2(c_{LS}^*) = \frac{1}{\int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f(\lambda)} \int_{-\pi}^{\pi} f(\lambda) \, d\varphi(\lambda)}.$$

PROOF. We shall carry out the proof in two stages, first proving (4) for any disturbance of the autoregressive type, and then extending it to any process in the class Y.

Suppose that the disturbance is generated by the autoregressive scheme

$$\alpha_0 y_{\nu+p} + \alpha_1 y_{\nu+p-1} + \cdots \alpha_p y_\nu = \epsilon_\nu$$

with $E\epsilon_\nu = 0$ and $E\epsilon_\nu\bar{\epsilon}_\mu = \delta_{\nu\mu}$: As usual we shall suppose that the roots of the characteristic equation

$$\alpha_0 z^p + \alpha_1 z^{p-1} + \cdots + \alpha_p = 0$$

are all of modulus less than one. We then know that $Ey_\mu\bar{\epsilon}_\nu = 0$ for $\mu < \nu + p$. The spectral density of the y-process is

$$f(\lambda) = 1/2\pi \left| \sum_{r=0}^{p} \alpha_r \epsilon^{ir\lambda} \right|^2. \tag{5}$$

There is no linear relation between the y's. We can then use the Gram-Schmidt procedure to orthonormalize the sequence $y_0, y_1, y_2 \cdots y_{p-1}$ and get

$$\begin{aligned} &c_{00}y_0 &&= \eta_0 \\ &c_{11}y_1 + c_{10}y_0 &&= \eta_1 \\ &\cdots &&\cdots \\ &c_{p-1\,p-1}y_{p-1} + c_{p-1\,p-2}y_{p-2} + \cdots + c_{p-1\,0}y_0 &&= \eta_{p-1} \end{aligned}$$

where

$$E\eta_\nu\bar{\eta}_\mu = \delta_{\nu\mu}; \qquad \nu, \mu = 0, 1, \cdots p - 1.$$

As ϵ_ν is orthogonal to $\eta_0, \eta_1, \cdots \eta_{p-1}$ for $\nu \geqq 0$ we see that putting $\eta_\nu = \epsilon_{\nu-p}$ for $\nu \geqq p$, the sequence $(\eta_0, \eta_1, \eta_2, \cdots$ ad inf.) is orthonormal. Every y_ν, $\nu \geqq 0$, is a linear combination of $\eta_\nu, \eta_{\nu-1}, \cdots \eta_0$.

We have the relation

$$x_\nu = c\varphi_\nu + y_\nu, \qquad \nu \geqq 0,$$

and, having observed a sample $x_0, x_1, \cdots x_N$, we want to consider an estimate c^* of c

$$c^* = \sum_{\nu=0}^{N} c_\nu x_\nu = c \sum_{\nu=0}^{N} c_\nu \varphi_\nu + \sum_{\nu=0}^{N} c_\nu y_\nu .$$

Let us call L the linear transformation that carried $(y_0, y_1, \cdots, y_N)$ over in $(\eta_0, \eta_1, \cdots, \eta_N)$. The corresponding matrix has then the rows $(c_{00}, 0, 0, \cdots, 0)$; $(c_{10}, c_{11}, \cdots, 0)$; $\cdots$; $(c_{p-1\,0}, c_{p-1\,1}, \cdots, c_{p-1\,p-1}, 0, \cdots, 0)$; $(\alpha_p, \alpha_{p-1}, \cdots, \cdots, \alpha_0, 0, \cdots, 0)$ and so on. Introduce

$$L(x_0, x_1, \cdots, x_N) = (\xi_0, \xi_1, \cdots, \xi_N),$$

$$L(\varphi_0, \varphi_1, \cdots, \varphi_N) = (\beta_0, \beta_1, \cdots, \beta_N).$$

As L is nonsingular, every linear estimate can be written as

$$c^* = \sum_{\nu=0}^{N} \gamma_\nu \xi_\nu = c \sum_{\nu=0}^{N} \gamma_\nu \beta_\nu + \sum_{\nu=0}^{N} \gamma_\nu \eta_\nu .$$

But now the η's are uncorrelated and all have unit variance. Because of (2) we then see that the variance of the B.L.U. estimate is

$$D^2[c^*_{BLU}] = 1 \Big/ \sum_{\nu=0}^{N} |\beta_\nu|^2. \tag{6}$$

To get an asymptotic expression for the minimum variance values for large values of N we note that the p first terms in the sum in the denominator of (6) contribute only a finite amount. Hence we need not take them into account as the sum diverges because of what has been said in Section 4. We have

$$\frac{1}{\Phi(N)} \sum_{\nu=p}^{N} |\beta_\nu|^2 = \sum_{t,u=0}^{p} \alpha_t \bar{\alpha}_u \frac{1}{\Phi(N)} \sum_{\nu=p}^{N} \varphi_{\nu-t} \bar{\varphi}_{\nu-u}.$$

Hence

$$\lim_{N\to\infty} \Phi^{-1}(N) D^{-2}[c^*_{BLU}] = \sum_{t,u=0}^{p} \alpha_t \bar{\alpha}_u R_{t-u}.$$

Using the spectral representation of the R's we get

$$\lim_{N\to\infty} \Phi^{-1}(N) D^{-2}[c^*_{BLU}] = \int_{-\pi}^{\pi} \left| \sum_{t=0}^{p} \alpha_t e^{it\lambda} \right|^2 d\varphi(\lambda).$$

Combined with (5), this gives

$$D^2[c^*_{BLU}] \sim \frac{2\pi}{\Phi(N)} \frac{1}{\displaystyle\int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f(\lambda)}}. \tag{7}$$

Let us now derive an asymptotic expression for the variance of the L.S. estimate. We have

$$c^*_{LS} = \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \bar{\varphi}_\nu x_\nu$$

and hence

$$D^2[c^*_{LS}] = \frac{1}{\Phi^2(N)} \sum_{\nu,\mu=0}^{N} \bar{\varphi}_\nu \varphi_\mu \rho_{\mu-\nu}.$$

