Chin, Ann. of Math.
13B: 2 (1992), 205—214.

THE ESTIMATE OF THE RANK FOR REGRESSION COEFFICIENT MATRIX IN A MEDIAN REGRESSION MODEL**

MIAO BAIQI (缪柏其)*

Abstract

It is discussed to infer the rank of regression coefficient matrix in a multivariate linear regression model. If the zero median vector is unique and the design matrices satisfy some weaker conditions, it is derived that the estimators of the rank of regression coefficient matrix under the minimum distance criterion by using model selection method is strongly consistent.

§1. Introduction

The technique of multivariate regression analysis are much useful and general in the multivariate data analysis. In this area one important object is to determine the rank of the regression matrix (RM).

The assumption that a series of errors of a model is independent normal is impressed in order to test the ranks of RM. When the error covariance matrix is known, Tintner (1945) derived the likelihood ratio test (LRT) statistic for the rank of RM. Anderson (1951) derived the expression for the LRT statistic to test the hypothesis H_r that the rank of RM is r. Fujikoshi (1974, 1977, 1978), Krishnaiah, Lin and Wang (1985) also investigated the expression of the asymptotic distribution when the underlying distribution is elliptially symmetric and similar test problem. For other results one is referred to [10, 11, 14, 7, 13].

Bai, Zhao and Krishnaiah (1985) further investigated this problem using model selection method. Only assuming that errors are iid. and satisfy some moment conditions, they derived that the estimators of the rank of coefficient matrix are strongly consistent.

In this paper we only assume that error vectors are iid., then the strongly consistent estimators can be derived by using model selection method under the minimum distance criterion.

Manuscript received April 6, 1990. Revised April 8, 1991.

* Department of Mathematics, University of Science and Technology of China, Hefei, Anhui 230026, China.

** Projects supported by the National Science Foundation of China.

§ 2. Statement of the Main Result

Consider the following model

$$Y_n = BX_n + \mathscr{E}_n,\ n=1, 2, \cdots, \tag{2.1}$$

where $Y_n=(y_1, y_2, \cdots, y_n)$, $n\geq 1$, are $p\times n$ observation matrices, X_n, $n\geq 1$, are $q\times n$ known design matrices, B is a $p\times q$ unknown regression matrix and $\mathscr{E}_n=(e_1, \cdots, e_n)$, $n\geq 1$, are $p\times n$ error matrices, where $e_1, e_2, \cdots, e_n$ are iid. $p\times 1$ random vectors. We want to estimate the rank of the coefficient matrix B by observation matrices and design matrices X_n. Notice that (2.1) can be rewritten as the following form

$$y_i = Bx_i + e_i,\ \text{for } i=1, 2, \cdots, n,\quad n=1, 2, \cdots. \tag{2.2}$$

Set

$$\hat{B}_{LS}^n \in \left\{B: \sum_{i=1}^n \|y_i - Bx_i\|^2 = \min\right\},$$

where $\|\cdot\|$ denotes the spectral norm of a matrix or vector.

$\hat{B}_{LS}^n$ is a LS estimator of B. In general, constructing a statistic by $\hat{B}_{LS}^n$, one can estimate and test the rank of B. Ingeniously using model selection technique and impressing various conditions on error vectors, Bai, Zhao and Krishnaiah investigated this problem and established that some statistics derived by $\hat{B}_{LS}^n$ are strongly consistent estimators of the rank of B. On the other hand, if we set

$$\hat{B}_n \in \left\{B: \sum_{i=1}^n \|y_i - Bx_i\| = \min\right\}, \tag{2.3}$$

that is, $\hat{B}_n$ is the estimators of B under the minimum distance criterion. We can also construct a statistic by $\hat{B}_n$ to estimate the rank of B. It is beyond doubt that this is an important technique, but it was not noticed for a long time in history. One reason is there was some difficulty in computation of $\hat{B}_n$ which has now been overcome with the advance of modern computing facilities, for detail one refers to Charnes et al. (1955). Another reason is the lack of an adequate asymptotic theory for $\hat{B}_n$, which has been overcome by Bai, Chen, Wu and Zhao (1990) in the case $p=1$ and by Bai, Miao and Rao (1988) in the case $p>1$. It offers a strong tool for us to estimate the rank of B by using adequate statistic constructed by $\hat{B}_n$. Now we establish the main theorem.

Assume that the common distribution F of iid. error vectors satisfies the following conditions:

(i) There is a $\delta>0$ such that F has a bounded density $f(u)$ for $\|u\|\leq\delta$. And for each non-zero vector c, $c\in R^p$, we have

$$P(c'e_1=0)<1.$$

(ii)

$$\int \frac{u}{\|u\|} dF(u)=0.$$

Here and hereafter the transpose of a matrix M is denoted by M'.

For design matrices X_n, $n\geqslant 1$, we impose these conditions:

(iii) There is a positive inter n_0 such that the rank of X_{n_0} is q.

