View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

JOURNAL OF MULTIVARIATE ANALYSIS 9, 343–361 (1979)

Strong Consistency of Least Squares Estimates in Multiple Regression II*

T. L. LAI, HERBERT ROBBINS, AND C. Z. WEI

Columbia University

Communicated by E. J. Hannan

The strong consistency of least squares estimates in multiple regression models is established under minimal assumptions on the design and weak dependence and moment restrictions on the errors.

1. INTRODUCTION AND SUMMARY

In this paper we establish the strong consistency of the least squares estimates for the parameters β_j of the multiple regression model

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \qquad (i = 1, 2,...) \tag{1.1}$$

under minimal assumptions on the design constants x_{ij} and very weak conditions on the random variables ϵ_i. Specifically, we shall assume that

$$\sum_1^\infty c_i \epsilon_i \text{ converges a.s. for all real sequences } \{c_i\} \text{ such that } \sum_1^\infty c_i^2 < \infty. \tag{1.2}$$

In particular, if the ϵ_i are i.i.d. with $E\epsilon_1 = 0$ and $E\epsilon_1^2 < \infty$, then (1.2) holds. More generally, by the martingale convergence theorem, (1.2) holds if $\{\epsilon_i\}$ is an L^2-bounded martingale difference sequence; i.e.,

$$E(\epsilon_{i+1} \mid \epsilon_1, ..., \epsilon_i) = 0 \qquad \text{for all } i \geqslant 1 \text{ and } \sup_i E\epsilon_i^2 < \infty. \tag{1.3}$$

The condition (1.2) also includes a large class of other important dependence structures for $\{\epsilon_i\}$, as will be shown in Section 4. It is interesting to note that

Received February 22, 1979.

AMS 1970 subject classifications: Primary 62J05; Secondary 60F15.

Key words and phrases: Multiple regression, strong consistency, martingale difference sequence, weakly multiplicative sequence, generalized inverse.

* Research supported by the National Science Foundation, the National Institute of Medical Sciences, and the Office of Naval Research.

0047-259X/79/030343-19$02.00/0

Copyright © 1979 by Academic Press, Inc.
All rights of reproduction in any form reserved.

even when the ϵ_i are nonrandom constants such that $\sum_1^\infty \epsilon_i^2 < \infty$, (1.2) still holds by the Schwarz inequality.

Throughout the sequel we shall let X_n denote the design matrix $(x_{ij})_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}$, and let $Y_n = (y_1, ..., y_n)'$ and $\beta = (\beta_1, ..., \beta_p)'$, where a prime denotes transpose. For $n \geqslant p$, the leastsquares estimate $b_n = (b_{n1}, ..., b_{np})'$ of the vector β based on the design matrix X_n and the response vector Y_n is given by

$$b_n = (X_n'X_n)^{-1}X_n'Y_n \tag{1.4}$$

provided that $X_n'X_n$ is nonsingular. Assuming $X_n'X_n$ to be nonsingular and the ϵ_i to be uncorrelated random variables with zero means and common variance σ^2, the Gauss–Markov theorem says that b_n is the best linear unbiased estimate of β, with $\mathrm{Cov}(b_n) = \sigma^2(X_n'X_n)^{-1}$. Therefore, for b_n to converge as $n \to \infty$ to β in quadratic mean, and hence in probability, it is sufficient that

$$(X_n'X_n)^{-1} \to 0 \qquad \text{as} \quad n \to \infty. \tag{1.5}$$

For the Gauss–Markov model (with $\sigma > 0$), the condition (1.5) is also necessary for b_n to converge to β in probability, as has been shown by Drygas [3].

The question whether (1.5) implies that b_n converges a.s. to β, however, is much harder, even under the assumption that the ϵ_i are i.i.d. with $E\epsilon_1 = 0$ and $E\epsilon_1^2 < \infty$. When the ϵ_i are i.i.d. $N(0, \sigma^2)$, Anderson and Taylor [1] have shown that (1.5) indeed implies the strong consistency of b_n. Without the assumption of normality, they have also shown in [2] that b_n converges a.s. to β under the assumption that the ϵ_i are i.i.d. generalized Gaussian random variables and that

$$\mathrm{tr}[(X_n'X_n)^{-1}] = o(1/\log n) \qquad \text{as } n \to \infty. \tag{1.6}$$

The latter assumption on the design is much stronger than (1.5), and the generalized Gaussian condition ($E \exp(\lambda\epsilon_1) \leqslant \exp(A\lambda^2)$ for some $A > 0$ and all real λ) is rather restrictive. Earlier, Drygas [3] considered independently distributed errors ϵ_i that satisfy (1.3) and obtained the strong consistency of b_n under the alternative assumption that

$$\sum_{n=1}^{\infty} x_{nj}^2 = \infty \qquad \text{for} \quad j = 1, ..., p,$$

and

$$\left\| (X_n'X_n)^{-1} \operatorname{diag}\left\{\sum_{i=1}^n x_{i1}^2, ..., \sum_{i=1}^n x_{ip}^2\right\}\right\| = O(1), \tag{1.7}$$

where $\| A \|$ denotes the maximum of the entries of the matrix A, and $\operatorname{diag}\{a_1, ..., a_p\}$ denotes the diagonal matrix with $a_1, ..., a_p$ as its successive diagonal elements. Although the condition (1.7) reduces to (1.5) when $p = 1$, it is much stronger than (1.5) when $p > 1$. In [6], Lai and Robbins considered the simple linear model $y_i = \beta_1 + \beta_2 t_i + \epsilon_i$, where the ϵ_i are i.i.d. with $E\epsilon_1 = 0$ and $E[\epsilon_1^2(\log^+ |\epsilon_1|)^r] < \infty$ for some $r > 1$, and established the strong consistency of the slope estimate b_{n2} under the minimal assumption

$$\sum_1^n (t_i - \bar{t}_n)^2 \to \infty \qquad \left(\bar{t}_n = n^{-1} \sum_1^n t_i\right) \tag{1.8}$$

on the design. Their method, based on an embedding technique to reduce the problem to the normal case, does not extend to the general multiple regression model (1.1), and therefore an alternative approach is needed for general p.

We have recently announced in [7], without giving details of the proof, that for the multiple regression model (1.1) the least squares estimate b_n indeed converges a.s. to β, under the assumption that (1.3) holds and that the design satisfies the minimal condition (1.5). The main purpose of the present paper is to give a complete proof of this theorem and to show that the method actually extends to the much more general case where the ϵ_i satisfy (1.2). Specifically, we shall prove the following.

THEOREM 1. *Suppose that in* (1.1), $\{x_{ij}\}$ $(i = 1, 2,\ldots; j = 1,\ldots, p)$ *is a double array of constants and* $\epsilon_1, \epsilon_2,\ldots,$ *are random variables satisfying* (1.2). *Assume that* $X_m'X_m$ *is nonsingular for some* m *(so that* $X_n'X_n$ *is nonsingular for all* $n \geqslant m$*). For* $n \geqslant m$, *let* $b_n = (b_{n1},\ldots, b_{np})'$ *be the least squares estimate defined by* (1.4), *and let*

$$V_n = (v_{ij}^{(n)})_{1\leqslant i,j\leqslant p} = (X_n'X_n)^{-1}. \tag{1.9}$$

Fix $j = 1,\ldots, p$. *If* $\lim_{n\to\infty} v_{jj}^{(n)} = 0$, *then for every* $\delta > 0$, *with probability* 1,

$$b_{nj} - \beta_j = o(\{v_{jj}^{(n)} |\log v_{jj}^{(n)}|^{1+\delta}\}^{1/2}) \qquad \text{as} \quad n \to \infty. \tag{1.10}$$

In Section 2 we shall establish some lemmas. Using these lemmas, we shall prove Theorem 1 in Section 3, where we shall also prove the following closely related result which is itself of independent interest.

