A maximal moment inequality for long range dependent time series with applications to estimation and model selection

(Running Head: A maximal moment inequality with applications)

Ching-Kang Ing

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan

Department of Economics, National Taiwan University, Taipei 100, Taiwan

Fax: 886-2-27831523

cking@stat.sinica.edu.tw

Ching-Zong Wei

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan

Fax: 886-2-27831523

czw@stat.sinica.edu.tw

Abstract. We establish a maximal moment inequality for the weighted sum of a long-range dependent process. An extension to Hájek-Rény and Chow's type inequality is then obtained. It enables us to deduce a strong law for the weighted sum of a stationary long-range dependent time series. To illustrate its usefulness, applications of the inequality to estimation and model selection in multiple regression models with long-range dependent errors are given.

AMS 1991 subject classifications. Primary 60E15; secondary 62M10.

Key words and phrases. Autoregressive fractionally integrated moving average, long range dependence, maximal inequality, model selection, convergence system, strong consistency.

1. INTRODUCTION

The study of long range dependent time series has recently become a rapidly developing subject (cf. Beran, 1994, for a survey). As applications have become broader, the involved functionals have become increasingly complicated. In this paper, we give a unified result for the related maximal inequality for the weighted sum of a long range dependent time series. Applications to estimation and model selection in multiple linear regressions with long range dependent errors are then discussed.

To fix ideas, let $\{\varepsilon_t\}$ be a zero-mean covariance stationary process with

$$\sup_{0\leq k<\infty} |\gamma(k)|(k+1)^{\alpha} < \infty, \tag{1.1}$$

where $0 < \alpha < 1$ and $\gamma(k) = E(\varepsilon_1\varepsilon_{1+k})$. The process $\{\varepsilon_t\}$ is said to be long range dependent if there is a real number $0 < \alpha < 1$ and constant $C_0 > 0$ such that

$$\lim_{k\to\infty} \frac{\gamma(k)k^{\alpha}}{C_0} = 1; \tag{1.2}$$

see, also, Beran (1994, Chapter 2). Therefore, condition (1.1) is fulfilled not only by most stationary short memory time series (eg, autoregressive moving average model), but also by long range dependent time series.

The first general moment inequality for the weighted sum of ε_t was given by Yajima (1988, pages 796 and 806). In particular, he showed that if (1.1) holds then for some constant $k > 0$,

$$E\left\{\max_{1\leq i\leq n}\left|\sum_{t=m+1}^{m+i} c_t\varepsilon_t\right|^2\right\} \leq k\left(\frac{\log 4n}{\log 2}\right)^2 (2n)^{1-\alpha}(\sum_{t=m+1}^{m+n} c_t^2). \tag{1.3}$$

Although (1.3) enabled Yajima to develop his asymptotic results for the least squares estimate in multiple linear regression models with long range dependent errors, this inequality does not seem entirely satisfactory. This can be seen by observing that the term on the right-hand side of (1.3) goes to infinity as n does. This is apparently not sharp when $\{c_t\}$ are exponentially small. As shown in (2.10) of Section 2, the maximal moment of (1.3) is bounded. A related maximal probability inequality which can be applied to the change-point estimation problem is from Lavielle and Moulines (2000, Theorem 1). They showed that for any $n \geq 1$, any $\delta > 0$, and any positive and nonincreasing sequence $b_1 \geq b_2 \geq \ldots \geq b_n > 0$,

$$P\left(\max_{1\leq k\leq n} b_k\left|\sum_{t=1}^{k}\varepsilon_t\right| > \delta\right) \leq \frac{C_1 n^{1-\alpha}}{\delta^2}\sum_{t=1}^{n} b_t^2, \tag{1.4}$$

where C_1 represents some positive number independent of n and b_t. Recall that when $\{\varepsilon_t\}$ are independent random variables with $E(\varepsilon_t) = 0$ for all $1 \leq t \leq n$ and $\max_{1\leq t\leq n} E(\varepsilon_t^2) < \infty$, (1.4) was established by Hájek and Rényi (1955) with the exponent $1-\alpha$ on the right-hand side replaced by 0. Chow (1960) subsequently extended Hájek and Rényi's result to submartingale differences. For related extensions of Hájek-Rényi-Chow's type inequality to short memory linear processes, see Bai (1994). On the other hand, it should be noted that the term on the right-hand side of (1.4) goes to infinity regardless of how fast b_t decreases. This is obviously not a desirable property especially for a probability inequality.

In view of the above discussion, this paper attempts to provide sharper bounds for the left-hand sides of (1.3) and (1.4) through unified theory. In Section 2, utilizing inequalities due to Móricz (1976) and Hardy, Littlewood and Pólya (1952), we establish a maximal moment inequality (2.1) for weighted sums of random variables having finite second moments. By a monotone inequality of Shorack and Smythe (1976), this inequality is generalized to a Hájek-Rényi-Chow's type maximal moment inequality; see Corollary 2.4. When this result is applied to $\{\varepsilon_t\}$ (which satisfies (1.1)), sharper bounds for the left-hand sides of (1.3) and (1.4) are given in Remark 2. In addition, almost sure behaviors of the weighted sum of $\{\varepsilon_t\}$ are obtained in Corollary 2.6. As will be shown later, Corollary 2.6 plays a very important role in investigating asymptotic properties of least squares estimates in regression models with long range dependent errors.

Sections 3-4 are devoted to various applications of the results obtained in Section 2. In Section 3, we first develop a strong consistency result for the least squares estimate in a multiple regression model with the assumption that the error structure is a "contracted" convergence system. We then apply the general theory to multiple regression models with error terms satisfying (1.1). It is shown that our convergence result requires less stringent conditions than those of Yajima (1988); see Remark 5 after Corollary 3.4. In addition, the strong consistency of the residual mean squared error is also established under rather mild assumptions.

When the model considered in Section 3 contains some possibly redundant variables, dropping these variables can increase estimation precision and yield more efficient statistical inferences. This motivates us to study model selection problems in Section 4. A model selection criterion is said to be strongly consistent if it can ultimately choose the most parsimonious correct model with probability 1; see (4.3) for a more precise definition. Note that strong consistency selections for multiple regression models with martingale difference or short memory time series errors have been obtained by several authors, including Wei (1992) and Chen and Ni (1989). For long range dependent errors, however, similar results still seem to be lacking. To fill this gap, we show in Theorem 4.1 that an information criterion with a penalty for larger models stronger than that of BIC (Schwarz, 1978) is strongly consistent in multiple regression models with errors satisfying (1.1). Since, as mentioned previously, (1.1) is fulfilled by both short and long memory time series, this result is especially useful in situations where the strength of dependence of $\{\varepsilon_t\}$ is unknown; see discussion given at the end of Section 4 for more details. Moreover, Remark 9 of Section 4 also provides a counterexample to illustrate that BIC is not consistent when the special case (1.2) is fulfilled by the error terms.

2. Maximal moment inequalities.

In this section, we begin with a maximal moment inequality for weighted sums of random variables having finite second moments.

