LIMIT THEOREMS FOR REGRESSIONS WITH UNEQUAL AND DEPENDENT ERRORS

FRIEDHELM EICKER
UNIVERSITY OF FREIBURG IM BREISGAU
and
COLUMBIA UNIVERSITY, NEW YORK

1. Summary

This paper deals with the asymptotic distribution of the vectorial least squares estimators (LSE) for the parameters in multiple linear regression systems. The regression constants are assumed to be known; the errors are assumed (a) to be independent but not necessarily identically or normally distributed (section 3), or (b) to constitute a generalized linear discrete stochastic process (section 4). The latter part includes the case of regression for time series. Conditions are studied under which the joint distribution functions (d.f.'s) of the vectorial LSE's tend to a multivariate normal d.f. as the sample size increases. In the proof a central limit theorem (CLT) for weighted averages of independent random variables is used. In case (a), a theorem for large classes of linear regressions is proved (theorem 3.2), whose conditions are in a certain sense also necessary. The theorem simultaneously permits consistent estimation of the limiting covariance matrix of the LSE's. The results in case (b) are contained in theorems 4.2, 4.3, 4.4, 4.5, 4.6. They are not naturally of as closed a form as those pertinent to case (a) because of the more complicated nature of the problem. Some use of spectral theory is made. Several examples are discussed (section 3.3). The assumptions made in this paper are weaker than those of results published earlier in the literature. (For a more recent survey, compare [6].) Their structure is quite simple so that they ought to be useful in applications. Section 4.4 contains some remarks on multivariate regression equations.

2. Introduction (notations)

There exists a considerable number of publications dealing with the asymptotic normality of parameter estimates for linear regressions, many of which deal with specific cases, however, or are unnecessarily narrow in the assumptions made. The most general paper among these, and the one closest to

Research supported in part by National Science Foundation Grant NSF-GP-3694 at Columbia University.

the present note, seems to be [9]. In that paper vectorial regression equations are considered while we are predominantly interested in scalar ones. For that case, however, the assumptions of [9] are more restrictive than those of the present note.

For the individual components of the vectorial LSE the asymptotic normality was already proved under general assumptions in [1] for the case of independent errors.

The system of (scalar) linear regression equations is denoted by

$$y_t = x_{t1}\beta_1 + \cdots + x_{tq}\beta_q + \epsilon_t, \qquad t = 1, \cdots, n, \tag{2.1}$$

$n \geq q$ being the sample size. In matrix notation this becomes

$$y(n) = X_n\beta + \epsilon(n), \tag{2.2}$$

where $y(n)$ is the (column) vector of observations (n-dimensional), $X_n = (x_{tj})$, the $(n \times q)$-matrix of known regression constants assumed to be of full rank throughout, $\beta = (\beta_1, \cdots, \beta_q)'$ is the vector of unknown regression parameters ($'$ denotes the transpose), and $\epsilon(n) = (\epsilon_1, \cdots, \epsilon_n)'$ is the n-vector of error random variables (r.v.'s) about which we assume throughout that

$$E\epsilon_t = 0, \qquad 0 < E\epsilon_t^2 < \infty, \qquad \text{for all } t. \tag{2.3}$$

All quantities are real.

Let $P_n = X_n'X_n$. Then the vectorial LSE for β, denoted by $b(n) = (b_1(n), \cdots, b_q(n))'$, become

$$b(n) = P_n^{-1}X_n'y(n) = \beta + P_n^{-1}X_n'\epsilon(n). \tag{2.4}$$

The row vectors of X_n will be denoted by $r_1', \cdots, r_n'$, and the column vectors by $x_1(n), \cdots, x_q(n)$. By F, we denote a (nonempty) set of d.f.'s whose elements G have the properties

$$\int x\, dG(x) = 0, \qquad 0 < \int x^2\, dG(x) < \infty. \tag{2.5}$$

3. Independent nonidentically distributed errors

3.1. *The asymptotic normality of the* $b(n)$. In order to find a limiting d.f., the vectors

$$b(n) - \beta = P_n^{-1}X_n'\epsilon(n) \tag{3.1}$$

have to be normalized by premultiplication by certain matrices B_n. Let

$$\Sigma_n = \operatorname{cov} \epsilon(n)\epsilon'(n) = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2), \qquad \sigma_k^2 = \operatorname{var} \epsilon_k, \tag{3.2}$$

be the covariance matrix of the error vector, and write

$$B_n^2 = P_n^{-1}X_n'\Sigma_nX_nP_n^{-1}. \tag{3.3}$$

If B_n is the unique positive definite square root of B_n^2, the q-vectors

$$B_n^{-1}P_n^{-1}X_n'\epsilon(n) \tag{3.4}$$

all have expectation zero and covariance matrix I_q (= q-dimensional identity matrix).

Let $\mathfrak{F}(F)$ be the set of all sequences $\epsilon \equiv \{\epsilon_1, \epsilon_2, \cdots\}$ of independent error r.v.'s ϵ_t (independent within each sequence) whose d.f.'s belong to some set F subject to (2.5). For a given sequence ϵ, the vectors $\epsilon(n)$ have as components the first n members of ϵ, $n = 1, 2, \cdots$.

Theorem 1 of [4] then applies without further ado and yields the following theorem.

THEOREM 3.1. *The d.f.'s of* $B_n^{-1}P_n^{-1}X_n'\epsilon(n)$ *tend to the q-dimensional normal d.f.* $N(0, I_q)$ *and the summands of* $B_n^{-1}P_n^{-1}X_n'\epsilon(n)$, *are infinitesimal both uniformly for all sequences* $\epsilon \equiv \{\epsilon_1, \epsilon_2, \cdots\} \in \mathfrak{F}(F)$, *if and only if the following three conditions are satisfied:*

$$\text{(I)} \qquad \max_{k=1,\cdots,n} r_k'P_n^{-1}r_k \to 0,$$

$$\text{(II)} \qquad \sup_{G\in F} \int_{|x|>c} x^2\, dG(x) \to 0, \qquad \text{as } c \to \infty,$$

$$\text{(III)} \qquad \inf_{G\in F} \int x^2\, dG(x) > 0.$$

(All limits throughout the paper hold for $n \to \infty$ unless otherwise stated.)

The fact that the assertion of the theorem holds for all sequences $\{\epsilon_t\} \in \mathfrak{F}(F)$ makes it particularly useful in practice, since one usually does not know the error d.f.'s if they are not identical. It may also be pointed out that condition (I) on the regression matrices does not necessitate any knowledge about the error sequence present in a particular regression. Analogously, (II) and (III) concern only the set F of admissible error d.f.'s. If the ordinary CLT were applied, one would obtain conditions concerning, simultaneously, the error sequence and the regression sequences. It is interesting that the consideration of the whole class $\mathfrak{F}(F)$ implies the necessity of the conditions (I)–(III). Condition (II) means *uniform integrability* of the variance integrals with respect to the class F.

As it stands, theorem 3.1 is still of limited practical use since the normalizing matrices require the knowledge of the usually unknown error variances σ_k^2. Applying a law of large numbers for nonnegative random variables ([5], p. 143), this defect can be removed by replacing σ_k^2 by the square of the k-th residual

$$e_k(n) = y_k - r_k'b(n) = \epsilon_k - r_k'P_n^{-1}X_n'\epsilon(n), \qquad k = 1, \cdots, n. \tag{3.5}$$

The matrix $D_n^2 \equiv P_n^{1/2}B_n^2P_n^{1/2}$ is then replaced by

$$C_n^2 = P_n^{-1/2}X_n'S_nX_nP_n^{-1/2} \tag{3.6}$$

with $S_n = \text{diag}\,(e_1^2(n), \cdots, e_n^2(n))$. This replacement amounts to an estimation of the matrix D_n^2 in the sense of (3.9), as will be shown below. After this substitution the estimator

$$C_n^{-1/2}P_n^{-1}X_n'y(n) \tag{3.7}$$

no more contains any unknown quantity. Without any new assumptions we then obtain the next theorem.

THEOREM 3.2. *Under the assumptions (I), (II), (III) of the preceding theorem, the d.f.'s of*

$$C_n^{-1} P_n^{-1/2}(b(n) - \beta) \tag{3.8}$$

tend to $N(0, I_q)$ *uniformly for all error sequences* $\epsilon \in \mathfrak{F}(F)$.

For the proof we need a lemma on matrices whose entries are random variables.

LEMMA 3.1. *A sequence of symmetrical random* $q \times q$*-matrices* $A_n \to I_q$, *i.p. if and only if* $c'A_nc \to 1$, *i.p. for all unimodular constant* q*-vectors* c.

PROOF. The "only if" part follows from Slutsky's theorem. To show the converse, first take $c = v_k$ [= k-th unit vector], $k = 1, \cdots, q$, and then c proportional to $v_k + v_j$, any pair $k \neq j$.

PROOF OF THEOREM 3.2. We show first

$$D_n^{-1}C_n^2D_n^{-1} \to I_q, \qquad \text{i.p.} \tag{3.9}$$

We introduce the vectors

$$(u_1(n), \cdots, u_n(n))' = c'D_n^{-1}P_n^{-1/2}X_n', \qquad n = q, q + 1, \cdots, \tag{3.10}$$

with some unimodular constant q-vector c. Then

$$c'D_n^{-1}C_n^2D_n^{-1}c = \sum_{k=1}^{n} u_k^2(n)e_k^2(n). \tag{3.11}$$

Since $\sum_k u_k(n)\epsilon_k = c'D_n^{-1}P_n^{-1/2}\,(b(n) - \beta)$ is a sum of independent infinitesimal r.v.'s whose d.f. for $n \to \infty$ tends to $N(0, 1)$ as a consequence of theorem 3.1, we have

$$\sum_k u_k^2(n)\epsilon_k^2 \to 1, \qquad \text{i.p.} \tag{3.12}$$

by theorem 4 of ([5], p. 143).