THEOREM 2. *Let* $\{\epsilon_i\}$ *be a sequence of random variables satisfying* (1.2). *Let* k *be a positive integer. For each* $n \geqslant 1$, *let* T_n *be a* k*-dimensional vector of constants and let* $H_n = \sum_1^n T_i T_i'$. *Assume that* H_m *is positive definite for some* m *(so that* H_n *is positive definite for all* $n \geqslant m$*). Let* $\{c_n\}$ *be a sequence of constants such that*

$$\sum_{m+1}^{\infty} c_i^2(1 + T_i'H_{i-1}^{-1}T_i) < \infty. \tag{1.11}$$

Then with probability 1,

$$\sum_{i=m+1}^{n} c_i T_i' H_{i-1}^{-1} \left(\sum_{j=1}^{i-1} T_j \epsilon_j \right) \quad \text{converges as } n \to \infty. \tag{1.12}$$

Theorem 2, which we announced without proof in [7] for the case where (1.3) holds, plays a key role in the proof of Theorem 1. Putting $k = 1$ and $T_i = 1$ in Theorem 2, we obtain

COROLLARY 1. *Let* $\{\epsilon_i\}$ *be a sequence of random variables satisfying* (1.2), *and let* $\{c_i\}$ *be a sequence of constants such that* $\sum_1^\infty c_i^2 < \infty$. *Then* $\sum_2^\infty c_i \bar{\epsilon}_{i-1}$ *converges a.s., where* $\bar{\epsilon}_n = n^{-1} \sum_1^n \epsilon_i$.

Thus Theorem 2 can be regarded as a multivariate generalization of this interesting result. Some other corollaries of Theorems 1 and 2 will be given in Section 4.

2. BASIC LEMMAS

An important tool in proving the strong consistency of b_n for error structures satisfying (1.2) is the following.

LEMMA 1. *Let* $\{a_n\}$, $\{c_n\}$, $n \geqslant m$, *be two sequences of real numbers such that* $a_m \neq 0$ *and*

$$\sum_{n=m}^{\infty} (c_n^2 A_{n+1}/A_n) < \infty, \tag{2.1}$$

where $A_n = A + \sum_{i=m}^n a_i^2$ *and* $A \geqslant 0$. *Then*

$$\sum_{n=m}^{\infty} | c_n a_{n+1}/A_n | < \infty, \tag{2.2}$$

$$\sum_{n=m}^{\infty} \left(\sum_{i=n}^{\infty} c_i a_{i+1}/A_i \right)^2 a_n^2 < \infty, \tag{2.3}$$

and

$$\sum_{n=m}^{\infty} (c_n^2 A_{n+1}/A_n) = \sum_{n=m}^{\infty} \left\{ \left(\sum_{i=n}^{\infty} c_i a_{i+1}/A_i \right) a_n - c_{n-1} \right\}^2 + A \left\{ \sum_{i=m}^{\infty} c_i a_{i+1}/A_i \right\}^2, \tag{2.4}$$

where $c_{m-1} = 0$.

Proof. We first establish (2.4) by showing that both sides of the identity are equal to the square of the l^2-norm of the same square-summable sequence. For $\mathbf{u} = \{u_n\}_{n \geqslant m-1} \in l^2$, $\mathbf{v} = \{v_n\}_{n \geqslant m-1} \in l^2$, we shall write $\mathbf{u} \perp \mathbf{v}$ to denote that they are orthogonal, i.e., $\sum_{n=m-1}^{\infty} u_n v_n = 0$. We let $\| \mathbf{u} \|$ denote the l^2-norm $(\sum_{n=m-1}^{\infty} u_n^2)^{1/2}$. For $i \geqslant m$, define $\mathbf{u}(i) = \{u_n(i)\}_{n \geqslant m-1}$ by

$$\begin{aligned} u_n(i) &= A^{1/2} a_{i+1}/A_i && \text{if} \quad n = m-1, \\ &= a_n a_{i+1}/A_i && \text{if} \quad m \leqslant n \leqslant i, \\ &= -1 && \text{if} \quad n = i+1, \\ &= 0 && \text{if} \quad n \geqslant i+2. \end{aligned}$$

Then $\| c_i \mathbf{u}(i) \|^2 = c_i^2 A_{i+1}/A_i$ and $\mathbf{u}(i) \perp \mathbf{u}(j)$ if $i \neq j$. Hence in view of (2.1),

$$\sum_{i=m}^{\infty} c_i \mathbf{u}(i) \in l^2 \qquad \text{and} \qquad \left\| \sum_{i=m}^{\infty} c_i \mathbf{u}(i) \right\|^2 = \sum_{i=m}^{\infty} c_i^2 A_{i+1}/A_i \,.$$

Since $\sum_{i=m}^{\infty} c_i u_{m-1}(i) = A^{1/2} \sum_{i=m}^{\infty} c_i a_{i+1}/A_i$, and since for fixed $n \geqslant m$,

$$\sum_{i=m}^{\infty} c_i u_n(i) = \left(\sum_{i=n}^{\infty} c_i a_{i+1}/A_i \right) a_n - c_{n-1} \,,$$

we obtain (2.4). Replacing c_i by $| c_i |$ in (2.4), (2.2) then follows. Moreover, using (2.4) and the inequality $x^2 \leqslant 2(x-y)^2 + 2y^2$, we obtain that

$$\begin{aligned} \sum_{n=m}^{\infty} \left(\sum_{i=n}^{\infty} c_i a_{i+1}/A_i \right)^2 a_n^2 &\leqslant 2 \sum_{n=m}^{\infty} c_{n-1}^2 + 2 \sum_{n=m}^{\infty} (c_n^2 A_{n+1}/A_n) \\ &= 2 \sum_{n=m}^{\infty} c_n^2 (1 + A_{n+1}/A_n) < \infty, \qquad \text{by (2.1)}. \end{aligned}$$

Hence (2.3) holds. ∎

Remark. In the special case $A = 0$ and $a_n \equiv 1$, (2.4) becomes

$$\sum_{n=m}^{\infty} c_n^2 (n+1)/n = \sum_{n=m}^{\infty} \left\{ \left(\sum_{i=n}^{\infty} c_i/i \right) - c_{n-1} \right\}^2 \qquad (c_{m-1} = 0).$$

Letting $B_n = \sum_{i=n}^{\infty} c_i/i$, we obtain from the above identity that

$$\begin{aligned} \sum_{n=m}^{\infty} B_n^2 &= \sum_{n=m}^{\infty} c_n^2/n + 2 \sum_{n=m}^{\infty} c_{n-1} B_n \leqslant 2 \sum_{n=m}^{\infty} c_n \{ (c_n/n) + B_{n+1} \} \\ &= 2 \sum_{n=m}^{\infty} c_n B_n \leqslant 2 \left(\sum_{n=m}^{\infty} c_n^2 \right)^{1/2} \left(\sum_{n=m}^{\infty} B_n^2 \right)^{1/2}. \end{aligned}$$

This therefore implies that $\sum_{n=m}^{\infty}(\sum_{i=n}^{\infty} c_i/i)^2 \leqslant 4\sum_{n=m}^{\infty} c_n^2$, which is the Copson–Hardy inequality (cf. [4, p. 246]). Thus Lemma 1 also gives a new proof of this classical result.

As an application of Lemma 1, we obtain the following.

