Statistica Sinica **16**(2006), 721-740

A MAXIMAL MOMENT INEQUALITY FOR LONG RANGE DEPENDENT TIME SERIES WITH APPLICATIONS TO ESTIMATION AND MODEL SELECTION

Ching-Kang Ing[1,2] and Ching-Zong Wei[2]

[1]*National Taiwan University and* [2]*Academia Sinica*

Abstract: We establish a maximal moment inequality for the weighted sum of a sequence of random variables with finite second moments. An extension to the Hájek-Rény and Chow inequalities is then obtained. When certain second-moment properties are fulfilled, it enables us to deduce a strong law for the weighted sum of a time series having long-range dependence. Applications to estimation and model selection in multiple regression models with long-range dependent errors are also given.

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

1. Introduction

The study of long range dependent time series has been developing rapidly (cf., Beran (1994) for a survey). As applications have become broader, the functionals involved have become increasingly complicated. In this paper, we give a unified 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

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

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 a constant $C_0 > 0$ such that

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

see Beran(1994, Chap.2). Condition (1.1) is fulfilled not only by most stationary short memory time series (e.g., an autoregressive moving average model), but also by long range dependent time series.

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

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

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, the inequality has shortcomings. Note, for example, that the term on the right-hand side of (1.3) goes to infinity as n does. And, 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 that of Lavielle and Moulines (2000, Theorem 1). They showed that for any $n \ge 1$, any $\delta > 0$, and any positive and nonincreasing sequence $b_1 \ge \cdots \ge b_n > 0$,

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

where C_1 is 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 $1 \le t \le n$ and $\max_{1\le t\le n} E(\varepsilon_t^2) < \infty$, (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 the Hájek-Rényi-Chow 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 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 a 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 type maximal moment inequality; see Corollary 2.4. This result is then applied to $\{\epsilon_t\}$ at the end of Section 2, where $\{\epsilon_t\}$ is a sequence of random variables whose second moments satisfy, for some $\alpha \in (0, 1)$,

$$S(\alpha,\epsilon) = \sup_{k\ge 0}(1+k)^{\alpha}\sup_{i,j\ge 1,|i-j|=k}|\mathrm{cov}(\epsilon_i,\epsilon_j)| < \infty. \tag{1.5}$$

(Note that (1.5), including (1.1) as a special case, allows more flexibility for practical applications.) Based on Corollary 2.4, Corollary 2.5 and Remark 2 give sharper bounds for the left-hand sides of (1.3) and (1.4) with $\{\varepsilon_t\}$ replaced by $\{\epsilon_t\}$. In addition, almost sure behaviors of the weighted sum of $\{\epsilon_t\}$ are obtained in Corollary 2.6. As will be shown later, Corollary 2.6 plays an important role in investigating asymptotic properties of least squares estimates in regression models.

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 (1.5). 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 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 ultimately chooses 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 (1.5). Some examples that help gain a better understanding of Theorem 4.1 are given at the end of Section 4.

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 $1 \leq i \leq n$. Assume $0 < \alpha < 1$. Then, for any sequence of real numbers $c_1, \ldots, c_n$,*

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

where k_α is a constant depending on α only, and $\gamma_n(k) = \max_{\substack{|i-j|=k \\ 1\leq i\leq j\leq n}} |E(f_i f_j)|$.

To show (2.1), two auxiliary inequalities are needed. The first one is due to Hardy, Littlewood and Pólya (1952, Thm. 381).

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

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

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 equence of random variables. Assume that there are nonnegative constants a_j satisfying $E|\sum_{j=1}^{i} Z_j|^p \leq (\sum_{j=1}^{i} a_j)^q$, for all $1 \leq i \leq n$. Then $E(\max_{1 \leq i \leq n} |\sum_{j=1}^{i} Z_j|^p) \leq C_{p,q}(\sum_{j=1}^{n} a_j)^q$ for some positive constant $C_{p,q}$ depending only on p and q.*

Proof of theorem 2.1. Fix $1 \leq i \leq n$. By observing that for $0 \leq k \leq n-1$, $1 \leq j_1 \leq j_2 \leq n$ and $|j_1 - j_2| = k$, $\gamma_n(k) \geq |E(f_{j_1} f_{j_2})|$, one gets