Thus

$$\Phi(N) D^2[c^*_{LS}] = \sum_{n=0}^{N} \frac{\rho_n}{\Phi(N)} \sum_{\nu=0}^{N-n} \bar{\varphi}_\nu \varphi_{\nu+n} + \sum_{n=-1}^{-N} \frac{\rho_n}{\Phi(N)} \sum_{n=-1}^{N} \bar{\varphi}_\nu \varphi_{\nu+n}. \tag{8}$$

As ρ_n is dominated by some $Ka^{|n|}$ with $0 < a < 1$, and as

$$\frac{1}{\Phi(N)} \left| \sum_{\nu=0}^{N-n} \bar{\varphi}_\nu \varphi_{\nu+n} \right| \leqq \sqrt{\frac{1}{\Phi(N)} \sum_{\nu=0}^{N-n} |\varphi_\nu|^2 \frac{1}{\Phi(N)} \sum_{\nu=n}^{N} |\varphi_\nu|^2} \leqq 1$$

we can perform the limit operation in each term of (8). We then get

$$\lim_{N\to\infty} \Phi(N) D^2[c^*_{LS}] = \sum_{n=-\infty}^{\infty} \rho_n R_{-n} = \sum_{n=-\infty}^{\infty} \rho_n \int_{-\pi}^{\pi} e^{-in\lambda} d\varphi(\lambda).$$

As we have

$$f(\lambda) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \rho_n e^{-n\lambda i},$$

where the sum is uniformly convergent, we get

$$D^2[c_{LS}^*] \sim \frac{2\pi}{\Phi(N)} \int_{-\pi}^{\pi} f(\lambda)\, d\varphi(\lambda). \tag{6}$$

Combining (7) and (9) we have

$$e^2(c_{LS}^*) = \frac{1}{\displaystyle\int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f(\lambda)} \int_{-\pi}^{\pi} f(\lambda)\, d\varphi(\lambda)}$$

which proves the theorem for any autoregressive scheme.

To extend this relation to an arbitrary process in the class Y we shall use a method due to G. Szegö, who suggested it in a discussion with the author, to prove the asymptotic efficiency of the equidistributed estimate of the mean of a process in the class Y. In statistical terminology, it consists in approximating the process by an appropriate autoregressive scheme.

As $f(\lambda)$ has been supposed to be positive and continuous, it is seen (using a well known argument) that it is possible for any $\delta > 0$ to find a trigonometric polynomial $P = \sum_0^p \alpha_r e^{ir\lambda}$, where p is a sufficiently large number, such that

$$f(\lambda) \leqq 1/\mid s(\lambda) \mid^2 \quad \text{and} \quad 1/\mid s(\lambda) \mid^2 - f(\lambda) \leqq \delta.$$

As we are interested only in the modulus of $s(\lambda)$, we can if necessary change the α's leaving $\mid s(\lambda) \mid$ unchanged in such a way that all the roots of the characteristic equation are less than one in absolute value.

In this way we can find two spectral densities of the type (5) such that

$$f_1(\lambda) \leqq f(\lambda) \leqq f_2(\lambda), \tag{10}$$

$$f_2(\lambda) - f_1(\lambda) \leqq \delta. \tag{11}$$

For any linear estimate c^* it follows from (10) that its variances calculated under the three different hypotheses satisfy

$$D_{f_1}^2[c^*] \leqq D_f^2[c^*] \leqq D_{f_2}^2[c^*].$$

This gives us

$$\frac{2\pi}{\displaystyle\int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f_2(\lambda)}} \leqq \varliminf_{N\to\infty} \Phi(N) D_f^2[c_{BLU}^*] \leqq \varlimsup_{N\to\infty} \Phi(N) D_f^2[c_{BLU}^*] \leqq \frac{2\pi}{\displaystyle\int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f_2(\lambda)}}.$$

Using (11) we see that

$$\lim_{N\to\infty} \Phi(N) D_f^2[c_{BLU}^*] = \frac{2\pi}{\displaystyle\int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f(\lambda)}}.$$

Similarly (9) is extended to hold for all processes in the class Y. This completes the proof of Theorem 3.

COROLLARY. *In order that the L.S. estimate shall be asymptotically efficient whatever be the spectral intensity, it is necessary and sufficient that the spectrum* $S(\varphi)$ *consist of only one point.*

In other words, the knowledge of the true covariance matrix gives us no additional information of relevance to our problem if and only if $S(\varphi)$ consists of only one point.

PROOF. Supposing the asymptotic efficiency of the L.S. estimate to be e, we get from (4) using Schwarz' inequality

$$\frac{1}{e^2} = \int_{-\pi}^{\pi} \frac{d\varphi(\lambda)}{f(\lambda)} \int_{-\pi}^{\pi} f(\lambda)\, d\varphi(\lambda) \geqq \left[\int_{-\pi}^{\pi} d\varphi(\lambda)\right]^2 = 1.$$

The equality sign holds if and only if $f(\lambda) \equiv$ const. for all $\lambda \in S(\varphi)$. This identity is true for every spectral intensity in Y if and only if $S(\varphi)$ reduces to a single point. This condition looks at first very restrictive but, as we shall see later, it is satisfied in the most important cases of analytical regression.

6. Possible extensions. The extension of Theorem 3 to several regression variables is studied in the next section, so will not be discussed here.

We have demanded that the disturbance process have an absolutely continuous spectrum with a positive continuous spectral density. This is likely to be a much more stringent condition than what is needed for the theorem to be true. Consider the case $\varphi_\nu \equiv 1$. Then $\Phi(N) = N + 1$ and $R_n \equiv 1$, so that $S(\varphi)$ consists only of the point $\lambda = 0$. The theorem states in this case that the equidistributed estimate $\bar{x} = \sum_{\nu=0}^{N} x_\nu/(N + 1)$ of the mean value of a process belonging to the class Y is asymptotically efficient.

