The Rate of Convergence of the Least Squares Estimator in a Non-Linear Regression Model with Dependent Errors

B. L. S. Prakasa Rao

Indian Statistical Institute, New Delhi

Communicated by E. J. Hannan

The rate of convergence of the least squares estimator in a non-linear regression model with errors forming either a φ-mixing or strong mixing process is obtained. Strong consistency of the least squares estimator is obtained as a corollary.

1. Introduction

Following the work of Jennrich [6], strong consistency properties of the least squares estimator in a non-linear regression model with dependent errors have been investigated by Hannan [3] and Nelson [7] among others. Our approach is similar to that of Ivanov [5], who discussed asymptotic expansions for the distribution of the least squares estimator when the errors form an i.i.d. sequence.

2. Preliminaries

Let \( \{ \varepsilon_j, -\infty < j < \infty \} \) be a stochastic process defined on a probability space \((\Omega, \mathcal{F}, P)\). The process is said be \( \phi \)-mixing if

\[
\phi(n) = \sup_k \sup_{A \in \mathcal{F}_k, B \in \mathcal{F}_{k+n}} \frac{1}{P(A)} |P(A \cap B) - P(A) P(B)|
\]

decreases to zero as \( n \to \infty \) and it is said to be strong mixing if

\[
a(n) = \sup_k \sup_{A \in \mathcal{F}_k, B \in \mathcal{F}_{k+n}} |P(A \cap B) - P(A) P(B)|
\]
decreases to zero as \( n \to \infty \), where \( \mathcal{F}_a^b \) denotes the \( \sigma \)-algebra generated by \( \varepsilon_j \), \( a \leq j \leq b \).

It is clear that if the process is \( \phi \)-mixing, then it is strong mixing.

**Lemma 1.** Let \( \{ \varepsilon_j, -\infty < j < \infty \} \) be \( \phi \)-mixing process and \( \xi \) and \( \eta \) be measurable with respect to \( \mathcal{F}^{-\infty}_k \) and \( \mathcal{F}_{k+1}^\infty \), respectively. If \( E|\xi|^p < \infty \) and \( E|\eta|^q < \infty \) where \( p > 1, q > 1, 1/p + 1/q = 1 \), then

\[
|E\xi\eta - E\xi E\eta| \leq 2(\phi(n))^{1/p} |E|\xi|^p|^{1/p} |E|\eta|^q|^{1/q}.
\]

**Lemma 2.** Let \( \{ \varepsilon_j, -\infty < j < \infty \} \) be \( \phi \)-mixing process satisfying the following conditions:

(A1) \( E\varepsilon_i = 0 \) and \( \sup_i E|\varepsilon_i|^4 \leq M < \infty \),

(A2) \( \sum_{i=1}^\infty (i+1)|\phi(i)|^{1/4} < \infty \).

Then, for every sequence of real numbers \( \{ a_k \} \) and for every integer \( n \),

\[
E\left[ \sum_{i=b+1}^{b+n} a_i \varepsilon_i \right]^4 \leq C \left[ \sum_{i=b+1}^{b+n} a_i^2 \right]^2
\]

for all \( b \geq 0 \) and \( n \geq 1 \), where \( C \) is an absolute constant.

Lemma 1 is proved in Ibragimov and Linnik [4] and Lemma 2 follows from Theorem 1 in Yoshihara [10]. It is easy to see that

\[
E \left( \sum_{j=1}^n \varepsilon_j^2 \right)^2 = O(n^2)
\]

under (A1) by Cauchy–Schwarz inequality.

Analogous lemmas for strong mixing processes can be proved under the condition that there exists \( \delta > 0 \) such that

(A1)' \( E\varepsilon_i = 0 \) and \( \sup_i E|\varepsilon_i|^{4+2\delta} \leq M < \infty \), and

(A2)' \( \sum_{i=1}^\infty (i+1)|\phi(i)|^{3/(4+\delta)} < \infty \)

using results in Davydov [1] and Yoshihara [10]. We omit them.