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

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

$$\Big(\sum_{j=1}^{i} c_j^2 \Big) \gamma_n(0) \leq \Big(\sum_{j=1}^{i} |c_j|^{\frac{2}{(2-\alpha)}} \Big)^{2-\alpha} \gamma_n(0). \qquad (2.3)$$

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

$$\Big(\sum_{j \neq l} \frac{|c_j c_l|}{|j-l|^{\alpha}} \Big) \leq M_\alpha \Big(\sum_{j=1}^{i} |c_j|^{\frac{2}{(2-\alpha)}} \Big)^{2-\alpha}. \qquad (2.4)$$

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

$$E\Big(|\sum_{j=1}^{i} c_j f_j|^2 \Big) \leq (M_\alpha + 1)\{ \max_{0 \leq k \leq n-1} \gamma_n(k)(k+1)^{\alpha} \} \Big(\sum_{j=1}^{i} |c_j|^{\frac{2}{(2-\alpha)}} \Big)^{2-\alpha}. \quad (2.5)$$

Since $2 - \alpha > 1$, apply Lemma 2.3 with $p = 2$, $q = 2 - \alpha$, $Z_j = c_j f_j$ and $a_j^{2-\alpha} = (M_\alpha + 1)\{\max_{0 \leq k \leq n-1} \gamma_n(k)(k+1)^{\alpha}\} c_j^2$ to finish the proof.

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

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

$$E\Big(\max_{1\leq i\leq n}\frac{|\sum_{j=1}^{i} c_j f_j|^2}{a_i^2}\Big) \leq 2k_\alpha\{\max_{0\leq k\leq n-1}\gamma_n(k)(k+1)^\alpha\}\Big(\sum_{i=1}^{n}|\frac{c_i}{a_i}|^{\frac{2}{(2-\alpha)}}\Big)^{2-\alpha}. \quad (2.6)$$

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

$$\max_{1\leq k\leq n}|\sum_{j=1}^{k}\nu_j|/a_k \leq 2\max_{1\leq k\leq n}|\sum_{j=1}^{k}\frac{\nu_j}{a_j}|. \quad (2.7)$$

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:

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

Applying (2.6) to a sequence of zero mean random variables $\{\epsilon_t\}$ with second moments satisfying (1.5), Corollary 2.5 brings (1.3) and (1.4) together in more general settings. 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 below.

Corollary 2.5. *Assume that* $\{\epsilon_t\}$ *is a sequence of zero mean random variables that satisfies* (1.5). *Then, for any* $m \geq 0$, *any* $n \geq 1$, *any sequence of real numbers* $c_j, j \geq 1$, *and any sequence of nondecreasing positive numbers* $a_j, j \geq 1$, *there is a positive constant* K_α *depending only on* α *such that*

$$E\Big(\max_{1\leq i\leq n}\frac{|\sum_{j=m+1}^{m+i} c_j\epsilon_j|^2}{a_{m+i}^2}\Big) \leq K_\alpha\Big(\sum_{i=m+1}^{m+n}|\frac{c_i}{a_i}|^{\frac{2}{(2-\alpha)}}\Big)^{2-\alpha}, \quad (2.8)$$

and for any $\delta > 0$,

$$\delta^2 P\Big(\max_{1\leq i\leq n}\frac{|\sum_{j=m+1}^{m+i} c_j\epsilon_j|^2}{a_{m+i}^2} > \delta\Big) \leq K_\alpha\Big(\sum_{i=m+1}^{m+n}|\frac{c_i}{a_i}|^{\frac{2}{(2-\alpha)}}\Big)^{2-\alpha}. \quad (2.9)$$

Remark 2. Assume that $a_i = 1$ for all $i \geq 1$. Then (2.8) becomes

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

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

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

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 $b_1 \geq \cdots \geq b_n > 0$,

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

where C_1 and C_2 are positive constants independent of n, m, and $\{b_j\}$. 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 the sequence $\{b_j\}$ is. 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 made smaller than any positive number, provided m is sufficiently large.