This result was first proved in [2] in the case of a continuous time parameter. The same proof can be used also in the considerably simpler case of a discrete parameter. The conditions imposed upon the process were weaker than those defining the class Y, and in [3] the result was extended to spectra containing also discrete and singular parts. This makes it at least plausible that Theorem 3 and its corollary hold for a class of disturbances considerably larger than Y. However, the method of proof used in this paper does not seem to lend itself easily to such an extension.

7. The case with several regression variables. In the treatment of this case we shall deal only with questions which did not appear when we had just one regression variable. The remaining part of the proof consists of a straightforward generalization of the procedure in Section 5.

In order to describe the spectral properties of an s-dimensional sequence $(\varphi_\nu^{(1)}, \varphi_\nu^{(2)}, \cdots \varphi_\nu^{(s)})$ we consider the expressions

$$\Phi^{(i)}(N) = \sum_{\nu=0}^{N} |\varphi_\nu^{(i)}|^2, \qquad i = 1, 2, \cdots s,$$

which are supposed to be unbounded and slowly increasing as $N \to \infty$. We suppose that the following limits exist

(12) $$R_{jk}(n) = \lim_{N\to\infty} \frac{1}{\sqrt{\Phi^{(j)}(N)\Phi^{(k)}_{(N)}}} \sum_{\nu=0}^{N} \varphi_\nu^{(j)}\bar{\varphi}_{\nu+n}^{(k)}, \qquad n \geqq 0.$$

Using the same argument as in Section 3 one can show that these limits are invariant against translations and that they have the properties of correlation matrices between s stationary and stationarily correlated stochastic processes.

From Cramér's extension (cf. [1]) of Khintchine's theorem on the representation of a correlation function we see that there exist s^2 complex-valued functions of bounded variation $F_{jk}(\lambda)$, $j, k = 1, 2, \cdots s$, defined in the interval $(-\pi, \pi)$ such that

$$R_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda}\, dF_{jk}(\lambda).$$

For the increments of $F_{jk}(\lambda)$ over an arbitrary interval in $(-\pi, \pi)$, the matrix $\{\Delta F_{jk}(\lambda); j, k = 1, 2, \cdots, s\}$ is Hermitian and nonnegative. From this it follows that the F_{jj} are distribution functions.

Of course we want to exclude the case when the regression variables are linearly dependent. We do this by assuming that the matrix $\{R_{jk}(0); j, k = 1, 2, \cdots s\}$ is nonsingular.

The introduction of the spectral measure is not quite as straightforward as for $s = 1$. Consider a rectangle r with sides (a_1, b_1), (a_2, b_2), $\cdots$, (a_s, b_s), in the s-cube with sides $(-\pi, \pi)$. Denoting the difference operators corresponding to these intervals by $\Delta_1, \Delta_2, \cdots, \Delta_s$, we form the nonnegative matrices

$$N_1 = \{\Delta_1 F_{jk}\}, \qquad N_2 = \{\Delta_2 F_{jk}\}, \qquad \cdots \qquad N_s = \{\Delta_s F_{jk}\}.$$

Let P be a permutation $(1, 2, \cdots s) \to (i_1, i_2, \cdots i_s)$. To P we associate the determinant

(13) $$D_P = |\, \Delta_{i_j} F_{jk}; j, k = 1, 2, \cdots s \,|.$$

We define the spectral measure of r as

(14) $$\varphi(r) = \frac{1}{s!} \sum_P D_P$$

where the summation is extended over all the permutations. We have to show that φ is nonnegative. As the value of φ for the whole cube is finite we can extend φ to a bounded measure in this cube.

As is well known there are Hermitian matrices $f_1, f_2, \cdots f_s$ so that $N_\nu = f_\nu^2$, $\nu = 1, 2, \cdots, s$, that is

$$\Delta_\nu F_{jk} = \sum_{n=1}^{s} f_{jn}^{(\nu)} f_{nk}^{(\nu)}.$$

Then

$$D_P = \sum_{n_1, n_2, \cdots, n_s = 1}^{s} | f_{i n_i}^{(i_j)} f_{n_j k}^{(i_j)} | = \sum_{n_1, n_2, \cdots, n_s = 1}^{s} f_{1 n_1}^{(i_1)} f_{2 n_2}^{(i_2)} \cdots f_{s n_s}^{(i_s)} | f_{n_j k}^{(i_j)} | .$$

After permutation of the rows of the determinants

$$D_P = \sum_{l_1, l_2, \cdots, l_s = 1}^{s} (-1)^P f_{k_1 l_1}^{(1)} f_{k_2 l_2}^{(2)} \cdots f_{k_s l_s}^{(s)} | f_{l_j k}^{(j)} |$$

where $(-1)^P$ is $+1$ or -1 according to whether P is an even or odd permutation. The permutation $(1, 2, \cdots, s) \to (k_1, k_2, \cdots, k_s)$ is the inverse of P and has the same order. But from the Hermitian property of $F^{(\nu)}$,

$$\sum_P D_P = \sum_{l_1, l_2, \cdots, l_s = 1}^{s} | f_{k l_j}^{(j)} | \, | f_{l_j k}^{(j)} | \geqq 0.$$

The measure φ obtained in this way will be called the *spectral measure* of the regression variables. The set of points in s-space such that every cube containing one of them in its interior has positive spectral measure is called the spectrum of the regression variables and denoted by $S(\varphi)$. From $| R_{jk}(0) | \neq 0$ it follows that $S(\varphi)$ is not empty. It is clearly symmetric with respect to permutations of the coordinate axes.

THEOREM 4. *The joint asymptotic efficiency of the L.S. estimates is given by*

$$e^{2s} = \frac{\left[\int d\varphi \right]^2}{\int f(u_1) \cdots f(u_s) \, d\varphi \int \frac{1}{f(u_1) \cdots f(u_s)} \, d\varphi} \tag{15}$$

where the integrations are carried out over $S(\varphi)$.

PROOF. It is sufficient to deal with the case when y_ν is an autoregressive process, as then we can apply the same approximation procedure as in Section 5.