Taking account of the second terms of $e_k(n)$ as given by (3.5), we have

$$E(r_k'P_n^{-1}X_n'\epsilon(n))^2 \leq Mr_k'P_n^{-1}r_k \tag{3.13}$$

where the existence of $M = \sup_G \int x^2\, dG(x) < \infty$ is implied by (II). Putting $m = \inf_G \int x^2\, dG(x)[> 0]$ and denoting by $\|\cdot\|$ the Euclidean norm, we have

$$\sum_k u_k^2(n) = c'D_n^{-1}c \leq \frac{1}{m}. \tag{3.14}$$

Hence, by (I),

$$E(\sum_k u_k^2(n)(r_k'P_n^{-1}X_n'\epsilon(n))^2) \leq (M/m) \max_k r_k'P_k^{-1}r_k \to 0, \tag{3.15}$$

and consequently, $\sum_k u_k^2(n)(r_k'P_n^{-1}X_n'\epsilon(n))^2 \to 0$, i.p. Finally, $\sum_k u_k^2(n)e_k^2(n) \to 1$, i.p. for all unimodular vectors c. Because of lemma 3.1, this proves (3.9).

We now prove

$$C_n - D_n \to 0, \qquad \text{i.p.} \tag{3.16}$$

Put $E_n = C_n^2 - D_n^2$. By (3.9) there exists a sequence of events Ω_n with $P\Omega_n \to 1$ such that $\sup_{\Omega_n} \|E_n\| \to 0$ where $\|E_n\| \equiv \max_{i,j=1,\cdots,q} |(E_n)_{ij}|$. In the following all quantities, and equations in quantities, showing the index n are considered only on the event Ω_n, $n = q, q + 1, \cdots$.

Putting $(D_n^2 + E_n)^{1/2} = D_n + E_n^*$, we then have to show $\|E_n^*\| \to 0$, that is, according to our convention, $\sup_{\Omega_n} \|E_n^*\| \to 0$. Now $E_n = D_n E_n^* + E_n^* D_n + E_n^{*2}$; E_n^*, being a real symmetric matrix, has only real characteristic values (c.r.), to be denoted by $\lambda_{1n}(\omega) \geq \lambda_{2n}(\omega) \geq \cdots \geq \lambda_{qn}(\omega)$, $\omega \in \Omega_n$, and possesses q orthogonal real characteristic vectors (c.v.). Suppose $\sup_{\omega \in \Omega_n} \lambda_{1n}(\omega) \equiv \Lambda_n > 0$ for an infinite set Γ of naturals. We assume for simplicity that there exists a matrix $E_n^*(\omega_n)$, $\omega_n \in \Omega_n$, that actually possesses Λ_n as a c.r. (otherwise we can always find an E_n^* whose maximum c.r. differs arbitrarily little from Λ_n). Let $v_n \in S_q$ (the unit sphere $\subset R_q$) be a c.v. of $E_n^*(\omega_n)$ associated with Λ_n. Then at $\omega = \omega_n$ respectively,

$$v_n E_n v_n = 2\Lambda_n v_n' D_n v_n + \Lambda_n^2 \to 0, \qquad n \in \Gamma, \tag{3.17}$$

since $\|E_n\| \to 0$ and the c.v. of D_n are bounded between the finite positive constants m and M.

On the other hand, suppose $\inf_{\omega \in \Omega_n} \lambda_{qn}(\omega) = \lambda_n < 0$ on some infinite set Γ' of integers, and let $u_n \in S_q$ be a c.v. associated with λ_n and a suitable matrix E_n^*. Then again $u_n' E_n u_n = \lambda_n(\lambda_n + 2u_n' D_n u_n) \to 0$, $n \in \Gamma'$, and hence either $\lambda_n \to 0$ or $\lambda_n + 2u_n' D_n u_n \to 0$ for some infinite sequence $\Gamma'' \subset \Gamma'$. But $0 < u_n'(D_n^2 + E_n)^{1/2} u_n = u_n' D_n u_n(1 - 2) + o(n)$ for $n \in \Gamma''$ which is impossible. Thus all c.r.'s of E_n^* tend to zero, which implies (3.16).

Finally, (3.16) implies $C_n^{-1} D_n \to I_q$ i.p., and premultiplication of

$$D_n^{-1} P_n^{-1/2} X_n' \epsilon(n)$$

yields the assertion.

Uniformity in $\epsilon \in \mathfrak{F}(F)$ of (3.8) follows from the fact that the preceding proof remains valid if instead of one and the same ϵ for each n, we take for each n an arbitrary $\epsilon(n) \in \mathfrak{F}(F)$. Thus, (3.8) holds for all sequences of sequences $\epsilon(n)$, and this is equivalent with uniformity in ϵ.

3.2. *Remarks.* (1) In practice, for finite n, one uses theorem 3.2 in the form

$$\text{d.f. } (b(n)) \sim N(\beta, C_n^2). \tag{3.18}$$

In certain situations this relation may save the trouble of computing the inverse square root of C_n^2.

(2) Theorem 2 of [1] states the asymptotic normality of the single components $b_j(n)$, after suitable normalization. We remark without proof that this theorem also remains valid under unchanged assumptions if, as in section 3.1, the unknown variances σ_k^2 in the normalizing factor are replaced by the squares of the residuals (3.5). Thus, the additional assumptions given in theorem 3 of [1] are in fact superfluous.

The progress of the present paper over [1] lies essentially in the determination of the joint asymptotic d.f. of the vectorial LSE $b(n)$ which was not possible by the method used in [1]. Besides that, condition (I) of the present paper is simpler than the corresponding condition in [1].

(3) If F contains only one element, say G, then (II) and (III) reduce to $0 < \int x^2\, dG(x) < \infty$. The errors are, in this case, identically distributed.

(4) The assumptions (I), (II), and (III) are no longer necessary in theorem 3.2. However, since they are necessary in theorem 3.1, the necessary and sufficient assumptions of theorem 3.2 presumably do not differ very much from (I)–(III).

(5) Concerning the admissible sequences of regression matrices we prove the following lemma.

LEMMA 3.2. *Condition* (I) *implies*

$$\lambda_{\min}(P_n) \to \infty \tag{3.19}$$

and

$$\max_{k=1,\cdots,n} |x_{k,j}|/||x_j(n)|| \to 0, \qquad j = 1, \cdots, q. \tag{3.20}$$

Here $\lambda_{\min}$ denotes the minimum characteristic value and $||\cdot||$ the Euclidean norm. Regression vectors $x_j(n)$ satisfying (3.20) are called *slowly increasing* (compare [7], p. 233).

PROOF. We introduce the $q \times q$ diagonal matrices

$$D_n = \operatorname{diag}(||x_1(n)||, \cdots, ||x_q(n)||), \qquad n = 1, 2, \cdots. \tag{3.21}$$

Since $\operatorname{tr}(D_n^{-1}P_nD_n^{-1}) = q$, we have

$$\sup_n \lambda_{\max}(D_n^{-1}P_nD_n^{-1}) \le q. \tag{3.22}$$

Now

$$r_k'P_n^{-1}r_k = r_k'D_n^{-1}(D_n^{-1}P_nD_n^{-1})^{-1}D_n^{-1}r_k \ge q^{-1}||D_n^{-1}r_k||^2. \tag{3.23}$$

By (I), $\max_k ||D_n^{-1}r_k|| \to 0$ and thus (3.20) follows. Equation (3.20) implies $||x_j(n)|| \to \infty$, for all j, which in turn implies (3.19).

(6) Sufficient for (I) is the relation (3.20) together with

$$\inf_n \lambda_{\min}(D_n^{-1}P_nD_n^{-1}) > 0, \tag{3.24}$$

as can be seen from an inequality similar to (3.23). However, conditions (3.24) plus (3.20) are, in general, not necessary. In particular, (3.24) is satisfied if

$$D_n^{-1}P_nD_n^{-1} \to R \tag{3.25}$$

where R is some positive definite $q \times q$ matrix.

3.3. *Examples.* We now discuss some examples of regression matrices and check whether they possess property (I) or not.

(1) *Polynomial regression.* Let $x_{k,j} = k^{c_j}$, $c_1 > \cdots > c_q > -\frac{1}{2}; j = 1, \cdots, q$, $k = 1, 2, \cdots$. The c_j need not be integers. Then (compare [2], p. 469) the $x_j(n)$ are slowly increasing and $D_n^{-1}P_nD_n^{-1} \to H$ where H is a positive definite submatrix of the Hilbert matrix. Hence, (3.25), and consequently (I), is satisfied.

(2) *Trigonometric regression.* Let

$$x_{k,2j-1} = \cos w_jk, \qquad x_{k,2j} = \sin w_jk, \qquad j = 1, \cdots, q, \qquad k = 1, 2, \cdots \tag{3.26}$$

where the w_j are such that rank $X_n = 2q$. Then (see [2], p. 477) the $x_j(n)$ are slowly increasing and $n^{-1}P_n \to I_q^*$ where I_q^* is a diagonal matrix with diagonal elements 1 or $\frac{1}{2}$. Again (3.25), and thus (I), is satisfied.

