Journal of Mathematical Research & Exposition
Sept., 2009, Vol. 29, No. 5, pp. 882–888
DOI:10.3770/j.issn:1000-341X.2009.05.014
Http://jmre.dlut.edu.cn

Generalized Ridge and Principal Correlation Estimator of the Regression Parameters and Its Optimality

GUO Wen Xing, ZHANG Shang Li, XUE Xiao Wei
(School of Science, Beijing Jiaotong University, Beijing 100044, China)
(E-mail: guowenxing998@163.com; shlzhang@bjtu.edu.cn)

Abstract In this paper, we propose a new biased estimator of the regression parameters, the generalized ridge and principal correlation estimator. We present its some properties and prove that it is superior to LSE (least squares estimator), principal correlation estimator, ridge and principal correlation estimator under MSE (mean squares error) and PMC (Pitman closeness) criterion, respectively.

Keywords linear regression model; generalized ridge and principal correlation estimator; mean squares error; Pitman closeness criterion.

Document code A
MR(2000) Subject Classification 62C15; 62F10
Chinese Library Classification O212.1

1. Introduction and preliminaries

In order to conquer the drawbacks of LSE, many biased estimators were proposed, among which the principal component estimator is popular but it still has some drawbacks. The principal correlation estimator was proposed and its superiorities to principal component estimator were introduced in [1]. The ridge and principal correlation estimator was proposed in [2]. In this paper, we propose the generalized ridge and principal correlation estimator and prove that generalized ridge and principal correlation estimator is respectively superior to LSE, principal correlation estimator and ridge and principal correlation estimator under some certain conditions and ordinary criterion.

Consider the linear regression model

$$Y = X\beta + e, \quad E(e) = 0, \quad \mathrm{Cov}(e) = \sigma^2 I, \tag{1}$$

where $Y_{n\times 1}$ is an observable random vector, $X_{n\times p}$ is a matrix with $\mathrm{rank}(X) = p$, $\beta_{p\times 1}$ is an unknown parameter vector and e denotes random error.

As we know, the LSE of β is $\widehat{\beta} = (X'X)^{-1}X'Y$. Let $X'X = P\Lambda P'$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$, $P_{p\times p}$ is an orthogonal matrix. Let $\Lambda_1 = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$. Consider the correlation coefficient order. The principal correlation was given in [1]. ρ_i is a correlation

Received date: 2007-05-28; **Accepted date**: 2009-03-10
Foundation item: the National Natural Science Foundation of China (Nos. 60736047; 10671007; 60772036); the Foundation of Beijing Jiaotong University (Nos. 2006XM037; 2007XM046).

coefficient about $\phi_i' x$ and y. We consider using the $\widehat{\rho}_i = \frac{\phi_i' X' Y}{\sqrt{\lambda_i}\sqrt{\widehat{\sigma}}}$, $i = 1, 2, \ldots, p$, to estimate ρ_i, and using $\widehat{\rho}_i$ to measure the effect degree of $\phi_i' x$ to ρ_i, where $\widehat{\sigma}$ is an estimate value of $\sigma = \sqrt{\mathrm{Var}(y)}$. It is easy to see that we only need to use the order $|\widehat{\rho}_i|$. So we only need to consider the order of $\frac{\phi_i' X' Y}{\sqrt{\lambda_i}}$. Given $|\widehat{\rho}_{i_1}| \geq |\widehat{\rho}_{i_2}| \geq \cdots \geq |\widehat{\rho}_{i_p}|$, let U_1 denote the matrix composed of $i_1, i_2, \ldots, i_r$ columns of I_p, and U_2 denote the matrix composed of the rest $(p-r)$ columns of I_p. Then $\widetilde{U} = (U_1 \cdots U_2)$ is an orthogonal matrix, and $\Lambda = (U_1 \ \ldots, U_2)\widetilde{\Lambda}(U_1, \ldots, U_2)'$, where $\widetilde{\Lambda} = \mathrm{diag}(\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_r}, \lambda_{i_{r+1}}, \ldots, \lambda_{i_p})$. Let $\widetilde{P} = P\widetilde{U} = P(U_1, \ldots, U_2)$. So $X'X = \widetilde{P}\widetilde{\Lambda}\widetilde{P}'$. Given $\widetilde{\Lambda} = \mathrm{diag}(\widetilde{\Lambda}_1, \widetilde{\Lambda}_2)$ where $\widetilde{\Lambda}_1 = \mathrm{diag}(\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_r})$, $\widetilde{P} = (\widetilde{P}_1 \ \ldots \ \widetilde{P}_2)$, $\widetilde{P}_1$ is a $n \times r$ column orthogonal matrix.

