Convergence Systems and Strong Consistency of Least Squares Estimates in Regression Models

Chen Gui-Jing,* T. L. Lai,† and C. Z. Wei‡

Anhwei University, Columbia University, and University of Maryland

Communicated by E. J. Hannan

A recent theorem of T. L. Lai, H. Robbins, and C. Z. Wei (J. Multivariate Anal. 9 (1979), 343–362) is extended to a more general form which unifies previous results in the literature on the strong consistency of least squares estimates in multiple regression models with nonrandom regressors. In particular the issue of strong consistency of the least squares estimate in the Gauss–Markov model, in the i.i.d. model with infinite second moment, and in general time series models is examined. In this connection, some basic properties of convergence systems are also obtained and are applied to the strong consistency problem.

1. INTRODUCTION

Consider the multiple regression model

\[ y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \quad (i = 1, 2, \ldots), \]  

where \( x_{ij} \) are known constants, \( \beta_1, \ldots, \beta_p \) are unknown parameters, and \( \varepsilon_i \) are unobservable random variables. Throughout the sequel we shall let \( X_n \) denote the design matrix \((x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}\), and let \( Y_n = (y_1, \ldots, y_n)' \), and \( \beta = (\beta_1, \ldots, \beta_p)' \), where a prime denotes transpose. For \( n \geq p \), the least squares estimate \( b_n = (b_{n1}, \ldots, b_{np})' \) of the vector \( \beta \) based on the design matrix \( X_n \) and the response vector \( Y_n \) is given by

\[ b_n = (X_n'X_n)^{-1}X_n'Y_n, \]
 PROVIDED THAT $X_n'X_n$ IS NONSINGULAR. THE FOLLOWING THEOREM OF LAI ET AL. [11] GIVES THE STRONG CONSISTENCY OF $b_n$ UNDER MINIMAL ASSUMPTIONS ON THE DESIGN CONSTANTS $x_{ij}$ AND VERY WEAK CONDITIONS ON THE RANDOM ERRORS $\varepsilon_i$. IT IMPLIES THE EARLIER RESULTS ON THE CONSISTENCY OF $b_n$ DUE TO DRYGAS [8], ANDERSON AND TAYLOR [2, 3], AND LAI AND ROBBINS [10].

**THEOREM 1.** *Suppose that in (1.1), \( \{x_{ij}\} \ (i = 1, 2, \ldots; \ j = 1, \ldots, p) \) is a double array of constants such that $X_n'X_n$ is nonsingular for all $n \geq m$. For $n \geq m$, let $b_n = (b_n, \ldots, b_{np})'$ be the least squares estimate defined by (1.2) and let

$$V_n = (v_{ij}^{(n)})_{1 \leq i, j \leq p} = (X_n'X_n)^{-1}. \quad (1.3)$$

Fix $j = 1, \ldots, p$. If $\lim_{n \to \infty} v_{jj}^{(n)} = 0$ and the random variables $\varepsilon_i$ satisfy the condition

$$\sum_{i=1}^{\infty} a_i \varepsilon_i \text{ converges a.s. for all real sequences } \{a_i\}$$

such that $\sum_{i=1}^{\infty} a_i^2 < \infty$, \quad (1.4)

then for every $\delta > 1$,

$$b_n - \beta_j = o(\{v_{jj}^{(n)}|\log v_{jj}^{(n)}|^{\delta/2}\}^{1/2}) \quad \text{a.s.} \quad (1.5)$$

A sequence of random variables $\varepsilon_i$ satisfying condition (1.4) is called a convergence system [4, 9]. In particular, if the errors $\varepsilon_i$ are i.i.d. with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2 < \infty$, then $\{\varepsilon_i\}$ is a convergence system. In this case, if $\sigma \neq 0$ and $X_n'X_n$ is nonsingular for all large $n$, then $\text{cov}(b_n) = \sigma^2(X_n'X_n)^{-1}$ and

$$(X_n'X_n)^{-1} \to 0 \quad \text{as } n \to \infty \quad (1.6)$$

is a necessary condition for $b_n$ to converge to $\beta$ in probability (cf. [8]), while Theorem 1 implies that this minimal condition (1.6) on the design is also sufficient for $b_n$ to be strongly consistent. As pointed out in [11], more general error structures which are convergence systems include $L^2$-bounded martingale difference sequences, stationary Gaussian sequences with absolutely summable autocorrelations, and certain types of weakly multiplicative sequences. Some interesting properties of convergence systems will be presented in Sections 3 and 4.

Although the general condition (1.4) covers a wide range of error structures, it does not necessarily hold when the $\varepsilon_i$ are only assumed to be uncorrelated with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2 < \infty$ (as in the Gauss–Markov
LEAST SQUARES ESTIMATES

model) or when the $\varepsilon_i$ are independent with zero means and $\sup_i E|\varepsilon_i| < \infty$ for some $1 \leq r < 2$. These two classes of random errors $\varepsilon_i$ have recently been considered by Chen [5, 6], who has established the strong consistency of $b_n$ for these error structures under stronger assumptions on the design than (1.6). In Section 2, we give an extension of Theorem 1 which also implies these results of Chen. This extension unifies all the known results on the strong consistency of $b_n$ in the literature. Some of its applications will be discussed in Section 4.

