Statistica Sinica **9**(1999), 289-296

STRONG CONSISTENCY OF LEAST SQUARES ESTIMATE IN MULTIPLE REGRESSION WHEN THE ERROR VARIANCE IS INFINITE

Jin Mingzhong and Chen Xiru

GuiZhou National College and Graduate School, Chinese Academy

Abstract: Let $Y_i = x_i'\beta + e_i$, $1 \le i \le n$, $S_n = \sum_{i=1}^n x_i x_i'$. Suppose that the random errors $e_1, e_2, \ldots$ are i.i.d., with a common distribution F belonging to the class $\mathcal{F}_r = \{F : \int_{-\infty}^{\infty} x dF = 0, 0 < \int_{-\infty}^{\infty} |x|^r dF < \infty\}$ for some $r \in [1, 2)$. In this paper we obtain a sufficient condition for the strong consistency of the Least Sequares Estimate (LSE) $\hat{\beta}_n$ of β. The condition is necessary in the following sense: If the condition is not satisfied, then for some $F \in \mathcal{F}_r, \hat{\beta}_n$ rails to converge a.s. to β.

Key words and phrases: Least squares estimate, linear models, strong consistency.

1. Introduction and Main Results

Consider the linear model

$$Y_i = x_i'\beta + e_i, \qquad 1 \le i \le n, n \ge 1. \tag{1.1}$$

In this paper we assume that $x_1, x_2, \ldots$ are known non-random p-vectors, and e_i is the random error in the ith observation, $i = 1, 2, \ldots$ The LSE of β will be denoted by $\hat{\beta}_n$.

Many statisticians have considered the problem of strong consistency of $\hat{\beta}_n$. In earlier days the problem was studied under the assumption that the e_i possess a finite variance. This case was finally solved in an important paper of Lai, Robbins and Wei (1979) in which they showed that if $e_1, e_2, \ldots$ are i.i.d., $Ee_1 = 0$ and $0 < Ee_1^2 < \infty$, a sufficient condition for the strong consistency of $\hat{\beta}_n$ is that

$$S_n^{-1} = \Big(\sum_{i=1}^n x_i x_i'\Big)^{-1} \to 0, \quad \text{as } n \to \infty. \tag{1.2}$$

Since the necessity of (1.2) was earlier proved by Drygas (1976), (1.2) is both necessary and sufficient. Later, a number of authors considered the case where e_i does not have a finite variance. A standard formulation is as follows: $e_1, e_2, \ldots$ are i.i.d. with a common distribution F belonging to the family

$$\mathcal{F}_r = \Big\{F : \int_{-\infty}^{\infty} x dF = 0,\ 0 < \int_{-\infty}^{\infty} |x|^r dF < \infty\Big\}, \quad 1 \le r < 2. \tag{1.3}$$

Under this condition Chen (1981) showed that for general p, if $S_n^{-1} = O(n^{-(2-r)/r}$ $(\log n)^{-a})$ for some $a > 1$ and some additional conditions are met, then $\hat{\beta}_n \to \beta$ a.s.. Almost at the same time Chen, Lai and Wei (1981) showed that strong consistency holds under the condition $S_n^{-1} = O(n^{-(2-r)/r}(\log n)^{-2/r-\varepsilon})$ for some $\varepsilon > 0$. In his unpublished doctoral dissertation, Zhu (1989) made an essential improvement: for general p the condition $S_n^{-1} = O(n^{-(2-r)/r})$ is already sufficient for $\hat{\beta}_n \to \beta$ a.s.. This conditon, though far weaker the earlier-mentioned conditions, is still not the best (see Remark 2 below). The best condition for the case of $p = 1$ was obtained by Chen, Zhu and Fang (1996) as a by-product of a more general result. The purpose of this paper is to give a solution for general p.