Zhang proposed the principal correlation estimator $\widetilde{\beta} = \widetilde{P}_1 \widetilde{\Lambda}_1^{-1} \widetilde{P}_1' X' Y$ in [1]. Some principal components that have minor effect on the dependent variable are deleted. Accordingly, its dimension degenerates to r. In [2], Yan defined the ridge and principal correlation estimator $\widetilde{\beta}(t) = \widetilde{P}_1(\widetilde{\Lambda}_1 + tI)^{-1}\widetilde{P}_1' X' Y$ that improved the principal correlation estimator in case that λ_i is close to zero. In this paper, we present the generalized ridge and principal correlation estimator, which is defined as follows:

Definition 1 *The generalized ridge and principal correlation estimator is* $\widetilde{\beta}(K) = \widetilde{P}_1(\widetilde{\Lambda}_1 + K)^{-1}\widetilde{P}_1' X' Y$, *where* $K = \mathrm{diag}(k_{i_1}, k_{i_2}, \ldots, k_{i_r}), k_{i_j} > 0, j = 1, 2, \ldots, r$.

Remark Some principal components which have minor effect on the dependent variable are deleted in generalized ridge and principal correlation estimator, moreover, the generalized ridge and principal correlation estimator improves the principal correlation estimator in case that λ_i is close to zero. And it becomes the ridge and principal correlation estimator if $k_{i_j} = t$, so the ridge and principal correlation estimator is the special case of the generalized ridge and principal correlation estimator.

In this paper we use $\widehat{\beta}$, $\widetilde{\beta}$, $\widetilde{\beta}(t)$, $\widetilde{\beta}(K)$ to denote LSE, principal correlation estimator, ridge and principal correlation estimator, generalized ridge and principal correlation estimator of β, respectively.

2. Properties of generalized ridge and principal correlation estimator

Property 1 $\widetilde{\beta}(K)$ *is a linear biased estimator of* β.

Proof Since $\widetilde{\beta}(K) = \widetilde{P}_1(\widetilde{\Lambda}_1 + K)^{-1}\widetilde{P}_1' X' Y = \widetilde{P}_1 M(K)\widetilde{P}_1'\widehat{\beta}$, we have $E(\widetilde{\beta}(K)) = \widetilde{P}_1 M(K)\widetilde{P}_1'\beta$, where $M(K) = \mathrm{diag}(\frac{\lambda_{i_1}}{\lambda_{i_1}+k_{i_1}}, \frac{\lambda_{i_2}}{\lambda_{i_2}+k_{i_2}}, \ldots, \frac{\lambda_{i_r}}{\lambda_{i_r}+k_{i_r}})$. Therefore, while $k_{i_j} > 0$, $\widetilde{\beta}(K)$ is a linear biased estimator of β.

Property 2 There exists $\|\widetilde{\beta}(K)\| < \|\widehat{\beta}\|$, *which shows that* $\widetilde{\beta}(K)$ *is a shrunken estimator of* $\widehat{\beta}$, *where* $\|\cdot\|$ *denotes Euclidean norm.*

Proof Since $\|\widetilde{\beta}(K)\| = \|\widetilde{P}_1 M(K)\widetilde{P}_1'\widehat{\beta}\| = \|M(K)\widetilde{P}_1'\widehat{\beta}\| < \|\widetilde{P}_1'\widehat{\beta}\| = \|\widehat{\beta}\|$, the property is true.