### 3. Main Result

Consider the non-linear regression model

\[
X_n = g_n(\theta) + \varepsilon_n, \quad n \geq 1,
\]

where \( \{ g_n(\theta) \} \) is a sequence of continuous functions possibly non-linear in
Suppose \( \{ \varepsilon_n \} \) is a \( \phi \)-mixing process satisfying conditions (A1)–(A2). Let

\[
Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} w_i^2 (X_i - g_i(\theta))^2,
\]

where \( \{ w_i \} \) is a known sequence of positive numbers. An estimator \( \hat{\theta}_n \) is said to be a least squares estimator of \( \theta \) if it minimizes \( Q_n(\theta) \) over \( \theta \in \Theta \).

Note that \( Q(x_1, \ldots, x_n; \theta) = Q_n(\theta) \) is defined on \( \mathbb{R}^n \times \Theta \), where \( \Theta \) is compact. Further \( Q(x; \theta) \), where \( x = (x_1, \ldots, x_n) \) is a Borel measurable function of \( x \) for any fixed \( \theta \in \Theta \) and continuous function of \( \theta \) for any fixed \( x \in \mathbb{R}^n \).

Lemma 3.3 in Schmetterer [8, p. 307] shows that there exists a Borel-measurable map \( \theta_n : \mathbb{R}^n \to \Theta \) such that \( Q(x; \theta_n(x)) = \inf(\theta \in \Theta : Q(x; \theta)) \).

Hereafter we consider this measurable version as the least squares estimator \( \theta_n \).

Let \( \theta_0 \) be the true parameter and suppose \( \theta_0 \in \text{Interior of } \Theta \). Define

\[
\psi_n(\theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^{n} w_i^2 (g_i(\theta_1) - g_i(\theta_2))^2.
\]

**Theorem.** Suppose \( \{ \varepsilon_i \} \) is a \( \phi \)-mixing process satisfying conditions (A1)–(A2). In addition suppose that

(A3) there exists constants \( k_1 > 0, k_2 < \infty \) such that

\[
k_1 (\theta_1 - \theta_2)^2 \leq \psi_n(\theta_1, \theta_2) \leq k_2 (\theta_1 - \theta_2)^2
\]

for all \( n \geq 1 \) and \( \theta_1, \theta_2 \) in \( \Theta \).

and

(A4) \( \sup_n \{ w_n \} = O(1) \).

Then, there exists a constant \( c > 0 \) such that

\[
P(\sqrt{n} | \theta_n - \theta_0 | > \rho) \leq c \rho^{-4}
\]

for every \( \rho > 0 \) and for all \( n \geq 1 \).

**Proof.** For simplicity, we will assume that \( w_i = 1 \) for all \( i \). The general case follows from similar arguments in view of (A4). Observe that \( \psi_n(\theta, \theta_0) > 0 \) for every \( n \geq 1 \) if \( \theta \neq \theta_0 \) by (A3) and \( \psi_n(\theta_0, \theta_0) = 0 \).
Let
\[ A_{nc} = \left( |\theta_n - \theta_0| > \varepsilon \right), \]
\[ U_n(\theta) = \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \left( \frac{g_j(\theta) - g_j(\theta_0)}{\psi_n(\theta, \theta_0)} \right), \quad \theta \neq \theta_0, \]
\[ V_n(\theta) = \frac{1}{n^{1/2}} \sum_{j=1}^{n} \varepsilon_j \{ g_j(\theta) - g_j(\theta_0) \}. \]

Note that, for \( \theta \neq \theta_0 \),
\[ U_n(\theta) = (n^{1/2} \psi_n(\theta, \theta_0))^{-1} V_n(\theta). \]