(iv) Set $S_n=X_nX_n'$, then

$$d_n=\max_{1\leqslant i\leqslant n}\|S_n^{-1/2}x_i\|=o(\log^{-1/2}n). \tag{2.4}$$

When F is known, the rank of B may be any integer between 0 and min $(p, q)=s$. Set

$$\Theta_r=\{B: \text{rank}(B)=r\}, \tag{2.5}$$
$$r\in\{0, 1, \cdots, s\}.$$

Next, we select the true one using model selection technique from these $(s+1)$ models.

Assume that $\hat{B}_n$ is determined by (2.3), and $\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_p\geqslant 0$ are eigenvalues of $\hat{B}_n\hat{B}_n'$. Let H_n be a sequence of numbers increasing to infinite such that

$$\lim_{n\to\infty} d_nH_n^3\log^{1/2}n=0. \tag{2.6}$$

Write

$$I(r, H_n)=-\sum_{j=r+1}^{p}\mu_j-rH_n^{-1}. \tag{2.7}$$

Take $\hat{r}$ as the estimator of the rank r of the coefficient matrix B. Then it follows:

Theorem. *Under the model* (2.1), *if $\mathscr{E}_n$ and X_n satisfy conditions* (i)—(iv), *then $\hat{r}$ defined by* (2.7) *is the strongly consistent estimator of r_0 when Θ_{r_0} is true.*

Before proving the theorem, we make some remarks.

(1) Conditions (i) and (ii) ensure that the median vector of the distribution F is unique. Therefore zero vector is the unique median vector of F under our conditions.

(2) Set

$$A=\int\frac{1}{\|u\|}\left(I_p-\frac{uu'}{\|u\|^2}\right)dF(u),$$
$$D=\int\frac{uu'}{\|u\|^2}\,dF(u).$$

It iseasy to verify that the matrices A and D are positive definite. And the condition that F is known can be substitutd by a weaker condition that positive definite matrices A and D are known which can be seen from the main theorem.

(3) The condition (iii) imposed on the design matrix X_n is general, the condition (iv) ensures that the smallest positive eigenvalues of $S_n=X_nX_n'$ tend to infinite with the rate higher than $\log n$. To see this let $\lambda_{n1}\geqslant\cdots\geqslant\lambda_{nq}>0$ be q positive eigenvalues of S_n, where $n\geqslant n_0$. By one result of Neuman (1937), we have

$$\text{tr}(S_{n_0}S_n^{-1})\geqslant\sum_{i=1}^{q}\lambda_{n_0i}/\lambda_{ni} \quad \text{for } n\geqslant n_0. \tag{2.8}$$

Then it follows from the condition (iv):

$$\operatorname{tr}(S_{n_0}S_n^{-1})=\sum_{j=1}^{n_0}\operatorname{tr}(x_jx_j'S_n^{-1})=\sum_{j=1}^{n_0}\operatorname{tr}(x_j'S_n^{-1}x_j)=\sum_{j=1}^{n_0}x_j'S_n^{-1}x_j$$
$$\leqslant n_0\max_{1\leqslant i\leqslant n_0}x_i'S_n^{-1}x_i=O(d_n^2). \tag{2.9}$$

But $\lambda_{n_0q}>0$, so we have $\lim_{n\to\infty}\lambda_{nq}/\log n=\infty$ by the inequatities (2.8) and (2.9). By the way we obtain

$$\lambda_{nq}^{-1/2}=o(d_n). \tag{2.10}$$

It is quite weaker that the smallest positive eigenvalues of S_n tend to infinite with rate higher than $\log n$. This condition is equivalent to the condition derived by Bai, Zhao and Krishnaiah (1985) when error vectors are normal.

(4) If the decreasing rate of d_n is higher than $\log^{-1/2}n$, say $d_n=o(\log^{-3/2}n)$, then H_n can be chosen as $\log n$. This can be seen from the following proof.

§ 3. Some Lemmas

For the sake of convenience, we need some lemmas.

Lemma 1. (Bennett) *Let $\xi_1, \cdots, \xi_n$ be independent variables with $E(\xi_j)=0$ and $|\xi_i|\leqslant b$ for $1\leqslant i\leqslant n$, where b is a positive constant. Set $B_n^2=\sum_{i=1}^n E\xi_i^2$. Then for any $\epsilon>0$, we have*

$$P\left(\left\|\sum_{i=1}^n\xi_i^2\right\|\geqslant\epsilon\right)\leqslant 2\exp\left\{-\frac{\epsilon^2}{2(B_n^2+b\epsilon)}\right\}.$$

For proving our theorem, we continue to rewrite the model as (2.2). In the following the Kroneker product is denoted by $\otimes$. Write the transpose B' of the coefficient matrix B defined by (2.1) as $(b_1, \cdots, b_p)$, where b_j, $1\leqslant j\leqslant p$, are $q\times 1$ column vectors. The $pq\times 1$ straightening vector $(b_1', \cdots, b_p')'$ is denoted by β_0. Also, the $pq\times 1$ straightening vector of $\hat{B}_n'$ is denoted by $\tilde{\beta}_n$.