We will prove that $\widetilde{\beta}(K)$ has some dispersion optimum properties in the class of generalized ridge and shrunken dimension estimators. In [3] the class of shrunken dimension estimators was given. We define the class of generalized ridge and shrunken dimension estimators as follows:

$$\Omega = \{\widetilde{\beta}(U)|\widetilde{\beta}(U) = U(U'X'XU + Q'KQ)^{-1}U'X'Y\},$$

where U is a $p \times r$ matrix and $\text{rank}(U) = r$, where $U = \widetilde{P}_1\widetilde{\Delta}\widetilde{Q}$, Q is a $r \times r$ orthogonal matrix. $\widetilde{\Delta} = \text{diag}(\mu_{i_1}, \mu_{i_2}, \ldots, \mu_{i_r})$, obviously, $\widetilde{\beta}(K) \in \Omega$. For any $\widetilde{\beta}(U) \in \Omega$, we have

$$\begin{aligned}\text{Cov}(\widetilde{\beta}(U)) &= \widetilde{P}_1(\widetilde{\Lambda}_1 + K\widetilde{\Delta}^{-2})^{-1}\widetilde{\Lambda}_1(\widetilde{\Lambda}_1 + K\widetilde{\Delta}^{-2})^{-1}\widetilde{P}_1'\sigma^2 \\ &= \widetilde{P}_1\text{diag}\Big(\frac{\lambda_{i_1}}{(\lambda_{i_1} + \mu_{i_1}^{-2}k_{i_1})^2}, \frac{\lambda_{i_2}}{(\lambda_{i_2} + \mu_{i_2}^{-2}k_{i_2})^2}, \ldots, \frac{\lambda_{i_r}}{(\lambda_{i_r} + \mu_{i_r}^{-2}k_{i_r})^2}\Big)\widetilde{P}_1'\sigma^2 \qquad (2)\end{aligned}$$

and

$$\begin{aligned}\text{Cov}(\widetilde{\beta}(K)) &= \widetilde{P}_1(\widetilde{\Lambda}_1 + K)^{-1}\widetilde{\Lambda}_1(\widetilde{\Lambda}_1 + K)^{-1}\widetilde{P}_1'\sigma^2 \\ &= \widetilde{P}_1\text{diag}\Big(\frac{\lambda_{i_1}}{(\lambda_{i_1} + k_{i_1})^2}, \frac{\lambda_{i_2}}{(\lambda_{i_2} + k_{i_2})^2}, \ldots, \frac{\lambda_{i_r}}{(\lambda_{i_r} + k_{i_r})^2}\Big)\widetilde{P}_1'\sigma^2 \qquad (3)\end{aligned}$$

Property 3 *If $\mu_{i_j} \geq 1$ $(j = 1, 2, \ldots, r)$, for any shrunken dimension estimator $\widetilde{\beta}(U) \in A$, then $\text{Cov}(\widetilde{\beta}(K)) \leq \text{Cov}(\widetilde{\beta}(U))$, the equality holds if and only if $\mu_{i_j} = 1$.*

Proof If $\mu_{i_j} \geq 1$ $(j = 1, 2, \ldots, r)$, then $\frac{\lambda_{i_j}}{(\lambda_{i_j}+\mu_{i_j}^{-2}k_{i_j})^2} \geq \frac{\lambda_{i_j}}{(\lambda_{i_j}+k_{i_j})^2}$. When $\mu_{i_j} = 1$, the equality holds. From (2) and (3), we obtain the conclusion.

Property 4 *Generalized ridge and principal correlation estimator satisfies the following inequalities*

a) $\text{trCov}(\widetilde{\beta}(K)) \leq \text{trCov}(\widetilde{\beta}(U))$
b) $\lambda_{i_j}\text{Cov}(\widetilde{\beta}(K)) \leq \lambda_{i_j}\text{Cov}(\widetilde{\beta}(U))$, $j = 1, 2, \ldots, r$
c) $\|\text{Cov}(\widetilde{\beta}(K))\| \leq \|\text{Cov}(\widetilde{\beta}(U))\|$.