2. EXTENSION OF THEOREM 1

In Theorem 1, condition (1.4) on the errors $\varepsilon_i$ is totally unrelated to the design constants $x_{ij}$. Obviously,

\[
\{\varepsilon_n\} \text{ is a convergence system and } \{c_n\} \text{ is a bounded sequence of real constants} \\
\Rightarrow \{c_n \varepsilon_n\} \text{ is a convergence system. (2.1)}
\]

Instead of requiring the random errors $\varepsilon_n$ to form a convergence system, the following theorem shows that it suffices to assume that $\{c_n \varepsilon_n\}$ is a convergence system for some “contraction” constants $c_n$, provided that these contraction constants are not too small compared with $v_{jj}^{(n)}$ (see condition (2.2) below).

**THEOREM 2.** With the same notation as in Theorem 1, if $\lim_{n \to \infty} v_{jj}^{(n)} = 0$ and there exists a positive nondecreasing function $g$ on $(0, \infty)$ such that $\{g(v_{jj}^{(n)}) \varepsilon_n\}$ is a convergence system and

\[
g(t)/t \uparrow \infty \text{ as } t \downarrow 0 \quad \text{and} \quad \int_0^A dt/g^2(t) < \infty \text{ for some } A > 0, \quad (2.2)
\]

then $b_{nj} \to \beta_j$ a.s.

In view of (2.1), there is no loss of generality in the requirement that the function $g$ in Theorem 2 be positive. Moreover, it is of interest to consider the minimal order of magnitude for $g$ allowed by (2.2). In this connection, note that the function $g(t) = \{t|\log t|^{\delta}\}^{1/2}$, $0 < t < e^{-18}$, satisfies (2.2) if $\delta > 1$, but does not satisfy the integrability condition in (2.2) if $\delta \leq 1$.

Theorem 2 follows from the following more general result with $f(t) = g^2(t)$ and $g_n = g(v_{jj}^{(n)})$, noting that $g_n$ is positive and non-increasing by the monotonicity of $g$ and (3.8) in Section 3.
THEOREM 3. With the same notation as in Theorem 1, if \( \lim_{n \to \infty} v^{(n)}_{jj} = 0 \) and \( \{ g_n \varepsilon_n \} \) is a convergence system for some constants \( g_n \) such that \( |g_n| \) is positive and non-increasing, then

\[
b_{nj} - \beta_j = o\left( (f(v^{(n)}_{jj}))^{1/2} / |g_n| \right) \quad \text{a.s.}
\]

(2.3)

for every positive function \( f \) on \((0, \infty)\) such that

\[
\int_0^A dt / f(t) < \infty \quad \text{for some } A > 0
\]

(2.4a)

and

\[
f(t) / t^2 \uparrow \infty \quad \text{as } t \downarrow 0.
\]

(2.4b)

3. Properties of Convergence Systems and Proof of Theorem 3

In this section, we prove Theorem 3 by modifying the ideas developed in [11] for the proof of Theorem 1. We also extend and re-interpret these ideas in terms of convergence systems in the following two lemmas and thereby obtain certain basic properties of convergence systems.

**Lemma 1.** Let \( \{ g_n \} \) be a sequence of constants such that \( |g_n| \) is positive and non-increasing. Let \( \{ \varepsilon_n \} \) be a sequence of random variables such that \( \{ g_n \varepsilon_n \} \) is a convergence system.

(i) Let \( a_n, \tilde{a}_n \) be real constants such that \( a_m \neq 0 \) and there exists \( C > 0 \) for which \( |\tilde{a}_n| \leq C |a_n| \) for all \( n \geq 1 \). Let \( A_n = \sum_n a_i^2 \). Then

\[
\left\{ g_n a_{n+1} (A_{n+1} A_n)^{-1/2} \left( \sum_{i=1}^n \tilde{a}_i \varepsilon_i \right), n \geq m \right\}
\]

(3.1)

is a convergence system.