Remark 3. Kokoszka and Leipus (1998, Thm. 3.1) gave an inequality that is closely related to (1.4). Under (1.5), their inequality implies that, for any $\delta > 0$,

$$\delta^2 P\left(\max_{1\leq k\leq n} b_k\left|\sum_{j=1}^{k}\epsilon_j\right| > \delta\right) \leq \sum_{j=1}^{n-1}|b_{j+1}^2 - b_j^2| j^{2-\alpha} + C_4\sum_{j=1}^{n-1} b_{j+1}^2 j^{\frac{(2-\alpha)}{2}} + C_5\sum_{j=0}^{n-1} b_{j+1}^2, \tag{2.13}$$

where C_4 and C_5 are some positive constants independent of n and $\{b_t\}$, where $\{b_t\}$ is a sequence of positive numbers. As observed, 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 the Kokoszka and Leipus inequality is that the b_k are not necessarily 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 $\{\epsilon_j\}$ (with general weights), which seems indispensable for 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 $\sum_{i=1}^n c_i\epsilon_i$ and its order of magnitude in case of divergence.

Corollary 2.6. *Adopt the assumptions of Corollary* 2.5.
(i) *If* $\sum_{i=1}^\infty |c_i|^{2/(2-\alpha)} < \infty$, *then* $\sum_{i=1}^n c_i\epsilon_i$ *converges almost surely* (a.s.).
(ii) *If* $\sum_{i=1}^\infty |c_i|^{2/(2-\alpha)} = \infty$, *then for any* $\delta > 1-(\alpha/2)$,

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

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

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

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

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

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

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

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

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

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

Since

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

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 $1/n \sum_{i=1}^{n} \epsilon_i = o(1/n^{1-\theta})$ a.s. This result provides an almost sure convergence rate for the first sample moment of a sequence of zero mean random variables whose second moments obey (1.5).

3. Strong Consistency of Least Squares Estimates

Consider the multiple regression model

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

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

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

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

To study the strong consistency of $\mathbf{b}_n$ under (3.1) with ϵ_t satisfying (1.5), we first introduce the concept of a "convergence system". A sequence of random variables $\{\epsilon_t\}$ is said to be a convergence system if $\sum_{i=1}^{n} c_i \epsilon_i$ 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 $\{\epsilon_t\}$ 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 $\{g_i \epsilon_i\}$ 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), $V_n = (\nu_{ij}^{(n)})_{1\le i,j\le p} = (\sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j')^{-1}$ *exists for all* $n \ge m$, $\lim_{n\to\infty} \nu_{jj}^{(n)} = 0$, *and* $\{g_i\epsilon_i\}$ *is a convergence system for some constants* g_i *such that* $|g_i|$ *is positive and nonincreasing. Then,*

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

if

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

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* $\{h_i\}$, $i = 1, 2, \ldots$, *be sequences of nonnegative numbers with* $0 < h_i \le h_{i+1}$ *for all* $i \ge 1$. *Define* $\mu_i = \sum_{j=i}^{\infty} l_j$. *Then, for any* $k \ge 1$,

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

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

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

Since all terms involved are positive, the lemma obviously holds if $\sum_{i=k}^{\infty} h_i l_i = \infty$. If $\sum_{i=k}^{\infty} h_i l_i < \infty$, then $o(1) = \sum_{i=n+1}^{\infty} h_i l_i \ge h_n \sum_{i=n+1}^{\infty} l_i = h_n\mu_{n+1}$. In view of this and (3.6), (3.5) follows.

Proof of theorem 3.1. Without loss of generality, take $j = 1$. For $n > m$, let $T_n = (x_{n2}, \ldots, x_{np})'$ 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

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

$$\rho_n = \rho_m + \sum_{i=m+1}^{n} d_i \Big\{ \epsilon_i - T_i' H_{i-1}^{-1} \Big(\sum_{k=1}^{i-1} T_k \epsilon_k \Big) \Big\}. \tag{3.8}$$

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 $\{g_i \epsilon_i\}$ is a convergence system yield that

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

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

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

Therefore, by (3.9),

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

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

Next we apply Theorem 3.1 to the case where $\{\epsilon_t\}$ satisfies (1.5). We start with a lemma that gives a sufficient condition on g_i under which $\{g_i \epsilon_i\}$ is a convergence system.

Lemma 3.3. *Under the same assumptions of Corollary* 2.5 *and*

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

$\{g_i\epsilon_i\}$ 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 (e.g., Hardy, Littlewood and Pólya (1952, p.22)) with $p = 2-\alpha$ and $q = (2-\alpha)/(1-\alpha)$, one obtains that

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

In view of (3.11), $\sum_{i=1}^{\infty} a_i^2 < \infty$ ensures that $\sum_{i=1}^{\infty} |a_i g_i|^{2/(2-\alpha)} < \infty$.