Property 5 *Generalized ridge and principal correlation estimator satisfies the following equality* $\min_U \max_C \frac{\text{Var}(C_1'\widetilde{\beta}(U))}{C_1'C} = \max_C \frac{\text{Var}(C_1'\widetilde{\beta}(K))}{C_1'C} = M\,\text{Cov}\widetilde{\beta}(K)$, *where* $M = \max(\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_r})$.

3. Optimalities of generalized ridge and principal correlation estimator

Let

$$M = \max(\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_r}), \quad m = \min(\lambda_{i_1}, \lambda_{i_2}, dots, \lambda_{i_r})$$
$$N = \max(k_{i_1}, k_{i_2}, \ldots, k_{i_r}), \quad n = \min(k_{i_1}, k_{i_2}, \ldots, k_{i_r}).$$

Theorem 1 *If $\beta'C\beta < \sigma^2$, then $\text{MSE}(\widetilde{\beta}(K)) < \text{MSE}(\widetilde{\beta})$, where*

$$C = \begin{pmatrix} C_1 & O \\ O & O \end{pmatrix}, \quad C_1 = \frac{M^2(M+N)^2N}{(m+n)^2(m^2+2mn)}I_r, \quad \beta = \begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} \end{pmatrix}, \quad \beta_{(1)} : r \times 1.$$

Proof Since

$$\text{MSE}(\widetilde{\beta}(K)) = \sum_{j=1}^{r} \frac{\sigma^2\lambda_{i_j}}{(\lambda_{i_j} + k_{i_j})^2} + \sum_{j=1}^{r} \frac{k_{i_j}^2\beta_j^2}{(\lambda_{i_j} + k_{i_j})^2} + \sum_{j=r+1}^{p} \beta_j^2$$
$$\text{MSE}(\widetilde{\beta}) = \sum_{j=1}^{r} \frac{\sigma^2}{\lambda_{i_j}} + \sum_{j=r+1}^{p} \beta_j^2.$$

Therefore

$$\begin{aligned}\mathrm{MSE}(\widetilde{\beta}(K)) < \mathrm{MSE}(\widetilde{\beta}) &\iff \sum_{j=1}^{r}\frac{\sigma^2\lambda_{i_j}}{(\lambda_{i_j}+k_{i_j})^2}+\sum_{j=1}^{r}\frac{k_{i_j}^2\beta_j^2}{(\lambda_{i_j}+k_{i_j})^2}<\sum_{j=1}^{r}\frac{\sigma^2}{\lambda_{i_j}}\\ &\iff \sum_{j=1}^{r}\frac{k_{i_j}^2\beta_j^2}{(\lambda_{i_j}+k_{i_j})^2}<\sum_{j=1}^{r}\frac{(2\lambda_{i_j}k_{i_j}+k_{i_j}^2)\sigma^2}{(\lambda_{i_j}+k_{i_j})^2\lambda_{i_j}}. \qquad (4)\end{aligned}$$

Note that if

$$\frac{N^2}{(m+n)^2}\sum_{j=1}^{r}\beta_j^2<\frac{n^2+2mn}{(M+N)^2M}\sigma^2,$$

then

$$\beta_{(1)}'\frac{N^2(M+N)^2M}{(m+n)^2(n^2+2mn)}I_r\beta_{(1)}<\sigma^2 \iff \beta_{(1)}'C_1\beta_{(1)}<\sigma^2$$

and (4) holds.

Let $\widetilde{\beta}(K)=A\widehat{\beta}$, $\widetilde{\beta}(t)=B\widehat{\beta}$, where

$$A=\begin{pmatrix}A_{(1)} & O\\ O & O\end{pmatrix},\quad A_{(1)}=(\widetilde{\Lambda}_1+K)^{-1}\widetilde{\Lambda}_1,\quad B=\begin{pmatrix}B_{(1)} & O\\ O & O\end{pmatrix},\quad B_{(1)}=(\widetilde{\Lambda}_1+tI)^{-1}\widetilde{\Lambda}_1.$$

Theorem 2 *If $\beta'C\beta<\sigma^2$, then $\mathrm{MSEM}(\widetilde{\beta}(K))<\mathrm{MSEM}(\widehat{\beta})$, where $C=(I-A)\widetilde{\Lambda}_1(I+A)^{-1}$.*