(ii) Let \( k \) be a positive integer. For each \( n \geq 1 \), let \( T_n \) be a \( k \)-dimensional vector of constants and let \( H_n = \sum_i T_i T_i^\top \). Assume that \( H_n \) is positive definite for all \( n \geq m \). Then

\[
\left\{ g_n T_{n+1} H_n^{-1} \left( \sum_{i=1}^n T_i \varepsilon_i \right) (1 + T_{n+1} H_n^{-1} T_{n+1})^{1/2}, n \geq m \right\}
\]

(3.2)

is a convergence system.
Proof. To prove (i), let $a_n = a_n/g_n$, $\bar{a}_n = \bar{a}_n/g_n$, and $G_n = \sum^n_i (a_i/g_i)^2$. Since $|g_n|$ is non-increasing,
\[
|a_{n+1}|(G_{n+1}G_n)^{-1/2} \geq |g_n a_{n+1}|(A_n A_{n+1})^{-1/2}.
\]
(3.3)
We note that Lemma 2 of [11] can be restated as
\[
\left\{ a_{n+1}(G_{n+1}G_n)^{-1/2} \left( \sum_{i=1}^n \bar{a}_i \zeta_i \right), n \geq m \right\}
\]
is a convergence system.
Putting $\zeta_n = g_n \epsilon_n$ in (3.4), we obtain from (2.1), (3.3), and (3.4) that (3.1) forms a convergence system. To prove (ii), we note that for the case $k = 1$, $T_n$ and $H_n$ are scalars, $H_n = \sum^n_1 T_i^2$, and therefore (3.2) reduces to (3.1). Induction on $k$ by an argument as in the proof of Theorem 2 of [11] then completes the proof.

Remark. In Lemma 1(ii), let $T_n = (t_{n1}, \ldots, t_{nk})'$ and consider the multiple regression model
\[
e_i = \theta_1 t_{i1} + \cdots + \theta_k t_{ik} + e_i \quad (i = 1, 2, \ldots).
\]
The term $H_n^{-1}(\sum^n_{i=1} T_i e_i)$ in (3.2) is the least squares estimate $(\hat{\theta}_1, \ldots, \hat{\theta}_k)'$ of the parameter vector $(\theta_1, \ldots, \theta_k)'$ based on the $n$-dimensional vector $(e_1, \ldots, e_n)'$ and the $n \times k$ design matrix $(t_{ij})_{1 \leq i \leq n, 1 \leq j \leq k}$. Hence the term $T_n^{-1} H_n^{-1}(\sum^n_{i=1} T_i e_i)$ can be regarded as the "least squares predictor" $\sum_{i=1}^k \hat{\beta}_i t_{i1} / n_{i+1}$ of $e_{n+1}$.

Lemma 2. Let $\{ g_n \}$ be a sequence of constants such that $|g_n|$ is positive and non-increasing. Let $\{ e_n \}$ be a sequence of random variables such that $\{ g_n e_n \}$ is a convergence system. Then for every sequence of real constants $a_n$ such that $A_n = \sum^n_i a_i^2 \to \infty$ as $n \to \infty$,
\[
\sum^n_{i=1} a_i e_i = o((h(A_n))^{1/2} / |g_n|) \quad \text{a.s.}
\]
(3.5)
for any given positive function $h$ on $(0, \infty)$ such that
\[
h(x) \uparrow \infty \text{ as } x \uparrow \infty \quad \text{and} \quad \int_c^\infty dx / h(x) < \infty \text{ for some } c > 0.
\]
(3.6)

Proof. In view of (3.6) and the integral comparison test, $\sum_m^\infty a_i^2 / (h(A_i)) < \infty$, and therefore
\[
\sum_m^\infty a_i g_i e_i / (h(A_i))^{1/2} \text{ converges a.s.}
\]
(3.7)
Since $(h(A_n))^{1/2}/|g_n| \uparrow \infty$ by (3.6) and the monotonicity of $|g_n|$, (3.5) follows from (3.7) and Kronecker's lemma.

**Proof of Theorem 3.** Without loss of generality, we shall assume that $j = 1$. Let $h(x) = x^2f(1/x)$, $x > 0$. Then a change of variable $t = 1/x$ in (2.4a) and (2.4b) shows that $h$ satisfies (3.6). In view of Lemma 2, we only consider the case $p \geq 2$. For $n > m$, let $T_n = (x_{n1}, ..., x_{np})'$ and write

$$X_n'X_n = \begin{pmatrix} \sum_{i=1}^{n} x_{ni}^2 & K_n \\ K_n' & H_n \end{pmatrix}, \quad d_n = x_{n1} - K_nH_n^{-1}T_n.$$

By Lemma 3 of [11], $b_n - \beta_1 = u_n/s_n$, where

$$s_n = 1/v_{11}^{(n)} = s_m + \sum_{m+1}^{n} d_i^2(1 + T_{i}H_{i-1}^{-1}T_{i}), \quad (3.8)$$

$$u_n = u_m + \sum_{m+1}^{n} \left\{ d_i \varepsilon_i - d_i T_{i}H_{i-1}^{-1} \left( \sum_{k=1}^{i-1} T_{k}\varepsilon_{k} \right) \right\}. \quad (3.9)$$

Therefore the desired conclusion (2.3) can be expressed as

$$u_n = o((h(s_n))^{1/2}/|g_n|) \quad \text{a.s.} \quad (3.10)$$

Since $\sum_{m+1}^{n} d_i^2 \leq s_n$ by (3.8) and since $\{g_n\varepsilon_n\}$ is a convergence system, we obtain by Lemma 2 that

$$\sum_{m+1}^{n} d_i \varepsilon_i = o((h(s_n))^{1/2}/|g_n|) \quad \text{a.s.} \quad (3.11)$$