Lemma 2. *Let α and β (or with subscripts) be real numbers and A be a matrix. Then we have*

1. $(A_1\otimes B_1)(A_2\otimes B_2)=A_1A_2\otimes B_1B_2$,
2. $(A\otimes B)'=A'\otimes B'$,
3. $(\alpha_1A_1+\alpha_2A_2)\otimes(\beta_1B_1+\beta_2B_2)=\sum_{i,j=1}^{2}\alpha_i\beta_j(A_i\otimes B_j)$,
4. $\overrightarrow{ABC}=(A\otimes C')B$,
5. $\operatorname{tr}(A\otimes B)=(\operatorname{tr}A)(\operatorname{tr}B)$,
6. $\|A\otimes B\|\leqslant\|A\|\cdot\|B\|$,

solong as the previous operations are reasonable, where $\|\cdot\|$ denotes the spectral norm. The proof is referred to Rao (1973, pp. 29—30).

Throughout this paper we always assume that $n\geqslant n_0$. Set

$$S_n=X_nX_n'=\sum_{i=1}^n x_ix_i',$$

$$U_n=\sum_{i=1}^{n}(I_p\otimes x_i)D(I_p\otimes x_i')=D\otimes S_n,$$

$$T_n=\sum_{i=1}^{n}(I_p\otimes x_i)A(I_p\otimes x_i')=A\otimes S_n,$$

$$T_n^*=U_n^{-1/2}T_nU_n^{-1/2}=D^{-1/2}AD^{-1/2}\otimes I_q, \tag{3.1}$$

$$x_{ni}=U_n^{-1/2}(I_p\otimes x_i)=D^{1/2}\otimes S_n^{-1/2}x_i, \tag{3.2}$$

$$\beta_{n0}=U_n^{1/2}\beta_0,$$

$$\hat{\beta}_n=U_n^{1/2}\tilde{\beta}_n=U_n^{1/2}\hat{B}_n',$$

$$\beta_n^*=U_n^{1/2}(\tilde{\beta}_n-\beta_0). \tag{3.3}$$

Now the relations (2.2) and (2.3) can be rewritten as

$$y_i=x_{ni}'\beta_{n0}+e_i,\ 1\leqslant i\leqslant n, \tag{3.4}$$

$$\hat{\beta}_n=\left\{\beta:\beta\in R^{pq},\ \sum_{i=1}^{n}\|y_i-x_{ni}'\beta\|=\min\right\}. \tag{3.5}$$

According to the definition of S_n and (3.2) we have $\sum_{i=1}^{n}x_{ni}Dx_{ni}'=I_{pq}$. From above definition and Lemma 2 it follows that

$$\sum_{i=1}^{n}x_{ni}'x_{ni}=\sum_{i=1}^{n}(D^{-1/2}\otimes S_n^{-1/2}x_i)(D^{-1/2}\otimes S_n^{-1/2}x_i)'=D^{-1}\otimes\sum_{i=1}^{n}(x_i'S_n^{-1}x_i)$$
$$=D^{-1}\otimes\operatorname{tr}\left(\sum_{i=1}^{n}S_n^{-1/2}x_ix_i'S_n^{-1/2}\right)=q\cdot D^{-1}. \tag{3.6}$$

Here and hereafter $\lambda_1(M)\geqslant\lambda_2(M)\geqslant\cdots\geqslant\lambda_p(M)$ denote the positive eigenvalues of a symmetric matrix M.

Lemma 3. *Let E be an open convex subset of R^p and let $f_1, f_2, \cdots$, be a sequence of random convex functions on E such that for any $x\in E$, $f_n(x)\to f(x)$, a.s. (or in pr.) as $n\to\infty$, where f is some real function on E. Then f is also convex and for all compact $D\subset E$,*

$$\lim_{n\to\infty}\sup_{x\in D}|f_n(x)-f(x)|=0,\ \text{a. s. (or in pr.)}.$$

Furthermore, if f has a unique minimum at $\hat{x}\in E$ and $\hat{x}_n$ minimize f_n, we have $\lim_{n\to\infty}\hat{x}_n=\hat{x}$, a.s. (*or in pr.*).

Refer to Anderson and Gill (1982, Theorem II. 1, Corollary II. 2).