Theorem 2.1. *Let $\{f_i\}$ be a sequence of random variables with $E(f_i^2) < \infty$ for $1 \le i \le n$. Assume $0 < \alpha < 1$. Then, for any sequence of real numbers $c_1, \ldots, c_n$,*

$$E(\max_{1\le i\le n} |\sum_{j=1}^{i} c_j f_j|^2) \le k_\alpha \{\max_{0\le k\le n-1} \gamma_n(k)(k+1)^\alpha\} (\sum_{i=1}^{n} |c_i|^{2/(2-\alpha)})^{2-\alpha}, \tag{2.1}$$

where k_α is a constant depending on α only and

$$\gamma_n(k) = \max_{\substack{|i-j|=k \\ 1\le i\le j\le n}} |E(f_i f_j)|.$$

To show (2.1), two auxiliary inequalities are needed. The first one is due to Hardy, Littlewood, and Pólya (1952).

Lemma 2.2. *Given two sequences of real numbers $a_i \ge 0$ and $b_i \ge 0, i = 1, \ldots, n$, if $p > 1, q > 1, \frac{1}{p} + \frac{1}{q} > 1$, and $\delta = 2 - \frac{1}{p} - \frac{1}{q}$, then*

$$\sum_{i\ne j} \left(\frac{a_i b_j}{|i-j|^\delta} \right) \le k_{p,q} \left(\sum_{i=1}^{n} a_i^p \right)^{1/p} \left(\sum_{i=1}^{n} b_i^q \right)^{1/q},$$

where $k_{p,q}$ is a positive constant depending only on p and q.

The second auxiliary inequality is a moment inequality from Móricz (1976).

Lemma 2.3. *Let $p > 0$ and $q > 1$ be two positive real numbers and Z_i be a sequence of random variables. Assume that there are nonnegative constants a_j satisfying*

$$E|\sum_{j=1}^{i} Z_j|^p \le (\sum_{j=1}^{i} a_j)^q,$$

for all $1 \le i \le n$. Then

$$E(\max_{1\le i\le n} |\sum_{j=1}^{i} Z_j|^p) \le C_{p,q} (\sum_{j=1}^{n} a_j)^q,$$

for some positive constant $C_{p,q}$ depending only on p and q.

PROOF OF THEOREM 2.1. Fix $1 \le i \le n$. By observing that for $0 \le k \le n-1$, $1 \le j_1 \le j_2 \le n$ and $|j_1 - j_2| = k$, $\gamma_n(k) \ge |E(f_{j_1} f_{j_2})|$, one gets

$$
\begin{aligned}
E|\sum_{j=1}^{i} c_j f_j|^2 &\le \sum_{j=1}^{i}\sum_{l=1}^{i} |c_j c_l||E(f_j f_l)| \\
&\le (\sum_{j=1}^{i} c_j^2)\gamma_n(0) + \left(\max_{1\le k\le n-1} \gamma_n(k)k^{\alpha}\right)(\sum_{j\ne l} |c_j c_l|/|j-l|^{\alpha}).
\end{aligned} \tag{2.2}
$$

Using the fact that $(\sum \nu_j^p) \le (\sum \nu_j)^p$ for $p \ge 1$ and $\nu_j > 0$, we have

$$
(\sum_{j=1}^{i} c_j^2)\gamma_n(0) \le (\sum_{j=1}^{i} |c_j|^{2/(2-\alpha)})^{2-\alpha}\gamma_n(0). \tag{2.3}
$$

Applying Lemma 2.2 with $p = q = 2/(2-\alpha)$ and $\delta = \alpha$, we have for some $M_\alpha > 0$ that

$$
(\sum_{j\ne l} |c_j c_l|/|j-l|^{\alpha}) \le M_\alpha (\sum_{j=1}^{i} |c_j|^{2/(2-\alpha)})^{2-\alpha}. \tag{2.4}
$$

In view of (2.2)-(2.4), we obtain that

$$
E(|\sum_{j=1}^{i} c_j f_j|^2) \le 2(M_\alpha + 1)\{\max_{0\le k\le n-1} \gamma_n(k)(k+1)^{\alpha}\}(\sum_{i=1}^{i} |c_i|^{2/(2-\alpha)})^{2-\alpha}. \tag{2.5}
$$

Since $2-\alpha > 1$, applying Lemma 2.3 with $p = 2$, $q = 2-\alpha$, $Z_j = c_j f_j$, and $a_j^{2-\alpha} = 2(M_\alpha + 1)\{\max_{0\le k\le n-1} \gamma_n(k)(k+1)^{\alpha}\}c_j^2$, Theorem 2.1 is proved. $\square$

An immediate extension of Theorem 2.1 is the following Hájek-Rény-Chow's type maximal moment inequality.

Corollary 2.4. *Let the same assumptions of Theorem 2.1 hold. Then, for $0 < a_1 \le a_2 \le \cdots \le a_n$,*

$$
E(\max_{1\le i\le n} |\sum_{j=1}^{i} c_j f_j|^2/a_i^2) \le 4k_\alpha\{\max_{0\le k\le n-1} \gamma_n(k)(k+1)^{\alpha}\}(\sum_{i=1}^{n} |c_i/a_i|^{2/(2-\alpha)})^{2-\alpha}. \tag{2.6}
$$

PROOF. By a monotone inequality of Shorack and Smythe (1976) (see, also, Shorack and Wellner, 1986, page 844), for any sequence of real numbers ν_j and a_j, if $0 < a_1 \le a_2 \le \cdots \le a_n$, then

$$
\max_{1\le k\le n} |\sum_{j=1}^{k} \nu_j|/a_k \le 2\max_{1\le k\le n} |\sum_{j=1}^{k} \nu_j/a_j|. \tag{2.7}
$$

Consequently, (2.6) is guaranteed by (2.1) and (2.7). □

Remark 1. As a direct application of Corollary 2.4, a reverse sum analogue of (2.6) is given as follows:

$$E(\max_{1\le i\le n}|\sum_{j=i}^{n} c_j f_j|^2/a_{n-i+1}^2) \le 4k_\alpha\{\max_{0\le k\le n-1}\gamma_n(k)(k+1)^\alpha\}(\sum_{i=1}^{n}|c_i/a_i|^{2/(2-\alpha)})^{2-\alpha}.$$

□

Applying (2.6) to zero mean stationary time series with autocovariance function satisfying (1.1), we bring (1.3) and (1.4) together through Corollary 2.5 (see below). When n is sufficiently large, the inequalities induced by Corollary 2.5 are much sharper than those in (1.3) and (1.4). For more details, see Remarks 2 and 3 after this corollary.

