ASYMPTOTIC NORMALITY AND CONSISTENCY OF THE LEAST SQUARES ESTIMATORS FOR FAMILIES OF LINEAR REGRESSIONS[1]

By F. Eicker

University of Freiburg im Breisgau

1. Summary and introduction. This paper deals with linear regressions

$$y_k = x_{k1}\beta_1 + \cdots + x_{kq}\beta_q + \epsilon_k, \qquad k = 1, 2, \cdots \tag{1.1}$$

with given constants x_{km} and with error random variables ϵ_k that are (a) uncorrelated or (b) independent. Let $E\epsilon_k = 0$, $0 < E\epsilon_k^2 < \infty$ for all k. The individual error distribution functions (d.f.'s) are not assumed to be known, nor need they be identical for all k. They are assumed, however, to be elements of a certain set F of d.f.'s. Consider the family of regressions associated with the family of all the error sequences possible under these restrictions. Then conditions on the set F and on the x_{km} are obtained such that the least squares estimators (LSE) of the parameters $\beta_1, \cdots, \beta_q$ are consistent in Case (a) (Theorem 1) or astmptotically normal in Case (b) (Theorem 2) for every regression of the respective families. The motivation for these theorems lies in the fact that under the given assumptions statements based only on the available knowledge must always concern the regression family as a whole. It will be noticed moreover that the conditions of the theorems do not require any knowledge about the particular error sequence occurring in (1.1). Most of the conditions are necessary as well as sufficient, with the consequence that they cannot be improved upon under the limited information assumed to be available about the model. Since the conditions are very mild, the results apply to a large number of actual estimation problems.

We denote by $\mathfrak{F}(F)$ the set of all sequences $\{\epsilon_k\}$ that occur in the regressions of a family as characterized above. Thus, $\mathfrak{F}(F)$ comprises all sequences of uncorrelated (Case (a)) or independent (Case (b)) random variables whose d.f.'s belong to F but are not necessarily the same from term to term of the sequence. For each $G \varepsilon F$ the relations $\int x\, dG = 0$ and $0 < \int x^2\, dG < \infty$ hold. In this paper, $\mathfrak{F}(F)$ may be looked upon as a parameter space. A parameter point then is a sequence of $\mathfrak{F}(F)$. Correspondingly, we say that a statement holds on $\mathfrak{F}(F)$ (briefly on F) if it holds for all $\{\epsilon_k\} \varepsilon \mathfrak{F}(F)$. The statements of Theorems 1 and 2 are of this kind.

The proof of Theorem 1, as well as the proof of the sufficiency in Theorem

Received February 20, 1962; revised December 17, 1962.

[1] Part of this research was carried out at the Department of Statistics, University of North Carolina. It was supported by the Office of Naval Research under Contract No. Nonr-855(06). Partly the work was done under the sponsorship of the National Science Foundation at the Department of Statistics, Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

2, is elementary and straight forward. Theorem 2 is a special case of a central limit theorem (holding uniformly on $\mathfrak{F}(F)$) for families of random sequences [3].

Some similarity between the roles of the parameter spaces $\mathfrak{F}(F)$ in our theorems and of the parameter spaces that occur, e.g., in the Gauss-Markov and related theorems may be seen in the fact that these theorems remain true only as long as the conclusions in the theorems hold for every parameter point in the respective spaces. As is well known, the statements in the Gauss-Markov and related theorems hold for every parameter vector $\beta_1, \cdots, \beta_q$ in a q-dimensional vector space (see e.g. Scheffé 1959, p. 13, 14).

A result in the theory of linear regressions that bears some resemblance with the theorems of this paper has been obtained by Grenander and Rosenblatt (1957, p. 244). Let the error sequence $\{\epsilon_k\}$ in (1.1) be a weakly stationary random sequence with piecewise continuous spectral density, and let the regression vectors admit a joint spectral representation. Under these assumptions Grenander and Rosenblatt give necessary and sufficient conditions for the regression spectrum and for the family of admissible spectral densities in order that the LSE are asymptotically efficient for every density of the family.