Suppose the event \( A_{nc} \) occurs. Then \( \theta_n \neq \theta_0 \) and by the definition of \( \theta_n \),
\[ \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (X_i - g_i(\theta_0))^2 \]
\[ \geq \sum_{i=1}^{n} (X_i - g_i(\theta_n))^2 \]
\[ = \sum_{i=1}^{n} \varepsilon_i^2 + n\psi_n(\theta_n, \theta_0) - 2nU_n(\theta_n)\psi_n(\theta_n, \theta_0) \]
and hence
\[ \psi_n(\theta_n, \theta_0)(1 - 2U_n(\theta_n)) \leq 0. \quad (6) \]

Note that \( \psi_n(\theta_n, \theta_0) > 0 \). In view of (6), it follows that \( U_n(\theta_n) \geq \frac{1}{2} \). Therefore
\[ P(A_{nc}) = P( |\theta_n - \theta_0| > \varepsilon ) \]
\[ \leq P\{ U_n(\theta_n) \geq \frac{1}{2}, |\theta_n - \theta_0| > \varepsilon \} \]
\[ \leq P\{ \sup_{|\theta - \theta_0| > \varepsilon} U_n(\theta) \geq \frac{1}{2} \} \]
for every \( \varepsilon > 0 \). In particular, it follows that
\[ P\{ n^{1/2} |\theta_n - \theta_0| > \rho \} \]
\[ \leq P\{ \sup_{|\theta - \theta_0| > \rho} U_n(\theta) \geq \frac{1}{2} \} \]
\[ \leq P\{ \sup_{|\theta - \theta_0| > \rho} U_n(\theta) \geq \frac{1}{2} \} + P\{ \sup_{\rho^{-1/2} \leq |\theta - \theta_0| \leq \rho} U_n(\theta) \geq \frac{1}{2} \} \quad (7) \]
for every $\rho > 0$. Now

\[
P \left\{ \sup_{|\theta - \theta_0| > \rho} |U_n(\theta)| \geq \frac{1}{2} \right\} = P \left\{ \sup_{|\theta - \theta_0| > \rho} \frac{|V_n(\theta)|}{n^{1/2} \psi_n(\theta, \theta_0)} \geq \frac{1}{2} \right\} \leq P \left\{ \sup_{|\theta - \theta_0| > \rho} \frac{|V_n(\theta)|}{n^{1/2} \psi_n^{1/2}(\theta, \theta_0)} \geq \frac{1}{2} k_1^{1/2} \rho \right\}.
\]

But

\[
\left( \frac{V_n(\theta)}{n^{1/2} \psi_n^{1/2}(\theta, \theta_0)} \right)^2 = \left\{ \frac{1}{n} \sum_{j=1}^n \varepsilon_j \left[ \frac{g_j(\theta) - g_j(\theta_0)}{\psi_n^{1/2}(\theta, \theta_0)} \right] \right\}^2 \leq \frac{1}{n} \sum_{j=1}^n \varepsilon_j^2.
\]

Therefore

\[
P \left\{ \sup_{|\theta - \theta_0| > \rho} |U_n(\theta)| \geq \frac{1}{2} \right\} \leq P \left\{ \frac{1}{n} \sum_{j=1}^n \varepsilon_j^2 \geq \frac{1}{4} k_1 \rho^2 \right\} \leq \frac{4^2}{k_1^2 \rho^4} \frac{1}{n^2} E \left\{ \sum_{j=1}^n \varepsilon_j^2 \right\}^2 \leq \frac{4^2}{k_1^2 \rho^4} \frac{1}{n^2} n^2 c_1
\]

for some constant $c_1 > 0$ by (1). Let

\[m\theta = \theta_0 + \frac{\rho}{n^{1/2}} + \frac{m\rho}{|n^{1/2}|}, \quad \rho_m = m\theta - \theta_0,\]

for $m = 0, 1, \ldots, |n^{1/2}|$. Then

\[
P \left\{ \sup_{\rho_m \leq \rho \leq \rho_{m+1}} |U_n(\theta)| \geq \frac{1}{2} \right\} \leq \sum_{m=0}^{|n^{1/2}|-1} P \left\{ \sup_{\rho_m \leq \theta - \theta_0 \leq \rho_{m+1}} |U_n(\theta)| \geq \frac{1}{2} \right\} \leq \sum_{m=0}^{|n^{1/2}|-1} P \left\{ \sup_{\rho_m \leq \theta - \theta_0 \leq \rho_{m+1}} |V_n(\theta)| \geq \frac{1}{2} k_1 \rho_m^2 n^{1/2} \right\}.
\]

