应用概率统计 第十一卷
第一期 1995年2月

Chinese Journal of Applied Probability
and Statistics Vol. 11. No. Feb, 1995.

岭型组合主成分估计

徐文莉 林举干
(华东师范大学,上海,200062)

摘 要

本文提出了回归系数的一种新的改进估计——岭型组合主成分估计. 讨论了它的可容许性、约束条件下的可容许性和相合性问题. 分别在均方误差意义下和 Pitman 接近原则下,证明了在一定条件下,它优于最小二乘估计和岭估计,并且证明了它有比它们更好的抗干扰能力和稳健性.

关键词: 可容许性,优效性、抗干扰性、相合性、影响函数,Pitman 接近原则.

§1. 引 言

本文讨论线性模型

$$Y=X\beta+\varepsilon,\ E(\varepsilon)=0,\ \cos(\varepsilon)=\sigma^2 I. \tag{1.1}$$

的参数向量β的估计问题,其中Y为$n\times 1$观测向量,X为$n\times p$设计阵,$rk(X)=p$,β为$p\times 1$待估回归系数向量,ε为$n\times 1$随机误差向量,I为n阶单位阵.

当设计阵X含有多重共线性关系时,$X'X$接近奇异,它的某些特征根非常接近于0. 于是总存在$r<p$,使得$X'X$的特征根有,$\lambda_1\geqslant\cdots\geqslant\lambda_r\geqslant 1>\lambda_{r+1}\geqslant\cdots\geqslant\lambda_p>0$. (以下均作这样的假设.)因此,在设计阵$X$存在多重共线性关系时,$\beta$的最小二乘估计($OLS$),$\hat\beta=(X'X)^{-1}X'Y$,尽管是线性最佳无偏估计. 但却未必是个好的估计. 因为$\hat\beta$的均方误差,$MSE(\hat\beta)=\sigma^2\sum\limits_{i=1}^{p}\frac{1}{\lambda_i}$,会变得很大,且其平均模,$E(\hat\beta'\hat\beta)=MSE(\hat\beta)+\|\beta\|^2$,会过长. 这将导至$\hat\beta$的性能变得很坏. 因此,对$OLS$的改进无论是理论上或应用上一直都是引人关注的热门课题.

一种较好的改进方法是增大偏度以降低均方误差,并同时压缩OLS的长度. 为此人们提出了一些压缩型有偏估计,如 Hoerl & Kennard[1] 提出的岭估计(ORR),Massy[2] 提出的主成分估计(PCR),Baye & Parker[3] 提出的岭型主成分估计,又称r—k类估计,贾忠贞[5] 提出的组合主成分估计,杨虎[10] 提出的单参数组合主成分估计等.

本文提出的岭型组合主成分估计改进了ORR,PCR以及[3]、[5]中的估计. 它一方面改善了$X'X$的特征根接近于零的程度,另一方面几乎删除了X的列向量间存在的多重共线性关系. 避免了使用主成分估计,r—k类估计时要求确定取几个主成分的困挠. 同时还体现了对特征根大于1的部分不应作过多压缩的思想. 本文证明了一定条件下,它比OLS、ORR有

本文1993年7月15日收到,1994年3月22日收到修改稿

更小的均方误差，在 pitman 准则下优于它们，且有比它的更好的优效性，抗干扰性和稳健性。

§2. 定义、可容许性和相合性

考虑模型(1.1)的典型形式，

$$Y=Z\alpha+\varepsilon,\ E(\varepsilon)=0,\ \cos(\varepsilon)=\sigma^2 I, \tag{2.1}$$

其中 $Z=XQ$, $\alpha=Q'\beta$, Q 为 p 阶正交阵使得 $Q'X'XQ=\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_p)$. 记 $\Lambda_1=\mathrm{diag}(\lambda_1, \cdots, \lambda_r)$, $\Lambda_2=\mathrm{diag}(\lambda_{r+1}, \cdots, \lambda_p)$. 典则参数 α 的 OLS 为 $\hat{\alpha}=\Lambda^{-1}Z'Y$, 易见 $MSE(\hat{\alpha})=MSE(\hat{\beta})=\sigma^2\sum\limits_{i=1}^{p}\frac{1}{\lambda_i}$.