LEMMA 2. *Let $\{a_n\}$, $\{c_n\}$, $n \geqslant 1$, be two sequences of constants such that $a_m \neq 0$ and*

$$\sum_{n=m+1}^{\infty} (c_n^2 A_n/A_{n-1}) < \infty, \qquad \text{where} \quad A_n = \sum_{1}^{n} a_i^2 \text{ for } n \geqslant m. \tag{2.5}$$

Let $\{\tilde{a}_n\}$ be a sequence of constants such that for some $C > 0$,

$$|\tilde{a}_n| \leqslant C|a_n| \qquad \text{for all } n \geqslant 1. \tag{2.6}$$

If $\{\epsilon_n\}$ is a sequence of random variables satisfying (1.2), then

$$\sum_{i=m+1}^{n} c_i a_i A_{i-1}^{-1}\left(\sum_{j=1}^{i-1} \tilde{a}_j \epsilon_j\right) \qquad \text{converges a.s. (as } n \to \infty).$$

Proof. In view of (2.5) and (2.6), Lemma 1 implies that

$$\sum_{m+1}^{\infty} |c_n a_n/A_{n-1}| < \infty \qquad \text{and} \qquad \sum_{n=m}^{\infty}\left(\sum_{i=n+1}^{\infty} c_i a_i/A_{i-1}\right)^2 \tilde{a}_n^2 < \infty. \tag{2.7}$$

For $n \geqslant m$, let $p_n = \sum_{n+1}^{\infty}(c_i a_i/A_{i-1})$. Then

$$\begin{aligned}
&\sum_{i=m+1}^{n} c_i a_i A_{i-1}^{i1}\left(\sum_{j=1}^{i-1} \tilde{a}_j \epsilon_j\right) \\
&\quad = \left(\sum_{j=1}^{m} \tilde{a}_j \epsilon_j\right)\left(\sum_{i=m+1}^{n} c_i a_i A_{i-1}^{-1}\right) + \sum_{j=m+1}^{n-1} \tilde{a}_j \epsilon_j \left(\sum_{i=j+1}^{n} c_i a_i A_{i-1}^{-1}\right) \\
&\quad = \left(\sum_{j=1}^{m} \tilde{a}_j \epsilon_j\right)(p_m - p_n) + \sum_{j=m+1}^{n-1} p_j \tilde{a}_j \epsilon_j - p_n \sum_{j=m+1}^{n-1} \tilde{a}_j \epsilon_j .
\end{aligned} \tag{2.8}$$

Since $\sum_{m}^{\infty} p_n^2 \tilde{a}_n^2 < \infty$ by (2.7), the condition (1.2) implies that

$$\sum_{m+1}^{\infty} p_j \tilde{a}_j \epsilon_j \qquad \text{converges a.s.} \tag{2.9}$$

Clearly, (2.9) still holds if we replace p_j by $p_j^* = \sum_{j+1}^{\infty} |c_i a_i|/A_{i-1}$. Since $p_j^* \downarrow 0$, it then follows from the Kronecker lemma that $p_n^* \sum_{m+1}^{n} \tilde{a}_j \epsilon_j \to 0$ a.s.

Therefore as $n \to \infty$,

$$\left| p_n \sum_{m+1}^{n-1} \tilde{a}_j \epsilon_j \right| \leqslant p_{n-1}^* \left| \sum_{m+1}^{n-1} \tilde{a}_j \epsilon_j \right| \to 0 \qquad \text{a.s.} \tag{2.10}$$

From (2.8), (2.9), and (2.10), the desired conclusion follows. ∎

To prove Theorem 1, it suffices to consider only b_{n1} (i.e., $j = 1$ in (1.10)). For $p \geqslant 2$, defining the $(p-1)$-dimensional vector

$$T_n = (x_{n2}, \ldots, x_{np})' \tag{2.11}$$

and partitioning the matrix $X_n'X_n$ as

$$X_n'X_n = \begin{pmatrix} \sum_{i=1}^{n} x_{i1}^2 & K_n \\ K_n' & H_n \end{pmatrix} \tag{2.12}$$

so that H_n is a $(p-1) \times (p-1)$ matrix, we have the following representation of b_{n1}.

LEMMA 3. *Let $p \geqslant 2$. Assume that $X_n'X_n$ is positive definite for $n \geqslant m (\geqslant p)$. Define H_n, K_n, and T_n by* (2.11) *and* (2.12). *Then for $n \geqslant m$,*

$$b_{n1} = \beta_1 + \frac{\sum_{i=1}^{n} (x_{i1} - K_n H_n^{-1} T_i)\, \epsilon_i}{\sum_{i=1}^{n} (x_{i1} - K_n H_n^{-1} T_i)^2}.$$

Define for $n \geqslant m$

$$u_n = \sum_{i=1}^{n} (x_{i1} - K_n H_n^{-1} T_i)\, \epsilon_i, \qquad w_n = u_n - u_{n-1},$$

$$d_n = x_{n1} - K_n H_n^{-1} T_n. \tag{2.13}$$

Then for $n > m$,

$$\sum_{i=1}^{n} (x_{i1} - K_n H_n^{-1} T_i)^2 = \sum_{i=1}^{n-1} (x_{i1} - K_{n-1} H_{n-1}^{-1} T_i)^2 + d_n^2 (1 + T_n' H_{n-1}^{-1} T_n), \tag{2.14}$$

$$w_n = d_n \left\{ \epsilon_n - T_n' H_{n-1}^{-1} \left(\sum_{i=1}^{n-1} T_i \epsilon_i \right) \right\}. \tag{2.15}$$

Moreover, if the ϵ_i are uncorrelated and have zero mean and the same variance σ^2 $(0 \leqslant \sigma < \infty)$, then

$$E(w_l w_n) = 0 = E(u_m w_n) \qquad \text{for} \quad l > n > m. \tag{2.16}$$

The above lemma is due to Anderson and Taylor [1]. Note that if the ϵ_i are uncorrelated with zero mean and the same variance σ^2, then for $n > m$,

$$Ew_n = 0, \qquad Ew_n^2 = d_n^2(1 + T_n' H_{n-1}^{-1} T_n)\,\sigma^2. \tag{2.17}$$

Moreover, in view of (2.16) and the fact that $u_n = u_{n-1} + w_n$,

$$Eu_n^2 = Eu_{n-1}^2 + Ew_n^2 \qquad \text{for } n > m. \tag{2.18}$$

On the other hand, by (2.13),

$$Eu_n^2 = \sum_{i=1}^{n} (x_{i1} - K_n H_n^{-1} T_i)^2\,\sigma^2 \qquad \text{for } n > m. \tag{2.19}$$

Thus the matrix identity (2.14) has a probabilistic interpretation through the relations (2.17)–(2.19). Combining (2.13) and (2.15), we also obtain the following useful identity:

$$\sum_{i=m+1}^{n} (x_{i1} - K_n H_n^{-1} T_i)\,\epsilon_i$$
$$= \sum_{i=m+1}^{n} (x_{i1} - K_i H_i^{-1} T_i) \left\{ \epsilon_i - T_i' H_{i-1}^{-1} \left(\sum_{l=1}^{i-1} T_l \epsilon_l \right) \right\}. \tag{2.20}$$

For convenience in reference, we now restate the identities (2.14) and (2.20) in parts (i) and (ii) of the following lemma, which also contains some other useful matrix identities related to (1.12).