In Sections 3 and 6 we discuss some examples relevant to Theorems 1 and 2.

2. Notations. The (real) regression equations (1.1) for $k = 1, \cdots, n$ can be written in matrix notation as

$$y = X\beta + \epsilon \tag{2.1}$$

with $y = (y_1, \cdots, y_n)'$, $\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$, $\beta = (\beta_1, \cdots, \beta_q)'$ (the prime denotes the transpose), and $X = (x_{km})$, the (n, q)-regression matrix. The dependence on the sample size n (as of the symbols y, ϵ, X, and of various others later to come) is not always marked when confusion seems unlikely.

Let r_k' be the kth row and x_j the jth column of X_n, so that $X_n' = (r_1, \cdots, r_n)$, $X_n = (x_1, \cdots, x_q)$. Let the rank of X_q be q. Then the same applies to X_n, $n > q$. Putting $X'X = P$, the normal equations become

$$X'y = Pb \tag{2.2}$$

where b, more explicitly $b_n = (b_{n1}, \cdots, b_{nq})'$, is the vector of the (unbiased) LSE for $\beta_1, \cdots, \beta_q$. Henceforth, we restrict n to values $n \geqq q$. Then P^{-1} exists, and

$$b - \beta = P^{-1}X'\epsilon, \tag{2.3}$$

$$\text{cov}\,(b - \beta)(b - \beta)' = P^{-1}X'SXP^{-1}, \qquad S = \text{cov}\,\epsilon\epsilon'. \tag{2.4}$$

The following notations will be useful. Let $p_i(n)$ be the ith column of P_n^{-1}, and

$$u_{nk}^{(i)} = r_k'p_i(n). \tag{2.5}$$

As we shall deal with each component b_{ni} of the LSE b_n separately, it is superfluous to carry always the superscript i in (2.5). Thus e.g. the quantities

$$B_n^2 = \sum_{k=1}^{n} u_{nk}^2 \sigma_k^2, \qquad \sigma_k^2 = E\epsilon_k^2, \tag{2.6}$$

$$U_n^2 = \max_{k=1,\cdots,n} u_{nk}^2 \tag{2.7}$$

still depend on i.

We shall use the identity

$$\sum_{k=1}^{n} u_{nk}^2 = p_i' P p_i = (P^{-1})_{ii}, \tag{2.8}$$

where $(A)_{ij}$ denotes the (i, j) element of the matrix A.

3. A necessary and sufficient condition for the consistency of the least squares estimators. In this section the set F is submitted to the conditions stated in Section 1 and, moreover, is required to contain at least one normal d.f. $N(0, \sigma^2)$. With the notations of Section 2 one then has

THEOREM 1. *The least squares estimatores b_n, $n = q, q + 1, \cdots$ estimate β consistently on F if and only if $\lambda_{\min}(P_n) \to \infty$, where $\lambda_{\min}(P_n)$ is the smallest characteristic value of P_n.*

The theorem can be generalized immediately by allowing the d.f. of ϵ_k to be any element out of a set F_k of d.f.'s where in the sequence $F_1, F_2, \cdots$ each F_k is submitted to the same restrictions as were imposed on F, and where, moreover, the variances are bounded uniformly in k. The theorem holds also for convergence in quadratic mean instead of convergence in probability (the latter defines our consistency concept).

PROOF.

(1) Sufficiency: because of $E(P^{-1}X'\epsilon) = 0$ in (2.3) the b_n are unbiased. The variance of each component of the vector $P^{-1}X'\epsilon$ tends to zero if and only if $E(\epsilon' X P^{-2} X' \epsilon) \to 0$. Because var $\epsilon_k <$ const (const means whenever it occurs some finite real constant independent of k and n) the left hand side is

$$O(\operatorname{tr} XP^{-2}X') = O(\operatorname{tr} P^{-1}) = O(1/\lambda_{\min}(P)). \tag{3.1}$$

Because $\lambda_{\min}(P) \to \infty$ thus $b_n \xrightarrow[\text{i.p.}]{} \beta$ on F.