Consider a matrix of the form

$$M = \left\{ \int_{-\pi}^{\pi} k(\lambda) \, dF_{jk}(\lambda); \qquad j, k = 1, 2, \cdots, s \right\}$$

where $k(\lambda)$ is a continuous function. Then one can show easily that

$$| M | = \int_{S(\phi)} k(u_1) \cdots k(u_s) \, d\varphi(u_1, \cdots, u_s).$$

Denote by A, B, C, respectively, the matrices obtained in this way by putting $k(\lambda) = 1, f(\lambda), 1/f(\lambda)$. It is easily seen that $A = \{R_{jk}(0)\}$ and that B and C are nonsingular.

Proceeding in the same way as before we derive the relation

$$\lim_{n \to \infty} \frac{1}{\sqrt{\Phi^{(j)}(N) \Phi^{(k)}(N)}} \operatorname{cov} \left[\sum_{\nu=0}^{N} \bar{\varphi}_\nu^{(j)} x_\nu ; \sum_{\nu=0}^{N} \bar{\varphi}_\nu^{(k)} x_\nu \right] = 2\pi \int_{-\pi}^{\pi} f(\lambda) \, dF_{kj}(\lambda).$$

From this,

$$\lim_{N\to\infty} \frac{1}{\sqrt{\Phi^{(j)}(N)\Phi^{(k)}(N)}} \operatorname{cov}\,[c^*_{j,LS}\,;\, c^*_{k,LS}] = m_{jk}\,,$$

where $\{m_{jk}\,; j, k = 1, 2, \cdots s\} = A^{-1}B^*A^{-1}$. To find the asymptotic covariances we use the same linear transformation as before and obtain

$$\lim_{n\to\infty} \frac{1}{\sqrt{\Phi^{(j)}(N)\Phi^{(k)}(N)}} \operatorname{cov}\,[c^*_{j,BLU}\,;\, c^*_{k,BLU}] = n_{jk}$$

where $\{n_{jk}\,; j, k = 1, 2, \cdots s\} = C^{-1}$.

We define the (joint, linear) efficiency e of s estimates with moment matrix α as

$$e = \lim_{N\to\infty} e_N\,, \qquad e_N^{2s} = |\,\beta\,|/|\,\alpha\,|,$$

where β is the moment matrix corresponding to the B.L.U. estimates.

Combining the obtained relations we get

$$e^{2s} = \frac{|\,A\,|^2}{|\,B\,|\,|\,C\,|} = \frac{\left[\int d\varphi\right]^2}{\int f(u_1)\cdots f(u_s)\,d\varphi \int \frac{1}{f(u_1)\cdots f(u_s)}\,d\varphi}.$$

COROLLARY. *In order that the L.S. estimates shall be (jointly) asymptotically efficient for every disturbance of the class Y it is necessary and sufficient that the spectrum $S(\varphi)$ of the regression variables contain only one point and the symmetric images of it.*

PROOF. From Schwarz' inequality it follows immediately that

$$\left[\int d\varphi\right]^2 \leqq \int f(u_1)\cdots f(u_s)\,d\varphi \int \frac{1}{f(u_1)\cdots f(u_s)}\,d\varphi,$$

with equality if and only if $f(u_1)f(u_2)\cdots f(u_s) \equiv$ const. for almost all

$$(u_1\,, u_2\,, \cdots u_s)\ \varepsilon\ S(\varphi).$$

If the spectrum contains at least two points which are not symmetric, the asymptotic efficiency cannot be one for all residuals of type Y. The sufficiency of the condition is obvious.

8. The case of analytical regression. When the regression sequences are given a priori in the form of analytic expressions in ν, we shall speak of analytical regression as opposed to the case when $\{\varphi_\nu^{(1)}\}$, $\{\varphi_\nu^{(2)}\}$, $\cdots$ are obtained as measurements of certain variables which are of a nondeterministic structure.

Let us suppose that we have a pair of regression variables of the type

$$\varphi_\nu = \nu^p \int_{-\pi}^{\pi} e^{i\nu\lambda}\,dF(\lambda) \qquad \Psi_\nu = \nu^q \int_{-\pi}^{\pi} e^{i\nu\lambda}\,dG(\lambda)$$

where p and q are nonnegative integers. Both F and G are of bounded variation but not necessarily real-valued. To bar cases where no consistent estimate of the regression coefficients exists, we shall suppose that both F and G have at least one discontinuity. This condition is satisfied in the applications we are going to consider.

Separating the continuous and discontinuous parts of F and G we have

$$F(\lambda) = F_c(\lambda) + F_d(\lambda) \qquad G(\lambda) = G_c(\lambda) + G_d(\lambda).$$

Let us put

$$c_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF_c(\lambda); \qquad d_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF_d(\lambda);$$

$$\gamma_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dG_c(\lambda); \qquad \delta_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dG_d(\lambda).$$

We have

(16a)
$$\lim_{N\to\infty} \frac{1}{2N} \sum_{-N}^{N} | c_\nu |^2 = \lim_{N\to\infty} \frac{1}{2N} \sum_{-N}^{N} | \gamma_\nu |^2 = 0$$

and

(16b)
$$\begin{cases} \lim_{N\to\infty} \frac{1}{2N} \sum_{-N}^{N} | d_\nu |^2 = \sum_u | m_u |^2 > 0 \\ \lim_{N\to\infty} \frac{1}{2N} \sum_{-N}^{N} | \delta_\nu |^2 = \sum_u | \mu_u |^2 > 0. \end{cases}$$

The frequencies where either $F(\lambda)$ or $G(\lambda)$ has a saltus are denoted λ_n, with $\Delta F(\lambda_n) = m_n$, and $\Delta G(\lambda_n) = \mu_n$.