定义 2.1 称 $\tilde{\alpha}(k)=A\hat{\alpha}$ 为 α 的岭型组合主成分估计，若其 $A=\mathrm{diag}\left(\frac{\lambda_1}{\lambda_1+k}, \cdots, \frac{\lambda_r}{\lambda_r+k}, \frac{\lambda_r+1}{1+k}, \cdots, \frac{\lambda_p}{1+k}\right)$, $k\geqslant 0$ 为其岭参数.

我们知道，λ_i 越大，$\hat{\alpha}_i$ 对 $\hat{\alpha}$ 的均方误差影响越小，不宜作过多压缩；相反，λ_i 越接近于 0, $\hat{\alpha}_i$ 使 $MSE(\hat{\alpha})$ 增大的越多，应对 $\hat{\alpha}_i$ 尽可能压缩. $\lambda_i\geqslant 1$ 时，[3]中的压缩因子为 λ_i^{-1}, 因此 λ_i 很大时可能压缩过多，我们这里压缩因子为 $(\lambda_i+k)^{-1}\lambda_i$, λ_i 充分大后，基本上未对 $\hat{\alpha}_i$ 作什么压缩. $\lambda_i<1$ 时，ORR 的压缩因子为 $(\lambda_i+k)^{-1}\lambda_i$, 由于 λ_i 很小时，加之 k 通常也取很小的值，$(\lambda_i+k)^{-}\lambda_i$ 非常接近于 1, 即对 $\hat{\alpha}_i$ 没作什么压缩，而岭型组合主成分估计的压缩因子为 $(1+k)^{-1}\lambda_i$, 会[1]由于 λ_i 很小而接近于 0, 即几乎丢弃了这个主成分，但又用上了所有的主成分，它是 Massy 主成分估计的凸线性组合.

定理 2.1 设 $\alpha^*_{(j)}=B_{(j)}\hat{\alpha}=\begin{pmatrix} I_{p-j} & 0 \\ 0 & 0 \end{pmatrix}\hat{\alpha}$, 这里 I_{p-j} 为 $p-j$ 阶单位阵，$j=0, 1, 2, \cdots, p-1$, 约定 $\alpha^*_{(p)}=0$, 则

$$\begin{cases} \tilde{\alpha}(k)=\sum\limits_{j=0}^{p} w_j\alpha^*_{(j)} \\ w_j\geqslant 0,\ j=0, 1, \cdots, p,\ \sum\limits_{j=0}^{p} w_j=1, \end{cases} \tag{2.2}$$

证：由(2.2)可知

$$\begin{cases} A=\sum\limits_{j=0}^{p} w_jB_{(j)} \\ \sum\limits_{j=0}^{p} w_j=1,\ w_j\geqslant 0,\ j=0, 1, \cdots, p \end{cases}$$

故

$$w_j=\begin{cases} (1+k)^{-1}\lambda_p, & j=0 \\ (1+k)^{-1}(\lambda_{p-j}-\lambda_{p-j+1}), & 1\leqslant j\leqslant p-r-1 \\ (\lambda_r+k)^{-1}\lambda_r-(1+k)^{-1}\lambda_{r+1}, & j=p-r \\ (\lambda_{p-j}+k)^{-1}\lambda_{p-j}-(\lambda_{p-j+1}+k)^{-1}\lambda_{p-j+1}, & p-r+1\leqslant j\leqslant p-1 \\ 1-(\lambda_1+k)^{-1}\lambda_1, & j=p. \end{cases}$$

[证毕].

注意到 $\alpha^*_{(j)}$ 即为丢弃后 j 个主成分的 Massy 主成分估计，又 $\tilde{\alpha}(k)$ 中含有岭参数 k, 故称之为岭型组合主成分估计. 它只含一个岭参数，这样便于理论上讨论和应用.

由[6]，在均方误差意义下，显然 $\tilde{\alpha}(k)$ 在线性估计类中是可容许的. 由[7]的定理 2.1, 由

于 $0<A<I$, $\operatorname{rank}(I-A)=p>p-2$, 立得 $\varepsilon\sim N_n(0,\ \sigma^2 I_n)$ 时, $\tilde{\alpha}(k)$ 是可容许的. 从这意义上, 它改进了, OLS 在 $p>2$ 时不可容许, PCR 及 $r-k$ 类估计在删除的主成分个数超过两个时不可容许, 组合主成分估计当有两个以上的将征根 λ_i 为 1 时不可容许, 这些性质. 顺便指出, 当模型(2.1)带有约束 $H\alpha=0$, $\mu(H)\subset\mu(X')$ 时, 可相应定义约束岭型组合主成分估计, $\tilde{\alpha}_E(k)=G_E Z'Y$, 其中 $G_E=A\Lambda^{-1}-A\Lambda^{-1}H'[HA\Lambda^{-1}H']^{-1}H\Lambda^{-1}A$, 这时, 可由[8]中定理易得它在均方误差意义下可容许.