(3) *Mixed trigonometric and polynomial regression.* Let

$$(3.27)\qquad \begin{cases} x_{kj} = k^{c_j}, & j = 1, \cdots, q \qquad \text{with } c_1 > \cdots > c_q > -\frac{1}{2}, \\ x_{kj} = e^{ikw_j}, & j = q+1, \cdots, Q \\ & \text{with } 0 < w_j < 2\pi, \quad w_j \neq w_k \text{ for } j \neq k. \end{cases}$$

Again (3.25) is satisfied with a matrix R of the form

$$(3.28)\qquad R = \begin{pmatrix} H & 0 \\ 0 & I_{Q-q} \end{pmatrix}$$

where H is as in example 1. In order to prove this, we observe first that

$$(3.29)\qquad n^{-c-1} \sum_{k=1}^{n} k^c e^{ikw} \to 0 \qquad \text{for } n \to \infty$$

if $c = 0, 1, \cdots$ and $0 < w < 2\pi$. This can be seen by deriving $\sum_{k=1}^{n} e^{ikw} = e^{iw}(1 - e^{inw})/(1 - e^{iw})$ repeatedly with respect to iw. In order to prove (3.29) for nonintegers c, we derive the left-hand side with respect to c and obtain

$$(3.30)\qquad n^{-c-1} \sum_{k} k^c e^{ikw} \ln (k/n),$$

which remains bounded for $c > -\frac{1}{2}$ as $n \to \infty$, since

$$(3.31)\qquad \int_0^1 \left(\frac{t}{n}\right)^c \ln \frac{t}{n}\, d\left(\frac{t}{n}\right) = -(c+1)^{-2}.$$

and since we have proved (3.29) already for integers c, it holds also for nonintegers $c > -\frac{1}{2}$.

Finally, because the matrix R is positive definite, property (I) holds.

(4) *Analysis of variance case.* Consider a one-way classification with q classes having $N_1, \cdots, N_q$ observations respectively. The regression matrix is given by

(3.32)

$$X_n' = \begin{pmatrix} 1, \cdots, 1, & 0, \cdots, 0, & \cdots & & 0 \\ 0, \cdots, 0, & 1, \cdots, 1, & 0, \cdots, & & 0 \\ & \cdots & & & \\ 0, & \cdots & & 0, & 1, \cdots, 1 \end{pmatrix}_{q \times n}, \qquad n = N_1 + \cdots + N_q.$$

Then $P_n = \operatorname{diag}(N_1, \cdots, N_q)$. Condition (I) is satisfied if $\min_j N_j \to \infty$. Hence, in this case the LSE of the effects are asymptotically normally distributed for every error sequence $\epsilon \in \mathfrak{F}(F)$.

(5) *Exponential regression.* The regression vectors $x_j(n)$ with $x_{kj} = c_j^k$ where $c_1 > c_2 > \cdots > c_q > 1$, are not slowly increasing since

$$(3.33)\qquad c_j^{2n} \Big/ \sum_{k=1}^{n} c_j^{2k} > \frac{c_j}{c_j^2} > 0.$$

Therefore, by lemma 3.2, (I) cannot be satisfied.

(6) *Mixed polynomial-exponential regression*, $x_{kj} = k^{c_j} d_j^k$, $d_j > 1$, $c_j > 0$. Again the regression vectors are not slowly increasing, since $\|x_j(n)\|^2 = O(n^{2c_j} d_j^{2n})$.

4. Dependent errors (regression for time series)

4.1. *The case of one regression vector.* Since for dependent errors the results are not of as closed a form as for independent errors, we consider first the case of only one regression vector ($q = 1$). This case already shows the main deviations from the previous results.

System (2.1) reduces now to

$$y_t = x_t\beta + \epsilon_t, \qquad t = 1, 2, \cdots, n, \tag{4.1}$$

in vectorial notation $y(n) = x(n)\beta + \epsilon(n)$. We assume, as is typical also for the analysis of time series,

$$\epsilon_t = \sum_{j=-\infty}^{\infty} c_j \eta_{t+j} = \sum_{j=-\infty}^{\infty} c_{j-t}\eta_j, \qquad t = 1, 2, \cdots, \tag{4.2}$$

where the sequence $c \equiv \{c_j\}$ of real constants is square summable, that is, $c \in l^2$ (the Hilbert space of all square summable sequences of real numbers). In all of the previous publications the stronger assumption $\sum_{j=-\infty}^{\infty} |c_j| < \infty$ has been made (see, for example, [9]).

As is well known, the condition $c \in l^2$ is necessary in order that (4.2) holds as a limit in quadratic mean. The r.v.'s η_j are assumed to be independent with expectations zero, but they need not be identically distributed. Let their d.f.'s, as in section 3, lie in a set F where F satisfies (2.5) and conditions (II) and (III) of section 3.1. Then each ϵ_t is, in fact, defined as a limit in the mean of the sums $\sum_{j=-n}^{n} c_j \eta_{t+j}$ for every sequence $\{\eta_i\} \in \mathfrak{F}(F)$ (see section 3.1). Random sequences $\{\epsilon_t\}$ of this type have been called *generalized linear processes* ([3]). If the η_t are identically distributed, the ϵ_t form a strictly stationary linear stochastic process.

With $P_n = \|x(n)\|^2$, the (scalar) LSE's of β are

$$b(n) = \|x(n)\|^{-2} x'(n) y(n), \qquad n = 1, 2, \cdots. \tag{4.3}$$

In order to investigate the asymptotic normality of the sequence $\{b(n)\}$, put

$$\begin{aligned} \zeta_n &= (\operatorname{var} b(n))^{-1/2}(b(n) - \beta) \\ &= (\operatorname{var} b(n))^{-1/2}\|x(n)\|^{-2} x'(n)\epsilon(n) \\ &= (\operatorname{var} b(n))^{-1/2}\|x(n)\|^{-2} \sum_{j=-\infty}^{\infty} \left(\sum_{t=1}^{n} x_t c_{j-t}\right) \eta_j. \end{aligned} \tag{4.4}$$

We have $E\zeta_n = 0$, $\operatorname{var} \zeta_n = 1$ for all n. Put

$$\begin{aligned} A_{nj} &= \sum_{t=1}^{n} x_j c_{j-t}, \\ S_n &= \sum_{j=-\infty}^{\infty} A_{nj}^2. \end{aligned} \tag{4.5}$$

Clearly, always $S_n < \infty$. In ([3], p. 319) the following proposition, which also holds in a more general context, has been proved.

THEOREM 4.1. *Let* $S_n > 0$ *for all* n. *In order that* (A): d.f. $(\zeta_n) \to N(0, 1)$, *and* (B): *the contributions of the summands of* ζ_n *in the last expression of* (4.4) *are infinitesimal, both for every* $\{\eta_j\} \in \mathfrak{F}(F)$, *the conditions* (II) *and* (III) *of theorem* 3.1 *and*

$$\sup_{j=-\infty,\cdots,\infty} A_{nj}^2/S_n \to 0 \tag{4.6}$$

are jointly necessary and sufficient.

We do not emphasize the validity of the theorem for the whole class $\mathfrak{F}(F)$. We rather consider a sequence $\{\eta_i\} \in \mathfrak{F}(F)$ to be given and shall now analyze in detail the remaining condition (4.6).

Condition (4.6) may be verified directly for a given sequence $c \in l^2$ and a given sequence $x \equiv \{x_1, x_2, \cdots\}$ of regression constants. It would be more convenient, however, if for each c the class $\mathfrak{X}_c$ of all x's satisfying (4.6), or for each x the class $\mathfrak{C}_x$ of all c's satisfying (4.6) were known. It then remains only to be checked whether a given x belongs to $\mathfrak{X}_c$, or a given c belongs to $\mathfrak{C}_x$. Since in a regression problem x is known but c usually is not, the classes $\mathfrak{C}_x$ are of greater interest. We shall, therefore, direct our attention mainly on $\mathfrak{C}_x$. If we are unable to determine a class $\mathfrak{C}_x$ completely, we shall try to find as large a subclass as possible.

Let $c(\lambda) \in L^2$ (the space of the complex valued functions over $\Lambda = \{\lambda: -\frac{1}{2} \leq \lambda \leq \frac{1}{2}\}$ whose moduli are Lebesgue square integrable) be such that

$$c_j = \int_\Lambda e^{-2\pi i\lambda j} c(\lambda)\, d\lambda, \qquad c(\lambda) \sim \sum_j c_j e^{2\pi i j\lambda}, \tag{4.7}$$

and put

$$x_n(\lambda) = \sum_{t=1}^{n} x_t e^{2\pi i t\lambda}. \tag{4.8}$$

Then for sufficiently large n,

$$S_n = \int_\Lambda |x_n(\lambda) c(\lambda)|^2\, d\lambda > 0 \tag{4.9}$$

as is required in theorem 4.1; null sequences x and c are of course excluded.

Because of their importance in practical applications, one will not want to exclude all finite sequences c from any $\mathfrak{C}_x$. But if $\mathfrak{C}_x$ contains any finite nonnull sequence, then (4.6) implies

$$\max_{k=1,\cdots,n} |x_k|/\|x(n)\| \to 0, \tag{4.10}$$

as will be seen from lemma 4.3 (x is slowly increasing). In order to investigate the behavior of the left-hand side of (4.6) under this additional assumption, we prove the following lemma.