Consider the quantities

(17)
$$\begin{cases} r_N(s) = \frac{1}{N^{2r+1}} \sum_{\nu=0}^{N} \varphi_\nu \bar{\Psi}_{\nu+s} \\ r'_N(s) = \frac{1}{N^{2r+1}} \sum_{\nu=0}^{N} d_\nu \bar{\delta}_{\nu+s} \nu^p (\nu + s)^q \end{cases}$$

where $r = (p + q)/2$. We have for large values of N

$$\begin{aligned} | r_N(s) - r'_N(s) | &\leqq \frac{2}{N} \sum_{\nu=0}^{N} | c_\nu \gamma_{\nu+s} | + \frac{2}{N} \sum_{\nu=0}^{N} | c_\nu \delta_{\nu+s} | + \frac{2}{N} \sum_{\nu=0}^{N} | d_\nu \gamma_{\nu+s} | \\ &\leqq 2 \sqrt{\frac{1}{N} \sum_{\nu=0}^{N} | c_\nu |^2 \frac{1}{N} \sum_{\nu=0}^{N} | \gamma_{\nu+s} |^2} + 2 \sqrt{\frac{1}{N} \sum_{\nu=0}^{N} | c_\nu |^2 \frac{1}{N} \sum_{\nu=0}^{N} | \delta_{\nu+s} |^2} \\ &\quad + 2 \sqrt{\frac{1}{N} \sum_{\nu=0}^{N} | d_\nu |^2 \frac{1}{N} \sum_{\nu=0}^{N} | \gamma_{\nu+s} |^2}. \end{aligned}$$

Using (16) we see that $\lim_{N\to\infty} | r_N(s) - r'_N(s) | = 0$. But it is easy to calculate the limit of $r'_N(s)$. If the number of discontinuities is finite we get immediately

$$
(18)\qquad \begin{aligned}\lim_{N\to\infty} r_N'(s) &= \frac{1}{N^{2r+1}} \sum_{u,v} \sum_{\nu=0}^{N} \nu^p\, m_u\, e^{i\nu\lambda_u}(\nu + s)^q\, \bar{\mu}_v\, e^{-i(\nu+s)\lambda_v} \\ &= \frac{1}{N^{2r+1}} \sum_u m_u\, \bar{\mu}_u\, e^{-is\lambda_u}.\end{aligned}
$$

If the number of discontinuities is infinite, a slight modification of the previous inequality shows that (18) holds in this case, too.

It is now easy to deal with the case when the regression variables are of the form

$$
(19)\quad \varphi_\nu = \nu^p \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF_0(\lambda) + \nu^{p-1} \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF_1(\lambda) + \cdots + \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF_p(\lambda).
$$

For $\nu > 0$ we can write this as

$$
\varphi_\nu = \nu^p \left[\int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF_0(\lambda) + a_\nu^{(1)} + \cdots + a_\nu^{(p)} \right].
$$

Since

$$
\sum_{\nu=1}^{\infty} | a_\nu^{(i)} |^2 < \infty, \qquad\qquad i = 1, 2, \cdots p,
$$

it follows that $a_\nu^{(i)}$ have Fourier-Stieltjes representations with absolutely continuous weighting functions. Hence

$$
\varphi_\nu = \nu^p \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF(\lambda),
$$

where $F(\lambda)$ has the same discontinuous part as $F_0(\lambda)$. We shall call the frequencies corresponding to these discontinuities the *stressed frequencies* of the sequence $\{\varphi_\nu\}$.

THEOREM 5. *A pair of regression variables* (φ, Ψ) *where each one is of the type* (19) *is stationary with the spectral distribution functions*

$$
F_{11}(\lambda) = \frac{1}{\sum_u | m_u |^2} \sum_{-\lambda_u \le \lambda} | m_u |^2
$$

$$
F_{12}(\lambda) = \frac{\sqrt{(2p+1)(2q+1)}}{2r+1} \frac{1}{\sqrt{\sum_u | m_u |^2 \sum_u | \mu_u |^2}} \sum_{-\lambda_u \le \lambda} m_u\, \bar{\mu}_u
$$

$$
F_{21}(\lambda) = \frac{\sqrt{(2p+1)(2q+1)}}{2r+1} \frac{1}{\sqrt{\sum_u | m_u |^2 \sum_u | \mu_u |^2}} \sum_{-\lambda_u \le \lambda} \bar{m}_u\, \mu_u
$$

$$
F_{22}(\lambda) = \frac{1}{\sum_u | \mu_u |^2} \sum_{-\lambda_u \le \lambda} | \mu_u |^2.
$$

PROOF. Putting $\Psi = \varphi$ and $s = 0$ in (17) we get from (18)

$$
\lim_{N\to\infty} \frac{1}{N^{2p+1}} \sum_{\nu=0}^{N} | \varphi_\nu |^2 = \lim_{N\to\infty} \frac{\Phi(N)}{N^{2p+1}} = \frac{1}{p+1} \sum_u | m_u |^2.
$$

Hence

(20)
$$\begin{cases} \Phi(N) \sim \dfrac{N^{2p+1}}{2p+1} \sum_u |m_u|^2 \\ \Psi(N) \sim \dfrac{N^{2q+1}}{2q+1} \sum_u |\mu_u|^2. \end{cases}$$

These functions are slowly increasing and, as we have already seen, $\{\varphi_\nu\}$ and $\{\Psi_\nu\}$ are regular. It follows that the pair (φ, Ψ) of regression variables is stationary.

We get from (17) and (18) that

$$R_{11}(n) = \frac{1}{\sum_u |m_u|^2} \sum_u |m_u|^2 e^{-in\lambda_v}$$

$$R_{12}(n) = \frac{\sqrt{(2p+1)(2q+1)}}{2r+1} \frac{1}{\sqrt{\sum_u |m_u|^2 \sum_u |\mu_u|^2}} \sum_u m_u \bar{\mu}_u e^{-in\lambda_u}$$

$$R_{21}(n) = \frac{\sqrt{(2p+1)(2q+1)}}{2r+1} \frac{1}{\sqrt{\sum_u |m_u|^2 \sum_u |\mu_u|^2}} \sum_u \bar{m}_u \mu_u e^{-in\lambda_v}$$

$$R_{22}(n) = \frac{1}{\sum_u |\mu_v|^2} \sum_u |\mu_u|^2 e^{-in\lambda_u}.$$

These relations determine the functions $F_{ij}(\lambda)$. We note that

$$| R_{jk}(0); j, k = 1, 2 | = 1 - | R_{12}(0) |^2$$

$$= 1 - \frac{(2p+1)(2q+1)}{(2r+1)^2} \frac{|\sum_u m_u \bar{\mu}_u|^2}{\sum_u |m_u|^2 \sum_u |\mu_u|^2}.$$

It is easily seen that this is positive, as is required if one wants to apply Theorem 4, if and only if at least one of the two following conditions is satisfied:

i. $p \neq q$.

ii. There is no linear relation $Am_u + B\mu_u = 0$ for all u. We note that if there is one such linear relation this implies that the stressed frequencies of $\{\varphi_\nu\}$ and $\{\Psi_\nu\}$ coincide.