(2) Necessity: choosing all ϵ_k in (1.1) distributed identically with d.f. $N(0, \sigma^2) \varepsilon F$ one has

$$b_{in} - \beta_i = p_i' X' \epsilon \sim N(0, p_i' X' S X p_i) \tag{3.2}$$

where S is the nth unit matrix multiplied by σ^2. Hence var $(b_{in} - \beta_i) = \sigma^2 (P^{-1})_{ii}$ which, because of the consistency, must tend to zero for every $i = 1, \cdots, q$. Thus $\sum_i (P^{-1})_{ii} = \sum_i \lambda_i^{-1}(P)$ must tend to zero and therefore $\lambda_{\min}(P) \to \infty$.

APPLICATIONS AND REMARKS TO THEOREM 1.

(1) Since $(P)_{ii} \geqq \lambda_{\min}(P)$ it follows from $\lambda_{\min}(P) \to \infty$ for each $i = 1, \cdots, q$ that $x_i' x_i \to \infty$ as $n \to \infty$. One also observes that $\lambda_{\min}(P)$ is a sequence of positive numbers, nondecreasing in n.

(2) The following criterion (compare e.g. Bodewig (1959), p. 67) can be helpful in connection with Theorem 1: for $\lambda_{\min}(P) \to \infty$ it is sufficient that for $i = 1, \cdots, q$

$$x_i'x_i - \sum_{\substack{j=1 \\ j \neq i}}^{q} |x_i'x_j| \to \infty \tag{3.3}$$

where the x_j, $j = 1, \cdots, q$, are n vectors, and $n \to \infty$.

(3) *Polynomial regression*: if

$$x_{ki} = k^{c_i}, \qquad c_i > -\tfrac{1}{2}, \quad i = 1, \cdots, q, \qquad k = 1, 2, \cdots, \tag{3.4}$$

and $c_i \neq c_j$, then

$$\lambda_{\min}(P_n) = O(n^{2c_0+1}), \qquad c_0 = \min_{i=1,\cdots,q} c_i \tag{3.5}$$

(for a proof compare [4]). Hence estimators are obtained that are consistent on F, also for non-integers c_i, if and only if $c_0 > -\frac{1}{2}$. Similarly, regression vectors may be treated in which exponentials, or exponentials and polynomials occur.

(4) One obtains consistent estimators also in a *trigonometric regression*. Let

$$x_{k,2i-1} = \cos \omega_i k, \qquad x_{k,2i} = \sin \omega_i k, \qquad i = 1, \cdots, q, \tag{3.6}$$

where $\omega_i \neq \omega_j$ and $\omega_i \neq 2\pi - \omega_j$ for $i \neq j$, and $0 \leqq \omega_i < 2\pi$ for all i. In this case $n^{-1}P_n$ tends to a diagonal matrix whose non-zero elements are $O(1)$.

(5) If one is interested only in the consistency of one, say the ith component b_{in} of the vector estimators b_n then $(P^{-1})_{ii} \to 0$ is a necessary and sufficient condition for $b_{in} \to \beta_i$ on F.

(6) The system (1.1) is a special case of the following general system of linear regression equations familiar in time series analysis, where regression on lagged variables is included (*autoregressive schemes*)

$$y_k = \alpha_1 y_{k-1} + \cdots + \alpha_p y_{k-p} + \beta_1 x_{k1} + \cdots + \beta_q x_{kq} + \epsilon_k, \tag{3.7}$$

for $k = 1, 2, \cdots$. The α's are constants, the other quantities are as defined in Sections 1 and 2. The α's and β's can be estimated by means of least squares, and sufficient conditions for the ϵ_k's and the x_{ki} are known which allow for the consistency of the estimators. Theorem 1 turns out to be a specialization of some results concerning this general scheme [4].

(7) In Theorem 1 it is assumed that F contains a normal distribution. This certainly is not necessary. However, it seems to be more difficult if it is at all possible to derive "simple" necessary and sufficient conditions for $\{X_n\}$ and F in order that consistency holds on F. Here "simple" may e.g. mean that the conditions are separate for the X_n and for F (as in Theorems 1 and 2), or at least less complex than those which can be obtained from general theorems (see e.g., Gnedenko and Kolmogorov (1954), p. 116).