Now

\[
P \left\{ \sup_{\rho_m \leq \theta - \theta_0 \leq \rho_{m+1}} |V_n(\theta)| \geq \frac{1}{2} k_1 \rho_m^2 n^{1/2} \right\} \leq P \left\{ |V_n(\theta)| \geq \frac{1}{4} k_1 \rho_m^2 n^{1/2} \right\} + P \left\{ \sup_{m\theta \leq \theta_1, \theta_2 \leq m+1 \theta} |V_n(\theta_1) - V_n(\theta_2)| \geq \frac{1}{4} k_1 \rho_m^2 n^{1/2} \right\}
\]

(10)
Observe that

\[ \frac{V_n(\theta)}{\rho_m} = \frac{1}{n^{1/2}} \sum_{j=1}^{n} e_j \left\{ \frac{g_j(\theta) - g_j(\theta_0)}{\rho_m} \right\} \]

and hence

\[ E \left[ \frac{V_n(\theta)}{\rho_m} \right] \leq c_2 \left( \sum_{j=1}^{n} \left\{ \frac{g_j(\theta) - g_j(\theta_0)}{\rho_m n^{1/2}} \right\}^2 \right)^{1/2} \]

\[ = \frac{c_2}{\rho_m} \left( \psi_n(\theta, \theta_0) \right)^2 \]

for some absolute constant \( c_2 > 0 \) independent of \( n \) by Lemma 2. Therefore

\[ E \left[ \frac{V_n(\theta)}{\rho_m} \right] \leq \frac{c_2}{\rho_m} k_2^2 (\theta - \theta_0)^4 \leq c_3 k_2^2 \]  

(12)

and

\[ P \left\{ \frac{|V_n(\theta)|}{\rho_m} \geq \frac{1}{4} k_1 \rho_m n^{1/2} \right\} \leq \frac{4^4}{k_1^4 \rho_m^4 n^2} c_2 k_2^2. \]  

(13)

On the other hand,

\[ E \left[ V_n(\theta_1) - V_n(\theta_2) \right]^4 = \frac{1}{n^2} E \left\{ \sum_{j=1}^{n} e_j (g_j(\theta_1) - g_j(\theta_2)) \right\}^4 \]

\[ \leq \frac{c_2}{n^2} \left\{ \sum_{j=1}^{n} (g_j(\theta_1) - g_j(\theta_2))^2 \right\}^2 = c_3 \left( \psi_n(\theta_1, \theta_2) \right)^2 \]

\[ \leq c_3 k_2^4 (\theta_2 - \theta_1)^4 \]  

(14)

for every \( n \geq 1 \). Then \( \{V_n(\theta), \theta \in [m \theta, m+1 \theta]\}, n \geq 1 \) can be considered as a family of stochastic processes with continuous sample paths in \([m \theta, m+1 \theta]\). The above inequality implies that (cf. Gikhman and Skorokhod [2, p. 192] or Stroock and Varadhan [9, p. 49])

\[ P \left\{ \sup_{m \theta \leq \theta_1, \theta_2, m+1 \theta} \left| V_n(\theta_1) - V_n(\theta_2) \right| \geq \frac{1}{4} k_1 \rho_m^2 n^{1/2} \right\} \leq \frac{c_4}{\rho_m^4 n^2} \left( \frac{\rho}{n^{1/2}} \right)^4. \]  