A special class of g_n that satisfies (3.11) has $g_n^{-1} = (n\log n)^{(1-\alpha)/2}(\log n)^{\delta}$ 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 (1.5) is fulfilled by $\{\epsilon_t\}$, a set of sufficient conditions for strong consistency of $\mathbf{b}_n$ is given in the following corollary.

Corollary 3.4. *Under model* (3.1), *assume* (1.5) *and* $\lim_{n\to\infty} v_{jj}^{(n)} = 0$, *where* $v_{jj}^{(n)}$ *is defined as in Theorem* 3.1. *If, in addition, 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.

Remark 5. To obtain the strong consistency of $\mathbf{b}_n$, Yajima (1988) needed to replace $\{\epsilon_t\}$ in (3.1) with the stationary process $\{\varepsilon_t\}$ satisfying (1.1), and to assume that

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

where $\lambda_k = \lambda_{\min}(\sum_{j=1}^{k} \mathbf{x}_j\mathbf{x}_j')$ is the minimal eigenvalue of $\sum_{j=1}^{k} \mathbf{x}_j\mathbf{x}_j'$. Since for each j, $\lambda_k^{-1} \geq \nu_{jj}^{(k)}$, Corollary 3.4 can achieve the same goal under weaker assumptions on the design points as well as the noise process.

Assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E(\epsilon_k^2) = \gamma^*(0) \text{ exists.} \tag{3.13}$$

Corollary 3.5 below shows that under model (3.1), 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 $\gamma^*(0)$. Define $(l_{ij}^{(n)})_{1\le i,j\le p} = \sum_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j'$ and $\Lambda_n = (\rho_{ij}^{(n)})_{1\le i,j\le p}$, where $\rho_{ij}^{(n)} = l_{ij}^{(n)}/(l_{ii}^{(n)}l_{jj}^{(n)})^{1/2}$. The following assumptions are required in our analysis.

(C.1) $\liminf_{n\to\infty}\lambda_{\min}(\Lambda_n) > 0$.

(C.2) For any $\delta_1 > 0$, $\log(\bar{D}_n) = o(n^{\delta_1})$, where $\bar{D}_n = \max_{1\le i\le p} l_{ii}^{(n)}$.

(C.3) There exists $\theta > 0$ such that

$$S(\theta,\epsilon^2) = \sup_{k\ge 0}(1+k)^{\theta}\sup_{i,j\ge 1,|i-j|=k}|\mathrm{cov}(\epsilon_i^2,\epsilon_j^2)| < \infty.$$

Corollary 3.5. *Under model* (3.1), *assume that* (1.5), (3.13), *and* (C.1)−(C.3) *hold. Then*

$$\lim_{n\to\infty}\hat{\sigma}_n^2 = \gamma^*(0) \text{ a.s..} \tag{3.14}$$

To prove (3.14), an auxiliary lemma is needed.

Lemma 3.6. *Under model* (3.1), *assume that* (1.5) *and* (C.1) *are satisfied. Then, for any* $\delta > 2-\alpha$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'\epsilon_i V_n\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i = o\Big(\frac{\{\log n + \log(\bar{D}_n)\}^{\delta}}{n^{\alpha}}\Big) \text{ a.s..} \tag{3.15}$$

Proof. First note that

$$\sum_{j=1}^{n}\mathbf{x}_j'\epsilon_j V_n\sum_{j=1}^{n}\mathbf{x}_j\epsilon_j = \Big(D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\Big)' D_n V_n D_n\Big(D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\Big),$$

where $D_n = \mathrm{Diag}((l_{11}^{(n)})^{1/2},\ldots,(l_{pp}^{(n)})^{1/2})$ is a diagonal matrix with the ith diagonal element equal to $(l_{ii}^{(n)})^{1/2}$. By (C.1) and Corollary 2.6, one has for any $\delta > 2-\alpha$,

$$\Big(D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\Big)' D_n V_n D_n\Big(D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\Big) = O(1)\|D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\|^2$$