LEMMA 4. *Let $k \geqslant 2$. For each $i \geqslant 1$, let $T_i = (t_{i1}, \ldots, t_{ik})'$ be a k-dimensional vector and let $H_n = \sum_{i=1}^{n} T_i T_i'$. Partition the matrix H_n as*

$$H_n = \begin{pmatrix} \sum_{i=1}^{n} t_{i1}^2 & P_n \\ P_n' & Q_n \end{pmatrix} \tag{2.21}$$

so that Q_n is a $(k-1) \times (k-1)$ matrix. Let $\hat{T}_i = (t_{i2}, \ldots, t_{ik})'$. Assume that H_n is positive definite for all $n \geqslant m$ ($\geqslant k$). Let $\{e_n\}$ be a sequence of real numbers. Define for $n \geqslant m$

$$z_n = \sum_{i=1}^{n} (t_{i1} - P_n Q_n^{-1} \hat{T}_i)\, e_i,$$

$$s_n = \sum_{i=1}^{n} (t_{i1} - P_n Q_n^{-1} \hat{T}_i)^2.$$

Then for $n > m$,

(i) $s_n = s_{n-1} + (t_{n1} - P_n Q_n^{-1} \hat{T}_n)^2 (1 + \hat{T}'_n Q_{n-1}^{-1} \hat{T}_n)$;

(ii) $z_n = z_m + \sum_{i=m+1}^{n} (t_{i1} - P_i Q_i^{-1} \hat{T}_i) \left\{ e_i - \hat{T}'_i Q_{i-1}^{-1} \left(\sum_{j=1}^{i-1} \hat{T}_j e_j \right) \right\}$;

(iii) $Q_n = \sum_{i=1}^{n} \hat{T}_i \hat{T}'_i , \qquad P_n = \sum_{i=1}^{n} t_{i1} \hat{T}'_i$;

(iv) $$T'_n H_{n-1}^{-1} \left(\sum_{i=1}^{n-1} T_i e_i \right) = \hat{T}'_n Q_{n-1}^{-1} \left(\sum_{i=1}^{n-1} \hat{T}_i e_i \right) + (t_{n1} - P_{n-1} Q_{n-1}^{-1} \hat{T}_n)(z_{n-1}/s_{n-1});$$

(v) $t_{n1} - P_{n-1} Q_{n-1}^{-1} \hat{T}_n = (t_{n1} - P_n Q_n^{-1} \hat{T}_n)(1 + \hat{T}'_n Q_{n-1}^{-1} \hat{T}_n)$;

(vi) $1 + T'_n H_{n-1}^{-1} T_n = (s_n/s_{n-1})(1 + \hat{T}'_n Q_{n-1}^{-1} \hat{T}_n)$.

Proof. (iii) is obvious. To prove (iv), partition the matrix $A = H_{n-1}^{-1}$ as

$$H_{n-1}^{-1} = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

so that A_{22} is a $(k - 1) \times (k - 1)$ matrix. We shall make use of the following identities:

$$Q_{n-1}^{-1} = A_{22} - a_{11}^{-1} A_{21} A_{12} , \tag{2.22}$$

$$a_{11}^{-1} A_{12} = a_{11}^{-1} A'_{21} = -P_{n-1} Q_{n-1}^{-1} , \tag{2.23}$$

$$a_{11}^{-1} = \sum_{i=1}^{n-1} t_{i1}^2 - P_{n-1} Q_{n-1}^{-1} P'_{n-1} = s_{n-1} . \tag{2.24}$$

The last relation in (2.24) follows from (iii). Since $T'_i = (t_{i1} , \hat{T}'_i)$, we obtain that for $n > m$

$$T'_n H_{n-1}^{-1} \left(\sum_{i=1}^{n-1} T_i e_i \right)$$

$$= t_{n1} \left\{ a_{11} \sum_{1}^{n-1} t_{i1} e_i + A_{12} \sum_{1}^{n-1} \hat{T}_i e_i \right\} + \hat{T}'_n \left\{ A_{21} \sum_{1}^{n-1} t_{i1} e_i + A_{22} \sum_{1}^{n-1} \hat{T}_i e_i \right\}$$

$$= (t_{n1} + a_{11}^{-1} \hat{T}'_n A_{21}) \left(a_{11} \sum_{1}^{n-1} t_{i1} e_i + A_{12} \sum_{1}^{n-1} \hat{T}_i e_i \right)$$

$$+ \hat{T}'_n (A_{22} - a_{11}^{-1} A_{21} A_{12}) \sum^{n-1} \hat{T}_i e_i$$

$$= (t_{n1} - P_{n-1}Q_{n-1}^{-1}\hat{T}_n) \left\{ a_{11} \sum_1^{n-1} \left(t_{i1} - P_{n-1}Q_{n-1}^{-1}\hat{T}_i \right) e_i \right\}$$

$$+ \hat{T}_n' Q_{n-1}^{-1} \left(\sum_1^{n-1} \hat{T}_i e_i \right) \qquad \text{by (2.22) and (2.23)}$$

$$= (t_{n1} - P_{n-1}Q_{n-1}^{-1}\hat{T}_n)(z_{n-1}/s_{n-1}) + \hat{T}_n' Q_{n-1}^{-1} \left(\sum_1^{n-1} \hat{T}_i e_i \right) \qquad \text{by (2.24).}$$

To prove (v) and (vi), define for $n > m$

$$f_n = t_{n1} - P_n Q_n^{-1} \hat{T}_n, \qquad g_n = t_{n1} - P_{n-1} Q_{n-1}^{-1} \hat{T}_n. \tag{2.25}$$

Then

$$\begin{aligned} g_n - f_n &= (P_n Q_n^{-1} - P_{n-1}Q_{n-1}^{-1})\, \hat{T}_n \\ &= \{(P_n - P_{n-1}) - P_n Q_n^{-1}(Q_n - Q_{n-1})\} Q_{n-1}^{-1} \hat{T}_n \\ &= \{t_{n1}\hat{T}_n' - P_n Q_n^{-1} \hat{T}_n \hat{T}_n'\} Q_{n-1}^{-1} \hat{T}_n \qquad \text{by (iii)} \\ &= f_n \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n. \end{aligned}$$

Hence (v) follows.

For $1 \leqslant i, j \leqslant n-1$, let

$$e_{ij} = 1 \quad \text{if} \quad i = j \qquad \text{and} \qquad e_{ij} = 0 \quad \text{if} \quad i \neq j. \tag{2.26}$$

It follows from (iv) that for $j = 1, \ldots, n-1$,

$$T_n' H_{n-1}^{-1} \left(\sum_{i=1}^{n-1} T_i e_{ij} \right) = \hat{T}_n' Q_{n-1}^{-1} \left(\sum_{i=1}^{n-1} \hat{T}_i e_{ij} \right) + (g_n/s_{n-1}) \sum_{i=1}^{n-1} (t_{i1} - P_{n-1}Q_{n-1}^{-1}\hat{T}_i)\, e_{ij}. \tag{2.27}$$

We note that

$$\sum_1^{n-1} (t_{i1} - P_{n-1}Q_{n-1}^{-1}\hat{T}_i)\, \hat{T}_i = 0 \tag{2.28}$$

(cf. [1, Eq. (13)]). From (2.26), (2.27), and (2.28), it then follows that

$$\begin{aligned} T_n' H_{n-1}^{-1} T_n &= T_n' H_{n-1}^{-1} \left(\sum_1^{n-1} T_i T_i' \right) H_{n-1}^{-1} T_n \\ &= \sum_{j=1}^{n} \left\{ T_n' H_{n-1}^{-1} \left(\sum_{i=1}^{n-1} T_i e_{ij} \right) \right\}^2 \end{aligned}$$

$$= \sum_{j=1}^{n} \left\{ \hat{T}_n' Q_{n-1}^{-1} \left(\sum_{i=1}^{n-1} \hat{T}_i e_{ij} \right) \right\}^2 + (g_n/s_{n-1})^2 \sum_{i=1}^{n-1} (t_{i1} - P_{n-1} Q_{n-1}^{-1} \hat{T}_i)^2$$

$$= \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n + g_n^2/s_{n-1} \qquad \text{by the definition of } s_{n-1}$$

$$= \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n + (1 + \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n)^2 f_n^2/s_{n-1} \qquad \text{by (v).}$$

Hence

$$1 + T_n' H_{n-1}^{-1} T_n = (1 + \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n)\{1 + (1 + \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n)(f_n^2/s_{n-1})\}$$

$$= (1 + \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n)(s_n/s_{n-1}) \qquad \text{by (i).} \quad \blacksquare$$

3. Proof of Theorems 1 and 2

In this section we shall first prove Theorem 2 by induction on k and then use Theorem 2 to prove Theorem 1.