4. Conditions for the asymptotic normality of the least squares estimators. Henceforth ϵ_k and ϵ_j are assumed to be independent for $j \neq k$, and F is some non-empty set of d.f.'s with zero means and finite, positive variances. The assumption that the variances are positive does not mean any loss of generality because if there were an error with vanishing variance there would exist a linear relation between $\beta_1, \cdots, \beta_q$ holding with probability one, and the system (1.1) could be reduced so that it contained one β_i less.

The next theorem is an immediate consequence of Theorem 3 published elsewhere [3]. In order to apply this theorem we only have to observe that for every $n \geqq q$ there exists a $k \geqq 1$ such that $p_i'(n)r_k \neq 0$, given $i \geqq 1$, $i \leqq q$. This follows from the assumption that the rank of X_n is q, and therefore $p_i'(n)X_n'$ is not the null vector. With the notations of Sections 1 and 2 we have

THEOREM 2. *For the convergence of the d.f.'s of $B_n^{-1}(b_{in} - \beta_i)$ on a set F, subject to the above restrictions, to the standard normal law for $n \to \infty$ and in order that the contribution of every ϵ_k be infinitesimal,*

$$\max_{k=1,\cdots,n} |u_{nk}| \sigma_k B_n^{-1} \to 0,$$

it is necessary and sufficient that the following conditions on F and the sequence of regression matrices $\{X_n\}$ be fulfilled:

(I) *The rank of X_q is q. Furthermore*

$$U_n^2/(P_n^{-1})_{ii} \to 0 \qquad \text{as } n \to \infty. \tag{4.1}$$

(II)
$$\sup_{G \varepsilon F} \int_{|z| \geqq c} z^2\, dG(z) \to 0 \qquad \text{for } c \to \infty. \tag{4.2}$$

(III)
$$\inf_{G \varepsilon F} \int z^2\, dG(z) > 0. \tag{4.3}$$

One notices that it is not sufficient to have instead of (II) the boundedness from above of the variances of all d.f.'s in F.

The above theorem remains valid if "convergence in distribution on F" is replaced by "convergence in distribution on F uniform on the real axis and on $\mathfrak{F}(F)$", which can be written, recalling that the quantities B_n and b_{in} depend upon $\{\epsilon_k\}$,

$$\sup_{\substack{\{\epsilon_k\} \varepsilon \mathfrak{F}(F) \\ -\infty < x < \infty}} |P(B_n^{-1}(b_{in} - \beta_i) < x) - \phi(x)| \to 0, \tag{4.4}$$

where $\phi(x)$ is the standard normal d.f.

The following theorem which instead of (II) uses the Lyapunov type conditions (II′) of the central limit theorem gives only sufficient conditions for the asymptotic normality of the least squares estimators and can easily be proved:

COROLLARY TO THEOREM 2. *For the convergence of the d.f.'s of $B_n^{-1}(b_{in} - \beta_i)$ on a set F, subject to the above restrictions, to the standard normal law for $n \to \infty$ it is sufficient that Conditions* (I) *and* (III) *of Theorem 2 hold and that for some $\eta > 0$*

(II′)
$$\sup_{G \varepsilon F} \int |z|^{2+\eta}\, dG(z) < \text{const.} \tag{4.5}$$

PROOF. We only need to show that (II) follows from (II′): for $c > 0$ and with a constant independent of $G \varepsilon F$

$$(4.6) \qquad \int_{|z| \geqq c} z^2 \, dG(z) \leqq \int_{|z| \geqq c} z^2 \, |z/c|^\eta \, dG(z) < \text{const } c^{-\eta}$$

which tends to zero for $c \to \infty$.

5. An asymptotically normal statistic for the regression parameters. Theorem 2 makes a statement about the quantities $B_n^{-1}(b_{in} - \beta_i)$. Here B_n is a function of the variances σ_k^2 of the errors ϵ_k. In many cases, however, the σ_k are unknown. One then can replace B_n by the estimate C_n appearing in the next theorem.