Assume that S_i^{-1} exists for large i. Write $a_i = S_i^{-1}x_i$. For small i such that S_i^{-1} does not exist, a_i can be defined arbitrarily. Define

$$N(K) = \#\{i : i \geq 1, \|a_i\| \geq K^{-1}\}.$$

Let $\{(n,1),\ldots,(n,n)\}$ be a permutation of $\{1,2,\ldots,n\}$ satisfying $\|a_{(n,1)}\| \geq \cdots \geq \|a_{(n,n)}\|$. Put

$$V(n,j) = S_n^{-1}\sum_{i=1}^{n} x_i I(\|a_i\| \geq \|a_{(n,j)}\|), \quad 1 \leq j \leq n,$$
$$V(n) = \max_{1\leq j\leq n} \|V(n,j)\|.$$

Now we can formulate the main result of this paper.

Theorem. *Suppose in model* (1.1) *that the* $e_1, e_2, \ldots$ *are i.i.d., with a common distribution* F *belonging to* $\mathcal{F}_r$*. Then a sufficient condition for the strong consistency of* $\hat{\beta}_n$ *is:*

$$\text{For } 1 < r < 2 : (1,2) \text{ holds and } N(K) = O(K^r) \text{ as } K \to \infty. \tag{1.4}$$
$$\text{For } r = 1 : (1.2) \text{ holds } N(K) = O(K) \text{ and } V(n) = O(1). \tag{1.5}$$

The condition is also necessary in the following sense: if (1.3) and (1.4) are not satisfied, then for some $F \in \mathcal{F}_r, \hat{\beta}_n$ fails to converge to β almost surely.

Remark 1. Consider model (1.1). Assume that the random errors $e_1, e_2, \ldots$ are i.i.d., and F is their common distribution. As mentioned earlier, for $F \in \mathcal{F}_2$ the necessary and sufficient condition for the strong consistency of $\hat{\beta}_n$ is that $S_n^{-1} \to 0$. This is true for *every* $F \in \mathcal{F}_2$, and we may simply say that the condition $S_n^{-1} \to 0$ is a necessary and sufficient condition for the class $\mathcal{F}_2$.

For $1 \leq r < 2$ the situation is more complicated. The above theorem does not imply that the condition $C_r \equiv$ "(1.3) or (1.4)" is a necessary and sufficient condition for the class $\mathcal{F}_r$. For according to the above theorem, when condition

C_r is not satisfied, we can only assert that there exists *at least* one $F \in \mathcal{F}_r$ such that if F is the common distribution of the $e_i s$, then "$\hat{\beta}_n \to \beta$ a.s." does not hold.

Can we find a condition D_r(involving only $\{x_i\}$) which is a necessary and sufficient condition for the strong consistency of $\hat{\beta}_n$ for the whole class $\mathcal{F}_r$? Evidently this is impossible. For since $\mathcal{F}_2 \subset \mathcal{F}_r$ for any $r \in [1, 2)$, if such a condition D_r exists, it must be (1.2). But according to our theorem, (1.2) alone is evidently not sufficient: A simple example shows that the condition $N(K) = O(K^r)$ is not a consequence of (1.2).

The meaning of our theorem can also be understood in the following way: if one wants a condition depending solely upon $\{x_i\}$ which is sufficient for the strong consistency of $\hat{\beta}_n$ for the whole class $\mathcal{F}_r$, then the condition stated in the theorem is the best possible.

Remark 2. A simple example shows that Zhu's (1989) condition $S_n^{-1} = O(n^{-(2-r)/r})$ is stronger than ours. Consider model (1.1) in which β is one-dimensional and

$$x_i = \begin{cases} 1, & \text{when } i = 2^1, 2^2, 2^3, \dots \\ 0, & \text{otherwise.} \end{cases} \tag{1.6}$$

Then $S_n = O(\log n)$ and Zhu's condition fails. We cannot assert from Zhu's theorem that $\hat{\beta}_n \to \beta$, a.s., but $\hat{\beta}_n \to \beta$ a.s. according to Kolmogorov's strong law of large numbers.