Proof of Theorem 2. For the case $k = 1$, T_n and H_n are scalars and $H_n = \sum_1^n T_i^2$; moreover, the condition (1.11) becomes

$$\sum_{m+1}^{\infty} c_i^2 (1 + T_i^2/H_{i-1}) < \infty.$$

Hence $\sum_{i=m+1}^{\infty} c_i T_i H_{i-1}^{-1} (\sum_{j=1}^{i-1} T_j \epsilon_j)$ converges a.s. by Lemma 2.

Assume that the theorem holds for dimension $k - 1$ ($\geqslant 1$). We shall now show that it holds for dimension k. With the same notation as in Lemma 4, set $e_j = \epsilon_j$, and define f_n, g_n as in (2.25). Then by Lemma 4(iv),

$$\sum_{i=m+1}^{n} c_i T_i' H_{i-1}^{-1} \left(\sum_{j=1}^{i-1} T_j \epsilon_j \right)$$

$$= \sum_{i=m+1}^{n} (c_i g_i z_{i-1}/s_{i-1}) + \sum_{i=m+1}^{n} c_i \hat{T}_i' Q_{i-1}^{-1} \left(\sum_{j=1}^{i-1} \hat{T}_j \epsilon_j \right). \tag{3.1}$$

Note that $\hat{T}_n$ is a $(k - 1)$-dimensional vector and that Q_n is a $(k - 1) \times (k - 1)$ matrix. By Lemma 4(vi) and (1.11),

$$\sum_{m+1}^{\infty} c_i^2 (1 + \hat{T}_i' Q_{i-1}^{-1} \hat{T}_i) \leqslant \sum_{m+1}^{\infty} c_i^2 (1 + T_i' H_{i-1}^{-1} T_i) < \infty.$$

Hence by the induction hypothesis,

$$\sum_{i=m+1}^{\infty} c_i \hat{T}_i' Q_{i-1}^{-1} \left(\sum_{j=1}^{i-1} \hat{T}_j \epsilon_j \right) \qquad \text{converges a.s.}$$

Therefore, in view of (3.1), it remains to show that

$$\sum_{i=m+1}^{\infty} c_i g_i z_{i-1}/s_{i-1} \qquad \text{converges a.s.} \tag{3.2}$$

By Lemma 4(ii), for $n > m$, $z_n = z_m + \sum_{m+1}^{n} f_j(\epsilon_j - \xi_j)$, where

$$\xi_i = \hat{T}_i' Q_{i-1}^{-1} \left(\sum_{j=1}^{i-1} \hat{T}_j \epsilon_j \right), \qquad i > m. \tag{3.3}$$

Therefore,

$$\begin{aligned}
&\sum_{i=m+1}^{n} c_i g_i z_{i-1}/s_{i-1} \\
&\quad = \sum_{i=m+1}^{n} (c_i g_i/s_{i-1}) \left\{ z_m + \sum_{j=m+1}^{i-1} f_j \epsilon_j - \sum_{j=m+1}^{i-1} f_j \xi_j \right\} \\
&\quad = \left(\sum_{i=m+1}^{n} c_i g_i/s_{i-1} \right) z_m + \sum_{i=m+1}^{n} c_i g_i s_{i-1}^{-1} \left(\sum_{j=m+1}^{i-1} f_j \epsilon_j \right) \\
&\qquad - \sum_{i=m+1}^{n} c_i g_i s_{i-1}^{-1} \left(\sum_{j=m+1}^{i-1} f_j \xi_j \right).
\end{aligned} \tag{3.4}$$

By Lemma 4(v), for $i > m$,

$$\begin{aligned}
c_i g_i &= \{c_i(1 + \hat{T}_i' Q_{i-1}^{-1} \hat{T}_i)^{1/2}\}\{(1 + \hat{T}_i' Q_{i-1}^{-1} \hat{T}_i)^{1/2} f_i\} \\
&= \tilde{c}_i \tilde{f}_i ,
\end{aligned} \tag{3.5}$$

where

$$\tilde{c}_i = c_i(1 + \hat{T}_i' Q_{i-1}^{-1} \hat{T}_i)^{1/2}, \qquad \tilde{f}_i = (1 + \hat{T}_i' Q_{i-1}^{-1} \hat{T}_i)^{1/2} f_i . \tag{3.6}$$

By Lemma 4(i), for $n > m$,

$$s_n = s_m + \sum_{m+1}^{n} \tilde{f}_i^2. \tag{3.7}$$

By Lemma 4(vi) and (1.11),

$$\begin{aligned}
\sum_{m+1}^{\infty} \tilde{c}_i^2 s_i/s_{i-1} &= \sum_{m+1}^{\infty} c_i^2 (1 + \hat{T}_i' Q_{i-1}^{-1} \hat{T}_i)(s_i/s_{i-1}) \\
&= \sum_{m+1}^{\infty} c_i^2 (1 + T_i' H_{i-1}^{-1} T_i) < \infty.
\end{aligned} \tag{3.8}$$

In view of (3.5), (3.7), and (3.8), we can apply Lemma 2 to obtain that

$$\sum_{i=m+1}^{\infty} c_i g_i s_{i-1}^{-1} \left(\sum_{j=m+1}^{i-1} f_j \epsilon_j \right) = \sum_{i=m+1}^{\infty} \tilde{c}_i \tilde{f}_i s_{i-1}^{-1} \left(\sum_{j=m+1}^{i-1} f_j \epsilon_j \right) \tag{3.9}$$

converges a.s. Moreover, by Lemma 1,

$$\sum_{m+1}^{\infty} | c_i g_i / s_{i-1} | < \infty. \tag{3.10}$$

For $n \geqslant m$, let $p_n = \sum_{n+1}^{\infty} \tilde{c}_i \tilde{f}_i / s_{i-1} = \sum_{n+1}^{\infty} c_i g_i / s_{i-1}$. Then

$$\begin{aligned} \sum_{i=m+1}^{n} c_i g_i s_{i-1}^{-1} \left(\sum_{j=m+1}^{i-1} f_j \xi_j \right) &= \sum_{j=m+1}^{n-1} f_j \xi_j \left(\sum_{i=j+1}^{n} c_i g_i s_{i-1}^{-1} \right) \\ &= \sum_{j=m+1}^{n-1} p_j f_j \xi_j - p_n \sum_{j=m+1}^{n-1} f_j \xi_j . \end{aligned} \tag{3.11}$$