Moreover, in view of (2.1) and the fact that $|g_n| \ll |g_{n-1}|$ and $H_n = \sum_{i}^{n} T_iT_i'$, Lemma 1(ii) implies that $\{g_n\xi_n, n \geq m+1\}$ is a convergence system, where

$$\xi_n = T_nH_n^{-1} \left( \sum_{k=1}^{n-1} T_{k}\varepsilon_{k} \right)/(1 + T_nH_n^{-1}T_n)^{1/2}.$$ 

Therefore by Lemma 2 and (3.8),

$$\sum_{m+1}^{n} d_i T_{i}H_{i-1}^{-1} \left( \sum_{k=1}^{i-1} T_{k}\varepsilon_{k} \right) = \sum_{m+1}^{n} d_i(1 + T_{i}H_{i-1}^{-1}T_{i})^{1/2} \xi_i - o((h(s_n))^{1/2}/|g_n|) \quad \text{a.s.} \quad (3.12)$$

From (3.11) and (3.12) the desired conclusion (3.10) follows.
In the rest of this section, we consider some implications of Lemma 1. Putting $a_n = \bar{a}_n = 1$ in Lemma 1(i), we obtain the interesting property that successive averaging preserves convergence systems, i.e.,

$$\{\varepsilon_n\}$$ is a convergence system $\Rightarrow \{\bar{\varepsilon}_n\}$$ is a convergence system,

(3.13)

where we use the notation $\bar{a}_n$ to denote the arithmetic mean $n^{-1} \sum_{i}^{n} a_i$ of $a_1, \ldots, a_n$. More generally, if $\{g_n\}$ is a sequence of constants such that $|g_n|$ is positive and non-increasing, then by Lemma 1(i),

$$\{g_n \varepsilon_n\}$$ is a convergence system $\Rightarrow \{g_n \bar{\varepsilon}_n\}$$ is a convergence system.

(3.14)

In Lemma 1(ii), setting

$$c_{ni} = T_{n+1} \left( \frac{1}{n} \sum_{j=1}^{n} T_j T_j' \right)^{-1} T_i, \quad n \geq m \text{ and } i = 1, \ldots, n,$$

(3.15)

we can restate the result in the form

$$\{g_n \varepsilon_n\}$$ is a convergence system $\Rightarrow \{g_n \bar{\varepsilon}_n\}$$ is a convergence system.

(3.16)

It is easy to see that the triangular array $\{c_{ni}; n \geq m, i = 1, \ldots, n\}$ defined by (3.15) satisfies the equation

$$\sum_{i=1}^{n} c_{ni} c_{vi} = c_{v,n+1} \quad \text{for all } v > n \geq m.$$

(3.17)

Conversely, if $\{c_{ni}\}$ is a triangular array satisfying (3.17), then letting $\{T_1, \ldots, T_m\}$ be a basis of $\mathbb{R}^m$ and defining inductively $T_{n+1} = \sum_{i=1}^{n} c_{ni} T_i$ for $n \geq m$, it can be shown by induction on $n$ that (3.15) holds with this choice of $T_j$. Hence Lemma 1 can be restated in terms of the simple condition (3.17) on the triangular array $\{c_{ni}\}$ as follows:

**Corollary 1.** Let $\{g_n\}$ be a sequence of constants such that $|g_n|$ is positive and non-increasing. Let $m$ be a positive integer, and let $\{c_{ni}; n \geq m, i = 1, \ldots, n\}$ be a triangular array of constants satisfying (3.17). Then (3.16) holds. In particular, if $\sup_{n} \sum_{i=1}^{n} c_{ni}^2 < \infty$ and $\{\varepsilon_n\}$ is a convergence system, then $\{\sum_{i=1}^{n} c_{ni} \varepsilon_i, n \geq m\}$ is also a convergence system.
4. Special Classes of Convergence Systems and Some Applications of Theorem 3

Theorem 1 is a special case of Theorem 3 with $g_n = 1$ and $f(t) = t |\log t|^\delta$ ($0 < t < e^{-\delta}$). The following corollary of Theorem 3 improves the result of Chen [5] on the strong consistency of the least squares estimate in the Gauss–Markov model, and we shall give a counter-example to show that condition (4.2) on the design constants in the corollary is in some sense minimal for the Gauss–Markov model.