THEOREM 3. *Let*

$$(5.1) \qquad C_n^2 = \sum_{k=1}^{n} u_{nk}^2 e_k^2$$

where e_k is the kth *component of $e = y - Xb$. Then it is sufficient for $C_n^{-1}(b_{in} - \beta_i)$, $i = 1, \cdots, q$, to have asymptotically the standard normal distribution on F that besides Condition* (I) *of Theorem 2 the following conditions hold*

$$(5.2) \qquad \min_{G\varepsilon F} \int z^2 \, dG(z) > 0, \qquad \sup_{G\varepsilon F} \int z^4 \, dG(z) < \infty.$$

PROOF. It suffices to show that $(C_n/B_n)^2$ tends to one in probability (Cramér (1946), p. 254), or equivalently $C_n/B_n \to 1$ i.p., $C_n = |C_n|$. With $\beta - b = -P^{-1}X'\epsilon$ we have for any $n \geqq q$

$$(5.3) \qquad e = y - Xb = \epsilon - XP^{-1}X'\epsilon,$$

hence $e_k = \epsilon_k - v_k'\epsilon$, where

$$(5.4) \qquad v_k = XP^{-1}r_k, \qquad k = 1, \cdots, n.$$

As the ϵ_j are independent, and because of (5.2) and $v_k'v_k = v_{kk}$, all k, we obtain

$$(5.5) \qquad E(v_k'\epsilon)^2 < \text{const } v_k'v_k = \text{const } v_{kk},$$

$$(5.6) \qquad E(v_k'\epsilon)^4 = E\left(\sum_{j=1}^{n} v_{kj}\epsilon_j\right)^4 < \text{const}\left(\sum_j v_{kj}^4 + 3\sum_{\substack{j,m \\ j \neq m}} v_{kj}^2 v_{km}^2\right) < \text{const } v_{kk}^2,$$

all constants independent of n and k. Now

$$(5.7) \qquad \begin{aligned} E(e_k^2) &= E(\epsilon_k^2) + E(v_k'\epsilon)^2 - 2E(\epsilon_k v_k'\epsilon) \\ &\leqq \sigma_k^2 + \text{const } v_{kk} - 2\sigma_k^2 v_{kk} \leqq \sigma_k^2 + \text{const } v_{kk}, \end{aligned}$$

observing $v_{kk} = v_k'P^{-1}r_k \geqq 0$. Because of the identity

$$(5.8) \qquad \sum_{k=1}^{n} v_k' v_k = \sum_{k=1}^{n} v_{kk} = \sum_{k=1}^{n} r_k' P^{-1} r_k = \text{tr } XP^{-1}X = q,$$

$$(5.9) \qquad E(C_n^2) < B_n^2 + \text{const } U_n^2, \qquad U_n^2 = \max_k u_{nk}^2.$$

From (2.8) and (5.2) there follows $B_n^2 > \text{const } (P^{-1})_{ii}$. Hence by (4.1), $B_n^{-2}E(C_n^2) \to 1$.

The variance is, using Minkovski's inequality for three summands,

$$\begin{aligned}\operatorname{var}(B_n^{-2}C_n^2) &\leqq B_n^{-4}\, E(C_n^2 - B_n^2)^2 \\ &= B_n^{-4}\, E\left[\sum_{k=1}^{n} u_{nk}^2(\epsilon_k^2 - \sigma_k^2 + (v_k'\epsilon)^2 - 2\epsilon_k v_k'\epsilon)\right]^2 \\ &\leqq B_n^{-4}[\{E(\sum_k u_{nk}^2(\epsilon_k^2 - \sigma_k^2))^2\}^{\frac{1}{2}} \\ &\quad + \{E(\sum_k u_{nk}^2(v_k'\epsilon)^2)^2\}^{\frac{1}{2}} + 2\,\{E(\sum_k u_{nk}^2\epsilon_k v_k'\epsilon)^2\}^{\frac{1}{2}}]^2.\end{aligned} \tag{5.10}$$