On the other hand it is easy to verify that sequence (1.5) satisfies (1.2)-(1.4). So the strong consistency of $\hat{\beta}_n$ follows from our theorem.

2. The Necessity of (1.2)

This follows directly from the following lemma of Chen (1981): if $\{e_1, e_2, \dots\}$ is a sequence of independent random variables containing no asymptotically degenerate subsequence (i.e. a subsequence $\{e_{n_i}\}$ such that $e_{n_i} - c_i \to 0$ in pr. for some constant sequence $\{c_i\}$), and $\{c_{ni}, 1 \le i \le n, n \ge 1\}$ is an array of constants, then $\sum_{i=1}^n c_{ni} e_i \to 0$ in pr. entails $\sum_{i=1}^n c_{ni}^2 \to 0$.

3. The Necessity of $N(K) = O(K^r)$

For a matrix $A = (a_{ij})$, define the matrix norm $|A| = \max_{i,j} |a_{ij}|$. Then it can easily be shown that $|S_n^{-1} S_{n-1}| = O(1)$. If $S_n^{-1} \to 0$, then

$$\lim_{i \to \infty} a_i = 0, \ \max_{1 \le i \le n} \|S_n^{-1} x_i\| \to 0.$$

Remember $a_i = S_i^{-1} x_i$. Now suppose $N(K)$ is not $O(K^r)$. We shall find a sequence $\{e_1, e_2, \dots\}$ of i.i.d. $r.v'.s$ with common distribution belonging to the family $\mathcal{F}_r$ such that

$$S_n^{-1}\sum_{i=1}^n x_i e_i \not\to 0, \quad a.s. \tag{3.1}$$

We can find a sequence of positive integers $n_1 < n_2 < \cdots$ such that $N(n_k)/n_k^r \to \infty$ as $k \to \infty$. Therefore there exists $\{p_k, k \geq 1\}$, such that

$$p_k > 0, \sum_{k=1}^\infty p_k = 1, \sum_{k=1}^\infty n_k^r p_k < \infty, \text{ and } \sum_{k=1}^\infty p_k N(n_k) = \infty.$$

Let $\{e_1, e_2, \ldots\}$ be an i.i.d. sequence with a common distribution F:

$$P(e_1 = n_k) = p(e_1 = -n_k) = p_k/2, \qquad k \geq 1.$$

Then F belongs to $\mathcal{F}_r$. Since $a_n \to 0$, we can rearrange $\{\|a_i\|, i \geq 1\}$ in a decreasing order: $\|a_{(1)}\| \geq \|a_{(2)}\| \geq \cdots$. Note that $\{(1), (2), \ldots\}$ is a permutation of $\{1, 2, \ldots\}$ and, by definition of $N(K)$, it follows that $\{|e_1| \geq \|a_{(i)}\|^{-1}\} \Rightarrow \{N(|e_1|) \geq (i)\}$.

Therefore

$$\begin{aligned}\infty &= \sum_{k=1}^\infty p_k N(n_k) = E(N(|e_1|)) = \sum_{i=1}^\infty P(N(|e_1|) \geq i) = \sum_{i=1}^\infty P(N(|e_1|) \geq (i)) \\ &\geq \sum_{i=1}^\infty P(|e_1| \geq \|a_{(i)}\|^{-1}) = \sum_{i=1}^\infty P(|e_1| \geq \|a_i\|^{-1}) = \sum_{i=1}^\infty P(\|a_i e_i\| \geq 1), \end{aligned} \tag{3.2}$$

which entails $P(\|a_i e_i\| \geq 1, i.o.) = 1$. Then (3.1) is proved and hence the necessity of $N(K) = O(K^r)$ follows.