**Corollary 2.** With the same notation as in Theorem 1, let $\{\varepsilon_i\}$ be a sequence of random variables such that $\sup_k E |\varepsilon_i|^2 < \infty$ and $E(\varepsilon_i \varepsilon_k) = 0$ for all $k \neq n$. Then $\{\varepsilon_n/|\log n|\}$ is a convergence system. Moreover, if $\lim_{n \to \infty} v_{jj}^{(n)} = 0$, then for every $\delta > 1$,

$$b_n - \beta_j = o\left(\frac{|\log v_{jj}^{(n)}|^{\delta/2} |\log n|}{\log \log n}\right) \quad \text{a.s.} \quad (4.1)$$

Hence $b_n \to \beta_j$ a.s. if

$$v_{jj}^{(n)} = O((|\log n|^{-\delta} |\log \log n|^{-1})) \quad \text{for some } \delta > 1. \quad (4.2)$$

**Proof.** For every real sequence $c_n$ such that $\sum_{i=1}^\infty c_n^2 < \infty$, since $\sum_{i=1}^\infty (|\log n|)^2 E(\varepsilon_n \varepsilon_n/|\log n|) \leq \infty$, it follows from the convergence theorem for orthonormal random variables [13, p. 201] that $\sum_{i=1}^\infty (c_n \varepsilon_n/|\log n|)$ converges a.s. Hence $\{\varepsilon_n/|\log n|\}$ is a convergence system, and therefore (4.1) follows from Theorem 3 with $f(t) = t |\log t|^\delta$. □

**Counter-Example.** We now show that condition (4.2) in Corollary 2 cannot be weakened to

$$v_{jj}^{(n)} = O((|\log n|^{-2} |\log \log n|^{-1})), \quad (4.3)$$

and therefore, a fortiori, the condition $\lim_{n \to \infty} v_{jj}^{(n)} = 0$ is not sufficient for the strong consistency of $b_n$ in the Gauss–Markov model where the errors are assumed to be uncorrelated and to have zero means and common variance $\sigma^2 > 0$. Consider the case $p = 1$ so that (1.1) reduces to $y_i = \beta x_i + \varepsilon_i$. Assuming that $x_i \neq 0$, the least squares estimate $b_n$ of $\beta$ can be expressed as

$$b_n = \beta + \left(\sum_{i=1}^n x_i \varepsilon_i\right)/s_n, \quad \text{where} \quad s_n = \sum_{i=1}^n x_i^2. \quad (4.4)$$

Suppose that $|x_n|$ is decreasing and that $x_n^2 \sim (|\log n|)(|\log \log n|)/n$ as $n \to \infty$. Then $s_n \sim (|\log n|)^2 (|\log \log n|)/2$, and therefore (4.3) holds since $s_n = 1/v_{jj}^{(n)}$. Moreover, $\sum_{i=1}^\infty (x_i/s_i)^2 (|\log i|^2 = \infty$, and hence by a theorem of Tandori
there exists a sequence \( \{e_n\} \) of orthogonal random variables such that \( Ee_n^2 = 1 \) for all \( n \) and

\[
\sum_{i=1}^{\infty} (x_i e_i / s_i) \text{ is everywhere divergent.} \quad (4.5)
\]

By Lemma 3 below, (4.5) implies that \( \left( \sum_{i=1}^{n} x_i e_i \right) / s_n \) diverges a.s., and therefore in view of (4.4), \( b_n \) diverges a.s.

**Lemma 3.** Let \( \{u_n\} \) be a sequence of random variables such that \( E|u_n| < \infty \) and \( E(u_n u_k) = 0 \) for all \( n \) and for all \( k \neq n \). Let \( U_n = \sum_{i=1}^{n} u_i \) and \( s_n = \sum_{i=1}^{n} E u_i^2 \). Assume that

\[
s_m > 0 \quad \text{and} \quad \lim_{n \to \infty} s_n = \infty. \quad (4.6)
\]

Then

\[
\sum_{n=m}^{\infty} (s_n^{-1} - s_{n+1}^{-1}) U_n \text{ converges a.s.} \quad (4.7)
\]

Consequently, there exists an event \( \Omega_0 \) such that \( P(\Omega_0) = 1 \) and

\[
\left\{ \sum_{n=m}^{\infty} (u_n / s_n) \text{ converges} \right\} \cap \Omega_0 = \{ U_n / s_n \to 0 \} \cap \Omega_0, \quad (4.8a)
\]

\[
\left\{ \sum_{n=m}^{\infty} (u_n / s_n) \text{ diverges} \right\} \cap \Omega_0 = \{ U_n / s_n \text{ diverges} \} \cap \Omega_0. \quad (4.8b)
\]

**Proof.** Equation (4.7) follows from

\[
E \left\{ \sum_{n=m}^{\infty} (s_n^{-1} - s_{n+1}^{-1}) |U_n| \right\}
\leq \sum_{n=m}^{\infty} \{(s_n^{-1} - s_{n+1}^{-1}) s_n^{1/2}\}, \quad \text{since } EU_n^2 = s_n,
\]

\[
= \sum_{n=m}^{\infty} \{(s_{n+1} - s_n) / (s_{n+1} s_n^{1/2})\}
\]

\[
< \infty, \quad \text{by Pringsheim's theorem.}
\]

Letting \( \Omega_0 = \{ \sum_{i=m}^{\infty} (s_i^{-1} - s_{i+1}^{-1}) U_i \text{ converges} \} \), \( P(\Omega_0) = 1 \) by (4.7). On \( \Omega_0 \), since

\[
\sum_{i=m}^{n} u_i / s_i = \frac{U_n}{s_n} - \frac{U_{m-1}}{s_m} + \sum_{i=m}^{n-1} (s_i^{-1} - s_{i+1}^{-1}) U_i,
\]
the convergence of $\sum_{i=m}^{\infty} u_i/s_i$ is equivalent to the convergence of $U_n/s_n$. In this case, we must also have $U_n/s_n \to 0$ by Kronecker's lemma.