Proof Since $\mathrm{MSEM}(\widetilde{\beta}(K))=\sigma^2A^2\widetilde{\Lambda}_1^{-1}+(I-A)\beta\beta'(I-A)'$, $\mathrm{MSEM}(\widehat{\beta})=\sigma^2\widetilde{\Lambda}_1^{-1}$, we have

$$\begin{aligned}\mathrm{MSEM}(\widetilde{\beta}(K))<\mathrm{MSEM}(\widehat{\beta}) &\iff \sigma^2A^2\widetilde{\Lambda}_1^{-1}+(I-A)\beta\beta'(I-A)'<\sigma^2\widetilde{\Lambda}_1^{-1}\\ &\iff (I-A)\beta\beta'(I-A)'<\sigma^2(I-A^2)\widetilde{\Lambda}_1^{-1}. \qquad (5)\end{aligned}$$

Note that if $\beta'X'NX\beta<\sigma^2$, then $N^{1/2}X\beta\beta'X'N^{1/2}<\sigma^2I_n$. As $\beta'C\beta<\sigma^2$, (5) holds.

Lemma 1 *Let the function $f(k)=\frac{\sigma^2\lambda+a^2k^2}{(\lambda+k)^2}$. If $k=\frac{\sigma^2}{a^2}$, then $f(k)$ has minimum value, where $k>0,\lambda>0$.*

Proof Since $f'(k)=\frac{2\lambda(a^2k-\sigma^2)}{(\lambda+k)^3}$, we can easily obtain the conclusion.

Theorem 3 *If $K=\mathrm{diag}(k_{i_1},k_{i_2},\ldots,k_{i_r})$, then $\mathrm{MSE}(\widetilde{\beta}(K))\le\mathrm{MSE}(\widetilde{\beta}(t))$ holds.*

Proof When $K>0$, let $k_{i_1}=\frac{\sigma^2}{a^2}$, $k_{i_2}=\cdots=k_{i_r}=t>0$. Then

$$\mathrm{MSE}(\widetilde{\beta}(K))-\mathrm{MSE}(\widetilde{\beta}(t))=\sum_{j=1}^{r}f_j(k_{i_j})-\sum_{j=1}^{r}f_j(t)=f_1(k_{i_1})-f_1(t).$$

By Lemma 1, we have $f_1(k_{i_1})\le f_1(t)$. Thus $\mathrm{MSE}(\widetilde{\beta}(K))\le\mathrm{MSE}(\widetilde{\beta}(t))$.

Let $L(\widetilde{\theta},\theta)=(\widetilde{\theta},\theta)'D(\widetilde{\theta},\theta)\triangleq\|\widetilde{\theta}-\theta\|_D^2$, where $D>0$, and $R(\widetilde{\theta},\theta)=E\|\widetilde{\theta}-\theta\|_D^2$.

Definition 2 *A linear estimator $\widetilde{\theta}$ is said to be admissible for θ if there does not exist $\widetilde{\theta}^*$ such that the inequality $R(\widetilde{\theta}^*,\theta)\le R(\widetilde{\theta},\theta)$ holds for every pair (θ,σ^2) and is strict for at least one such pair.*

Note that $D=X'X$, $L(\widetilde{\theta},\theta)=(\widetilde{\theta},\theta)'X'X(\widetilde{\theta},\theta)$ is a loss function under the Fisher's loss measure of closeness. So the Fisher's loss is a special case of PMC loss.

Theorem 4 *The generalized ridge and principal correlation estimator $\widetilde{\beta}(K)$ is an admissible estimator of β.*