Let us study the case of s regression variables of the type described above. If $\lambda_1, \lambda_2, \cdots$ is the set of all the stressed frequencies we will denote the saltus at λ_ν corresponding to the ith variable by $\mu_i^{(\nu)}$. The value of p for the ith variable will be denoted by p_i. Then it is easily seen that the matrix $A = \{R_{jk}(0)\}$ is given by $A = D\Lambda D$ where

$$\Lambda = \left\{ \sum_{\nu=1}^{\infty} \frac{\mu_j^{(\nu)} \bar{\mu}_k^{(\nu)}}{p_j + p_k + 1}; \; j, k = 1, 2, \cdots s \right\}$$

and D is the diagonal matrix

$$D = \left[\sqrt{\frac{(2p_j + 1)}{\sum_{\nu=1}^{\infty} | \mu_j^{(\nu)} |^2}} ; \quad j = 1, 2, \cdots, s \right],$$

so that A and Λ are singular or nonsingular at the same time. To get a criterion for the nonsingularity of Λ we consider the quadratic form

$$z^*\Lambda z = \sum_{\nu=1}^{\infty} \int_0^1 \left| \sum_{j=1}^{s} \mu_j^{(\nu)} x^{p_j} z_j \right|^2 dx$$

where z is a column vector with components $z_1, z_2, \cdots z_s$. If Λ is singular there exists a nonzero vector z for which the above quadratic form vanishes. But from this follows

$$\sum_{j=1}^{s} \mu_j^{(\nu)} x^{p_j} z_j = 0$$

for all ν and x. We will assume that we have labeled our regression variables in such a way that

$$p_j = \begin{cases} \pi_1 ; & j = 1, \cdots \alpha_1 \\ \pi_2 ; & j = \alpha_1 + 1, \cdots \alpha_2 \\ \cdots & \\ \pi_t ; & j = \alpha_{t-1} + 1, \cdots \alpha_t . \end{cases}$$

Then we get

$$\sum_{j=1}^{\alpha_1} \mu_j^{(\nu)} z_j = 0, \cdots, \sum_{j=\alpha_{j-1}+1}^{\alpha_t} \mu_j^{(\nu)} z_j = 0$$

for all ν, and as $z \neq 0$ at least one of these relations is nontrivial. The converse of this is shown in the same way. Hence a necessary and sufficient condition for A to be nonsingular is that in none of the t classes there exists a nontrivial linear relation between the saltuses corresponding to regression variables in such a class. For $s = 2$ this reduces to conditions 1 and 2 above.

Assuming that A is nonsingular, we will study the spectrum $S(\varphi)$. It is clear that it is discrete. Consider a point $l = (l_1, l_2, \cdots l_s)$. If $\varphi(l) > 0$ it is necessary that $l_1, l_2, \cdots l_s$ coincide with some of the stressed frequencies, but this is not always sufficient. Say that $l = (\lambda_{\nu_1}, \lambda_{\nu_2}, \cdots \lambda_{\nu_s})$ where the ν's are not necessarily distinct. Then

$$\varphi(l) = \int_0^1 \cdots \int_0^1 | D(x) |^2 dx_1 dx_2 \cdots dx_s$$

where $D(x) = | \mu_j^{(\nu_k)} x_k^{p_j} ; j, k = 1, 2, \cdots s |$.

To show this, consider a permutation $P: (1, 2, \cdots s) \rightarrow (i_1, i_2, \cdots i_s)$ and put $i_j = P_j$. Then for $j, k = 1, 2, \cdots s$,

$$\frac{\mu_j^{(\nu_{P_j})} \bar{\mu}_k^{(\nu_{P_j})}}{p_j + p_k + 1} = \int_0^1 \cdots \int_0^1 \Delta_P(x) \, dx_1 dx_2 \cdots dx_s$$

where

$$\Delta_P(x) = | \mu_j^{(\nu_{p_j})} \bar{\mu}_k^{(\nu_{p_j})} x_{p_j}^{p_j+p_k} |.$$

But

$$\Delta_P(x) = \prod_{n=1}^{s} \mu_n^{(\nu_{p_n})} x_{p_n}^{p_n} (-1)^P \, | \bar{\mu}_k^{(\nu_j)} x_j^{p_k} |$$

and hence $\sum_P \Delta_P(x) = | D(x) |^2$.

We are especially interested in the case when $S(\varphi)$ reduces to a single point (and its symmetric images). Let us treat two important cases.

First assume that all p_1 are different. If $\varphi(l) = 0$ we see that $D(x) \equiv 0$, implying that

$$\mu_{i_1}^{(\nu_1)} \mu_{i_2}^{(\nu_2)} \cdots \mu_{i_s}^{(\nu_s)} = 0$$

for all permutations $(1, 2, \cdots s) \to (i_1, i_2, \cdots i_s)$. If one regression variable has two stressed frequencies, say $\varphi^{(1)}$ has λ' and λ'', we choose

$$l' = (\lambda', l_2, l_3, \cdots l_s) \qquad l'' = (\lambda'', l_2, l_3, \cdots l_s)$$

where l_2 is a stressed frequency of $\varphi_\nu^{(2)}$ and so on. Taking $(i_1, i_2, \cdots i_s) = (1, 2, \cdots s)$ we see that $\varphi(l') > 0$ and $\varphi(l'') > 0$. But l' cannot be obtained by permutations from l''. On the other hand, if each variable has only one stressed frequency it follows that $S(\varphi)$ consists of only one point because if $\varphi(l) > 0$ then there is at least one permutation $(i_1, i_2, \cdots i_s)$ so that

$$\mu_{i_1}^{(\nu_1)} \neq 0, \qquad \mu_{i_2}^{(\nu_2)} \neq 0, \qquad \cdots \qquad \mu_{i_s}^{(\nu_s)} \neq 0.$$