Lemma 4. *Under conditions of the theorem, we have*

$$\lim_{n\to\infty}H_n^{-2}\log^{-1}n\sup_{\|\beta\|\leqslant v_n}\sum_{i=1}^{n}\{(\|e_i-x_{ni}'\beta\|-\|e_i\|)-E(\|e_i-x_{ni}'\beta\|-\|e_i\|)\}=0,\ \text{a.s.}$$

If β_n^* minimize $\sum_{i=1}^{n}(\|e_i-x_{ni}'\beta\|-\|e_i\|)$, then

$$\|\beta_n^*\|\leqslant H_n\log^{1/2}n,\ \text{a. s}$$

where

$$v_n=H_n\log^{1/2}n. \tag{3.7}$$

Proof. It is easy to verify that for any $e\neq 0$ and $e\neq a$ where e and a belong to R^p, by Taylor expansion one has

$$\left| \|e-a\|-\|e\|+\frac{a'e}{\|e\|}\right| \leqslant 2\cdot \|a\|^2/\|e\|, \tag{3.8}$$

$$\left| \|e-a\|-\|a\|+\frac{a'e}{2\|e\|}-\frac{a'}{2\|a\|}\left(I_p-\frac{e'e}{\|e\|^2}\right)a\right| \leqslant (p+1)\cdot \|a\|^2/2\|e\|, \tag{3.9}$$

$$\left\| \frac{e-a}{\|e-a\|}\right\| \leqslant 2\|a\|/\|e\|. \tag{3.10}$$

By the definition of x_{ni} and the condition (iv), it follows that for all i, $i\leqslant n$, $\|x'_{ni}\beta\| \leqslant \|x_{ni}\|\cdot\|\beta\| \leqslant \lambda_p^{-1/2}(D)d_n v_n \leqslant H_n^{-2} \to 0$ uniformly when $\|\beta\|\leqslant v_n$. And the condition (i) implies

$$\int \frac{1}{\|u\|}\, dF(u) \leqslant C < \infty. \tag{3.11}$$

So we have

$$E\sum_{i=1}^{n}(\|e_i-x'_{ni}\beta\|-\|e_i\|) = \frac{1}{2}\sum_{i=1}^{n}\beta' x_{ni}(A+o(1))x_{ni}\beta$$
$$=\frac{1}{2}\beta' T_n^*\beta + o(\|\beta\|^2), \tag{3.12}$$

by (3.9) and the condition (ii). Since

$$|\,\|e_i-x'_{ni}\beta\|-\|e_i\|\,| \leqslant \|x'_{ni}\beta\| \leqslant \lambda_p^{-1/2}(D)d_n v_n, \tag{3.13}$$

$$\sum_{i=1}^{n} E(\|e_i-x'_{ni}\beta\|-\|e_i\|)^2 \leqslant \sum_{i=1}^{n}\|x'_{ni}\beta\|^2 = \beta'(D^{-1}\otimes I_q)\beta \leqslant pq\lambda_p^{-1}(D)v_n^2, \tag{3.14}$$

where the smallest positive eigenvalue $\lambda_p(D)$ of D is only dependent on F, it follows from Lemma 1 that

$$P\left(\left|\sum_{i=1}^{n}\{(\|e_i-x'_{ni}\beta\|-\|e_i\|)-E(\|e_i-x'_{ni}\beta\|-\|e_i\|)\}\right| \geqslant \epsilon v_n^2\right)$$
$$\leqslant 2\exp\{-\epsilon^2 v_n^4/2[pq\lambda_p^{-1}(D)v_i^2+(2\lambda_p^{-1/2}(D)d_n v_n)(\epsilon v_n^2)]\}.$$
$$\leqslant 2\exp\{-\epsilon^2\lambda_p(D)v_n^2/3pq\} \leqslant 2\exp\{-2\log n\}.$$

So by Borel-Cantelli lemma and (3.12) we have for $\|\beta\|\leqslant v_n$,

$$\lim_{n\to\infty} v_n^{-2}\left(\sum_{i=1}^{n}\|e_i-x'_{ni}\beta\|-\|e_i\| - \frac{1}{2}\beta' T_n^*\beta\right)=0, \quad \text{a.s.} \tag{3.15}$$

Set $\beta=v_n\tilde\beta$, where $\|\tilde\beta\|\leqslant 1$, and

$$f_n(\tilde\beta)=v_n^{-2}\sum_{i=1}^{n}(\|e_i-x'_{ni}\beta\|-\|e_i\|).$$

Then by (3.15) we have for $\|\tilde\beta\|\leqslant 1$,

$$f_n(\tilde\beta)\to\frac{1}{2}\tilde\beta' T_n^*\tilde\beta \doteq f(\tilde\beta), \text{ a. s.}$$

Because $f_n(\tilde\beta)$ is convex, it follows by Lemma 3 that

$$\lim_{n\to\infty}\sup_{\|\beta\|\leqslant 1}\|f_n(\tilde\beta)-f(\tilde\beta)|=0, \quad \text{a. s.}$$

Especially take $\|\tilde\beta\|=1$ and notice that $\beta=v_n\tilde\beta$, we have for n large enough

$$v_n^{-2}\left(\sum_{i=1}^{n}\|e_i\| - \inf_{\|\beta\|=v_n}\sum_{i=1}^{n}\|e_i-x'_{ni}\beta\|\right) \leqslant -(1/4)\inf_{\|\beta\|=v_n} v_n^{-2}\beta' T_n^*\beta$$
$$\leqslant -(1/4)\lambda_p(AD^{-1})<0. \quad \text{a. s.}$$

The convexity of $f_n(\tilde\beta)$ implies that

$$\|\beta_n^*\| \leqslant v_n, \text{ a. s.}$$

where β_n^* minimize $f_n(v_n^{-1}\tilde{\beta})$.