LEMMA 4.1. *Equation* (4.10) *implies*

$$\sup_j |A_{n,j}|/||x(n)|| \to 0 \tag{4.11}$$

for all $c \in l^2$.

Presumably, (4.11) (for fixed c) also implies (4.10), but we do not prove it here, except for finite c-sequences (lemma 4.3).

PROOF. Choose $m_n < n$, $m_n \to \infty$ such that

$$m_n \max_k |x_k|/||x(n)|| \to 0. \tag{4.12}$$

There exists always an integer j_n such that $|A_{n,j_n}| = \sup_j |A_{n,j}|$. Put $J_n = \{j_n - n, \cdots, j_n - 1\}$ and split the sum

$$A_{n,j_n} = \sum_{t=1}^{n} x_t c_{j_n - t} = \sum_{t \in J_n} x_{j_n - t} c_t \tag{4.13}$$

into the sum $\alpha_n = \sum_{t \in J_n \cap K_n} x_{j_n - t} c_t$ and into the remainder β_n; here K_n is the index set $\{t : t| < [m_n/2]\}$.
Now

$$|A_{n,j_n}| \leq |\alpha_n| + |\beta_n| \leq m_n \sup_j |c_j| \max_k |x_k| + ||x(n)|| (\sum_{|t| \geq [m_n/2]} c_t^2)^{1/2}. \tag{4.14}$$

After division by $||x(n)||$, this tends to zero for $n \to \infty$.

THEOREM 4.2. *Let* (4.10) *be true and*

$$\operatorname*{ess\,inf}_{\lambda \in \Lambda} |c(\lambda)| > 0. \tag{4.15}$$

Then (4.6) *holds, and consequently, statement* (A) *is valid.*

PROOF. The proof follows from the preceding lemma and

$$S_n \geq ||x(n)||^2 \operatorname*{ess\,inf}_{\lambda} |c(\lambda)|^2. \tag{4.16}$$

For a large class of slowly increasing regression vectors, condition (4.15) is not necessary. Assume, besides (4.10), that

$$\lim_{n \to \infty} \sum_{t=1}^{n-h} x_{t+h} x_t / ||x(n)||^2 = \tilde{R}_h, \qquad h = 1, 2, \cdots, \tag{4.17}$$

exists. Put $\tilde{R}_{-h} = \tilde{R}_h$. Then $\{\tilde{R}_h\}$ is a positive definite sequence, and there exists a d.f. $M(\lambda)$, $\lambda \in \Lambda$, of finite variation such that

$$\tilde{R}_h = \int_\Lambda e^{2\pi i h \lambda} \, dM(\lambda) \tag{4.18}$$

([7], p. 233). We have $M(\frac{1}{2}) - M(-\frac{1}{2}) = 1$ and

$$||x(n)||^{-2} \int_{-1/2}^{\lambda} |x_n(\mu)|^2 \, d\mu \equiv M_n(\lambda) \to M(\lambda) \tag{4.19}$$

at continuity points of $M(\lambda)$.

Now let $\{I_1, \cdots, I_K\}$ be any partition of Λ into disjoint intervals, whose end points are not jump points of $M(\lambda)$. Let

$$\operatorname*{ess\,inf}_{I_k} |c(\lambda)|^2 \tag{4.20}$$

denote the essential infimum of $|c(\lambda)|^2$ over I_k. Then

$$S_n/\|x(n)\|^2 \geq \sum_{k=1}^{K} \operatorname*{ess\,inf}_{I_k} |c(\lambda)|^2 \Delta M_n(I_k) \tag{4.21}$$

where $\Delta M_n(I_k)$ is the variation of $M_n(\lambda)$ over I_k. Because of (4.19),

$$\liminf_n S_n/\|x(n)\|^2 \geq \sum_{k=1}^{K} \operatorname*{ess\,inf}_{I_k} |c(\lambda)|^2 \Delta M(I_k). \tag{4.22}$$

The relation remains valid if we take on the right-hand side the supremum with respect to all admissible partitions with arbitrary K.

DEFINITION. *The function $|c|$ on Λ is called essentially positive at λ if*

$$\sup_{I \ni \lambda} \operatorname*{ess\,inf}_{I} |c| > 0. \tag{4.23}$$

where I denotes on interval $\subset \Lambda$.

We now deduce theorem 4.3.

THEOREM 4.3. *Let* (4.10) *and* (4.17) *be true, and let $|c(\lambda)|$ be essentially positive on at least one point of the spectrum of $M(\lambda)$. Then the right-hand side of* (4.22) *is positive, and consequently* (4.6) *and statement* (A) *are valid.*

For some sequences of regression constants it is possible to obtain a complete characterization of the class $\mathcal{C}_x$.

EXAMPLE. Let $x_t = 1$, for all t. Then $\mathcal{C}_{\{1\}}$ consists of all $c \in l^2$ with only a small exceptional class characterized by $\lim_{n\to\infty} \sum_{j=-n}^{n} c_j = 0$ and convergence of $\sum_{j=0}^{n} c_j$ to T, say. The c's of this subclass satisfy (4.6) if and only if $\sum_{j=-n}^{n} c_j$ does not converge too fast to zero, namely, if and only if

$$\sum_{n=1}^{N} \left(\left(\sum_{j=0}^{n} c_j - T \right)^2 + \left(\sum_{j=-n}^{-1} c_j + T \right)^2 \right) \to \infty \tag{4.24}$$

for $N \to \infty$ ([3], p. 325). If $c_{-1} = c_{-2} = \cdots = 0$, (4.24) reduces to

$$\sum_{n=1}^{N} \left(\sum_{j=0}^{n} c_j \right)^2 \to \infty. \tag{4.25}$$

We conclude this section with a remark concerning the convergence properties of the sequence of functions

$$x_n(\lambda)c(\lambda) \sim \sum_{j=-\infty}^{\infty} A_{n,j} e^{2\pi i j \lambda}. \tag{4.26}$$

Let $d_n(\lambda)$, $n = 1, 2, \cdots$, $-\frac{1}{2} \leq \lambda \leq \frac{1}{2}$, be any sequence of functions with $d_n(\lambda) \in L^2$,

$$d_n(\lambda) \sim \sum_{j=-\infty}^{\infty} d_{nj} e^{2\pi i j \lambda}. \tag{4.27}$$

Then

$$\int_\Lambda \overline{d_n(\lambda)} x_n(\lambda) c(\lambda)\, d\lambda = \sum_j \overline{d_{n,j}} A_{n,j}. \tag{4.28}$$

LEMMA 4.2. *Let* $\sum_j |d_{n,j}|^2 = 1$, $\sum_j |d_{n,j}|^2$ *converge uniformly in* n. *Then* (4.6) *implies*

$$S_n^{-1/2} \int_\Lambda \overline{d_n(\lambda)} x_n(\lambda) c(\lambda)\, d\lambda \to 0 \tag{4.29}$$

uniformly with respect to the functions $d_n(\lambda)$ *out of the considered class.*

PROOF. Choose $m_n < n$, $m_n \to \infty$ such that

$$m_n \sup_j |A_{nj}|/S_n^{1/2} \to 0. \tag{4.30}$$

Then

$$S_n^{-1/2} \Big|\sum_j \overline{d_{n,j}} A_{n,j}\Big| \leq 2m_n \sup_j |A_{n,j}| S_n^{-1/2} + \Big(\sum_{|j| \geq m_n} |d_{n,j}|^2\Big)^{1/2} \to 0. \tag{4.31}$$

Let us now put $\gamma_n(\lambda) = S_n^{-1/2} x_n(\lambda) c(\lambda)$, and assume $\gamma_n(\lambda)$ converges boundedly in measure to an integrable limiting function $\gamma(\lambda)$ on Λ. Let $\mathrm{Re}\gamma(\lambda)$ be of one sign in some interval $\subset \Lambda$ and take all $d_n(\lambda) = d(\lambda)$, the characteristic function of this interval. Then by the bounded convergence theorem,

$$\int_\Lambda d(\lambda) \gamma_n(\lambda)\, d\lambda \to \int_\Lambda d(\lambda) \gamma(\lambda)\, d\lambda. \tag{4.32}$$

By the above lemma the integrals on the left tend to zero under (4.6), so that $\mathrm{Re}\gamma(\lambda) = 0$ [a.e.] on the considered interval. Repeating the argument for $\mathrm{Im}\gamma(\lambda)$, we obtain $\gamma(\lambda) = 0$ [a.e.]. This, however, is in contradiction with the fact that

$$\int_\Lambda |\gamma_n(\lambda)|^2\, d\lambda = 1 \qquad \text{for all} \quad n, \tag{4.33}$$

which implies

$$\int_\Lambda |\gamma(\lambda)|^2\, d\lambda = 1. \tag{4.34}$$

The same argument holds if any infinite subsequence of $\{\gamma_n(\lambda)\}$ is taken or any measurable subset of Λ is considered instead of Λ.