But then l must be a permutation of $(\lambda_1, \lambda_2, \cdots \lambda_s)$ where λ_1 is the only stressed frequency of $\varphi_\nu^{(1)}$ and so on. This proves:

COROLLARY 1. *If all p_i are different, it is necessary and sufficient for the L.S. estimates to be asymptotically efficient, whatever be the spectral density $f(\lambda)$, that each regression variable have only one stressed frequency.*

Assume now instead that all $p_i = p$. Then

$$D(x) = | \mu_j^{(\nu_k)}; j, k = 1, 2, \cdots s | (x_1 x_2 \cdots x_s)^p.$$

If $\varphi(l) > 0$ it is clear that the ν's must be different. If each regression variable has only one stressed frequency $S(\varphi)$ must consist of one point only. We are going to show that the converse of this is true.

Let l be the only point in $S(\varphi)$ and suppose we have labeled our variables so that $l = (\lambda_1, \lambda_2, \cdots \lambda_s)$. Then the column vectors

$$\begin{aligned} v_1 &= [\mu_1^{(1)}, \mu_2^{(1)}, \cdots \mu_s^{(1)}] \\ v_2 &= [\mu_1^{(2)}, \mu_2^{(2)}, \cdots \mu_s^{(2)}] \\ \cdots &\quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ v_s &= [\mu_1^{(s)}, \mu_2^{(s)}, \cdots \mu_s^{(s)}] \end{aligned}$$

are linearly independent and span R^s. If there is one more stressed frequency, say λ_{s+1}, then

$$v_{s+1} = [\mu_1^{(s+1)}, \mu_2^{(s+1)}, \cdots \mu_s^{(s+1)}] \neq 0,$$

and there are constants β_j so that

$$v_{s+1} = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_s v_s .$$

As the point

$$l^{(i)} = (\lambda_1, \lambda_2, \cdots \lambda_{i-1}, \lambda_{i+1}, \cdots \lambda_s, \lambda_{s+1})$$

is not an image of l it should have spectral mass zero. Then the determinant formed from the corresponding column vectors should vanish

$$| v_1, v_2, \cdots, v_{i-1}, v_{i+1}, \cdots, v_s, v_{s+1} | = \sum_{j=1}^{s} \beta_j | v_1 \cdots, v_j |$$

$$= \beta_i | v_1, v_2, \cdots, v_{i-1}, v_{i+1}, \cdots, v_s, v_i | = 0$$

so that $\beta_i = 0$. As this would hold for all i we have obtained a contradiction.

COROLLARY 2. *If all $p_i = p$ then a necessary and sufficient condition for the L.S. estimates to be asymptotically efficient, whatever be the spectral density $f(\lambda)$, is that each regression variable should have one stressed frequency only.*

COROLLARY 3. *In the case of parabolic regression, $\varphi_\nu^{(p)} = \nu^p$ and*

$$x_\nu = \sum_p c_p \nu^p + y_\nu,$$

the L.S. estimates are asymptotically efficient.

This follows immediately as $| A | > 0$ because the p's corresponding to different regression variables are different and each component has $\lambda = 0$ as the only stressed frequency.

COROLLARY 4. *In the case of trigonometric regression,*

$$\varphi_\nu^{(n)} = e^{i\nu\lambda_n} \qquad \text{and} \qquad x_\nu = \sum_n c_n e^{i\nu\lambda_n} + y_\nu,$$

the L.S. estimates are asymptotically efficient if the λ's are different.

In this case $p = 1$ for each component but as the nth regression variable has only one stressed frequency, λ_n, it follows that $| A | > 0$ and the corollary holds.

Other types of analytical regression could be studied in a similar way. In order that the same method shall be applicable, the regression variables must not be too small at infinity (in which case $\Phi(\infty) < \infty$ and no consistent estimate exists) or too large (so that $\Phi(N)$ is not slowly increasing and the sequence is not stationary). The first case does not seem to be of much interest but perhaps this can not be said about the second one.

Consider the case of exponential regression, that is $\varphi_\nu = a^\nu$, and let $a > 1$.

Then for $n \geqq 0$

$$R_n = \lim_{N\to\infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} \varphi_\nu \bar{\varphi}_{\nu+n} = a^n \lim_{N\to\infty} \frac{1}{\Phi(N)} \sum_{\nu=0}^{N} |\varphi_\nu|^2 = a^n$$

so that φ_ν is a regular sequence. As

$$\lim_{N\to\infty} \frac{\Phi(N+n)}{|\Phi(N)|} = a^{2n} > 1$$

the sequence is not stationary and one verifies easily that it has no spectrum. One could still in a natural way attribute certain stressed frequencies to functions of this type of growth and in this case we would get $\lambda = 0$ as the only stressed frequency. We shall not pursue this question but only show that the L.S. estimate of the corresponding regression coefficient is still not asymptotically efficient for all disturbances of the class Y.