$$= O(1)\sum_{i=1}^{p}\left(\frac{1}{(l_{ii}^{(n)})^{\frac{1}{2}}}\sum_{j=1}^{n}x_{ji}\epsilon_j\right)^2 = o\left(\sum_{i=1}^{p}\frac{E_{in}^{2-\alpha}}{l_{ii}^{(n)}}(\log E_{in})^{\delta}\right) + O(1) \text{ a.s.,}$$

where $E_{in} = \sum_{j=1}^{n} |x_{ji}|^{2/(2-\alpha)}$. This result and the fact that $E_{in}^{2-\alpha} \le n^{1-\alpha} l_{ii}^{(n)}$ imply

$$\frac{1}{n}\Big(D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\Big)' D_n V_n D_n \Big(D_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_i\Big) = o\Big(\frac{\{\log n + \log(\bar{D}_n)\}^{\delta}}{n^{\alpha}}\Big) \text{ a.s.},$$

and hence (3.15) follows.

Proof of corollary 3.5. For $n \ge m$,

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 - \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j'\epsilon_j V_n \sum_{j=1}^{n}\mathbf{x}_j\epsilon_j. \tag{3.16}$$

By (C.2) and Lemma 3.6,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'\epsilon_i V_n \sum_{i=1}^{n}\mathbf{x}_i\epsilon_i = o(1) \text{ a.s..} \tag{3.17}$$

To deal with the first term on the right-hand side of (3.16), first assume that (C.3) holds for $0 < \theta < 1$. Applying Corollary 2.5 to $\{\epsilon_t^2 - E\epsilon_t^2\}$, one has for $1 \le N_1 < N_2$ and some $C > 0$ (independent of N_1 and N_2),

$$E\Big(\max_{N_1\le j\le N_2}\Big(\sum_{j=N_1}^{N_2}\frac{\epsilon_j^2 - E\epsilon_j^2}{j}\Big)^2\Big) \le C\Big(\sum_{j=N_1}^{N_2} j^{\frac{-2}{(2-\theta)}}\Big)^{2-\theta}, \tag{3.18}$$

which, together with Kronecker's lemma and (3.13), yields

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 = \gamma^*(0) \text{ a.s..} \tag{3.19}$$

Now, assume that (C.3) holds for some $\theta \ge 1$. Then, for any $0 < \theta' < 1$, $S(\theta', \epsilon^2) < \infty$. This and Corollary 2.5 ensure that (3.18) holds with θ replaced by any θ'. Hence, (3.19) is still valid for the case of $\theta \ge 1$. As a result, (3.14) follows from (3.16), (3.17) and (3.19).

Remark 6. If (3.13) is replaced with

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}E(\epsilon_k^2) > 0, \tag{3.20}$$

then the same argument as in Corollary 3.5 yields

$$\liminf_{n\to\infty}\hat{\sigma}_n^2 > 0 \text{ a.s..} \tag{3.21}$$

Remark 7. (C.3) is easily fulfilled in practice. For example, if

$$\epsilon_t = \sum_{j=0}^{\infty} w_j \nu_{t-j}, \tag{3.22}$$

where $\{\nu_j\}$ is a sequence of independent random variables with $E(\nu_j) = 0$ and

$$\sup_j E\nu_j^4 < \infty, \tag{3.23}$$

and $\{w_j\}$ is a sequence of real numbers with $w_0 = 1$ and $w_j = O(j^{-\iota})$ for some $1/2 < \iota < 1$, then straightforward calculations yield (1.5) with $\alpha = 2\iota - 1$ and (C.3) with $\theta = 2\alpha$. In addition, when (1.5) is assumed and $\{\epsilon_t\}$ is a Gaussian process, it is also not difficult to see that (C.3) holds for $\theta = 2\alpha$.

4. Strongly Consistent Model Selection

As mentioned in Section 1, when some $\beta_j, 1 \le j \le p$, in model (3.1) vanishes, adopting a subset model may improve estimation and prediction efficiency. Let $\bar{M} = \{x_1, \ldots, x_p\}$ denote the full model described in (3.1). Based on $\{y_1, \ldots, y_n, x_{1j}, \ldots, x_{nj}, j = 1, \ldots, p\}$, 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 $M_0 = \{x_1, \ldots, x_q\}$, where $1 \le q \le p$. Hence, (3.1) can be rewritten as

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

where $\beta_i \neq 0$ for all $1 \le i \le q$. For $M \in \mathcal{M}$, define a loss function

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

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 $\hat{M}_n \in \mathcal{M}$ satisfy $L_n(\hat{M}_n) = \min_{M \in \mathcal{M}} L_n(M)$. In situations where (1.5) is fulfilled by $\{\epsilon_t\}$, Theorem 4.1 below shows that

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

provided P_n is suitably chosen. To achieve (4.3), we require a stronger condition than (C.1).