Proof $\widetilde{\beta}(K)$ is an admissible estimator of β

$$\begin{aligned}&\Longleftrightarrow A(X'X)^{-1}A' \leq A(X'X)^{-1}\\&\Longleftrightarrow \widetilde{P}_1(\widetilde{\Lambda}_1+K)^{-2}\widetilde{P}_1'X'X \leq \widetilde{P}_1(\widetilde{\Lambda}_1+K)^{-1}\widetilde{P}_1'\\&\Longleftrightarrow \widetilde{P}_1(\widetilde{\Lambda}_1+K)^{-2}\widetilde{\Lambda}_1\widetilde{P}_1' \leq \widetilde{P}_1(\widetilde{\Lambda}_1+K)^{-1}\widetilde{P}_1'\\&\Longleftrightarrow \widetilde{P}_1\mathrm{diag}\Big(\frac{\lambda_{i_1}}{(\lambda_{i_1}+k_{i_1})^2},\ldots,\frac{\lambda_{i_r}}{(\lambda_{i_r}+k_{i_r})^2}\Big)\widetilde{P}_1' \leq \widetilde{P}_1\mathrm{diag}\Big(\frac{1}{(\lambda_{i_1}+k_{i_1})},\ldots,\frac{1}{(\lambda_{i_r}+k_{i_r})}\Big)\widetilde{P}_1'.\end{aligned}$$

Since $\lambda_{i_j}+k_{i_j}\geq\lambda_{i_j}$, $\frac{\lambda_{i_j}}{(\lambda_{i_j}+k_{i_j})^2}\leq\frac{1}{\lambda_{i_j}+k_{i_j}}$ $(j=1,2,\ldots,r)$, we obtain the conclusion.

Pitman(1937) gave a closeness criterion which can discriminate the optimality of two estimators in [5]. Let $\widetilde{\theta}_1,\widetilde{\theta}_2$ be two different estimators of parameter θ, and $L(\widetilde{\theta},\theta)$ be a loss function. PMC criterion is defined as follows:

Definition 3 *If $\widetilde{\theta}_1$, $\widetilde{\theta}_2$ satisfy*

$$P(L(\widetilde{\theta}_1,\theta)\leq L(\widetilde{\theta}_2,\theta))\geq 0.5 \tag{6}$$

and there is at least one θ which makes the strict inequality in the right of (6) hold, then $\widetilde{\theta}_1$ is better than $\widetilde{\theta}_2$ under PMC.

Let $\xi=\frac{\widehat{\beta}}{\sigma}=\binom{\xi_{(1)}}{\xi_{(2)}}{}^{r}_{p-r}$, then $\xi\sim N_p(\frac{\beta}{\sigma},\widetilde{\Lambda}^{-1})$, $\xi_{(1)}\sim N_r(\frac{\beta_{(1)}}{\sigma},\widetilde{\Lambda}_1^{-1})$, $\xi_{(2)}\sim N_{m-r}(\frac{\beta_{(2)}}{\sigma},\widetilde{\Lambda}_2^{-1})$. According to [6], let

$$\begin{aligned}W(k,t)&=\|\widetilde{\beta}(K)-\beta\|^2-\|\widetilde{\beta}(t)-\beta\|^2=\sigma^2\Big[\|A\xi-\frac{\beta}{\sigma}\|^2-\|B\xi-\frac{\beta}{\sigma}\|^2\Big],\\W(k)&=\|\widetilde{\beta}(K)-\beta\|^2-\|\widehat{\beta}-\beta\|^2=\sigma^2\Big[\|A\xi-\frac{\beta}{\sigma}\|^2-\|\xi-\frac{\beta}{\sigma}\|^2\Big].\end{aligned}$$

Theorem 5 *If $\|\frac{\beta_{(1)}}{\sigma}\|\leq\frac{\delta^{(1)}_{\sqrt{0.5}}n(2m+n)(m+n)}{2N(M+N)^2}$, $\|\frac{\beta_{(2)}}{\sigma}\|\leq\frac{\delta^{(2)}_{\sqrt{0.5}}}{2}$, then $\widetilde{\beta}(K)$ is better than $\widehat{\beta}$ under PMC, where $\delta^{(i)}_{\sqrt{0.5}}$ satisfies $P(\|\xi_{(i)}\|\leq\delta^{(i)}_{\sqrt{0.5}})=\sqrt{0.5}$ $(i=1,2)$.*