The following corollary of Theorem 3 extends the results of Chen [6] dealing with independent errors $e_i$ such that $Ee_i = 0$ and $\sup_i E |e_i|^r < \infty$ for some $1 \leq r < 2$. It also simplifies his conditions on the design constants $x_{ij}$.

**Corollary 3.** With the same notation as in Theorem 1, let $\{e_n\}$ be a martingale difference sequence such that $\sup_n E |e_n|^r < \infty$ for some $1 \leq r < 2$. If

$$v_{jj}^{(n)} = O(n^{-(2-r)/r} (\log n)^{-\delta})$$  \hspace{1cm} (4.9)

for some $\delta > 2/r$, then $b_{nj} \to \beta_j$ a.s.; in fact,

$$b_{nj} - \beta_j = o(n^{(2-r)/r} |v_{jj}^{(n)}| \log v_{jj}^{(n)})^{1/2} \text{ a.s.}$$  \hspace{1cm} (4.10)

Corollary 3 follows from Theorem 3 and the following lemma with $g_n = n^{-(2-r)/2r} (\log n)^{-d/2}$ and $f(t) = t |\log t|^{d}$, where $d + \Delta = \delta$ and $d > (2 - r)/r$, $\Delta > 1$.

**Lemma 4.** Let $1 \leq r < 2$. Let $\{e_n\}$ be a martingale difference sequence such that $\sup_n E |e_n|^r < \infty$. Let $\{g_n\}$ be a sequence of constants such that

$$\sum_{i=1}^{\infty} |g_n|^{2r(2-r)} < \infty.$$

Then $\{g_n e_n\}$ is a convergence system.

**Proof.** Let $\{a_n\}$ be a sequence of constants such that $\sum_{i=1}^{\infty} a_n^2 < \infty$. By the H"older inequality,

$$\sum_{i=1}^{\infty} |a_n g_n|^{r} \leq \left( \sum_{i=1}^{\infty} a_n^2 \right)^{r/2} \left( \sum_{i=1}^{\infty} |g_n|^{2r(2-r)} \right)^{(2-r)/2} < \infty.$$

Hence $\sum_{i=1}^{\infty} E |a_n g_n e_n|^{r} < \infty$, and therefore $\sum_{i=1}^{\infty} a_n g_n e_n$ converges a.s. by the convergence property of martingales [7, Corollary 5].

In the case where the $e_n$ are i.i.d., the assumption $E |e_n| < \infty$ needed in Corollary 3 can be dropped; moreover, the requirement that $\delta > 2/r$ in Corollary 3 can be weakened to $\delta > 1$. This is the content of the more general result which deals with the strong consistency of the least squares estimate in the i.i.d. model with infinite second moment.

**Corollary 4.** Let $a, b > 0$ and let $F: [a, \infty) \to [b, \infty)$ be such that

$$F(x) \uparrow \infty \text{ and } x^2/F(x) \uparrow \infty \text{ as } x \uparrow \infty. \hspace{1cm} (4.11)$$
Let $G: [b, \infty) \rightarrow [a, \infty)$ denote the inverse $F^{-1}$ of the function $F$. Let $\varepsilon, \varepsilon_1, \ldots$ be i.i.d. such that $E\{F(|\varepsilon| \vee a)\} < \infty$, where $x \vee y = \max\{x, y\}$.

(i) If $\varepsilon$ is symmetric, then

$$\{n^{1/2}\varepsilon_n/G(n)\} \text{ is a convergence system.}$$

(ii) If $E \varepsilon = 0$ and $\sum_{b < k < n} k/G^2(k) = O(n^2/G^2(n))$, then (4.12) still holds.

(iii) If $\sum_{k=n}^{\infty} k/G^2(k) = O(n^2/G^2(n))$, then (4.12) still holds.