Corollary 2.5. *Assume that (1.1) holds. Then, for any $m \ge 0$, any $n \ge 1$, any sequence of real numbers $c_j, j \ge 1$, and any sequence of nondecreasing positive numbers $a_j, j \ge 1$, there is a positive constant K_α depending only on α such that*

$$E(\max_{1\le i\le n}|\sum_{j=m+1}^{m+i} c_j\varepsilon_j|^2/a_{m+i}^2) \le K_\alpha(\sum_{i=m+1}^{m+n}|c_i/a_i|^{2/(2-\alpha)})^{2-\alpha}, \tag{2.8}$$

and for any $\delta > 0$,

$$P(\max_{1\le i\le n}|\sum_{j=m+1}^{m+i} c_j\varepsilon_j|^2/a_{m+i}^2 > \delta) \le \frac{K_\alpha}{\delta^2}(\sum_{i=m+1}^{m+n}|c_i/a_i|^{2/(2-\alpha)})^{2-\alpha}. \tag{2.9}$$

Remark 2. Assume that $a_i = 1$ for all $i \ge 1$. Then, (2.8) yields that

$$E\left\{\max_{1\le i\le n}\left|\sum_{j=m+1}^{m+i} c_j\varepsilon_j\right|^2\right\} \le K_\alpha(\sum_{j=m+1}^{m+n}|c_j|^{2/(2-\alpha)})^{2-\alpha}. \tag{2.10}$$

Let $b_1 \ge b_2 \ge \ldots \ge b_n > 0$, $m = 0$, $c_i = 1$, and $1/a_i = b_i$. Then, by (2.9), we have for any $\delta > 0$,

$$P\left(\max_{1\le k\le n} b_k\left|\sum_{j=1}^{k}\varepsilon_j\right| > \delta\right) \le \frac{K_\alpha}{\delta^2}(\sum_{j=1}^{n} b_j^{2/(2-\alpha)})^{2-\alpha}. \tag{2.11}$$

It is worth noting that when $c_j = O(j^l)$ and $b_j = O(j^l)$ with $l < (\alpha-2)/2$, the right-hand sides of (2.10) and (2.11) are both bounded. Therefore, they are in sharp contrast to

(1.3) and (1.4), which provide upper bounds tending to infinity even if c_j and b_j decrease exponentially. Armed with (1.4), Lavielle and Moulines (2000) further obtained that for any $1 \leq m \leq n$ and $b_1 \geq b_2 \geq \ldots \geq b_n > 0$,

$$P\left(\max_{m\leq k\leq n} b_k \left|\sum_{j=1}^{k} \varepsilon_t\right| > \delta\right) \leq \frac{C_1 m^{2-\alpha} b_m^2}{\delta^2} + \frac{C_2(n-m)^{1-\alpha}}{\delta^2} \sum_{j=m+1}^{n} b_t^2, \qquad (2.12)$$

where C_1 and C_2 are positive constants independent of n, m, and b_t. Inequality (2.12) is still not sharp enough, because the second term on the right-hand side of (2.12) diverges (as $n \to \infty$), regardless of how large m is and how small $\{b_t\}$ are. However, according to (2.11), this term can be replaced by $C_3\{\sum_{m+1}^{n} b_t^{2/2-\alpha}\}^{2-\alpha}$, for some positive constant C_3 independent of n, m, and b_t. When $b_t = O(t^l)$ with $l < (\alpha-2)/2$, $C_3\{\sum_{m+1}^{n} b_t^{2/2-\alpha}\}^{2-\alpha}$ can be smaller than any positive number, provided m is sufficiently large. □

Remark 3. Kokoszka and Leipus (1998, Theorem 3.1) gave an inequality that is closely related to (1.4). Under (1.1), their inequality implies that for any $\delta > 0$,

$$\begin{aligned} \delta^2 P\left(\max_{1\leq k\leq n} b_k \left|\sum_{j=1}^{k} \varepsilon_j\right| > \delta\right) &\leq \sum_{j=1}^{n-1} |b_{j+1}^2 - b_j^2| j^{2-\alpha} + 2\gamma(0)^{1/2} \sum_{j=1}^{n-1} b_{j+1}^2 j^{(2-\alpha)/2} \\ &+ \gamma(0) \sum_{j=0}^{n-1} b_{j+1}^2, \end{aligned} \qquad (2.13)$$

where $\{b_k\}$ is a sequence of positive numbers. As observe, the right-hand side of (2.13) is bounded by a finite positive number if the b_k decays at an appropriate hyperbolic rate. One special feature of Kokoszka and Leipus's inequality is that the b_k is not necessary to be nonincreasing. On the other hand, when one focuses on the most-discussed case where b_k is nonincreasing, (2.11) is still more informative than (2.13). To see this, assume that $b_k = k^l, l < 0$. Straightforward calculations show that the right-hand side of (2.13) is bounded if and only if $l < -1 + (\alpha/4)$. However, to ensure that the right-hand side of (2.11) is bounded, only $l < -1 + (\alpha/2)$ is required. Another limitation (compared to (2.9)) of Kokoszka and Leipus's result is that they only considered the constant-weight case, $c_j = 1$ for all j. It remains unclear whether their result can be used to establish a sharp maximal probability inequality for the weighted sum of $\{\varepsilon_j\}$ (with general weights), which seems indispensable in exploring asymptotic properties of the least squares estimate in regression models; as detailed in Sections 3 and 4. □

The following corollary deals with the almost sure convergence of $\sum_{i=1}^n c_i\varepsilon_i$ and also its order of magnitude in case of divergence.

Corollary 2.6. *Assume that (1.1) holds.*

(i) If $\sum_{i=1}^\infty |c_i|^{2/(2-\alpha)} < \infty$, then $\sum_{i=1}^n c_i\varepsilon_i$ converges almost surely (a.s.).

(ii) If $\sum_{i=1}^\infty |c_i|^{2/(2-\alpha)} = \infty$, then for any $\delta > 1 - (\alpha/2)$,

$$\sum_{i=1}^n c_i\varepsilon_i = o(G_n^{(2-\alpha)/2}(\log G_n)^\delta) \ a.s., \tag{2.14}$$

where $G_n = \sum_{i=1}^n |c_i|^{2/(2-\alpha)}$.

PROOF. To show (i), define $S_n = \sum_{j=1}^n c_j\varepsilon_j$. By (2.10), we have for all $n \geq 1$,

$$E(\max_{m+1\leq j\leq m+n} |S_j - S_m|^2) \leq K_\alpha(\sum_{j=m+1}^\infty |c_j|^{2/(2-\alpha)})^{2-\alpha}, \tag{2.15}$$

where K_α is a positive constant depending only on α. (2.15) and Chebyshev's inequality yield for any $\delta > 0$,

$$\delta^2 P(\sup_{j\geq m+1} |S_j - S_m| > \delta) \leq K_\alpha(\sum_{j=m+1}^\infty |c_j|^{2/(2-\alpha)})^{2-\alpha}. \tag{2.16}$$

Now, the desired result follows from (2.16) and the hypothesis that $\sum_{i=1}^\infty |c_i|^{2/(2-\alpha)} < \infty$.