Thus we have the following: *under* (4.6), *no subsequence of* $\{\gamma_n(\lambda)\}$ *converges boundedly in measure to an integrable function.*

Let $c(\lambda) \equiv 1$. Then (4.6) is equivalent with (4.10). In addition, let (4.17) be true, so that (4.19) holds. Assume $M(\lambda)$ possesses a bounded derivative $M'(\lambda)$. Then the preceding proposition is somewhat surprising in view of the fact that

$$\int_{-1/2}^{\lambda} |\gamma_n(\mu)|^2\, d\mu = \|x(n)\|^{-2} \int_{-1/2}^{\lambda} |x_n(\mu)|^2\, d\mu \to \int_{-1/2}^{\lambda} M'(\mu)\, d\mu \tag{4.35}$$

for all λ. One may guess that $\gamma_n(\lambda)$ must be increasingly oscillatory for $n \to \infty$, which may be due to the fact that $\gamma_n(\lambda)$ is complex-valued.

Here the remark may be of interest that always

$$\sum_{t=1}^{n} |x_t| \Big/ \Big(\sum_{t=1}^{n} x_t^2\Big)^{1/2} \to \infty \tag{4.36}$$

if (4.10) holds so that, at least sometimes, the boundedness condition of the convergence will be violated.

4.2. *One regression vector* (*errors are finite moving averages*). For further study of condition (4.6), we now restrict ourselves to finite c-sequences. Let $k_1 < k_2$, $c_{k_1}, c_{k_2} \neq 0$, $c_j = 0$ for $j < k_1$, $j > k_2$. Then

$$A_{n,j} = 0 \quad \text{for} \quad j \leq k_1 \quad \text{and} \quad j > k_2 + n, \qquad n = 1, 2, \cdots, \tag{4.37}$$

$$A_{n,j} \equiv A_j = \sum_{k=k_1}^{k_2} c_k x_{j-k} \quad \text{if} \quad k_2 < j \leq k_1 + n, \qquad n > k_2 - k_1, \tag{4.38}$$

$$S_n = \sum_{j=k_1+1}^{k_1+n} A_j^2 + \sum_{j=k_1+n+1}^{k_2+n} A_{n,j}^2, \qquad n > k_2 - k_1, \tag{4.39}$$

with $A_j = x_1 c_{j-1} + \cdots + x_{j-k_1} c_{k_1}$ for $k_1 < j \leq k_2$. Since

$$\inf_n \sup_j |A_{n,j}| > 0 \tag{4.40}$$

(4.6) implies $\lim S_n = +\infty$, we have by (4.6)

$$\sum_{j=k_1+n+1}^{k_2+n} A_{n,j}^2 / S_n \to 0. \tag{4.41}$$

Thus (4.6) also implies

$$\sum_{j=k_1+1}^{n} A_j^2 \to \infty, \tag{4.42}$$

and it is equivalent to

$$\sup_j A_{n,j}^2 \Big/ \sum_{j=k_1+1}^{n} A_j^2 \to 0. \tag{4.43}$$

We have, moreover, the following lemma.

LEMMA 4.3. *If c is a finite nonnull sequence, then* (4.6) *implies* (4.10), *and* (4.10) *is equivalent to* (4.11).

PROOF. We have $\max_j A_{n,j}^2 \geq c_{k_2}^2 x_n^2$. Since $c(\lambda)$ is now continuous, we have $\sup_\lambda |c(\lambda)|^2 = \gamma < \infty$. Hence by (4.9),

$$S_n \leq \gamma \|x(n)\|^2. \tag{4.44}$$

Thus

$$\sup_j A_{n,j}^2 / S_n \geq \gamma' x_n^2 / \|x(n)\|^2, \qquad \gamma' > 0, \tag{4.45}$$

and (4.6) implies

$$|x_n| / \|x(n)\| \to 0. \tag{4.46}$$

As seen above, (4.6) implies $S_n \to \infty$; hence,

$$\|x(n)\| \to \infty. \tag{4.47}$$

However, this together with (4.46) is equivalent to (4.10), since otherwise, with k_n chosen such that

$$|x_{k_n}| = \max_{k=1,\cdots,n} |x_k|, \tag{4.48}$$

$|x_{k_n}| / \|x(k_n)\| \not\to 0$, in contradiction to (4.46).

The proof of the equivalence of (4.10) and (4.11) is similar. (End of proof.)

If $c(\lambda)$, which is now a Fourier polynomial, has no zeros, and if (4.10) is true, then we have asymptotic normality by theorem 4.2.

Condition (4.42) obviously is not satisfied (and thus we have no asymptotic normality) if the sequence $\{x_t\}$ satisfies the recursive system

$$\sum_{k=k_1}^{k_2} c_k x_{j-k} = A_j = 0, \qquad j > k_2. \tag{4.49}$$

If the equation

$$c_{k_1} t^{k_2-k_1} + \cdots + c_{k_2-1} t + c_{k_2} = 0 \tag{4.50}$$

possesses a root of modulus one, but no larger ones ($c(\lambda)$ then has a zero in $[-\frac{1}{2}, \frac{1}{2}]$), then $\{x_t\}$ satisfies (4.10), but (4.6) does not hold if (4.49) is true. Thus the converse of the first statement of lemma 4.3 does not hold. A criterion for the validity of (4.49) is that all the determinants

$$\begin{vmatrix} x_1, & \cdots, & x_n \\ & \cdots & \\ x_n, & \cdots, & x_{2n-1} \end{vmatrix} \tag{4.51}$$

vanish for $n > k_2 - k_1$.

There are many cases of pairs (x, c) of sequences incompatible with (4.6) that are of practical interest. For example, let $x_t = 1$, for all t, and $c_0 = -c_m = 1$, all other $c_j = 0$. Then $x_t - x_{t-m} = 0$, and $t = m + 1, m + 2, \cdots$ is a recursive system.

From this example we see that a theorem of the type "If x is any element of a class of regression sequences independent of c, and if c is any element of a class independent of x, then asymptotic normality holds," is not desirable because it excludes too many important cases. It is for this reason that we entered into a sharper analysis and tried to obtain the classes $\mathcal{C}_x$.

For the case that $c(\lambda)$ has zeros, we can state the following: if (4.10) is true, and if the sequence of functions

$$\left(\max_{j=1,\cdots,n} x_j^2\right)^{-1} \int_{-1/2}^{\lambda} |x_n(\mu)|\, d\mu, \qquad -\tfrac{1}{2} \le \lambda \le \tfrac{1}{2}, \tag{4.52}$$

or any subsequence thereof does not tend to a pure jump function whose jump points are all zeros of $c(\lambda)$, then asymptotic normality holds.

About the estimation of the normalizing factors used in (4.4) some remarks are made in section 4.4 (3).

4.3. *Multiple regression* (*dependent errors*). We use the notation of sections 2 and 3.1 with $q > 1$ and errors given by (4.2). Let $B_n^2 = \operatorname{cov}(b(n)b'(n))$. This matrix has no longer the simple structure (3.3) because $\sum_n$ in general is not diagonal. In order to derive the next theorem, we introduce the q-vectors

$$\zeta(n) = B_n^{-1} P_n^{-1} X_n' \epsilon(n) = B_n^{-1} P_n^{-1} \sum_{j=-\infty}^{\infty} \left(\sum_{t=1}^{n} r_t c_{j-t}\right) \eta_j, \tag{4.53}$$

which all have expectation zero and covariance matrix I_q. (The vector r_k was defined as the k-th column of X_n'. The matrices $P_n (n \geq q)$ are assumed non-singular.)

Now

$$\text{d.f.}\ (\zeta(n)) \to N(0, I_q) \tag{4.54}$$

if and only if

$$\text{d.f.}\ (d'(n)\zeta(n)) \to N(0, 1) \tag{4.55}$$

for all sequences of constant q-vectors $d(n)$ with $d(n) \in S_q$ (the unit sphere in R_q). For

$$d(n) = B_n P_n f(n) \| B_n P_n f(n) \|^{-1}, \qquad f(n) \in S_q, \tag{4.56}$$

we have

$$d'(n)\zeta(n) = \| B_n P_n f(n) \|^{-1} \sum_j \left(\sum_{t=1}^{n} f'(n) r_t c_{j-t} \right) \eta_j. \tag{4.57}$$

Here $\{\eta_j\} \in \mathfrak{F}(F)$, F subject to conditions (II) and (III) of theorem 3.1. We now have reduced the problem of finding conditions asserting (4.54) to the one-dimensional case. By theorem (4.1), (4.55) holds if

$$\sup_j A_{n,j}^2 / S_n \to 0, \qquad S_n = \sum_{j=-\infty}^{\infty} A_{n,j}^2, \tag{4.58}$$

where now

$$A_{n,j} = \sum_{t=1}^{n} f'(n) r_t c_{j-t} = f'(n) X_n' \begin{pmatrix} c_{j-1} \\ \cdot \\ \cdot \\ \cdot \\ c_{j-n} \end{pmatrix}. \tag{4.59}$$

Relation (4.54) holds if (4.58) holds for every sequence $\{f(n)\}$, $f(n) \in S_q$. That is the case if and only if

$$\sup_{f \in S_q} \left[\sup_j (f' X_n' c(j, n))^2 \Big/ \sum_{k=-\infty}^{\infty} (f' X_n' c(k, n))^2 \right] \to 0; \tag{4.60}$$

here $c(j, n) = (c_{j-1}, \cdots, c_{j-n})'$. Putting

$$\begin{gathered} \sum_{j=-\infty}^{\infty} c_{j+k} c_j = R_k = R_{-k}, \\ R(n) = \begin{pmatrix} R_0, & \cdots, & R_{n-1} \\ & \cdots & \\ R_{n-1}, & \cdots, & R_0 \end{pmatrix}, \end{gathered} \tag{4.61}$$

the denominator in (4.60) becomes

$$S_n = f' X_n' R(n) X_n f. \tag{4.62}$$

In analogy to (4.60), we also have, with

$$x_n(\lambda) = \sum_{t=1}^{n} f' r_t e^{2\pi i t \lambda}, \tag{4.63}$$

$$S_n = \int_\Lambda |x_n(\lambda) c(\lambda)|^2 \, d\lambda. \tag{4.64}$$

Suppose

$$\operatorname*{ess\,inf}_{\lambda} |c(\lambda)|^2 \equiv \gamma_0 > 0. \tag{4.65}$$

Then

$$S_n \geq \gamma_0 f' P_n f, \qquad P_n = X_n' X_n, \tag{4.66}$$

and (4.60) holds, according to lemma 4.1, if

$$\max_t \sup_{f \in S_q} (f' r_t)^2 / f' P_n f \to 0. \tag{4.67}$$

This is equivalent to

$$\max_t \sup_{f \in S_q} (f' P_n^{-1/2} r_t)^2 = \max_t r_t' P_n^{-1} r_t \to 0. \tag{4.68}$$

Hence, the following theorem holds.