The squared first term in the square bracket is bounded by

$$\text{const}\; U_n^2 \sum_k u_{nk}^2 = \text{const}\; U_n^2 (P^{-1})_{ii} \tag{5.11}$$

and tends to zero after division by B_n^4 which is $>$ const $\{(P^{-1})_{ii}\}^2$. By the Schwarz inequality and by (5.6), for all k and m

$$E[(v_k'\epsilon)^2(v_m'\epsilon)^2] < \text{const}\; v_{kk}v_{mm}\,. \tag{5.12}$$

Thus the squared second term is bounded by

$$\text{const} \sum_{k,m} u_{nk}^2 v_{kk} u_{nm}^2 v_{mm} < \text{const}\; U_n^4 \tag{5.13}$$

and tends to zero if multiplied by B_n^{-4}. Finally, with $v_{km} = r_m'P^{-1}r_k = v_{mk}$, $v_{mk}^2 \leqq v_{kk}v_{mm}$,

$$E(\epsilon_k v_k'\epsilon\,\epsilon_m v_m'\epsilon) < \begin{cases}\text{const}\;(v_{kk}v_{mm} + v_{km}^2) < \text{const}\; v_{kk}v_{mm}, & k \neq m \\ \text{const}\; v_{kk}, & k = m.\end{cases} \tag{5.14}$$

By (5.8) the squared third term in (5.10) multiplied by B_n^{-4} is now seen to be $O(U_n^4\{(P^{-1})_{ii}\}^2)$ and thus tends to zero.

6. Remarks and examples.

(1) Condition (I) of Theorem 2 is equivalent with

$$[(P^{-1})_{ii}]^{-\frac{1}{2}} p_i'X' \to (0), \qquad \text{as } n \to \infty, \tag{6.1}$$

where (0) is the zero vector. If in (6.1) only the first q components are considered, which amounts to replacing the X that appears explicitly in (6.1) by X_q, and recalling that X_q is nonsingular, we obtain $[(P^{-1})_{ii}]^{-\frac{1}{2}}p_i \to 0$ or, taking the jth component,

$$(P^{-1})_{ij}^2/(P^{-1})_{ii} \to 0, \qquad j = 1, \cdots, q. \tag{6.2}$$

Taking in particular $j = i$, (6.2) reduces to

$$(P^{-1})_{ii} \to 0, \tag{6.3}$$

and one notices that (6.3) is the assumption made in Theorem 1 (see Remark 5). Thus the asymptotic normality of $B_n^{-1}(b_{ni} - \beta_i)$ on F implies the consistency of b_{in} on F. Without applying Theorem 1 this conclusion may also be drawn simply from the fact that $B_n^2 < \text{const}\,(P^{-1})_{ii}$ tends to zero.

(2) Condition (I) is implied by

$$\text{(I}'\text{)} \qquad [(P^{-2})_{ii}/(P^{-1})_{ii}]\max_{k=1,\cdots,n} r_k'r_k \to 0 \qquad \text{for } n \to \infty$$

as can be seen with the help of the inequalities $\lambda_{\max}(p_ip_i') \leqq p_i'p_i = (P^{-2})_{ii}$. With $\lambda_{\max}(A) \geqq (A)_{ii} \geqq \lambda_{\min}(A)$ for every real symmetric matrix A and for all possible i, and denoting by $\lambda_j(P)$ the jth largest characteristic root of P, Condition (I$'$) is seen in turn to be implied by

$$\text{(I}''\text{)} \qquad [\lambda_1(P)/\lambda_q^2(P)]\max_{k=1,\cdots,n} r_k'r_k \to 0 \qquad \text{for } n \to \infty.$$

(3) Theorem 2 asserts the asymptotic normality of the single components of the vector estimators b_n of β. For the vectors b_n themselves, however, it seems not to be possible to state similarly a theorem concerning the asymptotic convergence on F to a joint d.f. if the set F contains at least two distributions with different variances.