4. Sufficiency: $1 < r < 2$

Lemma 1. *Let a_i be defined as earlier. The convergence of $\sum_{i=1}^\infty a_i e_i$ entails $S_n^{-1}\sum_{i=1}^n x_i e_i \to 0$.*

Proof. Write $T_o = 0$, $T_j = \sum_{i=1}^j a_i e_i$, and $T = \sum_{i=1}^\infty a_i e_i$. We have

$$\begin{aligned}\Big|S_n^{-1}\sum_{i=1}^n x_i e_i\Big| &= \Big|S_n^{-1}\sum_{j=1}^n x_j x_j'(T_n - T_{j-1})\Big| = \Big|S_n^{-1}\sum_{j=1}^n x_j x_j'(T_n - T - (T_{j-1} - T))\Big| \\ &\leq |T_n - T| + \sum_{j=1}^k \delta_{jn}|T_{j-1} - T| + \sum_{j=k+1}^n \delta_{jn}|T_{j-1} - T|, \end{aligned} \tag{4.1}$$

where $\delta_{jn} = x_j' S_n^{-1} x_j$ and k is a large integer for which $|T_n - T|$ and $|T_l - T|$ are small for $l > k$. Since $\delta_{jn} \to 0$ for fixed j (which follows from $S_n^{-1} \to 0$), and $\sum_{j=1}^n \delta_{jn} = p$, each term in the right hand side of (4.1) can be made arbitrarily small, and the lemma follows.

Now fix $j \in \{1, \ldots, p\}$. Denote by d_i the jth component of a_i. Define $e_i' = e_i I(|e_i| < |d_i|^{-1})$. To prove the strong consistency of $\hat{\beta}_n$, by Lemma 1 we need only show

$$\sum_{i=1}^{\infty} d_i e_i' \text{ converges a.s..} \tag{4.2}$$

Then, applying Kolmogorov's three series theorem, we need to verify

$$\sum_i P(|e_i d_i| \geq 1) < \infty, \tag{4.3}$$

$$\sum_i E d_i e_i I_{[|e_i d_i| < 1]} \text{ converges ,} \tag{4.4}$$

$$\sum_i E d_i^2 e_i^2 I_{[|e_i d_i| < 1]} < \infty. \tag{4.5}$$

Since

$$\begin{aligned} P(|e_i d_i| \geq 1) &= P(|e_i| \geq |d_i|^{-1}) = P(|e_i| I(|e_i| \geq |d_i|^{-1}) \geq |d_i|^{-1}) \\ &\geq |d_i| E(|e_i| I(|e_i| \geq |d_i|^{-1})), \end{aligned}$$

the proof of (4.3) follows from the argument below for (4.4). Also, the proof of (4.5) is similar to (4.4). So we proceed with (4.4). Let $q_i = P(i-1 \leq |e_1| < i)$, $i = 1, 2, \ldots$ Since $Ee_i = 0$, we have $Ee_i' = -E(e_i I(|e_i| \geq |d_i|^{-1}))$. If $k - 1 \leq |d_i|^{-1} < k$, we have

$$E|d_i e_i'| \leq |d_i| E|e_i I(|e_i| \geq |d_i|^{-1})| \leq (k-1)^{-1} \sum_{j=k-1}^{\infty} j q_j, \quad k \geq 2.$$

Further, noticing that $\#\{i : i \geq 1, k-1 < |d_i|^{-1} \leq k\} = \tilde{N}(k) - \tilde{N}(k-1)$, where $\tilde{N}(k) = \#\{i : i \geq 1, |d_i|^{-1} \leq k\}$, we have

$$\begin{aligned} \sum_{i=1}^{\infty} |E d_i e_i'| &\leq \tilde{N}(1) \sup_{i \geq 1} |d_i| E|e_1| + \sum_{k=2}^{\infty} (\tilde{N}(k) - \tilde{N}(k-1))(k-1)^{-1} \sum_{j=k-1}^{\infty} j q_j \\ &\equiv J_1 + J_2. \end{aligned} \tag{4.6}$$