(C.1') $\liminf_{n\to\infty} \lambda_{\min}(\Lambda_n) > 0$, and for some $r_1 > 1 - \alpha$,

$$\liminf_{n\to\infty} \frac{\min_{1\le i\le p} l_{ii}^{(n)}}{n^{r_1}} > 0. \tag{4.4}$$

Theorem 4.1. *Under models* (3.1) *and* (4.1)*, assume that* (1.5), (3.20), (C.1'), (C.2) *and* (C.3) *hold. Let* P_n *in* (4.2) *satisfy*

$$\lim_{n\to\infty} n^{1-\min\{1,r_1\}} P_n = 0, \tag{4.5}$$

and for some $0 < \eta < \alpha$,

$$\liminf_{n\to\infty} n^{\eta} P_n > 0. \tag{4.6}$$

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 $M \neq M_0$, then we are going to prove that

$$P(L_n(M) > L_n(M_0), \text{eventually}) = 1. \tag{4.7}$$

When $q < p$, there is a model $M_u \supseteq M_0$ which satisfies $\text{card}(M_u) < \text{card}(\bar{M})$. 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 (4.7), it suffices to show that

$$P(\log \hat{\sigma}_n^2(M_u) - \log \hat{\sigma}_n^2(M_u^*) - P_n < 0, \text{eventually}) = 1. \tag{4.8}$$

First note that

$$\log \hat{\sigma}_n^2(M_u) - \log \hat{\sigma}_n^2(M_u^*) \le (\hat{\sigma}_n^2(M_u) - \hat{\sigma}_n^2(M_u^*))/\hat{\sigma}_n^2(M_u^*). \tag{4.9}$$

The same reasoning that shows (3.21) yields

$$\liminf_{n\to\infty} \hat{\sigma}_n^2(M_u^*) > 0 \text{ a.s.}. \tag{4.10}$$

In view of these and (4.6), (4.9) and (4.10), (4.8) is guaranteed by showing that for any $0 < \eta_1 < \alpha$,

$$n\Big(\hat{\sigma}_n^2(M_u) - \hat{\sigma}_n^2(M_u^*)\Big) = o(n^{1-\eta_1}) \text{ a.s.}. \tag{4.11}$$

Now,

$$\begin{aligned} n\Big(\hat{\sigma}_n^2(M_u) - \hat{\sigma}_n^2(M_u^*)\Big) &= \Big(\sum_{i=1}^n \mathbf{u}_i^{*\prime} \epsilon_i\Big) V_n(M_u^*) \Big(\sum_{i=1}^n \mathbf{u}_i^* \epsilon_i\Big) \\ &\quad - \Big(\sum_{i=1}^n \mathbf{u}_i' \epsilon_i\Big) V_n(M_u) \Big(\sum_{i=1}^n \mathbf{u}_i \epsilon_i\Big), \end{aligned}$$

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.2) and the same argument used to obtain Lemma 3.6, we have for any $0 < \eta_1 < \alpha$,

$$\Big(\sum_{i=1}^n \mathbf{u}_i' \epsilon_i\Big) V_n(M_u) \Big(\sum_{i=1}^n \mathbf{u}_i \epsilon_i\Big) = o\left(n^{1-\eta_1}\right) \text{ a.s.,} \tag{4.12}$$

$$\Big(\sum_{i=1}^n \mathbf{u}_i^{*'} \epsilon_i\Big) V(M_u^*) \Big(\sum_{i=1}^n \mathbf{u}_i^* \epsilon_i\Big) = o\left(n^{1-\eta_1}\right) \text{ a.s..} \tag{4.13}$$

Consequently, (4.11) follows.