THEOREM 4.4. *If*

$$\max_t r_t' P_n^{-1} r_t \to 0 \tag{I}$$

and ess $\inf_\lambda |c(\lambda)| > 0$, *then* (4.54) *holds.*

As in section 4.1, assumption (4.65) may be weakened by assuming that the regression matrices X_n allow for a harmonic analysis. Generalizing (4.17), we assume

$$\max_{t=1,\cdots,n} x_{tj}^2 / \|x_j(n)\|^2 \to 0 \qquad \text{for} \quad j = 1, \cdots, q, \tag{4.69}$$

$$\sum_{t=1}^{n} x_{t+k,r} x_{t,s} / \|x_r(n)\| \, \|x_s(n)\| \to R_h^{(r,s)}, \qquad \begin{matrix} r, s = 1, \cdots, q, \\ h = 0, \pm 1, \cdots. \end{matrix} \tag{4.70}$$

Then $R_{-h}^{(s,r)} = R_h^{(r,s)}$, and the $(q \times q)$-matrices $\tilde{R}_h = \{R_h^{(r,s)}\}$ admit the spectral representation

$$\tilde{R}_h = \int_\Lambda e^{2\pi i h \lambda} \, dM(\lambda), \qquad h = 0, \pm 1, \cdots, \tag{4.71}$$

where the elements of the $(q \times q)$-matrix $M(\lambda)$ are functions of bounded variation and $M(\lambda_2) - M(\lambda_1)$ is positive semidefinite for every $\lambda_2, \lambda_1, -\frac{1}{2} \leq \lambda_1 < \lambda_2 \leq +\frac{1}{2}$ (see [7], p. 233). The set of all points λ with $M(\lambda_2) - M(\lambda_1)$ positive definite for all $\lambda_1 < \lambda < \lambda_2$ is called the *spectrum* of $M(\lambda)$.

We assume

$$\tilde{R}_0 = \int_\Lambda dM(\lambda) = M(\tfrac{1}{2}) - M(-\tfrac{1}{2}) \equiv M \tag{4.72}$$

to be nonsingular and put

$$D_n = \operatorname{diag}(\|x_1(n)\|, \cdots, \|x_q(n)\|). \tag{4.73}$$

Then ([7], p. 238)

$$D_n P_n^{-1} X_n' R(n) X_n P_n^{-1} D_n \to M^{-1} \int_\Lambda |c(\lambda)|^2 \, dM(\lambda) M^{-1} \equiv \tilde{M}. \tag{4.74}$$

Now from (4.62), for $n \to \infty$,

$$S_n = f' X_n' R(n) X_n f = f' P_n D_n^{-1} \tilde{M} D_n^{-1} P_n f \geq \lambda_q \|D_n^{-1} P_n f\|^2 \tag{4.75}$$

where we have put $\lambda_q = \lambda_{\min}(\tilde{M})$. Since by (4.70) $D_n^{-1}P_nD_n^{-1} \to \tilde{R}_0$, we have in the limit $S_n \geq \gamma\|D_nf\|^2$ with some new constant γ that is positive if $\lambda_q > 0$.

We now assume $\lambda_q > 0$ and have (4.60) if

$$\sup_j \sup_{f\in S_q} (f'D_nD_n^{-1}X_n'c(j, n))^2/\|D_nf\|^2 = \sup_j \sup_{f\in S_q} (f'D_n^{-1}X_n'c(j, n))^2 \to 0. \tag{4.76}$$

Choosing $f = D_n^{-1}X_n'c(j, n)/\|D_n^{-1}X_n'c(j, n)\|$, we see that

$$\sup_j \|D_n^{-1}X_n'c(j, n)\| \to 0, \tag{4.77}$$

or that

$$\sup_j \sum_{k=1}^{n} x_{tk}c_{j-k}|/\|x_k(n)\| \to 0 \qquad \text{for all} \quad k = 1, \cdots, q \tag{4.78}$$

is sufficient for (4.60). The latter, however, is true because of (4.69) and lemma (4.1).

A condition implying $\lambda_q > 0$ is that $\int_\Lambda |c(\lambda)|^2\, dM(\lambda)$ be nonsingular in case this integral is defined and is the limit of the corresponding finite approximations. It is the higher dimensional generalization of (4.22).

In summary, the following holds.

THEOREM 4.5. *Let* (4.69) *and* (4.70) *be true, and let* $\tilde{R}_0$ *be nonsingular. Let* $c(\lambda)$ *be essentially positive on at least one point of the spectrum of* $M(\lambda)$. *Then* (4.54) *holds.*

Concerning the estimation of the normalizing matrices B_n, the reader is referred to section 4.4 (3).

Finally, it may be noticed that $\operatorname{ess\,inf}_{\lambda\in\Lambda} |c(\lambda)| > 0$ can always be achieved by simply adding to the y_t's independent random variables of an artificial sequence $\{\rho_t\}$, which are also independent of the η_j. The function $c^*(\lambda)$, associated with the new combined error sequence $\rho_t + \epsilon_t$, always satisfies $\operatorname{ess\,inf}_{\lambda\in\Lambda} |c^*(\lambda)| > 0$. This method may be considered as a particular type of prewhitening. It always implies, however, an increase in variance of the LSE $b(n)$. A similar proposal has been made by Hannan [9].

4.4. *Concluding remarks.* (1) As is well known, the Gauss-Markov estimators

$$b_G(n) = (X_n'\textstyle\sum_n^{-1}X_n)^{-1}X_n'\sum_n^{-1}y(n) \tag{4.79}$$

are the minimum variance linear unbiased estimators for β, whether the ϵ_t are correlated or not. One may, therefore, try to use $b_G(n)$ instead of the above considered LSE $b(n)$. However, in the first place, the covariance matrix $\sum_n$ usually is unknown, and a useful estimate for $\sum_n^{-1}$ (or for the functions of $\sum_n^{-1}$ that occur in (4.79)) cannot be obtained from a single sequence of observations $\{y_t\}$. Instead, it appears to be more appropriate to use a distribution free method. In the second place, there is not much point in preferring $b_G(n)$ to $b(n)$, because it is known that both are equally efficient asymptotically for a rather large class of error sequences $\{\epsilon_t\}$ and of sequences of regression matrices $\{X_n\}$ [7].

(2) The results of the preceding sections can be extended to the case of *vectorial or multivariate regression equations*

$$y_t = x_{t1}\beta_1 + \cdots + x_{tq}\beta_q + \epsilon_t \tag{4.80}$$

where now $y_t = (y_t^1, \cdots, y_t^k)'$, $x_{tj} = (x_{tj}^1, \cdots, x_{tj}^k)'$, $\epsilon_t = (\epsilon_t^1, \cdots, \epsilon_t^k)'$ are k-vectors and the β_j are scalars as before ([9], for a review see [11]). Sometimes vectorial regression equations are expressed in the form

$$y_t = B\phi_t + \epsilon_t \tag{4.81}$$

where the vector $\phi_t = (\phi_{t1}, \cdots, \phi_{tp})'$ contains the known regression constants, and the $(k \times p)$-matrix B contains the unknown regression constants. However, (4.81) is a special case of (4.80), as is seen by putting

$$(x_{t1}, \cdots, x_{tq}) = \begin{pmatrix} \phi_t', 0, & \cdots 0 \\ 0, \ \phi_t', & \cdots 0 \\ \vdots & \vdots \\ 0, \ 0, & \phi_t' \end{pmatrix} \tag{4.82}$$

where the zeros represent zero row vectors of p dimensions. With $q = kp$, both sides are $(k \times kp)$-matrices. We get the row vector β by placing the rows of B one behind the other.