Since $a_n \to 0$, J_1 remains bounded as $n \to \infty$. On the other hand, we have

$$\begin{aligned} J_2 = &\sum_{j=1}^{\infty} j^{-1} \tilde{N}(j+1) j q_j - \sum_{j=1}^{\infty} \tilde{N}(1) j q_j + \sum_{j=1}^{\infty} \Big(\sum_{k=2}^{j} \tilde{N}(k)((k-1)^{-1} - k^{-1}) \Big) j q_j \\ &-H_1 - H_2 + H_3. \end{aligned}$$

From $\tilde{N}(j+1) \leq c(j+1)^r$ (for some c) and $E|e_1|^r < \infty$, it follows that $H_1 < \infty$. Likewise for H_2. As for H_3, $\tilde{N}(k) = O(k^r)$, $(k-1)^{-1} - k^{-1} = O(k^{-2})$, and $r > 1$, so $\sum_{k=2}^{j}\tilde{N}(k)((k-1)^{-1}-k^{-1})=O(j^{r-1})$ and $H_3<\infty$ follows from the fact that $E|e_1|^r<\infty$. Summing up, and noticing (4.6), we have (4.4), concluding this part of the proof.

5. Sufficiency: $r=1$

The above argument breaks down for the case $r=1$, since in this case we can only get $H_3=O(\log n)$ and not $O(1)$. This is the reason for imposing the additional condition $V(n)=O(1)$.

Under the condition $V(n)=O(1)$, apply the arguments in Section 4 to the sequence $\{e_i-Ee_iI_{[|e_i|<\|a_i\|^{-1}]}\}$ to set

$$S_n^{-1}\sum_{i=1}^{n}x_i(e_i''-Ee_i'')\to 0,\ \text{a.s. (where, } e_i''=e_iI(|e_i|<\|a_i\|^{-1})). \tag{5.1}$$

Therefore we need only show

$$S_n^{-1}\sum_{i=1}^{n}x_iEe_i''\to 0. \tag{5.2}$$

To this end, define

$$t_i=\int_{\|a(n,i)\|^{-1}\le|x|<\|a_{(n,i+1)}\|^{-1}}xdF,\quad 1\le i\le n,$$

with the convention $\|a_{(n,n+1)}\|^{-1}=\infty$. We have

$$\begin{aligned}-E(S_n^{-1}\sum_{i=1}^{n}x_ie_i'')&=S_n^{-1}\sum_{i=1}^{n}x(n,i)\int_{|x|\ge\|a_{(n,i)}\|^{-1}}xdF\\&=S_n^{-1}\sum_{i=1}^{n}x_{(n,i)}\sum_{j=i}^{n}t_j=S_n^{-1}\sum_{j=1}^{n}t_j\sum_{i=1}^{j}x_{(n,i)}.\end{aligned}$$

Two cases are possible: the first case is $\|a_{(n,j)}\|=\|a_{(n,j+1)}\|$. Then $t_j=0$ and $t_jS_n^{-1}\sum_{i=1}^{j}x_{(n,i)}=t_jV(n,j)$. The second case is $\|a_{(n,j)}\|>\|a_{(n,j+1)}\|$. Then we have $S_n^{-1}\sum_{i=1}^{j}x_{(n,i)}=V(n,j)$ by definition. Hence we always have $t_jS_n^{-1}\sum_{i=1}^{j}x_{(n,i)}=t_jV(n,j)$. It follows that

$$-E(S_n^{-1}\sum_{i=1}^{n}x_ie_i'')=\sum_{j=1}^{n}t_jV(n,j)=\sum_{j=1}^{h}t_jV(n,j)+\sum_{j=h+1}^{n}t_jV(n,j)\equiv J_1+J_2, \tag{5.3}$$