To deal with (ii), first note that if

$$\sum_{i=m+1}^n \frac{c_i\varepsilon_i}{G_i^{(2-\alpha)/2}(\log G_i)^\delta} \quad \text{converges a.s.}, \tag{2.17}$$

where m is the smallest positive integer such that $G_m > 1$, then (2.14) is an immediate consequence of Kronecker's lemma. By (i) of Corollary 2.6, (2.17) is guaranteed by

$$\sum_{i=m+1}^\infty \frac{|c_i|^{2/(2-\alpha)}}{G_i(\log G_i)^{2\delta/(2-\alpha)}} < \infty. \tag{2.18}$$

By observing

$$\sum_{i=m+1}^\infty \frac{|c_i|^{2/(2-\alpha)}}{G_i(\log G_i)^{2\delta/(2-\alpha)}} \leq \int_{G_m}^\infty \frac{1}{x(\log x)^{2\delta/(2-\alpha)}}dx$$

and $2\delta/(2-\alpha) > 1$, (2.18) follows. □

Remark 4. Since for $(2-\alpha)/2 < \theta < 1$, $\sum_{i=1}^\infty i^{-2\theta/(2-\alpha)} < \infty$, (i) of Corollary 2.6 and Kronecker's lemma imply

$$\frac{1}{n}\sum_{i=1}^n \varepsilon_i = o(\frac{1}{n^{1-\theta}}) \ a.s.$$

This result provides an almost sure convergence rate for the first sample moment of a stationary process with zero mean and autocovariance function satisfying (1.1). □

3. Strong consistency of least squares estimates.

Consider the multiple regression model

$$y_i = \beta_1 x_{i1} + \cdots, \beta_p x_{ip} + \varepsilon_i, i = 1, \ldots, n, \tag{3.1}$$

where n is the number of observations, p is a known positive integer, $x_{ij}, j = 1, \cdots, p$ are known constants, and $\beta_1, \cdots, \beta_p$ are unknown parameters. Throughout this section we let $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$, $\mathbf{y}_n = (y_1, \cdots, y_n)'$, and $\beta = (\beta_1, \cdots, \beta_p)'$. For $n \geq p$, the least squares estimate $\mathbf{b}_n = (\beta_{n1}, \cdots, \beta_{np})'$ of β based on $(y_i, \mathbf{x}_i')', i = 1, \cdots, n$ is given by

$$\mathbf{b}_n = (\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i')^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i, \tag{3.2}$$

provided $\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$ is nonsingular.

To study the strong consistency problem of $\mathbf{b}_n$ under model (3.1) with ε_t satisfying (1.1), we first introduce the concept of "convergence system". A sequence of random variables $\{\varepsilon_t\}$ is said to be a convergence system if $\sum_{i=1}^{n} c_i \varepsilon_i$ converges a.s. for every sequence $\{c_i\}$ with $\sum_{i=1}^{\infty} c_i^2 < \infty$. The strong consistency of $\mathbf{b}_n$ under the assumption that $\{\varepsilon_t\}$ constitutes a convergence system has been studied by Lai, Robbins, and Wei (1979). To cover a wider range of dependent error structures, we now generalize their result to the case where $\{g_i \varepsilon_i\}$ is a convergence system. Here, $\{g_i\}$ is a sequence of "contraction" constants. Its role will be clarified in the following theorem.

Theorem 3.1. *Suppose that in (3.1), $V_n = (\nu_{ij}^{(n)})_{1 \leq i,j \leq p} = (\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j')^{-1}$ exists for all $n \geq m$, $\lim_{n \to \infty} \nu_{jj}^{(n)} = 0$, and $\{g_i \varepsilon_i\}$ is a convergence system for some constants g_i such that $|g_i|$ is positive and nonincreasing. Then,*

$$\beta_{nj} - \beta_j = o(1) \text{ a.s.}, \tag{3.3}$$

provided

$$\sum_{i=m}^{\infty} \nu_{jj}^{(i)} (g_{i+1}^{-2} - g_i^{-2}) < \infty. \tag{3.4}$$

Before proceeding to the proof of Theorem 3.1, we need a lemma that provides a better understanding of the series in (3.4).

Lemma 3.2. *Let* $\{l_i\}$ *and* $\{h_i\}$, $i = 1, 2, \cdots$, *be sequences of nonnegative numbers with* $0 < h_i \leq h_{i+1}$ *for all* $i \geq 1$. *Define* $\mu_i = \sum_{j=i}^{\infty} l_j$. *Then, for any* $k \geq 1$,

$$\sum_{i=k}^{\infty} l_i h_i = h_k \mu_k + \sum_{i=k+1}^{\infty} \mu_i (h_i - h_{i-1}). \tag{3.5}$$

PROOF. First note that if $\mu_k = \sum_{i=k}^{\infty} l_i = \infty$, then $\sum_{i=k}^{\infty} l_i h_i \geq h_k \mu_k = \infty$. In this case, both side of (3.5) are infinite. Hence, without loss of generality, we can assume that $\mu_k < \infty$. Observe that for $n \geq k+1$,

$$h_k \mu_k + \sum_{i=k+1}^{n} \mu_i (h_i - h_{i-1}) = \sum_{i=k}^{n} h_i l_i + \mu_{n+1} h_n. \tag{3.6}$$

Since all terms involved are positive, the lemma obviously holds if $\sum_{i=k}^{\infty} h_i l_i = \infty$. We can further assume that $\sum_{i=k}^{\infty} h_i l_i < \infty$. This implies

$$o(1) = \sum_{i=n+1}^{\infty} h_i l_i \geq h_n \sum_{i=n+1}^{\infty} l_i = h_n \mu_{n+1}.$$

In view of this and (3.6), (3.5) follows. □

PROOF OF THEOREM 3.1. Without loss of generality, we assume that $j = 1$. For $n > m$, let $T_n = (x_{n2}, \cdots, x_{np})'$ and write $d_n = x_{n1} - K_n H_n^{-1} T_n$, where K_n and H_n satisfy

$$\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j' = \begin{pmatrix} \sum_{i=1}^{n} x_{i1}^2 & K_n \\ K_n' & H_n \end{pmatrix}.$$

By Lemma 3 of Lai, Robbins, and Wei (1979), $b_{n1} - \beta_1 = \rho_n / s_n$, where

$$s_n = \frac{1}{\nu_{11}^{(n)}} = s_m + \sum_{i=m+1}^{n} d_i^2 (1 + T_i' H_{i-1}^{-1} T_i), \tag{3.7}$$

and

$$\rho_n = \rho_m + \sum_{i=m+1}^{n} d_i \{\varepsilon_i - T_i' H_{i-1}^{-1} (\sum_{k=1}^{i-1} T_k \varepsilon_k)\}, \tag{3.8}$$

Let us first assume that

$$\sum_{i=m+1}^{\infty} \frac{d_i^2(1+T_i' H_{i-1}^{-1} T_i)}{s_i^2 g_i^2} < \infty. \tag{3.9}$$

Equation (3.9) and the assumption that $\{g_i \varepsilon_i\}$ is a convergence system yield that

$$\sum_{i=m+1}^{\infty} \frac{d_i \varepsilon_i}{s_i} = \sum_{i=m+1}^{\infty} \frac{d_i g_i \varepsilon_i}{s_i g_i} \textit{ converges a.s.} \tag{3.10}$$

But by Lemma 1 (ii) of Chen, Lai, and Wei (1981), $\{g_i \xi_i\}$ is also a convergence system, where

$$\xi_n = \frac{T_n' H_{n-1}^{-1}(\sum_{k=1}^{n-1} T_k \varepsilon_k)}{(1+T_n' H_{n-1}^{-1} T_n)^{1/2}}.$$

Therefore, by (3.9),

$$\sum_{i=m+1}^{n} \frac{d_i T_i' H_{i-1}^{-1}(\sum_{k=1}^{i-1} T_k \varepsilon_k)}{s_i} = \sum_{i=m+1}^{n} \frac{d_i(1+T_i' H_{i-1}^{-1} T_i)^{1/2} g_i \xi_i}{s_i g_i}$$

converges a.s. In view of this, (3.10), (3.7), (3.8), and Kronecker's lemma, $b_{n1} - \beta_1 = \rho_n/s_n = o(1)$ a.s.