(iv) With the same notation as in Theorem 1, if (4.12) holds and

$$v_{jj}^{(n)} = O \left( \frac{n}{G^2(n)} \left| \log \frac{n}{G^2(n)} \right|^{\delta} \right)$$

for some $\delta > 1$, then $b_{nj} \rightarrow \beta_j$ a.s.; in fact,

$$b_{nj} - \beta_j = o\left( |n^{-1}G^2(n) v_{jj}^{(n)} | \log v_{jj}^{(n)} |^{\delta/2} \right) \text{ a.s.} \quad (4.14)$$

Remark. In the case $F(x) = x^r$ (so that $E |\varepsilon|^r < \infty$) with $0 < r < 2$, (4.11) holds and $G(x) = x^{1/r}$. Therefore (4.13) reduces to (4.9) in this case; however, we only require $\delta > 1$ (instead of $\delta > 2/r$ in Corollary 3). Note that for the case $r < 1$ (so that $(2 - r)/r > 1$),

$$\sum_{k=n}^{\infty} k/G^2(k) = \sum_{k=n}^{\infty} k^{-(2-r)/r} = O(n^2/G^2(n)),$$

while for the case $r > 1$, we have

$$\sum_{k=1}^{n} k/G^2(k) = \sum_{k=1}^{n} k^{-(2-r)/r} = O(n^2/G^2(n)).$$

Another interesting application of Corollary 4 is the following.

**Example.** Let $\varepsilon$ have the Cauchy density $\psi(x) = |\pi(1 + x^2)|^{-1}$, $-\infty < x < \infty$. Then for every $p > 1$, $E\{F_p(|\varepsilon|)\} < \infty$, where

$$F_p(x) = x/|\log(1 + x)|^p, \ x > 0; \quad F_p(0) = 0.$$

Clearly $F_p$ satisfies (4.11). Since $\varepsilon$ is symmetric, it follows from Corollary 4(i) and (iv) that a sufficient condition for the strong consistency of $b_{nj}$ in the regression model (1.1) with i.i.d. Cauchy errors $\varepsilon_i$ is

$$v_{jj}^{(n)} = O(n^{-1}(\log n)^{-\delta}) \quad \text{for some } \delta > 3. \quad (4.15)$$
Proof of Corollary 4. Condition (4.11) can be expressed as
\[ G(y) \uparrow \infty \text{ and } y/G^2(y) \downarrow 0 \quad \text{as } y \uparrow \infty. \] (4.16)

Since \( \sum_{n \geq b} P[|\varepsilon_n| > G(n)] \leq \sum_{n \geq b} P[F(|\varepsilon| > n)] \leq E[F(|\varepsilon| \vee a)] < \infty \), it follows from the Borel–Cantelli lemma that \( P[|\varepsilon_n| > G(n) \text{ i.o.}] = 0 \). Therefore to prove that \( \{n^{1/2}\varepsilon_n/G(n)\} \) is a convergence system, it suffices to show that for \( m \geq b \),
\[
\sum_{n=m}^{\infty} a_n(n^{1/2}/G(n)) \varepsilon_n I_{[|\varepsilon_n| < G(n)]} \text{ converges a.s.} \quad (4.17)
\]
for every sequence of constants \( a_n \) such that \( \sum_{n=1}^{\infty} a_n^2 < \infty \). Defining \( G(x) = a \) for \( x < b \), we note that
\[
\sum_{n=m}^{\infty} a_n^2(n/G^2(n)) E\varepsilon_n^2 I_{[|\varepsilon_n| < G(n)]} \leq \sum_{n=m}^{\infty} a_n^2(n/G^2(n)) \sum_{k=1}^{n} G^2(k) P[k - 1 < F(|\varepsilon| \vee a) \leq k] = \sum_{k=1}^{\infty} G^2(k) P[k - 1 < F(|\varepsilon| \vee a) \leq k] \left\{ \sum_{n=k}^{\infty} a_n^2(n/G^2(n)) \right\} = \left( \sum_{n=1}^{\infty} a_n^2 \right) \sum_{k=1}^{\infty} O(kP[k - 1 < F(|\varepsilon| \vee a) \leq k]), \quad \text{by (4.16)}. \]

In view of the finiteness of the above series and Kolmogorov's three-series theorem, (4.17) holds if it can be shown that
\[
\sum_{n=m}^{\infty} a_n(n^{1/2}/G(n)) E\varepsilon I_{[|\varepsilon| < G(n)]} \text{ converges.} \quad (4.18)
\]

If \( \varepsilon \) is symmetric, (4.18) is trivial and therefore we have proved (i). To prove (ii), since \( E\varepsilon = 0 \), (4.18) follows from
\[
\sum_{n=m}^{\infty} |a_n|(n^{1/2}/G(n)) E|\varepsilon| I_{[|\varepsilon| > G(n)]} \leq \sum_{k=m}^{\infty} G(k + 1) P[k < F(|\varepsilon| \vee a) \leq k + 1] \left\{ \sum_{n=m}^{k} |a_n| n^{1/2}/G(n) \right\} < \infty,
\]
noting that \( \sum_{n=m}^{k} |a_n| n^{1/2}/G(n) \leq (\sum_{n=1}^{\infty} a_n^2)^{1/2} (\sum_{n=m}^{k+1} n/G^2(n))^{1/2} = O(k/G(k + 1)). \)
To prove (iii), we obtain (4.17) from

\[
\sum_{n=m}^{\infty} |a_n| \frac{n^{1/2}G(n)}{E|\varepsilon| I_{[n^{1/2}G(n)]}} \leq \sum_{k=1}^{\infty} G(k) P[k - 1 < F(|\varepsilon| \vee a) \leq k] \left\{ \sum_{n=k}^{\infty} |a_n| n^{1/2}G(n) \right\} < \infty,
\]
noting that \(\sum_{n=k}^{\infty} |a_n| n^{1/2}G(n) \leq (\sum_{n=k}^{\infty} a_n^2) \frac{1}{2} (\sum_{n=k}^{\infty} n/G^2(n))^{1/2} = O(k/G(k)).\) From (4.12) and Theorem 3, (iv) follows.