For the sake of convenience of notations, set

$$\phi_{ni}(\beta) = (e_i - x'_{ni}\beta)/\|e_i - x'_{ni}\beta\|.$$

We have the following result.

Lemma 5. *Under the conditions of the theorem, it holds that*

$$\lim_{n\to\infty} H_n^{-2}\log^{-1/2} n \sup_{\|\beta\|\leqslant v_n} \left\| \sum_{i=1}^n x_{ni}(\phi_{ni}(\beta) - \phi_{ni}(0)) \right\| = 0, \quad \text{a. s.}$$

Proof Let C be defined by (3.11), and let

$$\xi_{ni} = \|e_i\|^{-1} I(\|e_i\| > 2\|x_{ni}\|v_n), \ 1 \leqslant i \leqslant n.$$

Notice that $\|e_i\| > 2\|x_{ni}\|v_n$ implies that $e_i - x'_{ni}\beta \neq 0$, we have by (3.10)

$$\|\phi_{ni}(\beta) - \phi_{ni}(0)\| \leqslant 2\|x'_{ni}\beta\|/\|e_i\| \leqslant 2\|x_{ni}\|v_n/\|e_i\|. \tag{3.16}$$

Hence when $\|\beta\| \leqslant v_n$, we have from (3.10)

$$\left\| \sum_{i=1}^n x_{ni}(\phi_{ni}(\beta) - \phi_{ni}(0)) I(\|e_i\| > 2\|x_{ni}\|v_n) \right\|$$
$$\leqslant 2v_n \sum_{i=1}^n \|x_{ni}\|^2(\xi_{ni} - E\xi_{ni}) + 2v_n \sum_{i=1}^n \|x_{ni}\|^2 E\xi_{ni}.$$

Since the right of above inequality is independent of β, we have uniformly by (3.6)

$$\sup_{\|\beta\|\leqslant v_n} H_n^{-2}\log^{-1/2} n \left\| \sum_{i=1}^n x_{ni}(\phi_{ni}(\beta) - \phi_{ni}(0)) I(\|e_i\| > 2\|x_{ni}\|v_n) \right\|$$
$$\leqslant 2H_n^{-1} \sum_{i=1}^n \|x_{ni}\|^2(\xi_{ni} - E\xi_{ni}) + 2H_n^{-1}\cdot q\cdot\lambda_p^{-1}(D)\cdot C. \tag{3.17}$$

But $\xi_{ni} \leqslant 1/(2\|x_{ni}\|v_n)$, it follows that

$$\max_{i\leqslant n} \|x_{ni}\|^2/(2\|x_{ni}\|v_n) \leqslant \lambda_p^{-1/2}(D)d_n/(2v_n) \doteq \epsilon_1 d_n/v_n,$$

$$\sum_{i=1}^n \|x_{ni}\|^4 Var(\xi_{ni}) \leqslant \sum_{i=1}^n \|x_{ni}\|^4 E\xi_{ni}^2 \leqslant \sum_{i=1}^n \|x_{ni}\|^4 \cdot C/(2\|x_{ni}\|v_n)$$
$$\leqslant q\lambda_p^{-3/2}(D)\cdot(C/2)\cdot(d_n/v_n) \doteq \epsilon_2 d_n/v_n.$$

Take $\beta = \epsilon_1 d_n/v_n$ and $B_n^2 = \epsilon_2 d_n/v_n$ in Lemma 1 (notice that the right of the inequality in Lemma 1 is increasing in B_n), from the condition (iv) we have

$$P\left(\left| \sum_{i=1}^n \|x_{ni}\|^2(\xi_{ni} - E\xi_{ni}) \right| \geqslant \epsilon H_n \right) \leqslant 2\exp\{-(\epsilon/5\epsilon_1)(H_n v_n/d_n)\} \leqslant 2\exp\{-2\log n\}$$

for n large enough. The Borel-Cantelli lemma implies that

$$H_n^{-1} \sum_{i=1}^n \|x_{ni}\|^2(\xi_{ni} - E\xi_{ni}) \to 0, \text{ a. s.} \tag{3.18}$$

when n tends to infinite. By (3.17) and (3.18) and noticing $H_n \nearrow \infty$, we have

$$\lim_{n\to\infty} \sup_{\|\beta\|\leqslant v_n} (H_n v_n)^{-1} \left\| \sum_{i=1}^n x_{ni}(\phi_{ni}(\beta) - \phi_{ni}(0)) I(\|e_i\| > 2\|x_{ni}\|v_n) \right\| = 0, \quad \text{a. s.} \tag{3.19}$$