By (3.6), (3.8), and Lemma 1,

$$\begin{aligned} \sum_{m+1}^{\infty} (p_n f_n)^2 (1 + \hat{T}_n' Q_{n-1}^{-1} \hat{T}_n) &= \sum_{m+1}^{\infty} p_n^2 \tilde{f}_n^2 \\ &= \sum_{n=m+1}^{\infty} \left(\sum_{i=n+1}^{\infty} \tilde{c}_i \tilde{f}_i / s_{i-1} \right)^2 \tilde{f}_n^2 < \infty. \end{aligned}$$

Hence by (3.3) and the induction hypothesis, as $n \to \infty$,

$$\sum_{i=m+1}^{n} p_i f_i \xi_i = \sum_{i=m+1}^{n} (p_i f_i) \hat{T}_i' Q_{i-1}^{-1} \left(\sum_{j=1}^{i-1} \hat{T}_j \epsilon_j \right) \quad \text{converges a.s.} \tag{3.12}$$

Clearly (3.12) still holds if we replace p_j by $p_j^* = \sum_{j+1}^{\infty} | \tilde{c}_i \tilde{f}_i | / s_{i-1}$. Since $p_j^* \downarrow 0$, we then obtain by Kronecker's lemma that

$$\left| p_n \sum_{m+1}^{n-1} f_j \xi_j \right| \leqslant p_{n-1}^* \left| \sum_{m+1}^{n-1} f_j \xi_j \right| \to 0 \quad \text{a.s.} \tag{3.13}$$

From (3.4) and (3.9)–(3.13), the desired conclusion (3.2) follows. ∎

Proof of Theorem 1. We shall only consider b_{n1}. For the case $p = 1$, $b_{n1} - \beta = (\sum_1^n x_{i1} \epsilon_i)/(\sum_1^n x_{i1}^2)$. Since $\sum_1^n x_{i1}^2 = 1/v_{11}^{(n)} \to \infty$ a.s., the desired conclusion (1.10) follows easily from (1.2) and Kronecker's lemma.

Let $p \geqslant 2$. Define the $(p - 1)$-dimensional vector T_n by (2.11) and partition the matrix $X_n' X_n$ as in (2.12). Using the notation and results of Lemma 3, we obtain that $b_{n1} - \beta_1 = u_n / s_n$, where

$$s_n = \sum_{i=1}^{n} (x_{i1} - K_n H_n^{-1} T_n)^2 = 1/v_{11}^{(n)}. \tag{3.14}$$

(For the last equality in (3.14), see the identity (2.24).) Since $u_n = u_m + \sum_{i=m+1}^{n} w_i$, it suffices for the proof of (1.10) to show that

$$\left(\sum_{i=m+1}^{n} w_i\right) \Big/ \{s_n \mid \log s_n \mid^{1+\delta}\}^{1/2} \to 0 \qquad \text{a.s.} \tag{3.15}$$

By (2.14) and (3.14), for $n > m$,

$$s_n = s_m + \sum_{m+1}^{n} d_i^2(1 + T_i' H_{i-1}^{-1} T_i). \tag{3.16}$$

From (3.14) and (3.16), $s_n \uparrow \infty$. Therefore by the Kronecker lemma, (3.15) indeed holds if it can be shown that as $n \to \infty$

$$\sum_{i=m+1}^{n} (w_i/\{s_i \mid \log s_i \mid^{1+\delta}\}^{1/2}) \qquad \text{converges a.s.} \tag{3.17}$$

By (2.15), for $i > m$,

$$w_i = d_i \epsilon_i - d_i T_i' H_{i-1}^{-1} \left(\sum_{j=1}^{i-1} T_j \epsilon_j\right). \tag{3.18}$$

In view of (3.16) and the integral comparison test,

$$\sum_{m+1}^{\infty} \frac{d_i^2}{s_i \mid \log s_i \mid^{1+\delta}} \leqslant \sum_{m+1}^{\infty} \frac{d_i^2(1 + T_i' H_{i-1}^{-1} T_i)}{s_i \mid \log s_i \mid^{1+\delta}} < \infty. \tag{3.19}$$

Hence by (1.2), as $n \to \infty$,

$$\sum_{m+1}^{n} \frac{d_i \epsilon_i}{\{s_i \mid \log s_i \mid^{1+\delta}\}^{1/2}} \qquad \text{converges a.s.} \tag{3.20}$$

Moreover, in view of (3.19) and the fact that $H_n = \sum_1^n T_i T_i'$, we can apply Theorem 2 to obtain that

$$\sum_{m+1}^{n} \frac{d_i T_i' H_{i-1}^{-1} (\sum_1^{i-1} T_j \epsilon_j)}{\{s_i \mid \log s_i \mid^{1+\delta}\}^{1/2}} \qquad \text{converges a.s.} \tag{3.21}$$

From (3.18), (3.20), and (3.21), (3.17) follows. ∎

4. Some Corollaries

As we have indicated in Section 1, Theorem 1 provides a complete solution to the problem of strong consistency of b_n when the ϵ_i are i.i.d. with zero mean and finite variance.

COROLLARY 2. *Suppose that in the multiple regression model* (1.1) *the errors* $\epsilon_1, \epsilon_2, \ldots$ *are i.i.d. with* $E\epsilon_1 = 0$ *and* $0 < E\epsilon_1^2 < \infty$. *Moreover, assume that the design matrix* $X_n = (x_{ij})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p}$ *is of full rank* p *for some* n. *Then*

$$b_n \to \beta \text{ a.s.} \Leftrightarrow (X_n'X_n)^{-1} \to 0, \tag{4.1}$$

where b_n *is the least squares estimate of* β *defined by* (1.4).

In the Gauss–Markov model, the ϵ_i are assumed to be uncorrelated with mean 0 and variance σ^2. A refinement of the concept of orthogonal random variables is the notion of a multiplicative sequence. Let r be a positive even integer. A sequence of random variables $\{\epsilon_i\}$ is said to be *multiplicative of order* r if

$$E(\epsilon_{i_1}\epsilon_{i_2} \cdots \epsilon_{i_r}) = 0 \qquad \text{for all } 1 \leqslant i_1 < i_2 < \cdots < i_r. \tag{4.2}$$

When $r = 2$, this reduces to the case of orthogonal random variables. In [5], Komlós has shown that the condition (1.2) is satisfied when the ϵ_i have zero means and common variance σ^2 and form a multiplicative sequence of order $r \geqslant 4$ (r even) such that $\sup_i E\epsilon_i^4 < \infty$. Recently Longnecker and Serfling [8] extended this result of Komlós to the following three types of weakly multiplicative sequences.

DEFINITION. Let r be a positive even integer. Let $\{\epsilon_i\}$ be a sequence of random variables.