The remaining part of the proof focuses on the underspecified case. In particular, we show that for any pair of models M_u and M_v, with $M_u \supseteq M_0$, $M_v \subset M_u$, and $M_0 \neq M_v$,

$$P(\log \hat{\sigma}_n^2(M_v) - \log \hat{\sigma}_n^2(M_u) - (\text{card}(M_u) - \text{card}(M_v)) P_n > 0, \text{eventually}) = 1. \tag{4.14}$$

Without loss of generality, assume $M_u = \{x_1, \ldots, x_{q+s}\}$ for some $0 \le s \le p - q$ and $M_v \supseteq \{x_l, \ldots, x_q\}$ for some $2 \le l \le q$. By Wei (1992, Thm. 3.1),

$$\begin{aligned} &\hat{\sigma}_n^2(M_v) - \hat{\sigma}_n^2(M_u) \ge \hat{\sigma}_n^2(\underline{M}_v) - \hat{\sigma}_n^2(M_u) \ge \hat{\sigma}_n^2(\underline{M}_v) - \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 \\ &= \frac{1}{n}\Big(\beta_1^2 s_{n(u)} + 2\beta_1 \sum_{i=1}^n (x_{i1} - K_{n(u)} H_{n(u)}^{-1} T_{i(u)}) \epsilon_i - R_n\Big), \end{aligned} \tag{4.15}$$

where $\underline{M}_v = \{x_2, \ldots, x_{q+s}\}$, $s_{n(u)} = \sum_{i=1}^n (x_{i1} - K_{n(u)} H_{n(u)}^{-1} T_{i(u)})^2$, $T_{i(u)} = (x_{i2}, \ldots, x_{i(q+s)})'$, $R_n = (\sum_{i=1}^n T_{i(u)}' \epsilon_i) H_{n(u)}^{-1} (\sum_{i=1}^n T_{i(u)} \epsilon_i)$, and with $\mathbf{u}_j$ being the ith regressor corresponding to model M_u,

$$\begin{pmatrix} \sum_{i=1}^n x_{i1}^2 & K_{n(u)} \\ K_{n(u)}' & H_{n(u)} \end{pmatrix} = \sum_{j=1}^n \mathbf{u}_j \mathbf{u}_j'.$$

By the same reasoning as (4.12) and (4.13), we have for any $0 < \eta_1 < \alpha$,

$$R_n = o\left(n^{1-\eta_1}\right) \text{ a.s..} \tag{4.16}$$

Corollary 2.6 ensures that for any $\delta > 1 - (\alpha/2)$,

$$\sum_{i=1}^n \left(x_{i1} - K_{n(u)} H_{n(u)}^{-1} T_{i(u)}\right) \epsilon_i = O(1) + o\Big(F_n^{\frac{(2-\alpha)}{2}} (\log F_n)^\delta\Big) \text{ a.s.,} \tag{4.17}$$

where $F_n = \sum_{i=1}^n |x_{i1} - K_{n(u)} H_{n(u)}^{-1} T_{i(u)}|^{2/(2-\alpha)}$. According to (C.1'), (C.2), (4.15)$-$(4.17), using $F_n^{2-\alpha} \leq n^{(1-\alpha)} s_{n(u)}$ and

$$(\frac{1}{q})\lambda_{\min}\Big(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\Big) \leq s_{n(u)} \leq \|\mathbf{x}_1^{(n)}\|^2 \tag{4.18}$$

(see Lai and Wei (1982)), one has

$$\hat{\sigma}_n^2(M_v) - \hat{\sigma}_n^2(M_u) \geq \frac{s_{n(u)}\beta_1^2}{n}\Big(1 + o(1)\Big) \text{ a.s..} \tag{4.19}$$

This, (4.18), (C.1'), and the fact that $\liminf_{n\to\infty} \hat{\sigma}_n^2(M_u) > 0$ (which is guaranteed by the same reasoning as (4.10)) further yield that

$$\liminf_{n \to\infty}\Big(\log \hat{\sigma}_n^2(M_v) - \log \hat{\sigma}_n^2(M_u)\Big) n^{1-\min\{1,r_1\}} > 0 \text{ a.s..} \tag{4.20}$$

Consequently, (4.14) follows from (4.5) and (4.20).

Remark 8. The assumptions used in Theorem 4.1 are almost the same as those used in Corollary 3.5, except that (3.13) is replaced by (3.20) and (4.4) is added into (C.1). In fact, (4.4) is given to ensure that the signal in the model is strong enough so that a suitable P_n (which satisfies (4.5)) can be introduced to prevent $L_n(M)$ from choosing an underspecified model, as clarified in (4.15)$-$(4.20). On the other hand, (4.6) is used to avoid overfitting.