On the other hand the fact $\|e_i\| \leqslant 2\|x_{ni}\|v_n$ implies

$$\|\phi_{ni}(\beta) - \phi_{ni}(0)\| \leqslant 2. \tag{3.20}$$

Set $\eta'_{ni} = I(\|e_i\| \leqslant 2\|x_{ni}\|v_n)$. Then if $\|\beta\| \leqslant v_n$, we have from (3.20)

$$\begin{aligned}&(H_nv_n)^{-1}\Big\|\sum_{i=1}^n x_{ni}(\phi_{ni}(\beta)-\phi_{ni}(0))I(\|e_i\|\leqslant 2\|x_{ni}\|v_n)\Big\|\\&\leqslant (2/H_nv_n)\sum_{i=1}^n\|x_{ni}\|\eta_{ni}\\&=(2/H_nv_n)\sum_{i=1}^n\|x_{ni}\|(\eta_{ni}-E\eta_{ni})+(2/H_nv_n)\sum_{i=1}^n\|x_{ni}\|E\eta_{ni}\\&\doteq I_1+I_2.\end{aligned}\tag{3.21}$$

Since $\|x_{ni}\|v_n\leqslant\lambda_p^{-1/2}(D)d_nv_n\to 0$, it follows that $\|x_{ni}\|v_n\leqslant\delta$ uniformly for $i\leqslant n$ and for n large enough. By the condition (i),

$$\begin{aligned}E\eta_{ni}&=P(\|e_i\|\leqslant 2\|x_{ni}\|v_n)=\int_{\|u\|<2\|x_{ni}\|v_n}f_n(u)du\leqslant M\int_{\|u\|<2\|x_{ni}\|v_n}du\\&=MC_p(2\|x_{ni}\|v_n)^p,\end{aligned}\tag{3.22}$$

where M is the upper boundary of $f(u)$ and the constant C_p is only dependent on dimensional p. The relation (3.6) and $p>1$ implies that

$$\begin{aligned}I_2&=(2/H_nv_n)\sum_{i=1}^n\|x_{ni}\|E\eta_{ni}\leqslant(2/H_nv_n)\sum_{i=1}^n MC_p\cdot 2^p\cdot\|x_{ni}\|^{p+1}v_n^p\\&\leqslant 2^{p+1}C_p\cdot M\cdot\lambda_p^{-(p-1)}(D)d_n^{p-1}v_n^{p-1}\to 0.\end{aligned}\tag{3.23}$$

Next we estimate I_1. From the fact that $\|x_{ni}\|\eta_{ni}\leqslant\lambda_p^{-1/2}(D)d_n$ and (3.22),

$$\begin{aligned}\sum_{i=1}^n\|x_{ni}\|^2\mathrm{Var}(\eta_{ni})&\leqslant\sum_{i=1}^n\|x_{ni}\|^2E\eta_{ni}^2=\sum_{i=1}^n\|x_{ni}\|^2E\eta_{ni}\leqslant\|x_{ni}\|^2MC_p\cdot 2^p\|x_{ni}\|^pv_n^p\\&\leqslant MC_p2^pq\lambda_p^{-(p-1)}(D)(d_nv_n)^p\doteq\epsilon_1(d_nv_n)^p.\end{aligned}$$

Take $b=2\lambda_p^{-1/2}(D)d_n$ and $B_n^2=\epsilon_1(d_nv_n)^p$ in Lemma 1 and notice that $p>1$, we have

$$\begin{aligned}&P\Big(\Big|\sum_{i=1}^n\|x_{ni}\|(\eta_{ni}-E\eta_{ni})\Big|\geqslant\epsilon H_nv_n/2\Big)\\&\leqslant 2\exp\{-\epsilon^2H_n^2v_n^2/4\cdot 2(\epsilon_1(d_nv_n)^p+2\lambda_p^{-1/2}(D)d_n(\epsilon H_nv_n/2))\}.\end{aligned}$$

Using Borel-Cantelli lemma we get

$$\lim_{n\to\infty}(2/H_nv_n)\sum_{i=1}^n\|x_{ni}\|(\eta_{ni}-E\eta_{ni})=0,\ \text{a. s.}\tag{3.24}$$

Since the right of (3.21) is independent in β when $\|\beta\|\leqslant v_n$, it holds by (3.23) and (3.24) that

$$\lim_{n\to\infty}(H_nv_n)^{-1}\sup_{\|\beta\|\leqslant v_n}\Big\|\sum_{i=1}^n x_{ni}(\phi_{ni}(\beta)-\phi_{ni}(0))I(\|e_i\|\leqslant 2\|x_{ni}\|v_n)\Big\|=0.\ \text{a. s.}\tag{3.25}$$

Combining (3.19) and (3.25), we have this lemma.