(i) The sequence $\{\epsilon_i\}$ is said to be *weakly multiplicative of type* A_r if $E\epsilon_i^r < \infty$ for all i and there exists a symmetric function g with $r - 1$ arguments such that

$$| E(\epsilon_{i_1} \cdots \epsilon_{i_r})| \leqslant g(i_2 - i_1, i_3 - i_2, \ldots, i_r - i_{r-1}) \prod_{j=1}^{r} (E\epsilon_{i_j}^r)^{1/r}$$

for all $1 \leqslant i_1 < \cdots < i_r$ and

$$\sum_{k=1}^{\infty} \sum_{j_1=1}^{k} \cdots \sum_{j_{r-2}=1}^{k} g(j_1, \ldots, j_{r-2}, k) < \infty.$$

(ii) The sequence $\{\epsilon_i\}$ is said to be *weakly multiplicative of type* B_r if $E\epsilon_i^r < \infty$ for all i and there exists a symmetric function g with $\frac{1}{2}r$ arguments such that

$$| E(\epsilon_{i_1} \cdots \epsilon_{i_r})| \leqslant g(i_2 - i_1, i_4 - i_3, \ldots, i_r - i_{r-1}) \prod_{j=1}^{r} (E\epsilon_{i_j}^r)^{1/r}$$

for all $1 \leqslant i_1 < \cdots < i_r$ and

$$\sum_{k=1}^{\infty} \sum_{j_1=1}^{k} \cdots \sum_{j_{r/2-1}=1}^{k} g(j_1, \ldots, j_{r/2-1}, k) < \infty.$$

(iii) The sequence $\{\epsilon_i\}$ is said to be *weakly multiplicative of type* C_r if $E\epsilon_i^r < \infty$ for all i and there exists a function $f(j)$ and a function g with $\frac{1}{2}r - 1$ arguments such that

$$| E(\epsilon_{i_1} \cdots \epsilon_{i_r})| \leqslant \min\{f(i_2 - i_1), f(i_r - i_{r-1})\}$$
$$\times\, g(i_3 - i_2, i_5 - i_4, \ldots, i_{r-1} - i_{r-2}) \prod_{j=1}^{r} (E\epsilon_{i_j}^r)^{1/r}$$

for all $1 \leqslant i_1 < \cdots < i_r$, $\sum_{j=1}^{\infty} f(j) < \infty$, and $\sum_C g(j_1, \ldots, j_{r/2-1}) < \infty$, where C denotes the set of all $(\frac{1}{2}r - 1)$-tuples $(j_1, \ldots, j_{r/2-1})$ with $1 \leqslant j_\nu \leqslant j_m$ for $\nu \neq m$, $1 \leqslant j_m < \infty$, and $m = 1, \ldots, \frac{1}{2}r - 1$.

LEMMA 5 (Longnecker and Serfling [8, p. 17]). *Let $r \geqslant 4$ be an even integer. Let $\{\epsilon_i\}$ be a sequence of random variables such that* $\sup_i E\epsilon_i^r < \infty$, and let $\{c_i\}$ *be a sequence of constants such that $\sum_1^\infty c_i^2 < \infty$. Suppose that $\{\epsilon_i\}$ is weakly multiplicative of type A_r or B_r or C_r. Then $\sum_1^\infty c_i \epsilon_i$ converges a.s.*

This lemma is an extension of the result of Komlós for multiplicative sequences to weakly multiplicative sequences. Together with Theorem 1 it gives

COROLLARY 3. *Suppose that in the multiple regression model* (1.1), *$\{\epsilon_i\}$ is a weakly multiplicative sequence of type A_r or B_r or C_r, where r is an even integer $\geqslant 4$. Assume that* $\sup_i E\epsilon_i^r < \infty$, *and that the design matrix $X_n = (x_{ij})_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}$ is of full rank p for some n. Define V_n as in* (1.9). *Then for each fixed $j = 1, \ldots, p$,*

$$\lim_{n\to\infty} v_{jj}^{(n)} = 0 \Rightarrow \text{(1.10) holds a.s. for every } \delta > 0. \tag{4.3}$$

As shown in [8, p. 5, 18], the weakly multiplicative sequences in Corollary 3 include important classes of Gaussian time series and stationary mixing sequences. Thus Corollary 3 contains as a special case the following.

COROLLARY 4. *Suppose that in* (1.1), *$\{\epsilon_i\}$ is a stationary sequence with $E\epsilon_1 = 0$. Assume that the design matrix X_n is of full rank p for some n, and define V_n as in* (1.9).

(i) *Suppose that $\{\epsilon_i\}$ is a Gaussian sequence with covariance function $r(k) = E(\epsilon_1\epsilon_{k+1})$. If $|r(k)|$ is nonincreasing and $\sum_1^\infty |r(k)| < \infty$, then* (4.3) *holds.*

(ii) *Suppose that $\{\epsilon_i\}$ is a strongly mixing sequence with mixing coefficient $\varphi(k) = \sup\{|P(A \cap B) - P(A)P(B)|: A \in F_n, B \in G_{n+k}, n \geqslant 1\}$, where F_n is the σ-field generated by $\{\epsilon_1, \ldots, \epsilon_n\}$ and G_n is the σ-field generated by $\{\epsilon_n, \epsilon_{n+1}, \ldots\}$. If $\sum_1^\infty k\varphi(k) < \infty$ and $\{\epsilon_i\}$ is uniformly bounded (i.e., $|\epsilon_i| \leqslant C$ for some constant C and all i), then* (4.3) *still holds.*

Suppose that $\{\epsilon_i\}$ is a martingale difference sequence such that $E|\epsilon_i|^r < \infty$ for all i, where r is a positive even integer. Then $\{\epsilon_i\}$ is multiplicative of order r.

While Corollary 3 gives the strong consistency of b_{nj} under the assumption that $\sup_i E\epsilon_i^r < \infty$ for some $r \geqslant 4$, the martingale convergence theorem implies, however, that $\sup_i E\epsilon_i^2 < \infty$ suffices for (1.2) to hold in this case, and therefore we obtain from Theorem 1 the following result announced in [7].

COROLLARY 5. *Suppose that in* (1.1), $\{\epsilon_i\}$ *is a martingale difference sequence such that* $\sup_i E\epsilon_i^2 < \infty$. *Assume that the design matrix* X_n *is of full rank* p *for some* n, *and define* V_n *as in* (1.9). *Then* (4.3) *holds.*

In the above corollaries we have assumed that $X_n'X_n$ is nonsingular for some, and therefore for all, large n. We now consider the general case where $X_n'X_n$ may be singular for all n. Let $\mathscr{R}^p$ denote the p-dimensional Euclidean space of column vectors. For $\alpha \in \mathscr{R}^p$, $\alpha' b$ is unique for all solutions $b \in \mathscr{R}^p$ of the normal equation $X_n'X_n b = X_n'Y_n$ if and only if $\alpha \in \mathscr{L}$, where $\mathscr{L}$ is the linear space generated by the set of vectors $\{Z_i: i = 1, 2,...\}$ and

$$Z_i = (x_{i1}, ..., x_{ip})' \tag{4.4}$$

(cf. [9, p. 181]). A solution of the equation $X_n'X_n b = X_n'Y_n$ is

$$b = X_n^+ Y_n, \tag{4.5}$$

where X_n^+ denotes the Moore–Penrose generalized inverse of the matrix X_n (cf. [3]). This reduces to the unique solution (1.4) when $X_n'X_n$ is nonsingular. Even when $X_n'X_n$ is singular, $\alpha' X_n^+ Y_n$ is the unique least squares estimate of $\alpha'\beta$ for all large n if $\alpha \in \mathscr{L}$. Assume that in (1.1) the random variables ϵ_i are uncorrelated and have a common variance $\sigma^2 > 0$. Then for $\alpha \in \mathscr{L}$, $\alpha' X_n^+ Y_n$ is an unbiased estimate of $\alpha'\beta$ for all large n, with $\operatorname{Var}(\alpha' X_n^+ Y_n) = \sigma^2 \alpha' X_n^+ (X_n^+)' \alpha$, and a necessary and sufficient condition for $\alpha' X_n^+ Y_n$ to converge to $\alpha'\beta$ in probability is

$$\alpha' X_n^+ (X_n^+)' \alpha \to 0 \quad \text{as} \quad n \to \infty \tag{4.6}$$

(cf. [3]). By reducing the general case to the nonsingular case, we obtain from Theorem 1 the strong consistency of $\alpha' X_n^+ Y_n$ under the minimal assumption (4.6) on the design in the following.