This shall be demonstrated by the following example where F is taken to consist of the two normal d.f.'s $N(0, 1)$ and $N(0, 2)$ and where the regression and error vectors are chosen in such a way, that the covariance between the two components of the vector estimate b_n does not exist asymptotically. Then also a joint asymptotic d.f. does not exist although the assumptions of Theorem 2 are satisfied. This is clear for (4.2) and (4.3), and it will be shown for (4.1).[2]

Let $q = 2$. The covariance between $(b_{n1} - \beta_1)/B_n^{(1)}$ and $(b_{n2} - \beta_2)/B_n^{(2)}$ is proportional to the (1, 2)-element of the covariance matrix (2.4) and equals

$$\text{(6.4)} \qquad \sum \sigma_k^2 u_{nk}^{(1)} u_{nk}^{(2)} / B_n^{(1)} B_n^{(2)},$$

where $B_n^{(j)}$ is defined by (2.6) with $u_{nk} = u_{nk}^{(j)}$. The summation runs here and in the following sums over $k = 1, \cdots, n$ in case nothing else is stated. We choose as regression vectors the n-vectors $x_1' = (1, 1, \cdots, 1)$, $x_2' = (1, -1, 1, \cdots, \pm 1)$. Putting $c_n = [1 - (-1)^n]/2$, we obtain $x_1'x_2 = c_n$, $x_1'x_1 = x_2'x_2 = n$, $|P_n| = n^2 - c_n^2$, and

$$P_n^{-1} = |P_n|^{-1}\begin{pmatrix} x_1'x_1 & -x_1'x_2 \\ -x_1'x_2 & x_2'x_2 \end{pmatrix} = n\,|P_n|^{-1}\begin{pmatrix} 1 & -c_n/n \\ -c_n/n & 1 \end{pmatrix}.$$

Furthermore, with $r_k' = (1, (-1)^{k+1})$ and recalling $P_n^{-1} = (p_1, p_2)$,

$$u_{nk}^{(1)} = r_k'p_1 = n\,|P_n|^{-1}(1 + (-1)^k c_n/n) = (-1)^{k+1}u_{nk}^{(2)}.$$

Now $\max_k (u_{nk}^{(i)})^2/(P_n^{-1})_{ii} = (1 + c_n/n)^2 n/(n^2 - c_n^2) \to 0$ so that (4.1) is satisfied. Putting $[1 + (-1)^k c_n/n]^2 = d_{nk}$, (6.4) becomes

$$\text{(6.5)} \qquad \frac{\sum(-1)^{k+1}\,\sigma_k^2(u_{nk}^{(1)})^2}{\sum \sigma_k^2(u_{nk}^{(1)})^2} = \frac{\sum(-1)^{k+1}\,\sigma_k^2\,d_{nk}}{\sum \sigma_k^2\,d_{nk}}.$$

Since $c_n/n \to 0$, $d_{nk} \to 1$ uniformly in k. Thus (6.5) lies in an interval with the

[2] Some details of the construction of this example are due to Professor R. A. Wijsman.

endpoints

$$(1 \pm \delta)\{[1 + (c_n/n)^2] \sum (-1)^{k+1}\sigma_k^2 - 2(c_n/n) \sum \sigma_k^2\}/\sum \sigma_k^2$$

where $\delta > 0$ can be made arbitrarily small for large n. The center of the interval is, up to terms of order $O(c_n/n)$, equal to

$$t_n = \sum_{j=1}^{[n/2]} \delta_j/\sum_{k=1}^{n} \sigma_k^2, \qquad \delta_j = \sigma_{2j}^2 - \sigma_{2j-1}^2 .$$

Since $\sigma_k^2 = 1$ or 2, i.e. $\delta_j = 1$ or -1, the sequence $\{\delta_j\}$ can be chosen so that $\{t_n\}$ has the cluster points $+\frac{2}{3}$ and $-\frac{2}{3}$. This can be done by choosing alternately uninterrupted sequences of $+1$'s or -1's for the δ_j's whose respective lengths divided by n tend to one. Consequently, the covariance (6.4) does not converge asymptotically as was to be shown.