Lemma 6. (Courant-Fischer) *Let K be $p\times p$ symmetric matrix and $\mu_1,\cdots,\mu_p$ be the eigenvalues of K. Then*

$$\sum_{i=q+1}^p\mu_i=\min\{\mathrm{tr}(U'KU):U\in M(p,p-q),\ U'U=I_{p-q}\},$$

$$\mu_q=\max_{B\in M(p-q,p)}\min\{\alpha'K\alpha:\alpha'\alpha=1,\ B\alpha=0\},$$

where $M(p,q)$ is the set of all $p\times q$ real matrices.

This is referred to Rao (1973, pp. 62—63).

§ 4. The Proof of the Theorem

The model (3.4) implies that

$$\sum_{i=1}^{n}\|y_i-x'_{ni}\beta\|=\sum_{i=1}^{n}\|e_i-x'_{ni}(\beta-\beta_{n0})\|.$$

So by the definition (3.3) of β_n^* and Lemma 4, it follows that

$$\|\beta_n^*\|\leqslant H_n\log^{1/2}n=v_n, \text{ a. s.} \tag{4.1}$$

And by Lemma 5,

$$\lim_{n\to\infty}\sup_{\|\beta\|\leqslant v_n}\left(H_n^{-2}\log^{-1/2}n\left\|\sum_{i=1}^{n}x_{ni}E(\phi_{ni}(\beta)-\phi_{ni}(0))\right\|=0, \text{ a. s.}\right. \tag{4.2}$$

Therefore,

$$\lim_{n\to\infty}(H_n^{-2}\log^{1/2}n)\Big(\sum_{i=1}^{n}x_{ni}E(\phi_{ni}(\beta_n^*)-\phi_{ni}(0))\Big)I(\|\beta_n^*\|\leqslant H_n\log^{1/2}n)=0, \text{ a. s.} \tag{4.3}$$

By (3.9) the fact that $\|x'_{ni}\beta\|\leqslant\lambda_p^{-1/2}(D)d_nv_n\to 0$ when $\|\beta\|\leqslant v_n$ implies that

$$\sum_{i=1}^{n}x_{ni}E\phi_{ni}(\beta)=-T_n^*\beta+o(H_n\log^{1/2}n). \tag{4.4}$$

Combining (4.1) and (4.3) and noticing that $E\phi_{ni}(0)=Ee_i/\|e_i\|=0$, we get

$$T_n^*\beta_n^*=o(H_n^2\log^{1/2}n), \text{ a. s.}$$

That is, based on the definition of T_n^* and β_n^* and Lemma 2,

$$S_n^{1/2}(\hat{B}_n-B)'AD^{-1/2}=o(H_n^2\log^{1/2}n), \text{ a. s.} \tag{4.5}$$

Write

$$S_n^{1/2}(\hat{B}_n-B)'=R'_n, \tag{4.6}$$

we have from (4.5)

$$R_n=o(H_n^2\log^{1/2}n). \text{ a. s.} \tag{4.7}$$

For $p\times q$ coefficient matrix B there exist orthogonal matrices Q and K with degree p and q, respectively, such that

$$B=Q\Lambda K',$$

where

$$\Lambda_{p\times q}=\begin{pmatrix}\Lambda_{11} & 0\\ 0 & 0\end{pmatrix},\quad \Lambda_{11}=\mathrm{diag}(\sqrt{\lambda_1},\ \cdots,\ \sqrt{\lambda_r})$$

and $\lambda_1\geqslant\cdots\geqslant\lambda_r$ are non-zero eigenvalues of BB'. Write

$$\hat{\Gamma}_n=Q'\hat{B}_nK,\ G_n=Q'R_nS_n^{-1/2}K.$$

Then (4.6) can be rewritten as

$$\hat{\Gamma}_n=\Lambda+G_n, \tag{4.8}$$

and

$$\hat{\Gamma}_n\hat{\Gamma}'_n=Q'\hat{B}_n\hat{B}'_n=(\Lambda+G_n)(\Lambda'_n+G'_n). \tag{4.9}$$

By (2.10) and (4.7) it follows that

$$G_n=o(H_n^2\log^{1/2}n\cdot H_n^{-3}\log^{-1/2}n)=o(H_n^{-1}), \text{ a. s.} \tag{4.10}$$