If the k-vectors ϵ_t are independent for different t with possibly dependent components for each fixed t, and assuming $E\epsilon_t = 0$, we have almost the case considered in section 3, except that the error sequence $(\epsilon_1^1, \cdots, \epsilon_1^k, \epsilon_2^1, \cdots, \epsilon_2^k, \cdots)$ now is k-dependent. We might apply a CLT for k-dependent r.v.'s. However, it appears to be almost as easy to appeal directly to the CLT for independent r.v.'s. For this purpose, let

(i) R be a nonempty set of strictly positive definite $(k \times k)$-covariance matrices,
(ii) F be the set of d.f.'s defined in section 2,
(iii) $\mathcal{G}(F, R)$ be the set of all sequences $\{\epsilon_t\}$ of independent k-vectors ϵ_t with cov $(\epsilon_t\epsilon_t') \in R$ and d.f. $(\epsilon_t^j) \in F$ for all t, j,
(iv) $b(n)$ be the vectorial LSE for β,

$$b(n) = P_n^{-1}X_n'y(n). \tag{4.83}$$

Here now $X_n = (x_{tj}; t = 1, \cdots, n, j = 1, \cdots, q)$ is a $(kn \times q)$-matrix, $P_n = X_n', X_n$, and $y(n) = (y_1^1, \cdots, y_1^k, y_2^1, \cdots, y_2^k, \cdots, y_n^k)'$. Similarly, $\epsilon(n)$ is defined. Furthermore,

$$\begin{aligned} B_n^2 &= \text{cov}\,(b(n)b'(n)) = P_n^{-1}X_n'\textstyle\sum_n X_nP_n^{-1} \\ &= P_n^{-1}\sum_{j=1}^{n} x_j'\rho_jx_jP_n^{-1} \end{aligned} \tag{4.84}$$

with

$$\begin{aligned} x_j &= (x_{j1}, \cdots, x_{jq})_{k\times q}, \\ \rho_j &= E(\epsilon_j\epsilon_j')_{k\times k}. \end{aligned} \tag{4.85}$$

Let r_m be the m-th column vector of X_n', $m = 1, \cdots, kn$. Finally, let $\lambda_k(\rho)$ be the smallest characteristic root of $\rho \in R$. Then we have

THEOREM 4.6. *The d.f.* $(B_n^{-1}(b(n) - \beta) \to N(0, I_q)$ *for all* $\{\epsilon_t\} \in \mathcal{G}(F, R)$ *if*

(I) $$\max_{m=1,\cdots,kn} r_m' P_n^{-1} r_m \to 0,$$

(II) $$\sup_{G \in F} \int_{|x|>c} x^2 \, dG(x) \to 0 \quad for \quad c \to \infty,$$

(III) $$\inf_{\rho \in R} \lambda_k(\rho) > 0.$$

We indicate the *proof* only for $q = 1$. By (4.83) we have

$$B_n^{-1}(b(n) - \beta) = B_n^{-1} P_n^{-1} \sum_{j=1}^{n} x_j' \epsilon_j \tag{4.86}$$

where $x_j'\epsilon_j = \sum_{s=1}^{k} x_{j1}^s \epsilon_j^s$ are independent r.v.'s. Putting $\inf_{\rho \in R} \lambda_k(\rho) = \lambda$ and $\max_{s=1,\cdots,k} (x_{j1}^s)^2 P_n^{-1} = \kappa_{n,j}$, we obtain

$$B_n^2 \geq P_n^{-2} \lambda \sum_{j=1}^{n} \|x_j\|^2 = \lambda P_n^{-1}. \tag{4.87}$$

Now for any $\delta > 0$,

$$\begin{aligned} P(|B_n^{-1} P_n^{-1} x_j' \epsilon_j| > \delta) &\leq P(B_n^{-1} P_n^{-1} \|x_j\| \, \|\epsilon_j\| > \delta) \\ &\leq P(\|\epsilon_j\|^2 > \lambda \delta^2 / (k \kappa_{n,j})) \\ &\leq \sum_{j=1}^{k} P((\epsilon_j^s)^2 > \lambda \delta^2 / (k^2 \kappa_{n,j})) \\ &\leq \sum_{s=1}^{k} \frac{k^2 \kappa_{n,j}}{\lambda \delta^2} \int_{x^2 > \lambda \delta^2 / (k^2 \kappa_{n,j})} x^2 \, dG_j^s(x) \\ &\leq k^3 \kappa_{n,j} \lambda^{-1} \delta^{-2} \phi\left(\left(\frac{\lambda \delta^2}{k^2 \kappa_{n,j}}\right)^{1/2}\right) \end{aligned} \tag{4.88}$$

where $G_j^s = \text{d.f.}\ (\epsilon_j^s)$ and $\phi(c) = \sup_{G \in F} \int_{|x|>c} x^2 \, dG(x)$. Now $\sum_{j=1}^{n} \kappa_{n,j} \leq P_n / P_n = 1$. Putting $\kappa_n = \max_{j=1,\cdots,n} \kappa_{n,j}$, we finally have

$$\sum_{j=1}^{n} P(|B_n^{-1} P_n^{-1} x_j' \epsilon_j| > \delta) \leq k^3 \lambda^{-1} \delta^{-1} \phi\left(\left(\frac{\lambda \delta^2}{k^2 \kappa_n}\right)^{1/2}\right) \to 0 \tag{4.89}$$

by (II). Hence, the CLT holds.

The case where the random k-vectors ϵ_t are generated by a moving average process $\epsilon_t = \sum_{j=-\infty}^{\infty} A_j \eta_{t-j}$ with independent and identically distributed random vectors η_j of r components and with constant $(k \times r)$-matrices A_j has been considered by Hannan [9]. There it is assumed that

$$\sum_{j=-\alpha}^{\infty} (\lambda_{\max}(A_j' A_j))^{1/2} < \infty, \tag{4.90}$$

and that the regression matrices allow for a generalized harmonic analysis in order to derive the asymptotic normal distribution of the LSE. Although the results are likely to be true under more general conditions similar to those of sections 4.1 to 4.3, we do not discuss this possibility here.

(3) An estimation of the normalizing constants $B_n^2 = \text{var}\, b(n)$, or of the normalizing matrices B_n^2 in multiple regression, for an unknown sequence of dependent errors ϵ_t with unequally distributed residuals η_j (see (4.2)) seems to be impossible if only a single sequence $\{y_t\}$ of observations is given. By estimation of B_n^2 we mean a relation like (3.9) with a suitably adapted statistic C_n^2. The reason for the difficulty lies, first of all, in the admission of nonidentically distributed residuals η_j in (4.2) which, together with the unknown c_j's, introduces to many unknown parameters.

Therefore, we restrict ourselves in the following to strictly stationary error sequences $\{\epsilon_t\}$, that is, *we assume the η_j to be independently identically distributed with variance one* (besides $E\eta_j = 0$). In fact, it suffices to have only identical variances of the η_j not necessarily equal to one. We also restrict ourselves to simple regression ($q = 1$). Then

$$B_n^2 = \|x(n)\|^{-4}E(x'(n)\epsilon(n))^2 = \|x(n)\|^{-4}x'(n)R(n)x(n) \tag{4.91}$$

where

$$R(n) = E(\epsilon(n)\epsilon'(n)) = \begin{pmatrix} R_0, & R_1, & \cdots, & R_{n-1} \\ R_1, & R_0, & \cdots, & R_{n-2} \\ & & \cdots, & \\ R_{n-1}, & R_{n-2}, & \cdots, & R_0 \end{pmatrix} \tag{4.92}$$

is the covariance matrix of the ϵ_t with

$$R_k = R_{-k} = E(\epsilon_t\epsilon_{t+k}) = \sum_{j=-\infty}^{\infty} c_j c_{j+k}, \qquad k = 0, 1, \cdots. \tag{4.93}$$

Putting

$$\tilde{R}_{k,n} = \tilde{R}_{-k,n} = \|x(n)\|^{-2} \sum_{t=1}^{n-k} x_t x_{t+k}, \qquad k = 0, 1, \cdots, n-1, \tag{4.94}$$

we obtain

$$B_n^2 = \|x(n)\|^{-2} \sum_{k=-n+1}^{n-1} R_k \tilde{R}_{k,n} = \|x(n)\|^{-4} S_n \tag{4.95}$$

where S_n was defined in section 4.1.

In order to estimate B_n^2, we introduce first the sample covariances of $\{\epsilon_t\}$,

$$R_{k,n} = \frac{1}{n-|k|} \sum_{h=1}^{n-|k|} \epsilon_h \epsilon_{h+|k|}, \qquad k = 0, \pm 1, \cdots, \tag{4.96}$$

and, as an estimator for B_n^2,

$$B_{n,\alpha}^2 = \|x(n)\|^{-2} \sum_{k=-n+1}^{n-1} R_{k,n} \tilde{R}_{k,n} \tag{4.97}$$

(later we shall replace the ϵ_h by the known residuals $e_h(n)$). Since $ER_{k,n} = R_k$, then $EB_{n,\alpha}^2 = B_n^2$.

We now introduce the *additional assumption*

$$\sum_{k=-\infty}^{\infty} R_k^2 < \infty. \tag{4.98}$$

Then ([10], p. 16) with some $\gamma < \infty$,

$$\operatorname{var} R_{k,n} < \gamma/(n - |k|). \tag{4.99}$$

By Schwarz's inequality,

$$\begin{aligned} \operatorname{var} B^2_{n,\alpha} &\le \|x(n)\|^{-4} \sum_{k=-n+1}^{n-1} \tilde{R}^2_{k,n} \operatorname{var} R_{k,n} \\ &\le \frac{\gamma}{\|x(n)\|^4} \sum_{k=-n+1}^{n-1} \tilde{R}^2_{k,n}/(n - |k|). \end{aligned} \tag{4.100}$$

We have to divide var $B^2_{n\alpha}$ by B^4_n and need the ratio tending to zero. We assume

$$S_n > \gamma'\|x(n)\|^2, \qquad \text{for some} \quad \gamma' > 0. \tag{4.101}$$

This is, for instance, the case under the assumptions of theorems 4.2 and 4.3 where the asymptotic normality of the LSE $b(n)$ for the regression parameters is proved.