In time series analysis, a general class of random errors \(\varepsilon_i\) for the regression model (1.1) is defined by moving averages of the form \(\varepsilon_n = \sum i = -\infty c_{n-i} \varepsilon_i,\) where \(\{\varepsilon_i\}\) is a martingale difference sequence such that \(E\varepsilon_i = \sigma^2 < \infty\) for all \(i\) and \(\{c_i\}\) is a sequence of constants such that \(\sum_{i=-\infty}^{\infty} c_i^2 < \infty.\) This is called a linear process generated by the sequence \(\{\varepsilon_i\},\) and is wide-sense stationary. In [11, p. 358], it is shown that if \(\{\varepsilon_n\}\) is a stationary Gaussian sequence with zero means and covariance function \(\rho(k) = E\varepsilon_1 \varepsilon_{1+k}\) such that

\[
|\rho(k)| \text{ is non-increasing and } \sum_{k=1}^{\infty} |\rho(k)| < \infty, \quad (4.19)
\]
then \(\{\varepsilon_n\}\) is a convergence system and therefore Theorem 1 can be applied to show the strong consistency of \(b_{nj}\) under the assumption that \(v_{ij}^{(n)} \to 0.\) As is well known, such a Gaussian sequence \(\{\varepsilon_n\}\) can be expressed as a linear process generated by an i.i.d. sequence of standard normal random variables \(u_n,\) and \(\{\varepsilon_n\}\) has a continuous spectral density (cf. [12]). The following corollary of Theorem 3 replaces assumption (4.19) by a weaker condition on the spectral density of \(\{\varepsilon_n\}\) and also replaces the restrictive Gaussian model by a general linear process generated by a martingale difference sequence.

**Corollary 5.** With the same notation as in Theorem 1, let \(\{\varepsilon_n\}\) be a linear process generated by a martingale difference sequence ..., \(u_{-1}, u_0, u_1, \ldots\) such that \(E\varepsilon_n^2 = \sigma^2 (< \infty)\) for all \(n.\) Then \(\{\varepsilon_n\}\) is wide-sense stationary and has a spectral density \(\psi.\)

\(\psi(\theta) < \infty.\) If \(\lim_{n \to \infty} v_{ij}^{(n)} = 0,\) then \(b_{nj} \to \beta_j\) a.s.

**Assume that for some \(r > 1, \int_0^{2\pi} \psi^r(\theta) d\theta < \infty and**

\[
v_{ij}^{(n)} = O(n^{-1/2}(log n)^{-\delta}) \quad \text{for some } \delta > 1 + r^{-1}. \quad (4.20)
\]
Then \(b_{nj} \to \beta_j\) a.s.
The above result is an immediate consequence of Theorem 3 and the following lemma.

**Lemma 5.** With the same notation as in Corollary 5, let \( \{g_n\} \) be a sequence of real constants.

(i) If \( \sup_n E|u_n|^p < \infty \) for some \( p > 2 \) and \( \text{ess sup}_{0<\theta<2\pi} \psi(\theta) < \infty \), then \( \{e_n\} \) is a convergence system.

(ii) If for some \( r \geq 1 \), \( \int_0^{2\pi} \psi(\theta) \, d\theta < \infty \) and \( \sum_{i=1}^{\infty} |g_n|^{2r} < \infty \), then \( \{g_n e_n\} \) is a convergence system.

**Proof:** Part (i) has been established in [12]. To prove (ii), as has been shown in [12], the condition \( \int_0^{2\pi} \psi(\theta) \, d\theta < \infty \) implies that \( \sum_{i=1}^{\infty} c_n e_n \) converges a.s. for every sequence of constants \( c_n \) such that

\[
\sum_{i=1}^{\infty} |c_n|^{2r(r+1)} < \infty. \tag{4.21}
\]

Let \( a_n \) be a sequence of constants such that \( \sum_{i=1}^{\infty} a_n^2 < \infty \). By the Hölder inequality,

\[
\sum_{i=1}^{\infty} |a_n g_n|^{2r(r+1)} \leq \left( \sum_{i=1}^{\infty} a_n^2 \right)^{r/(r+1)} \left( \sum_{i=1}^{\infty} |g_n|^{2r} \right)^{1/(r+1)} < \infty,
\]

and therefore (4.21) holds with \( c_n = a_n g_n \). Hence \( \sum_{i=1}^{\infty} a_n g_n e_n \) converges a.s. \( \blacksquare \)

**References**