Remark 9. When (C.2) in Theorem 4.1 is strengthened to

(C.2') for some $r_3 \geq 1$, $\log(\bar{D}_n) = O((\log n)^{r_3})$,

and (4.6) is weakened to

$$\liminf_{n\to\infty} \frac{n^\alpha P_n}{(\log n)^{2r_3}} > 0, \tag{4.21}$$

(4.3) still follows.

The following examples help gain further insight into Theorem 4.1.

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

Example 2. When (C.1') and (C.2) are satisfied, to attain (4.5) and (4.6) (so that $L_n(M)$ is strongly consistent), P_n can be chosen to be C/n^{η_2}, where $C > 0$ and $1 - \min\{1, r_1\} < \eta_2 < \alpha$. If (C.2') is also satisfied, then by Remark 9, a milder penalty $P_n = C_1(\log n)^{2r_3}/n^{\alpha}, C_1 > 0$, would suffice to achieve the same purpose. However, these choices of P_n are not applicable in situations where α is unknown. Fortunately, in case $r_1 \geq 1$ (which occurs in many practical situations, as shown in Example 1), this dilemma can be alleviated by letting

$$P_n = \frac{C_2}{(\log n)^{C_3}}, \tag{4.22}$$

where C_2 and C_3 are some positive numbers. It is not difficult to see that (4.5) and (4.6) (or (4.5) and (4.21)) are attained by (4.22), without information about α.

Example 3. When (C.1') and (C.2') are assumed, $L_n(M)$ cannot be consistent even if (4.21) 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 $\{\epsilon_t\}$ is second-order stationary and satisfies (1.2). By some algebraic manipulations, it can be shown that

$$\liminf_{n\to\infty} \frac{E\{n(\hat{\sigma}_n^2(M_0) - \hat{\sigma}_n^2(\bar{M}))\}}{n^{1-\alpha}} > C_\alpha, \tag{4.23}$$

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 only on α. If we further assume that $\{\epsilon_t\}$ is a Gaussian process, then by Corollary 3.5 and (4.23),

$$\liminf_{n\to\infty} P\Big(\log \hat{\sigma}_n^2(M_0) - \log \hat{\sigma}_n^2(\bar{M}) > P_n\Big) > 0, \tag{4.24}$$

provided $P_n = O(n^{-\alpha})$. Inequality (4.24) 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. Chen and Ni (1989) have shown that BIC is strongly consistent, provided $\{\epsilon_t\}$ is a stationary short memory process with spectral density that is bounded and bounded away from zero.

Before leaving this section, we note that since (1.5) is satisfied by both short and long memory time series, Theorem 4.1 is especially useful in situations where the strength of dependence of $\{\epsilon_t\}$ is unknown. More specifically, assume that (C.1'), with $r_1 \geq 1$, and (C.2) (or (C.2')) hold, and that $\{\epsilon_t\}$ is an ARFIMA$(0, d, 0)$ process, where $-1/2 < d < 1/2$ is unknown. The latter ensures that ϵ_t has an infinite moving-average representation as given by (3.22). We also assume that ν_t in (3.22) satisfies (3.23). By Brockwell and Davis (1987, p.467), Theorem 4.1, Remark 7 and Example 2, $L_n(M)$ with $P_n = C_1/(\log n)^{C_2}$, for some $C_1, C_2 > 0$, is strongly consistent. On the other hand, since Brockwell and Davis (1987, p.467) also showed that (1.2) is fulfilled by an ARFIMA$(0, d, 0)$ model with $0 < d < 1/2$, according to Example 3, BIC is not consistent once the value of d falls in $(0, 1/2)$.

Acknowledgement

The authors would like to thank Dr. Hwai-Chung Ho for his valuable and constructive comments on previous versions of this paper. They would also like to thank the referee for many brilliant suggestions (in particular (1.5), (C.1) and (C.3)) that helped improve the original results substantially. Research of the first author was partially supported by the National Science Council of Taiwan under grant NSC 92-2118-M-001-015.

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.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.
Department of Economics, National Taiwan University, Taipei 100, Taiwan.

E-mail: cking@stat.sinica.edu.tw

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

E-mail: czw@stat.sinica.edu.tw

(Received June 2005; accepted March 2006)