REMARK. If instead of the ordinary least squares estimators the minimum variance linear estimators

$$\hat{b}_n = (X'S^2X)^{-1}X'S^{-2}y, \qquad S^2 = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2),$$

had been used, the asymptotic normality on a set F with at least two d.f.s having different variances again could not have been inferred. In this case the covariance matrix of $D(\hat{b}_n - \beta)$ equals

$$D(X'S^{-2}X)^{-1}D = [D^{-1}X'S^{-2}XD^{-1}]^{-1} \tag{6.6}$$

where $D^2 = \operatorname{diag}(d_1^2(n), \cdots, d_q^2(n))$ is the matrix consisting of the diagonal of $(X'S^{-2}X)^{-1}$ and zeros elsewhere. Let (6.6) have a (nonsingular) limiting matrix for $n \to \infty$. Then the sequence of the inverses also possesses a limiting matrix, so that the (i, i)-elements of the inverse matrices $\sum_{k=1}^{n} \sigma_k^{-2}x_{ki}^2/ d_i^2(n)$ converge for $i = 1, \cdots, q$.

Hence, since the sequence of the (i, j)-elements of the inverse matrices $\sum_k \sigma_k^{-2}x_{ki}x_{kj}/d_i(n)\, d_j(n)$ converges, also

$$\sum_k \sigma_k^{-2}x_{ki}x_{kj}/\{\sum_k \sigma_k^{-2}x_{ki}^2 \sum_k \sigma_k^{-2}x_{kj}^2\}^{\frac{1}{2}} \tag{6.7}$$

does. However, like above examples can be constructed for which (6.7) does not converge even if it does for identical σ's. To this end choose the x_{ki} as in the previous example. Then (6.7) is the same as t_n in that example, after replacing σ_k^2 by σ_k^{-2}. Thus, by choosing σ_k^2 to be 1 or $\frac{1}{2}$ according as the previous σ_k^2 was chosen 1 or 2, we produce the same divergent sequence $\{t_n\}$ as before.

EXAMPLES.

(1) If $q = 1$ then there is only one regression vector, denoted by x. As i assumes only the value one, we have $P^{-1} = (P^{-1})_{11} = p_1 = 1/x'x$, and Condition (I) reduces to $\max_{k=1,\cdots,n} |x_k|/\|x\| \to 0$, $\|x\| = (x'x)^{\frac{1}{2}}$, which is equivalent with the pair of conditions $x_n/\|x\| \to 0$ and $\|x\| \to \infty$ for $n \to \infty$. A regression vector of this kind is called *slowly increasing* [6].

(2) In the case $q = 2$ one may put $\rho_n = x_1'x_2/\|x_1\| \cdot \|x_2\|$, the "correlation coefficient" between the n-vectors x_1 and x_2. Then one easily computes that

(4.1) is equivalent with

$$[1/x_1'x_1(1 - \rho_n)] \max_{k=1,\cdots,n} [x_{k1} - \rho_n x_{k2}(x_1'x_1/x_2'x_2)^{\frac{1}{2}}]^2 \to 0$$

and a similar relation in which the subscripts 1 and 2 are interchanged.

Acknowledgment. I should like to express my warm thanks to Professor R. A. Wijsman for many valuable suggestions and criticisms by which he contributed substantially to the paper.

REFERENCES

[1] BODEWIG, E. (1959). *Matrix Calculus* (2nd ed.). Interscience, New York.
[2] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.
[3] EICKER, F. (1963). Central limit theorems for families of sequences of random variables. *Ann. Math. Statist.* **34** 439–446.
[4] EICKER, F. (1964). Über die Konsistenz von Parameterabschätzungen in einem gemischten Zeitreihenmodell. To be published.
[5] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables,* (translated from the Russian by K. L. CHUNG). Addison-Wesley, Cambridge.
[6] GRENANDER, ULF, and ROSENBLATT, MURRAY (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.
[7] SCHEFFÉ, HENRY (1959). *The Analysis of Variance.* Wiley, New York.