Proof Since

$$\begin{aligned}W(k)\leq 0 &\Longleftrightarrow \|A\xi-\frac{\beta}{\sigma}\|^2\leq\|\xi-\frac{\beta}{\sigma}\|^2\\&\Longleftrightarrow \|A_{(1)}\xi_{(1)}-\frac{\beta_{(1)}}{\sigma}\|^2+\|\vec{0}-\frac{\beta_{(2)}}{\sigma}\|^2\leq\|\xi_{(1)}-\frac{\beta_{(1)}}{\sigma}\|^2+\|\xi_{(2)}-\frac{\beta_{(2)}}{\sigma}\|^2 \qquad (7)\end{aligned}$$

Note that if

$$\|A_{(1)}\xi_{(1)}-\frac{\beta_{(1)}}{\sigma}\|^2\leq\|\xi_{(1)}-\frac{\beta_{(1)}}{\sigma}\|^2 \tag{8}$$

and

$$\|\vec{0}-\frac{\beta_{(2)}}{\sigma}\|^2\leq\|\xi_{(2)}-\frac{\beta_{(2)}}{\sigma}\|^2 \tag{9}$$

synchronously hold, then (7) holds.

(8) holds

$$\begin{aligned}&\Longleftrightarrow \sum_{j=1}^{r}(\frac{\lambda_{i_j}\xi_j}{\lambda_{i_j}+k_{i_j}}-\frac{\beta_j}{\sigma})^2\le\sum_{j=1}^{r}(\xi_j-\frac{\beta_j}{\sigma})^2\\&\Longleftrightarrow 2\sum_{j=1}^{r}\frac{k_{i_j}\xi_j\beta_j}{(\lambda_{i_j}+k_{i_j})\sigma}\le\sum_{j=1}^{r}\frac{k_{i_j}(2\lambda_{i_j}+k_{i_j})\xi_j^2}{(\lambda_{i_j}+k_{i_j})^2}. \qquad (10)\end{aligned}$$

Note that if

$$\frac{2N}{m+n}\sum_{j=1}^{r}\frac{\xi_j\beta_j}{\sigma}\le\frac{n(2m+n)}{(M+N)^2}\sum_{j=1}^{r}\xi_j^2, \qquad (11)$$

then (10) holds. According to Cauchy-Schwarz inequality, when $\|\xi_{(1)}\|\ge\frac{2N(M+N)^2}{n(2m+n)(m+n)}\|\frac{\beta_{(1)}}{\sigma}\|$, (8) holds. Thus $P((8)\text{ holds})\ge\sqrt{0.5}$. And (9) holds

$$\begin{aligned}&\Longleftrightarrow \sum_{j=r+1}^{p}(\frac{\beta_j}{\sigma})^2\le\sum_{j=r+1}^{p}(\xi_j-\frac{\beta_j}{\sigma})^2\\&\Longleftrightarrow 2\sum_{j=r+1}^{p}\frac{\beta_j\xi_j}{\sigma}\le\sum_{j=r+1}^{p}\xi_j^2\text{ holds.} \qquad (12)\end{aligned}$$

Similarly to the above proof, when $\|\frac{\beta_{(2)}}{\sigma}\|\le\frac{\delta^{(2)}_{\sqrt{0.5}}}{2}$, $P((9)\text{ holds})\ge\sqrt{0.5}$. We have

$$P(W(k)\le 0)\ge P((8)\text{ holds})\times P((9)\text{ holds})\ge 0.5.$$

Theorem 6 *If $n>t$, $\|\frac{\beta_{(1)}}{\sigma}\|\le\frac{\delta^{(1)}_{\sqrt{0.5}}(n-t)(m+t)(m+n)(2m+n+t)}{2(N-t)(M+t)^2(M+N)^2}$, then $\widetilde{\beta}(K)$ is better than $\widetilde{\beta}(t)$ under PMC, where $\delta^{(1)}_{\sqrt{0.5}}$ satisfies $P(\|\widetilde{\Lambda}_1\xi_{(1)}\|\le\delta^{(1)}_{\sqrt{0.5}}=0.5$.*