As φ_ν is increasing very fast, it seems probable that $c_0^* = x_N/\varphi_N = a^{-N}x_N$ would be an asymptotically good estimate. Its variance is

$$D^2[c_0^*] = a^{-2N}D^2[x_N] = a^{-2N}\int_{-\pi}^{\pi} f(\lambda)\, d\lambda.$$

For the estimate

$$c_{LS}^* = \left(\sum_{\nu=0}^{N} a^\nu x_\nu\right) \Big/ \left(\sum_{\nu=0}^{N} a^{2\nu}\right)$$

we have

$$D^2[c_{LS}^*] = \frac{1}{\left|\sum_{\nu=0}^{N} a^{2\nu}\right|^2} \int_{-\pi}^{\pi} \left| \frac{1 - a^{N+1}e^{i(N+1)\lambda}}{1 - ae^{i}} \right|^2 f(\lambda)\, d\lambda$$

$$= \frac{a^{2N+2}}{\left|\sum_{\nu=0}^{N} a^{2\nu}\right|^2} \int_{-\pi}^{\pi} \left| \frac{a^{-N-1} - e^{i(N+1)\lambda}}{1 - ae^{i}} \right|^2 f(\lambda)\, d\lambda.$$

Hence

$$D^2[c_{LS}^*] \sim \frac{(1-a^2)^2}{a^{2N+2}} \int_{-\pi}^{\pi} \frac{1}{|1 - ae^{i\lambda}|^2} f(\lambda)\, d\lambda,$$

which gives us

$$\lim_{N\to\infty} \frac{D^2[c_0^*]}{D^2[{}_{LS}^*]} = \frac{a^2}{(1-a^2)^2} \frac{\int_{-\pi}^{\pi} f(\lambda)\, d\lambda}{\int_{-\pi}^{\pi} \frac{f(\lambda)}{|1 - ae^{i\lambda}|^2}\, d\lambda}.$$

If the disturbance has most of its spectral mass concentrated in the neighborhood of $\lambda = 0$, the last expression is near $[a/(1 + a)]^2 < 1$. Thus the L.S. estimate is not asymptotically efficient for all spectral densities in this case.

9. The case of fixed variates. When nothing is known about the way in which the regression variables are generated, it is clearly impossible to make any general statement about the asymptotic efficiency of the L.S. estimates of the regression coefficients.

Let us consider a very simple model of generation which is not of the analytical type. For simplicity let us consider the case $s = 1$. Let $\{\varphi_\nu\}$ be a stochastic process independent of the residual process, stationary (in the strict sense), and ergodic. We denote its spectral distribution function by $P(\lambda)$. Then it is known that (with probability one) the sequence $\{\varphi_\nu\}$ has a spectrum and that the spectral mass in (α, β) is

$$\int_\alpha^\beta dP(\lambda) \Big/ \int_{-\pi}^{\pi} dP(\lambda).$$

(cf. [4]). It follows from Theorem 3 that the asymptotic efficiency of the L.S. estimate, regarding the φ's as fixed variates, is almost certainly

$$e^2 = \left[\int_{-\pi}^{\pi} dP(\lambda)\right]^2 \Big/ \int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, dP(\lambda) \int_{-\pi}^{\pi} f(\lambda)\, dP(\lambda).$$

If the spectrum of $\{\varphi_\nu\}$ contained more than one point we see that $e < 1$ (cf. the corollary of Theorem 3).

It would be interesting to know how generally this holds. A class of processes, generating the regression variables in a natural way, could be defined as follows. Let Δ be a finite difference operator and suppose that $\{\Delta\varphi_\nu\}$ forms a stationary process. One simple case is when $\Delta\varphi_\nu = \varphi_{\nu+1} - \varphi_\nu$ and $\Delta\varphi_\nu$ is purely random. Then φ_ν would be a temporally homogeneous differential process. Are the realizations of such processes regular and stationary in the sense of Section 3, and if so, what spectra do they have?

10. The information of the covariance matrix. If the regression sequence $\{\varphi_\nu\}$ has a spectrum consisting of a single point we have seen that the L.S. estimate is asymptotically efficient. We have then said that the knowledge of the true covariance matrix of the disturbance does not give us any information with respect to the problem of inference under consideration. In this section we shall make this statement more precise.

Suppose that we have two covariance matrices M_1 and M_2 for the residual $y_\nu \varepsilon Y$. Let the B.L.U. estimates calculated under the two hypotheses M_1 and M_2 be c_1^* and c_2^*. If

$$\lim_{N\to\infty} \frac{D[c_1^*]}{D[c_2^*]} = 1,$$

for all possible pairs M_1, M_2, we shall say that the covariance matrix gives us no information.

Let the spectral density under the two hypotheses be $f_1(\lambda)$ and $f_2(\lambda)$ and let the true one be $f(\lambda)$ corresponding to a covariance matrix M. The corollary of Theorem 1 implies that the correlation coefficient between c_{LS}^* and the B.L.U.

estimate tends to one as N tends to infinity (we are considering the case when the spectrum of φ_ν is reduced to a single point). Hence

$$D_{f_1}[c_1^* - c_{LS}^*]/D_{f_1}[c_{LS}^*] \to 0, \qquad N \to \infty.$$

But because of (3) we get

$$D_f[c_1^* - c_{LS}^*]/D_f \mid c_{LS}^* \mid \to 0, \qquad N \to \infty.$$

But the triangle inequality gives us

$$D_f[c_1^* - c_{LS}^*] \geqq \mid D_f[c_1^*] - D_f[c_{LS}^*] \mid$$

and hence

$$D_f[c_1^*]/D_f[c_{LS}^*] \to 1, \qquad N \to \infty.$$

Similarly we get

$$D_f[c_1^*]/D_f[c_{LS}^*] \to 1, \qquad N \to \infty,$$

so that

$$D_f[c_1^*]/D_f[c_2^*] \to 1, \qquad N \to \infty.$$

Thus we have shown that if the spectrum of the regression sequence contains only one point, then knowledge of the covariance matrix does not give us any information of interest with respect to estimating the regression coefficient.

REFERENCES

[1] H. Cramér, "On the theory of stationary random processes," *Ann. Math.*, Vol. 41 (1940), pp. 215–230.

[2] U. Grenander, "Stochastic processes and statistical inference," *Arkiv Mat.* (1950).

[3] U. Grenander, "On Toeplitz forms and stationary processes," *Arkiv Mat.* (1951).

[4] E. Hopf, "Ergodentheorie," *Ergebnisse der Math.* Vol. 5, No. 2, J. Springer, Berlin, 1937.

[5] H. Wold, "On least squares regression with autocorrelated variables and residuals," *Transactions of the International Institute of Statistics*, Berne, 1950.