It remains to prove (3.9). For this, in (3.5) let $k = m+1$, $h_i = g_i^{-2}$, and

$$l_i = \frac{d_i^2(1+T_i' H_{i-1}^{-1} T_i)}{s_i^2}.$$

Then, the series (3.9) is equivalent to

$$g_{m+1}^{-2} \mu_{m+1} + \sum_{i=m+2}^{\infty} \mu_i (g_i^{-2} - g_{i-1}^{-2}),$$

where for $i \geq m+1$

$$\mu_i = \sum_{j=i}^{\infty} l_j = \sum_{j=i}^{\infty} \frac{s_j - s_{j-1}}{s_j^2} \leq \frac{1}{s_{i-1}}.$$

Hence, $\sum_{i=m}^{\infty} \nu_{11}^{(i)} (g_{i+1}^{-2} - g_i^{-2}) = \sum_{i=m}^{\infty} s_i^{-1} (g_{i+1}^{-2} - g_i^{-2}) < \infty$ implies (3.9). This completes the proof. $\square$

Next we apply Theorem 3.1 to the case where $\{\varepsilon_t\}$ satisfies (1.1). We start with a lemma that gives a sufficient condition on g_i under which $\{g_i \varepsilon_i\}$ is a convergence system.

Lemma 3.3. *Assume that (1.1) holds. If*

$$\sum_{i=1}^{\infty} |g_i|^{2/(1-\alpha)} < \infty, \tag{3.11}$$

then $\{g_i\varepsilon_i\}$ is a convergence system.

PROOF. By Corollary 2.6, it suffices to show that if $\sum_{i=1}^{\infty} a_i^2 < \infty$, then $\sum_{i=1}^{\infty} |a_i g_i|^{2/(2-\alpha)} < \infty$. By Hölder's inequality with $p = 2 - \alpha$ and $q = (2-\alpha)/(1-\alpha)$, one obtains that

$$\sum_{i=1}^{\infty} |a_i g_i|^{2/(2-\alpha)} \le (\sum_{i=1}^{\infty} a_i^2)^{1/(2-\alpha)} (\sum_{i=1}^{\infty} |g_i|^{2/(1-\alpha)})^{(1-\alpha)/(2-\alpha)}.$$

Since $\sum_{i=1}^{\infty} |g_i|^{2/(1-\alpha)} < \infty$ is assumed, $\sum_{i=1}^{\infty} a_i^2 < \infty$ yields that $\sum_{i=1}^{\infty} |a_i g_i|^{2/(2-\alpha)} < \infty$. $\square$

A special class of g_n that satisfies (3.11) is given by $g_n^{-1} = (n \log n)^{(1-\alpha)/2} (\log n)^{\delta}$ for some $\delta > 0$. In this case,

$$\frac{1}{g_n^2} - \frac{1}{g_{n-1}^2} \sim (1-\alpha) n^{-\alpha} (\log n)^{1-\alpha+2\delta}. \tag{3.12}$$

Now, when (1.1) is fulfilled by $\{\varepsilon_t\}$, a set of sufficient conditions for strong consistency of $\mathbf{b}_n$ is given in the following corollary.

Corollary 3.4. *Let (1.1) and the hypotheses of Theorem 3.1 hold. Moreover, if for some $\delta > 0$*

$$\sum_{k=m}^{\infty} \frac{\nu_{jj}^{(k)} (\log k)^{1-\alpha+\delta}}{k^{\alpha}} < \infty,$$

then $\beta_{nj} - \beta_j = o(1)$ a.s.

PROOF. Since the value of δ in (3.12) can be arbitrary and

$$n^{-\alpha} (\log n)^{1-\alpha+\delta} / (n+1)^{-\alpha} (\log(n+1))^{1-\alpha+\delta}$$

converges to 1 as n tends to infinity, Corollary 3.4 follows from these observations, Theorem 3.1, and Lemma 3.3. $\square$

Remark 5. To obtain the strong consistency of $\mathbf{b}_n$, Yajima (1988) assumed that

$$\sum_{k=m}^{\infty} \frac{\lambda_k^{-1} \log^2 k}{k^{\alpha}} < \infty,$$

where $\lambda_k = \lambda_{\min}(\sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j')$ is the minimal eigenvalue of $\sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j'$. This requires a global condition on $\sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j'$. Corollary 3.4 can handle individual β_j. Moreover, since for each j, $\lambda_k^{-1} \geq \nu_{jj}^{(k)}$, Corollary 3.4 also gives a better result. □

When (1.1) is assumed, Corollary 3.5 (see below) shows that the residual mean squared error

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \mathbf{x}_i'\mathbf{b}_n)^2$$

is a strongly consistent estimate of $\gamma(0) = E(\varepsilon_1^2)$. To achieve this goal, we need the following conditions. Define $L_n = (l_{ij}^{(n)})_{1\leq i,j\leq p} = \sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j'$ and $\|\mathbf{x}_i^{(n)}\|^2 = l_{ii}^{(n)}$.

(C.1) $\rho_{ij} = \lim_{n\to\infty} l_{ij}^{(n)}/(\|\mathbf{x}_i^{(n)}\|\|\mathbf{x}_j^{(n)}\|)$ exists for all $1 \leq i, j \leq p$.

(C.2) $\lambda_{min}(\Lambda) > 0$, where $\Lambda = (\rho_{ij})_{1\leq i,j\leq p}$.

(C.3) For any $\delta > 0$,

$$\log(\lambda_{\max}(L_n)) = o(n^\delta),$$

where $\lambda_{\max}(L_n)$ denotes the maximal eigenvalue of L_n.