Proof Since

$$W(k,t)\le 0\Longleftarrow\begin{cases}\|A_{(1)}\xi_{(1)}-\frac{\beta_{(1)}}{\sigma}\|^2\le\|B_{(1)}\xi_{(1)}-\frac{\beta_{(1)}}{\sigma}\|^2\\ \|\vec{0}-\frac{\beta_{(2)}}{\sigma}\|^2\le\|\vec{0}-\frac{\beta_{(2)}}{\sigma}\|^2\end{cases} \qquad (13)$$

(13) holds

$$\begin{aligned}&\Longleftrightarrow \sum_{j=1}^{r}(\frac{\lambda_{i_j}\xi_j}{\lambda_{i_j}+k_{i_j}}-\frac{\beta_j}{\sigma})^2\le\sum_{j=1}^{r}(\frac{\lambda_{i_j}\xi_j}{\lambda_{i_j}+t}-\frac{\beta_j}{\sigma})^2\\&\Longleftrightarrow \sum_{j=1}^{r}\frac{2(k_{i_j}-t)\lambda_{i_j}\xi_j\beta_j}{(\lambda_{i_j}+t)(\lambda_{i_j}+k_{i_j})\sigma}\le\sum_{j=1}^{r}\frac{(k_{i_j}-t)(k_{i_j}+t+2\lambda_{i_j})\lambda_{i_j}^2\xi_j^2}{(\lambda_{i_j}+k_{i_j})^2(\lambda_{i_j}+t)^2} \qquad (14)\end{aligned}$$

Note that if $\frac{2(N-t)}{(m+t)(m+n)}\sum_{j=1}^{r}\frac{\lambda_{i_j}\xi_j\beta_j}{\sigma}\le\frac{(n-t)(n+t+2m)}{(M+N)^2(M+t)^2}\sum_{j=1}^{r}\lambda_{i_j}^2\xi_j^2$, then (14) holds.

According to Cauchy-Schwarz inequality, when

$$\|\widetilde{\Lambda}_1\xi_{(1)}\|\ge\frac{2(N-t)(M+t)^2(M+N)^2}{(n-t)(m+t)(m+n)(n+t+2m)}\|\frac{\beta_{(1)}}{\sigma}\|,$$

$P((13)\text{ holds})\ge 0.5$. So $P(W(k,t)\le 0)\ge P((13)\text{ holds})\ge 0.5$.

Similarly to Theorems 5 and 6, we have Theorem 7.

Theorem 7 *If $\|\frac{\beta_{(1)}}{\sigma}\|\le\frac{(2m+t)\delta^{(1)}_{\sqrt{0.5}}}{2(M+t)}$, then generalized ridge and principal correlation estimator is* *better than principal correlation estimator under PMC, where* $\delta^{(1)}_{\sqrt{0.5}}$ *satisfies* $P(\|\xi_{(1)}\| \leq \delta^{(1)}_{\sqrt{0.5}}) = 0.5$.

References

[1] ZHANG Wenwen. *Main correlation estimation of the regression parameters and its optimality* [J]. Acta Math. Appl. Sinica, 1996, **19**(4): 566–570. (in Chinese)

[2] YAN Li, CHEN Xia. *Ridge and principal correlation estimation of regression parameters and its optimality* [J]. J. Math. (Wuhan), 2006, **26**(3): 323–326. (in Chinese)

[3] WANG Songgui. *Optimality of principal components and a class of generalized principal component estimates* [J]. Chinese J. Appl. Probab. Statist., 1985, **1**(1): 23–30. (in Chinese)

[4] WANG Songgui. *Theory of Linear Model and Its Application* [M]. Hefei: Anhui Education Press, 1987. (in Chinese)

[5] PITMAN E J G. *The closet estimates of statistical parameters* [J]. Proc. Cambridge Philos. Soc., 1937, **33**: 212–222

[6] XU Wenli, LIN Jugan. *A combined principal component ridge estimator* [J]. Chinese J. Appl. Probab. Statist., 1995, **11**(1): 52–59.

[7] TIAN Baoguang, WANG Jianxin, CHANG Guijuan. *Variance optimality of combining generalized ridge and principal components estimate in the class of dimension-shrinking estimates* [J]. Sichuan Shifan Daxue Xuebao Ziran Kexue Ban, 2005, **28**(4): 426–428. (in Chinese)