where h is fixed. Without loss of generality assume $x_i\neq 0$ for all $i\ge 1$, and if S_i^{-1} does not exist choose $a_i\neq 0$. Then $a_i\neq 0$ for all $i\ge 1$. Since $\lim_{n\to\infty}a_n=0$, we have $\lim_{n\to\infty}(n,i)=(i)$. Hence, considering $S_n^{-1}\to 0$, we have

$$\limsup_{n\to\infty}\|V(n,j)\|\le\limsup_{n\to\infty}\sum_{i=1}^{j'}\|S_n^{-1}x_{(i)}\|=0, \tag{5.4}$$

where $j'=\max(l:\|a_{(l)}\|=\|a_{(j)}\|)$. From (5.4) we get $\lim_{n\to\infty}J_1=0$ for fixed h. Further, by assumption (1.4), $\|V(n,j)\|\le V(n)=O(1)$. Hence

$$\|J_2\|\le O(1)\sum_{j=h+1}^{n}t_j\le O(1)\int_{|x|\ge\|a_{(n,h+1)}\|^{-1}}|x|dF\to O(1)\int_{|x|\ge\|a_{(h+1)}\|^{-1}}|x|dF.$$

The last integral can be made arbitrarily small by cloosing h large enough. Summing up and noticing (5.3), we obtain (5.1). The result in (5.2) can be proved similarly (and without the assumption $V(n)=O(1)$).

6. The Necessity of $V(n)=O(1)$ for $r=1$

Suppose that $\hat\beta_n$ is strongly consistent, so (3.1) holds. From Section 2 and Section 3, we have $S_n^{-1}\to 0$ and $N(K)=O(K)$. As pointed out at the end of Section 5, these two facts entail (5.2). Therefore $S_n^{-1}\sum_{i=1}^n x_i(e_i''-Ee_i'')\to 0$, a.s.. These two facts, together with (3.1), entail $S_n^{-1}\sum_{i=1}^n x_ie_i''\to 0$, a.s.. Summing up, we get

$$S_n^{-1}\sum_{i=1}^n x_iEe_i''\to 0. \tag{6.1}$$

Therefore to prove the necessity of the condition $V(n)=O(1)$ we have to show that, if $V(n)$ is not bounded, we can construct a sequence of i.i.d. random variables $\{e_i\}$ with $Ee_1=0$ such that (6.1) is not true. This can be done as in Section 2 of Chen (1995); the details are omitted.

Acknowledgement

The authors very much appreciate the work of an anonymous referee who suggested simplified proofs for some points in the original text. Especially, he pointed out that an additional condition in the original version of the paper was not needed, thus providing an essential improvement of the original result.

The authors recieved support from Guizhou province and the NNSF of China.

References

Chen, X. R. (1981). Again on the consistency of least squares estimates in linear models. *Acta Math. Sinica* **24**, 36-44 (in Chinese).

Chen, X. R. (1981). On a problem of weak consistency of linear estimates in linear models. *Chin. Annals Math.* **2**, 131-138 (in Chinese).

Chen, X. R. (1995). Consistency of LS estimates of multiple regression under a lower order moment conditon. *Science in China* (*Ser. A*) **38**, 1420-1431.

Chen, X. R., Zhu, L. X. and Fang, K. T. (1996). Almost sure convergence of weighted sums. *Statist. Sinica* **6**, 499-507.

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

Drygas, H. (1976). Weak and strong consistency of the least squares estimates in regression models. *Z. Wahrsch. verw. Gebiete* **34**, 119-127.

Lai, T. L., Robbins, H. and Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9**, 343-362.

Zhu, L. X. (1989). Doctorial dissertation. Published by the Institute of Systems Science. The Chinese Academy (in Chinese).

Department of Mathematics, Guizhou National College, Huaxi, Guiyang 550025.

E-mail: wlcjw@xhinfo.com

Graduate School, Chinese Academy, Beijing

(Received December 1995; accepted April 1998)