Corollary 3.5. *Under model (3.1), assume that (1.1) and (C.1)-(C.3) hold. Further assume that $sup_{1\leq j<\infty} E(\varepsilon_j^4) < \infty$ and that for any integers $1 \leq i \neq j < \infty$, there are positive constants C and θ such that*

$$E\{(\varepsilon_i^2 - \gamma(0))(\varepsilon_j^2 - \gamma(0))\} \leq C|i-j|^{-\theta}. \tag{3.13}$$

Then,

$$\lim_{n\to\infty} \hat{\sigma}_n^2 = \gamma(0) \ a.s.$$

PROOF. First note that for $n \geq m$,

$$\begin{aligned}\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - \frac{1}{n}\sum_{j=1}^n \mathbf{x}_j'\varepsilon_j(\sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j')^{-1}\sum_{j=1}^n \mathbf{x}_j\varepsilon_j \\ &= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - \frac{1}{n}(D_n^{-1}\sum_{i=1}^n \mathbf{x}_i\varepsilon_i)'(D_nV_nD_n)(D_n^{-1}\sum_{i=1}^n \mathbf{x}_i\varepsilon_i), \qquad (3.14)\end{aligned}$$

where $D_n = Diag(\|\mathbf{x}_1^{(n)}\|, \cdots, \|\mathbf{x}_p^{(n)}\|)$ is a diagonal matrix with the ith diagonal element equal to $\|\mathbf{x}_i^{(n)}\|$. By (C.1), (C.2), and Corollary 2.6, one has for any $\delta > 2 - \alpha$,

$$
\begin{aligned}
&(D_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i)'(D_n V_n D_n)(D_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i) = O(1) \|D_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i\|^2 \\
= \;& O(1) \sum_{i=1}^{p} \left(\frac{1}{\|\mathbf{x}_i^{(n)}\|} \sum_{j=1}^{n} x_{ji} \varepsilon_j \right)^2 = o\left(\sum_{i=1}^{p} \frac{E_{in}^{2-\alpha}}{\|\mathbf{x}_i^{(n)}\|^2} (\log E_{in})^{\delta} \right) + O(1) \; a.s., \quad (3.15)
\end{aligned}
$$

where $E_{in} = \sum_{j=1}^{n} |x_{ji}|^{2/(2-\alpha)}$. This result, (C.3), and the fact that $E_{in}^{2-\alpha} \leq n^{1-\alpha} \|\mathbf{x}_i^{(n)}\|^2$ further imply

$$
\begin{aligned}
\frac{1}{n} (D_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i)'(D_n V_n D_n)(D_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i) &= o\left(\frac{\{\log n + \log \lambda_{\max}(L_n)\}^{\delta}}{n^{\alpha}} \right) \\
&= o(1) \; a.s. \qquad (3.16)
\end{aligned}
$$

To deal with the first term on the right-hand side of (3.14), first assume that in (3.13) $0 < \theta < 1$. By an argument similar to that used for showing Theorem 2.1, one has for $1 \leq N_1 < N_2$ and some $C > 0$ (independent of N_1 and N_2),

$$
E\left(\max_{N_1 \leq j \leq N_2} \left(\sum_{j=N_1}^{N_2} \frac{\varepsilon_j^2 - \gamma(0)}{j} \right)^2 \right) \leq C(\sum_{j=N_1}^{N_2} j^{-2/(2-\theta)})^{2-\theta},
$$

which together with Kronecker's lemma yields

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 = \gamma(0). \qquad (3.17)
$$

Assume that in (3.13) $\theta > 1$. Then, it is not difficult to see that for $1 \leq N_1 < N_2$, there is a constant $C > 0$ (independent of N_1 and N_2) such that

$$
E\left(\sum_{j=N_1}^{N_2} \frac{\varepsilon_j^2 - \gamma(0)}{j} \right)^2 \leq C \sum_{j=N_1}^{N_2} j^{-2} \leq C(\sum_{j=N_1}^{N_2} j^{-2/\delta})^{\delta}, \qquad (3.18)
$$

where δ is any number lying between 1 and 2. When $\theta = 1$, some algebraic manipulations also yield that for $1 \leq N_1 < N_2$, there is a constant $C > 0$ (independent of N_1 and N_2) such that

$$
E\left(\sum_{j=N_1}^{N_2} \frac{\varepsilon_j^2 - \gamma(0)}{j} \right)^2 \leq C \sum_{j=N_1}^{N_2} (\log j / j^2) \leq C(\sum_{j=N_1}^{N_2} (\log j / j^2)^{1/\delta})^{\delta}, \qquad (3.19)
$$

where $1 < \delta < 2$. According to Lemma 2.3, (3.18), (3.19), and Kronecker's lemma, (3.17) is still valid for the case of $\theta \geq 1$. As a result, Corollary 3.5 follows from (3.14), (3.16), and (3.17). □

Remark 6. In situations where (1.1) holds, (3.13) is quite easily fulfilled in practice. For example, if $\{\varepsilon_t\}$ is a linear or a Gaussian process, then (3.13) holds for $\theta = 2\alpha$. □

4. Strongly consistent model selection.

As mentioned in Section 1, when some $\beta_j, 1 \leq j \leq p$, in model (3.1) vanishes, adopting a subset model may improve estimation and prediction efficiency. Let $\bar{M} = \{x_1, \cdots, x_p\}$ denote the full model described in (3.1). Based on $\{y_1, \cdots, y_n, x_{1j}, \cdots, x_{nj}, j = 1, \cdots, p\}$, this section aims to select a subset model $M_0 \in \mathcal{M} = \{M : M \subseteq \bar{M}\}$, which is the correct model with fewest variables. Without loss of generality, we assume that $M_0 = \{x_1, \cdots, x_q\}$, where $1 \leq q \leq p$. Hence, (3.1) can be rewritten as

$$y_i = \beta_1 x_{i1} + \cdots, \beta_q x_{iq} + \varepsilon_i, \tag{4.1}$$

where $\beta_i \neq 0$ for all $1 \leq i \leq q$. For later reference, let $\mathbf{x}_{i0} = (x_{i1}, \cdots, x_{iq})'$. For $M \in \mathcal{M}$, define a loss function

$$L_n(M) = \log \hat{\sigma}_n^2(M) + P_n card(M), \tag{4.2}$$

where $\hat{\sigma}_n^2(M)$ represents the residual mean squared error obtained from fitting model M using least squares, $card(M)$ is the number of the regressor variables in model M, and P_n, depending on n, is a positive number to be determined later. Let $\hat{M}_n \in \mathcal{M}$ denote the model having the minimal loss function value, namely, $L(\hat{M}_n) = \min_{M \in \mathcal{M}} L(M)$. In situations where (1.1) is fulfilled by $\{\varepsilon_t\}$, Theorem 4.1 (see below) shows that

$$P(\hat{M}_n = M_0, eventually) = 1, \tag{4.3}$$

provided P_n is suitably chosen. To obtain this result, rates of divergence of $\lambda_{\min}(L_n)$ and $\lambda_{\max}(L_n)$ need to be imposed.