When the model Θ_{r0} is true, we have

$$\sum_{j=r+1}^{p}\mu_j=\min\{\mathrm{tr}U'(\Lambda+G_n)(\Lambda'+G'_n)U,\ U\in M(p,p-q),\ U'U=I_{p-q}\}$$

$$\leqslant \mathrm{tr}\{(0, I_{p-q})(A'+G_n)(A'+G'_n)(0', I_{p-q})'\}$$
$$= \sum_{j=r+1}^{r_0} \lambda_j + \mathrm{tr}\{(0, I_{p-q})(AG'_n+G_nA'+G_nG'_n)(0', I_{p-q})'\}$$
$$= \sum_{j=r+1}^{r_0} \lambda_j + o(H_n^{-1}), \text{ a. s.} \tag{4.11}$$

and

$$\mu_r = \max_{L\in M(p-r,p)} \min\{\alpha'(A+G_n)(A'+G'_n)\alpha : \alpha'\alpha=1, L\alpha=0\}$$
$$\geqslant \min\{(\alpha'_1, 0')(A+G_n)(A'+G'_n)(\alpha'_1, 0')' : \alpha_1\in R^r, \alpha'_1\alpha_1=1\}$$
$$= \min\{\alpha'_1 A_{11}^2\alpha_1 + (\alpha'_1, 0')(AG'_n+G_nA'+G_nG'_n)(\alpha'_1, 0')' : \alpha_1\in R^r, \alpha'_1\alpha_1=1\}$$
$$\geqslant \min\{\lambda_r + o(H_n^{-1}), \alpha'_1\alpha_1=1\} \geqslant \lambda_r/2, \text{ a. s.} \tag{4.12}$$

by Lemma 6, (4.10) and the fact that the eigenvalues of $\hat{\Gamma}_n\hat{\Gamma}'_n$ are the same as ones of $\hat{B}_n\hat{B}'_n$. If $r>r_0$, by (4.11), it follows that

$$I(r_0, H_n) - I(r, H_n) = (r-r_0)H_n^{-1} - \sum_{j=r_0+1}^{r} \mu_j$$
$$= (r-r_0)H_n^{-1} + o(H_n^{-1}) \geqslant (r-r_0)/(2H_n) > 0. \tag{4.13}$$

On the other hand, if $r<r_0$, by (4.12), it follows

$$I(r_0, H_n) - I(r, H_n) = \sum_{j=r+1}^{r_0} \mu_j - (r_0-r)H_n^{-1} \geqslant \mu_{r0} - (r_0-r)/H_n^{-1}$$
$$\geqslant \lambda_{r0}/4 > 0, \text{ a. s.} \tag{4.14}$$

Combining (4.13) and (4.14), we get the theorem.

References

[1] Anderson, P. K. & Gill, R. D., Cox's regression model for counting processes: A large sample study, *Ann. Statist.*, **10** (1982), 1100—1120.

[2] Anderson, T. W., Estimating linear restrictions on regression coefficient for multivariate normal distributions, *Ann. Math. Statist.*, **8** (1980), 1400.

[3] Bai, Z. D., Chen, X. R., Miao, B. Q. & Rao, C. R., Asymptotic theory of least distances estimate in multivariate linear models, *Statistics*, **21**:4 (1990), 503—519.

[4] Bai, Z. D., Chen, X. R., Wu, Y. H. & Zhao, L. C., Asymptotic normality of minimum L_1-norm estimates in linear models, *Science in China Series A*, **33**:11 (1990).

[5] Bai Z. D., Krishnaiah, P. R. & Zhao, L. C., Inference on the ranks of regression matrix and canonical correlation matrix using model selected methods, Center for Multivariate Analysis, University of Pittsburgh, Technical Report, No. 86—52, 1986.

[6] Bennett, C., Probability inequalities for the sums of independent random variables, *J. Amer. Statist. Assoc.*, **57** (1962), 33—45.

[7] Fang, C. & Krishnaiah, P. R., Asymptotic distributions of functions of the eigenvalues of some rendom matrices for non-normal populations. *J. Multivariate Anal.*, **12** (1982), 39—63.

[8] Fujikoshi, Y., The likelihood ratio tests for the demensionality of regression coefficients, *J. Multivariate Anal.* 4 (1974), 327—340.

[9] Fujikoshi, Y., Asymptotic expansions for the distributions of some multivariate tests, In Multivariate Analysis IV, P. R. Krishnaiah, ed., 51—71. North-Holland, Amsterdam, 1977.

[10] Fujikoshi, Y., Asymptotic expansions for the distributions of some functions of the latent roots of matrices in three situation, *J. Multivariate Anal.*, **12** (1978), 124—138.

[11] Krishnaiah, P. K. & Lee, J. C., On the asymptotic joint distributions of certain functions of the eigenvalues of somerandom matrices, *J. Multivariate Anal.*, **9** (1979), 248—258.

[12] Krishnaiah, P. K., Lin, J. & Wang, L., Tests for the demensionality of the regression matrices when the underlying distributions are elliptielly symmetric, Center for Multivariate Analysis, University of Pittsburgh, Technical Report, No. 85—36, 1985.