COROLLARY 6. *Suppose that in* (1.1) *the errors* $\epsilon_1, \epsilon_2, ...$ *satisfy* (1.2). *Let* $\mathscr{L}$ *be the linear subspace of* $\mathscr{R}^p$ *generated by the set of vectors* $\{Z_i, i = 1, 2,...\}$, *where* $Z_i = (x_{i1}, ..., x_{ip})'$, *and let* $\alpha \neq 0$ *belong to* $\mathscr{L}$. *Let* $\rho_n^2(\alpha) = \alpha' X_n^+ (X_n^+)' \alpha$. *Then the sequence* $\{\rho_n(\alpha)\}$ *is eventually nonincreasing and* $\rho_n(\alpha) > 0$ *for all large* n. *Moreover, if* $\lim_{n\to\infty} \rho_n(\alpha) = 0$, *then for every* $\delta > 0$, *with probability* 1,

$$\alpha' X_n^+ Y_n - \alpha'\beta = o\{\rho_n(\alpha) \mid \log \rho_n(\alpha)|^{(1+\delta)/2}\} \quad \textit{as} \quad n \to \infty. \tag{4.7}$$

Proof. Let $\| \alpha \|$ denote $(\alpha'\alpha)^{1/2}$. Let

$$\mathscr{L}_n = \{X_n\theta : \theta \in \mathscr{R}^p \quad \text{and} \quad \theta'\alpha = 0\}, \tag{4.8}$$

and let Π_n denote the projection matrix associated with the linear subspace $\mathscr{L}_n$ of $\mathscr{R}^n$, i.e., $\Pi_n x$ is the projection of x into $\mathscr{L}_n$ for every $x \in \mathscr{R}^n$. Let

$$W_n = (I_n - \Pi_n) X_n\alpha, \tag{4.9}$$

where I_n is the identity $(n \times n)$ matrix. Then as shown in [3, pp. 121–122], for all large n (say $n \geqslant n_0$), $W_n \neq 0$ and

$$\alpha' X_n{}^+ Y_n = \alpha'\beta + \| \alpha \|^2 \| W_n \|^{-2} W_n' E_n , \tag{4.10}$$

where $E_n = (\epsilon_1, \ldots, \epsilon_n)'$. We note that for $\theta \in \mathscr{R}^p$

$$X_n\theta = \begin{pmatrix} Z_1'\theta \\ \vdots \\ Z_n'\theta \end{pmatrix} \in \mathscr{R}^n. \tag{4.11}$$

From (4.8) and (4.11), it follows that dim $\mathscr{L}_n$ is nondecreasing in n. Since dim $\mathscr{L}_n \leqslant p - 1$, there exist $N \geqslant n_0$ and $m \leqslant p - 1$ such that

$$\dim \mathscr{L}_n = m \qquad \text{for all} \quad n \geqslant N. \tag{4.12}$$

Choose linearly independent vectors $\theta_1, \ldots, \theta_m \in \mathscr{R}^p$ such that $\theta_i'\alpha = 0$ $(i = 1, \ldots, m)$ and $\{X_N\theta_1, \ldots, X_N\theta_m\}$ is a basis of $\mathscr{L}_N$. In view of (4.11), $\{X_n\theta_1, \ldots, X_n\theta_m\}$ is a linearly independent set for $n \geqslant N$, and therefore by (4.12),

$$\{X_n\theta_1, \ldots, X_n\theta_m\} \qquad \text{is a basis of } \mathscr{L}_n \text{ for } n \geqslant N. \tag{4.13}$$

Since $(I_n - \Pi_n) X_n\alpha (= W_n) \neq 0$ for $n \geqslant N$, it then follows from (4.13) that $\{X_n\alpha, X_n\theta_1, \ldots, X_n\theta_m\}$ is a linearly independent set. Therefore, the matrix

$$\tilde{X}_n = \begin{pmatrix} Z_1'\alpha & Z_1'\theta_1 & \cdots & Z_1'\theta_m \\ Z_2'\alpha & Z_2'\theta_1 & \cdots & Z_2'\theta_m \\ \cdot & \cdot & \cdot & \cdot \\ Z_n'\alpha & Z_n'\theta_1 & \cdots & Z_n'\theta_m \end{pmatrix} = (X_n\alpha, X_n\theta_1, \ldots, X_n\theta_m)$$

is nonsingular for $n \geqslant N$. Define $\tilde{b}_n = (\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n'E_n$ and $\tilde{V} = (\tilde{v}_{ij}^{(n)}) = (\tilde{X}_n\tilde{X}_n)^{-1}$ for $n \geqslant N$. Then

$$\tilde{b}_{n1} = \left\{\sum_1^n \epsilon_i(Z_i'\alpha - u_{ni})\right\} \Big/ \left\{\sum_1^n (Z_i'\alpha - u_{ni})^2\right\}, \tag{4.14}$$

where $U_n = (u_{n1}, ..., u_{nn})'$ is the projection of $X_n\alpha$ into the linear subspace generated by $\{X_n\theta_1, ..., X_n\theta_m\}$. From (4.9), (4.13), and (4.14), it then follows that for $n \geqslant N$

$$\tilde{b}_{n1} = \| W_n \|^{-2} W_n' E_n . \tag{4.15}$$

By (3.14), $\tilde{v}_{11}^{(n)} = \| W_n \|^{-2}$, and as shown in [3, p. 122], $\| W_n \|^{-2} = \| \alpha \|^{-4} \rho_n^2(\alpha)$ is positive and nonincreasing in $n \geqslant N$. From (4.10), (4.15), and Theorem 1, the desired conclusion (4.7) follows. ∎

As we have mentioned above, Komlós' theorem implies that condition (1.2) is satisfied if

$$\begin{gathered} E\epsilon_i = 0 \text{ and } E\epsilon_i^2 = \sigma^2 \text{ for all } i, \sup_i E\epsilon_i^4 < \infty, \\ \text{and } \{\epsilon_i\} \text{ is multiplicative of order 4.} \end{gathered} \tag{4.16}$$

Hence by Corollary 6, (4.7) holds under the assumptions (4.6) and (4.16). In [7] we have shown by a simpler argument that (4.7) holds under (4.6), (4.16), and the additional assumption

$$E(\epsilon_i\epsilon_j) = E(\epsilon_i^3\epsilon_j) = E(\epsilon_i^2\epsilon_j\epsilon_k) = 0 \qquad \text{for any distinct } i, j, k. \tag{4.17}$$

However, this simpler argument depends heavily on (4.17) and does not generalize to weakly multiplicative sequences or L^2-bounded martingale difference sequences.

References

[1] Anderson, T. W., and Taylor, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.* 4 788–790.

[2] Anderson, T. W., and Taylor, J. B. (1976). *Conditions for Strong Consistency of Least Squares Estimates in Linear Models.* Tech. Report No. 213. Institute for Mathematical Studies in Social Sciences, Stanford University.

[3] Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 34 119–127.

[4] Hardy, G. H., Littlewood, J. E., and Polya, G. (1964). *Inequalities*, 2nd ed. Cambridge Univ. Press, London/New York.

[5] Komlós, J. (1972). On the series $\Sigma c_k \varphi_k$. *Studia Sci. Math. Hungar.* 7 451–458.

[6] Lai, T. L., and Robbins, H. (1977). Strong consistency of least squares estimates in regression models. *Proc. Nat. Acad. Sci. USA* 74 2667–2669.

[7] Lai, T. L., Robbins, H., and Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. USA* 75 3034–3036.

[8] Longnecker, M., and Serfling, R. J. (1978). Moment inequalities for S_n under general dependence restrictions with applications. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* 43 1–21.

[9] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.