(C.3') For some positive numbers r_1, r_2, and r_3 with $1 - \alpha < r_1 \leq r_2$ and $r_3 \geq 1$,

$$\liminf_{n\to\infty} \frac{\lambda_{\min}(L_n)}{n^{r_1}} > 0,$$

and

$$\lambda_{\max}(L_n) = O(exp\{(r_2 \log n)^{r_3}\}).$$

Theorem 4.1. *Under models (3.1) and (4.1), assume that (1.1),* $\sup_{1 \le j < \infty} E(\varepsilon_j^4) < \infty$*, (3.13), (C.1), (C.2), and (C.3') hold. Furthermore, assume that* P_n *in (4.2) satisfies*

$$\lim_{n \to \infty} n^{1-\min\{1, r_1\}} P_n = 0, \tag{4.4}$$

and

$$\liminf_{n \to \infty} \frac{P_n}{\frac{(\log n)^{2r_3}}{n^{\alpha}}} > 0, \tag{4.5}$$

where r_1 *and* r_3 *are defined in (C.3'). Then, (4.3) follows.*

PROOF. We first show that if $p > q$, then it is not possible to choose an overfitting model for $L_n(M)$ as n is sufficiently large. More precisely, if $M \in \mathcal{M}$ is a subset model with $M \supseteq M_0$ and $M - M_0 \neq \emptyset$, where $\emptyset$ denotes the empty set and $M - M_0$ denotes the difference of set M minus set M_0, then we are going to prove that

$$P(L_n(M) > L_n(M_0), eventually) = 1. \tag{4.6}$$

When $q < p$, there is a model $M_u \supseteq M_0$ which satisfies $card(M_u) < card(\bar{M})$. Let $\mathbf{u}_i$ denote the ith regressor corresponding to model M_u. Choose a variable x^* from $\bar{M} - M_u$ and add x^* into M_u. Denote this extended model by M_u^*. Then, $L_n(M_u^*) - L_n(M_u) = \log \hat{\sigma}_n^2(M_u^*) - \log \hat{\sigma}_n^2(M_u) + P_n$. To obtain (4.6), it suffices to show that

$$P(\log \hat{\sigma}_n^2(M_u) - \log \hat{\sigma}_n^2(M_u^*) - P_n < 0, eventually) = 1. \tag{4.7}$$

First note that

$$\log \hat{\sigma}_n^2(M_u) - \log \hat{\sigma}_n^2(M_u^*) \le (\hat{\sigma}_n^2(M_u) - \hat{\sigma}_n^2(M_u^*))/\hat{\sigma}_n^2(M_u^*).$$

In addition, since (C.3') implies (C.3), Corollary 3.5 yields that

$$\lim_{n \to \infty} \hat{\sigma}_n^2(M_u^*) = \gamma(0) \ a.s.$$

In view of these and (4.5), (4.7) is guaranteed by showing that

$$n(\hat{\sigma}_n^2(M_u) - \hat{\sigma}_n^2(M_u^*)) = o(n^{1-\alpha}(\log n)^{2r_3}) \ a.s. \tag{4.8}$$

Now,

$$n(\hat{\sigma}_n^2(M_u) - \hat{\sigma}_n^2(M_u^*)) = (\sum_{i=1}^n \mathbf{u}_i^{*'}\varepsilon_i)V_n(M_u^*)(\sum_{i=1}^n \mathbf{u}_i^*\varepsilon_i) - (\sum_{i=1}^n \mathbf{u}_i'\varepsilon_i)V_n(M_u)(\sum_{i=1}^n \mathbf{u}_i\varepsilon_i), \tag{4.9}$$

where $V_n(M_u^*) = (\sum_{i=1}^n \mathbf{u}_i^*\mathbf{u}_i^{*'})^{-1}$, $V_n(M_u) = (\sum_{i=1}^n \mathbf{u}_i\mathbf{u}_i')^{-1}$, and $\mathbf{u}_i^*$ is the ith regressor corresponding to model M_u^*. By (C.1), (C.2), (C.3'), Corollary 2.6, and arguments similar to those used for obtaining (3.15) and (3.16), we have

$$\begin{aligned}
& (\sum_{i=1}^n \mathbf{u}_i'\varepsilon_i)V_n(M_u)(\sum_{i=1}^n \mathbf{u}_i\varepsilon_i) \\
= \ & (D_n^{-1}(M_u)\sum_{i=1}^n \mathbf{u}_i'\varepsilon_i)(D_n(M_u)V_n(M_u)D_n(M_u))(D_n^{-1}(M_u)\sum_{i=1}^n \mathbf{u}_i\varepsilon_i) \\
= \ & o\left(n^{1-\alpha}(\log n)^{2r_3}\right) \ a.s.,
\end{aligned} \tag{4.10}$$

where $D_n(M_u)$ is a diagonal matrix with the ith diagonal element equal to the square root of the ith diagonal element of $V_n^{-1}(M_u)$. Similarly,

$$(\sum_{i=1}^n \mathbf{u}_i^{*'}\varepsilon_i)V(M_u^*)(\sum_{i=1}^n \mathbf{u}_i^*\varepsilon_i) = o\left(n^{1-\alpha}(\log n)^{2r_3}\right) \ a.s. \tag{4.11}$$

Consequently, (4.8) follows from (4.9)-(4.11).

The remaining part of this proof focuses on the underspecified case. In particular, we show that for any $M_v \subset M_0$ with $M_0 - M_v \neq \emptyset$,

$$P(\log \hat{\sigma}_n^2(M_v) - \log \hat{\sigma}_n^2(M_0) - (card(M_0) - card(M_v))P_n > 0, eventually) = 1. \tag{4.12}$$

Without loss of generality, assume $M_v = \{x_v, \cdots, x_q\}$ with $2 \leq v \leq q$. By Wei (1992, Theorem 3.1), it can be shown that

$$\begin{aligned}
& \hat{\sigma}_n^2(M_v) - \hat{\sigma}_n^2(M_0) \geq \hat{\sigma}_n^2(\underline{M}_0) - \hat{\sigma}_n^2(M_0) \geq \hat{\sigma}_n^2(\underline{M}_0) - \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \\
= \ & \frac{1}{n}(\beta_1^2 s_{n0} + 2\beta_1\sum_{i=1}^n (x_{i1} - K_{n0}H_{n0}^{-1}T_{i0})\varepsilon_i - R_n),
\end{aligned} \tag{4.13}$$

where $\underline{M}_0 = \{x_2, \cdots, x_q\}$, $s_{n0} = \sum_{i=1}^n (x_{i1} - K_{n0}H_{n0}^{-1}T_{i0})^2$, $T_{i0} = (x_{i2}, \cdots, x_{iq})'$, $R_n = (\sum_{i=1}^n T_{i0}'\varepsilon_i)H_{n0}^{-1}(\sum_{i=1}^n T_{i0}\varepsilon_i)$, and

$$\sum_{j=1}^n \mathbf{x}_{j0}\mathbf{x}_{j0}' = \begin{pmatrix} \sum_{i=1}^n x_{i1}^2 & K_{n0} \\ K_{n0}' & H_{n0} \end{pmatrix}.$$