Now with (4.95), (4.100), and (4.101),

$$\begin{aligned} B_n^{-4} \operatorname{var} B^2_{n,\alpha} &= \|x(n)\|^8 S_n^{-2} \operatorname{var} B^2_{n,\alpha} \\ &= O\left(\sum_{k=-n+1}^{n-1} \tilde{R}^2_{k,n}/(n - |k|) \right). \end{aligned} \tag{4.102}$$

If a regression sequence $\{x_t\}$ has the property that the last expression tends to zero, then $B_n^{-2} B^2_{n,\alpha}$ is a (strongly) consistent sequence of estimators of one. This is certainly true if

$$\sup_n \sum_{k=-n+1}^{n-1} \tilde{R}^4_{k,n} < \infty, \tag{4.103}$$

since because of the slowly increasing character of $\{x_t\}$ there exists a sequence of integers $m_n \to \infty$ such that

$$\sum_{k=m_n}^{n-1} \tilde{R}^2_{k,n}/(n - k) \to 0, \tag{4.104}$$

and also

$$\sum_{k=n-m_n}^{n} k^{-2} \to 0. \tag{4.105}$$

Now estimate the central part of

$$\sum_{k=-n+1}^{n-1} \tilde{R}^2_{k,n}/(n - |k|) \tag{4.106}$$

for $k = -m_n, \cdots, m_n$ by Schwarz's inequality. Because of (4.103) and (4.105), it tends to zero. The rest of the sum tends to zero by (4.104).

As pointed out before $B^2_{n,\alpha}$ is not yet an estimate of B^2_n, since $B^2_{n,\alpha}$ contains the error r.v.'s ϵ_t. We now proceed to replace them by the residuals

$$e_k(n) = y_k - x_k b(n) = \epsilon_k - \|x(n)\|^{-2} x_k x'(n) \epsilon(n). \tag{4.107}$$

Let

$$\hat{R}_{k,n} = \frac{1}{n - |k|} \sum_{h=1}^{n-|k|} e_h(n) e_{h+|k|}(n). \tag{4.108}$$

Putting

$$R'_{k,n} = \frac{(x'(n)\epsilon(n))^2}{(n - |k|)||x(n)||^4} \sum_{h=1}^{n-|k|} x_h x_{h+|k|} = \frac{(x'(n)\epsilon(n))^2}{(n - |k|)||x(n)||^2} \tilde{R}_{k,n}, \tag{4.109}$$

$$R''_{k,n} = \frac{x'(n)\epsilon(n)}{(n - |k|)||x(n)||^2} \sum_{h=1}^{n-|k|} (\epsilon_h x_{h+|k|} + \epsilon_{h+|k|} x_h), \tag{4.110}$$

we obtain

$$\hat{R}_{k,n} = R_{k,n} + R'_{k,n} - R''_{k,n}. \tag{4.111}$$

Consider now,

$$\hat{B}_n^2 = ||x(n)||^{-2} \sum_{k=-n+1}^{n-1} \hat{R}_{k,n}\tilde{R}_{k,n} = B_{n,\alpha}^2 + B_{n,\beta}^2 - B_{n,\gamma} \tag{4.112}$$

where

$$B_{n,\beta}^2 = ||x(n)||^{-2} \sum_{k=-n+1}^{n-1} R'_{k,n}\tilde{R}_{k,n} = ||x(n)||^{-4}(x'(n)\epsilon(n))^2 \sum_{k=-n+1}^{n-1} (n - |k|)^{-1}\tilde{R}_{k,n}^2, \tag{4.113}$$

$$B_{n,\gamma} = ||x(n)||^{-2} \sum_{k=-n+1}^{n-1} R''_{k,n}\tilde{R}_{k,n}. \tag{4.114}$$

Now

$$B_n^{-2}EB_{n,\beta}^2 = \sum_{k=-n+1}^{n-1} (n - |k|)^{-1}\tilde{R}_{k,n}^2 \rightarrow 0, \tag{4.115}$$

making use of our assumption concerning (4.102). Since $B_{n,\beta}^2 \geq 0$, this implies

$$B_n^{-2}B_{n,\beta}^2 \rightarrow 0, \qquad \text{i.p.} \tag{4.116}$$

Concerning $B_{n,\gamma}$, we proceed as follows. First,

$$R''_{k,n} = (n - |k|)^{-1}||x(n)||^{-2} \sum_{j,\ell=-\infty}^{\infty} A_{nj}(A_{n-|k|,\ell-|k|} + A_{n\ell} - A_{|k|\ell})\eta_j\eta_\ell. \tag{4.117}$$

After some straightforward computations, we obtain

$$\mathrm{E}(\sum_{j,\ell} A_{n,j}A_{n-|k|,1-|k|}\eta_j\eta_\ell)^2 < \text{const } S_n S_{n-|k|}, \tag{4.118}$$

and similar relations hold for the other terms of $E(R''_{k,n})^2$. Hence,

$$\begin{aligned} EB_{n,\gamma}^2 &< \text{const } ||x(n)||^{-8}S_n \sum_{|k|<n} \tilde{R}_{k,n}^2 \frac{1}{n - |k|} \sum_{k=1}^{n} (S_k + S_n - S_{n-k})k^{-1}, \\ B_n^{-4}EB_{n,\gamma}^2 &< \text{const} \sum_{|k|<n} R_{k,n}^2 \frac{1}{n - |k|} ||x(n)||^{-2} \sum_{k=1}^{n} k^{-1}(||x(k)||^2 + ||x(n)||^2 - ||x(n - k)||^2). \end{aligned} \tag{4.119}$$

This tends to zero, and hence $B_n^{-2}B_{n,\gamma} \rightarrow 0$, i.p., if we assume

$$(4.120)\qquad \|x(n)\|^{-2} \max_{t=1,\cdots,n} x_t^2 = O(n),$$

besides what we have assumed previously. In summary we have the following theorem.

THEOREM 4.7. *The d.f.'s of the statistics* $(\hat{B}_n^2)^{-1}(b(n) - \beta)$ *tend to* $N(0, 1)$ *if* $\{\epsilon_t\}$ *is a strictly stationary linear process, if* ess $\inf_\lambda |c(\lambda)| > 0$ *and* (4.98) *holds, and if the regression sequence satisfies* (4.120) *and*

$$(4.121)\qquad \sum_{k=-n+1}^{n-1} \tilde{R}_{k,n}^2 (n - |k|)^{-1} \to 0.$$

Condition (4.120) is satisfied, for example, for polynomial regression sequences $x_t = t^c$, $x > -\frac{1}{2}$. However, (4.121) is not satisfied at least for some polynomial sequences. A slightly weaker assumption than (4.120) that is sufficient is

$$(4.122)\qquad \|x(n)\|^{-2} \sum_{t=1}^{n} x_t^2 \ln (n/t) = O(1).$$

In this subsection it was not our aim to achieve the utmost in generality, we rather wanted to indicate one possibility of replacing the normalizing constant B_n^2 by a statistic. The assumptions on the regression sequence, in particular (4.121), can be weakened considerably, if stronger assumptions are imposed on the admissible error sequences $\{\epsilon_t\}$, such as

$$(4.123)\qquad \sum_{k=-\infty}^{\infty} |R_k| < \infty,$$

and if instead of $\hat{B}_n^2$, a different estimator is used, for instance,

$$(4.124)\qquad \hat{\hat{B}}_n^2 = \|x(n)\|^{-2} \sum_{|k| < \sqrt{n}} R_{k,n}\tilde{R}_{k,n}.$$

Clearly, (4.123) is satisfied, for example, for finite c-sequences.

REFERENCES

[1] F. EICKER, "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *Ann. Math. Statist.*, Vol. 34 (1963), pp. 447–456.

[2] ———, "Uber die Konsistenz von Parameterschätzfunktionen für ein gemischtes Zeitreihen-Regressionsmodell," *Z. Wahrscheinlichkeitstheorie*, Vol. 1 (1963), pp. 456–477.

[3] ———, "Ein Zentraler Grenzwertsatz für Summen von Variablen einer verallgemeinerten linearen Zufallsfolge," *Z. Wahrscheinlichkeitstheorie*, Vol. 3 (1965), pp. 317–327.

[4] ———, "A central limit theorem for linear vector forms with random coefficients," to be published.

[5] B. V. GNEDENKO and A. N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables*, Cambridge, Addison-Wesley, 1954.

[6] A. J. GOLDMAN and M. ZELEN, "Weak generalized inverses and minimum variance linear unbiased estimation," *J. of Research*, National Bureau of Standards, Washington, D.C., Vol. 68B (1964), pp. 151–172.

[7] U. GRENANDER and M. ROSENBLATT, *Statistical Analysis of Stationary Time Series*, New York, Wiley, 1957.

[8] E. J. HANNAN, *Time Series Analysis*, London, Methuen, 1960.
[9] ———, "A central limit theorem for systems of regressions," *Proc. Cambridge Philos. Soc.*, Vol. 57 (1961), pp. 583–588.
[10] Z. A. LOMNICKI and S. K. ZAREMBA, "On estimating the spectral density function of a stochastic process," *J. Roy. Statist. Soc. Ser. B*, Vol. 18 (1956), pp. 13–37.
[11] M. ROSENBLATT, "Statistical analysis of stochastic processes with stationary residuals," *Probability and Statistics* (Cramér volume), edited by U. Grenander, New York, Wiley, 1959.