(15)
LEAST SQUARES ESTIMATOR

Note that $p_0 = \rho n^{-1/2}$ and $\rho_m > m \rho / n^{1/2}$. Combining inequalities (9)–(15), we have

$$P \left( \sup_{\rho n^{-1/2} < \theta - \theta_0 < \rho} |U_n(\theta)| \geq \frac{1}{2} \right) \leq c_s \rho^{-4}$$  \hspace{1cm} (16)

for some constant $c_s > 0$ independent of $n$ and $\rho$. Similar inequality holds when $\rho n^{-1/2} < \theta_0 - \theta < \rho$. Inequalities (7), (8) and (16) prove that

$$P(\rho^{1/2} |\theta_n - \theta_0| > \rho) \leq c \rho^{-4}$$  \hspace{1cm} (17)

for some constant $c > 0$ independent of $n$ and $\rho$. This proves the main theorem.

Let $\rho = \rho_n$ where $\frac{1}{4} < \gamma < \frac{1}{2}$. As a consequence of relation (17), it follows that

$$\sum_{n=1}^{\infty} P(\rho^{1/2} |\theta_n - \theta_0| > \rho_n) < \infty.$$  \hspace{1cm} (18)

An application of Borel–Cantelli lemma proves that

$$\rho^{1/2} |\theta_n - \theta_0| \leq \rho_n = n^\gamma \text{ a.s.}$$

for large $n$ and hence $\theta_n \rightarrow \theta_0$ a.s. In fact

$$\theta_n - \theta_0 = o(n^{-1/4 + \epsilon}) \text{ a.s.}$$  \hspace{1cm} (19)

for every $0 < \epsilon < \frac{1}{4}$.

4. REMARKS

Assumption (A3) is not a strong restriction, for if $g_i(\theta)$ is linear in $\theta$ (say)

$$g_i(\theta) = \alpha_i + \beta_i(\theta),$$

then (A3) holds provided

$$0 < \inf \frac{1}{n} \sum_{i=1}^{n} w_i^2 \beta_i^2 \leq \sup \frac{1}{n} \sum_{i=1}^{n} w_i^2 \beta_i^2 < \infty$$

and in general if $g_i(\theta)$ is non-linear in $\theta$ and differentiable with respect to $\theta$

with derivative $g_i'(\theta)$, then (A3) holds if

$$0 < \inf \frac{1}{n} \sum_{i=1}^{n} w_i^2 \inf_{\theta \in \Theta} |g_i'(\theta)|^2$$

$$\leq \sup \frac{1}{n} \sum_{i=1}^{n} w_i^2 \sup_{\theta \in \Theta} |g_i'(\theta)|^2 < \infty$$
and the last relation holds in turn if

\[ 0 < K_1 \leq \inf_{i} \inf_{\theta \in \Theta} \{ g_i'(\theta) \}^2 \leq \sup_{i} \sup_{\theta \in \Theta} \{ g_i'(\theta) \}^2 < K_2 < \infty \]

and \( 0 < \beta_1 \leq w_i \leq \beta_2 < \infty \) for all \( i \geq 1 \).

Let us consider the model

\[ X_n = \beta g(n - \theta) + e_n, \quad n \geq 1, |\theta| \leq 1, \]

which is a special case of example of a time series discussed in Hannan [3]. Clearly (A3) holds for this model provided \( g(\cdot) \) is differentiable and

\[ 0 < \inf_{x} g'(x) < \sup_{x} g'(x) < \infty, \]

and \( 0 < \beta_1 \leq w_i \leq \beta_2 < \infty \) for all \( i \geq 1 \).

It can be checked easily that the main theorem holds if \( \{e_n\} \) is a strong-mixing process under conditions (A1)', (A2)', (A3) and (A4). We will not give the proof as it is similar to the case of \( \phi \)-mixing process. The result can be extended to the multiparameter case by analogous arguments. It would be interesting to obtain Berry–Esseen type bounds for the distribution of least squares estimator when the errors form a dependent process. We will come back to this problem later.

References