By (C.1), (C.2), (C.3'), and an argument similar to that used for obtaining (4.10), we have

$$R_n = o\left(n^{1-\alpha}(\log n)^{2r_3}\right) \ a.s. \tag{4.14}$$

By Corollary 2.6, one has for any $\delta > 1 - (\alpha/2)$,

$$\sum_{i=1}^{n}(x_{i1} - K_{n0}H_{n0}^{-1}T_{i0})\varepsilon_i = O(1) + o(F_n^{(2-\alpha)/2}(\log F_n)^{\delta}) \ a.s., \tag{4.15}$$

where $F_n = \sum_{i=1}^{n} |x_{i1} - K_{n0}H_{n0}^{-1}T_{i0}|^{2/(2-\alpha)}$. According to (C.3'), (4.13)-(4.15), and the facts that $F_n^{2-\alpha} \leq n^{(1-\alpha)}s_{n0}$ and

$$(1/q)\lambda_{\min}(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i') \leq s_{n0} \leq \|\mathbf{x}_1^{(n)}\|^2 \leq \lambda_{\max}(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i') \tag{4.16}$$

(see Lai and Wei, 1982), one has

$$\hat{\sigma}_n^2(M_v) - \hat{\sigma}_n^2(M_0) \geq \frac{s_{no}\beta_1^2}{n}(1 + o(1)) \ a.s.$$

This, (4.16), (C.3'), and the fact that $\lim_{n\to\infty} \hat{\sigma}_n^2(M_0) = \gamma(0)$ (which is guaranteed by Corollary 3.5) further yield that

$$\liminf_{n \to\infty}(\log \hat{\sigma}_n^2(M_v) - \log \hat{\sigma}_n^2(M_0))n^{1-\min\{1,r_1\}} > 0 \ a.s. \tag{4.17}$$

Consequently, (4.12) follows from (4.4) and (4.17). □

Remark 7. Let $\mathbf{x}_i = (1, \cos v_1 i, \sin v_1 i, \cdots, \cos v_l i, \sin v_l i, \cos v_{l+1} i)'$ be the ith regressor variable associated with the full model, where $l \geq 1$ is a positive integer, $0 < v_1 < v_2 < \cdots < v_l < \pi$ are real numbers, and $v_{l+1} = \pi$. Then, it can be shown (eg, Zygmund, 1959, Chapter I) that (C.1), (C.2), and (C.3') with $r_1 = r_2 = r_3 = 1$ hold. For the polynomial regression, the ith regressor variable associated with the full model can be assumed to be $\mathbf{x}_i = (1, i, \cdots, i^{p-1})'$ for some positive integer p. By Anderson (1971, pages 581-582) and Yajima (1988), (C.1), (C.2), and (C.3') with $r_1 = r_3 = 1$ and $r_2 = 2p - 1$ are satisfied. □

Remark 8. To fulfill (4.4) and (4.5), P_n can be chosen to be

$$\frac{C(\log n)^{2r_3}}{n^{\alpha}}, \tag{4.18}$$

for some $C > 0$. Unfortunately, (4.18) cannot be applied to situations where α is unknown. However, if $r_1 \geq 1$ (which occurs in many practical situations, as observed in Remark 7), then one may choose $P_n = C_1/(\log n)^{C_2}$ to satisfy (4.4) and (4.5) without information about α. Here, C_1 and C_2 are any positive numbers. □

Remark 9. It is worth noting that $L_n(M)$ cannot be consistent even if (4.5) is marginally violated. To see this, consider the polynomial regression model and assume that $\mathbf{x}_i = (1, i)'$ and $\mathbf{x}_{i0} = (1)$. Therefore, the full model contains a constant term and a linear trend, whereas the smallest true model contains only a constant term. We also assume that (1.2) is satisfied. By some algebraic manipulations, it can be shown that

$$\liminf_{n\to\infty} \frac{E\{n(\hat{\sigma}_n^2(M_0) - \hat{\sigma}_n^2(\bar{M}))\}}{n^{1-\alpha}} > C_\alpha, \tag{4.19}$$

where $\hat{\sigma}_n^2(M_0)$ and $\hat{\sigma}_n^2(\bar{M})$ are residual mean squared errors corresponding to $\mathbf{x}_{i0}$ and $\mathbf{x}_i$, respectively, and C_α is some positive constant depending upon α only. If we further assume that $\{\varepsilon_t\}$ is a Gaussian process, then by Corollary 3.5, (4.19), and Gaussianity,

$$\liminf_{n\to\infty} P\left(\log \hat{\sigma}_n^2(M_0) - \log \hat{\sigma}_n^2(\bar{M}) > P_n\right) > 0, \tag{4.20}$$

provided $P_n = O(n^{-\alpha})$. Inequality (4.20) shows that $L_n(M)$ with $P_n = O(n^{-\alpha})$ is no longer consistent. Since $\log n/n = O(n^{-\alpha})$, one important implication of this result is that BIC, that is, $L_n(M)$ with $P_n = \log n/n$, is **not** consistent in the regression model with long range dependent errors. This is a somewhat different situation from that encountered in the case of short memory errors, because Chen and Ni (1989) showed that BIC is strongly consistent, provided $\{\varepsilon_t\}$ is a short memory linear process with spectral density that is bounded and bounded away from zero. □

Before leaving this section, we note that since (1.1) is satisfied by both short and long memory time series, Theorem 4.1 is especially useful in situations where the strength of dependence of $\{\varepsilon_t\}$ is unknown. More specifically, assume that r_1 in (C.3') is not less than 1 and $\{\varepsilon_t\}$ is an ARFIMA$(0, d, 0)$ process, where $-1/2 < d < 1/2$ is an unknown real number. Then, by Brockwell and Davis (1987, page 467), $\{\varepsilon_t\}$ satisfies (1.1) for any $-1/2 < d < 1/2$, and hence $L_n(M)$ with $P_n = C_1/(\log n)^{C_2}$, for some $C_1, C_2 > 0$, is strongly consistent (see Remark 8). On the other hand, since Brockwell and Davis

(1987, page 467) also showed that (1.2) is fulfilled by an ARFIMA$(0, d, 0)$ model with $0 < d < 1/2$, Remark 9 yields that BIC is not consistent once the value of d falls in $(0, 1/2)$.

REFERENCES

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Series Anal.* **15** 453-472.

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman and Hall, New York.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Chen, G., Lai, T. L. and Wei, C. Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivariate Anal.* **11**, 319-333.

Chen, Z. G. and Ni, J. Y. (1989). Subset regression time series and its modeling procedures. *J. Multivariate Anal.* **31** 266-288.

Chow, Y. S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* **11** 107-111.

Hájek, J. and Rényi, A. (1955). Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hungar.* **6** 281-283.

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press, London.

Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statist. Prob. Letters* **40** 385-393.

Lai, T. L., Robbins, H. and Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression. II. *J. Multivariate Anal.* **9** 343-361.

Lai, T. L. and Wei, C. Z. (1982). Asymptotic properties of projections with applications to stochastic regression problems. *J. Multivariate Anal.* **12** 346-370.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.* **21** 33-59.

Móricz, F. (1976). Moment inequalities and the strong laws of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **35** 299-314.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461-464.

Shorack, G. R. and Smythe, R. (1976). Inequalities for $\max |S_k|/b_k$ where $k \in N^r$. *Proc. Amer. Math. Soc.* **54** 331-336.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* John Wiley and Sons, New York.

Wei, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20** 1-42.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791-807.

Zygmund, A. (1959). *Trigonometric Series*, 2nd ed. Cambridge Univ. Press, London.