另外, 当误差服从正态分布时, 在二次损失下, 岭型组合主成分估计是正态——逆 Γ 先验分布下的 Bayes 估计.

关于相合性是明显的. 这里想指出, 当随机误差序列 $\{\varepsilon_i\}_{i=1,\cdots,n}$ 不满足 Gauss-Markov 条件时, OLS 不再相合, 但对其作微小改进的估计, 如岭估计却是相合的, 本文所给出的岭型组合主成分估计也同样是相合的, 其 r 阶平均相合($0<r\leqslant 2$), 弱相合, 强相合的条件完全类似于岭估计相合性的条件[14].

§3. 均方误差意义下的局部改善及估计的优效性

定理 3.1 在椭球 $\alpha' C\alpha\leqslant\sigma^2$, $C=(I-A)(I+A)^{-1}\Lambda$ 内, $MSE(\tilde{\alpha}(k))\leqslant MSE(\hat{\alpha})$, 其中 A 如定义 2.1 中所给(以下皆如此).

证: $MSEM(\hat{\alpha})=\sigma^2\Lambda^{-1}$, $MSEM(\tilde{\alpha}(k))=\sigma^2A^2\Lambda^{-1}+(I-A))\alpha\alpha'(I+A)$,

$$\begin{aligned}MSE(\tilde{\alpha}(k))\leqslant MSE(\hat{\alpha})&\Leftrightarrow MSEM(\tilde{\alpha}(k))\leqslant MSEM(\hat{\alpha})\\&\Leftrightarrow(I-A)\alpha\alpha'(I-A)\leqslant\sigma^2(I-A^2)\Lambda^{-1}\\&\Leftrightarrow\alpha'(I-A)(I+A)^{-1}\Lambda\alpha\leqslant\sigma^2.\end{aligned}$$ [证毕].

定理 3.2 在椭球 $\alpha' C_1\alpha\leqslant a\sigma^2$, $a=1'(\Lambda_2+kI)^{-1}[(2k+1)(\Lambda_2-\Lambda_2^2)+\Lambda_2^2-\Lambda_2^3]1$, $C_1=\begin{pmatrix}0&0\\0&(\Lambda_2+kI)^{-2}[2k(1+k)(\Lambda_2-\Lambda_2^2)+(\Lambda_2-\Lambda_2^2)^2]\end{pmatrix}$ 内, $MSE(\tilde{\alpha}(k))\leqslant MSE(\hat{\alpha}_R(k))$, 其中 $\hat{\alpha}_R(k)=(\Lambda+kI)^{-1}\Lambda\hat{\alpha}\triangleq B_r\hat{\alpha}$ 为 α 的 ORR.

证: $$MSE(\tilde{\alpha}(k))=\sum_{i=1}^{r}\frac{\sigma^2\lambda_i+k\alpha_i^2}{(\lambda_i+k)^2}+\sum_{i=r+1}^{p}\left[\frac{\sigma^2\lambda_i}{(1+k)^2}+\frac{(1+k-\lambda_i)^2\alpha_i^2}{(1+k)^2}\right],$$

$$MSE(\hat{\alpha}_R(k))=\sum_{i=1}^{p}\frac{\sigma^2\lambda_i+k\alpha_i^2}{(\lambda_i+k)^2}.$$

$$MSE(\tilde{\alpha}(k))\leqslant MSE(\hat{\alpha}_R(k))\Leftrightarrow$$

$$\sum_{i=r+1}^{p}\frac{\lambda_i(1-\lambda_i)}{(\lambda_i+k)^2(1+k)^2}\left[(2k^2+2k+\lambda_i-\lambda_i^2)\frac{\alpha_i^2}{\sigma^2}-(2k+1+\lambda_i)\right]\leqslant\sigma^2.$$ [证毕].

注意到, $\dfrac{2k+1+\lambda_p}{2k^2+2k+\frac{1}{4}}\leqslant\dfrac{2k+1+\lambda_i}{2k^2+2k+\lambda_i-\lambda_i^2}$, $i=r+1,\ \cdots,\ p$, 立得

定理 3.3 $\max\limits_{r+1\leqslant i\leqslant p}\dfrac{\alpha_i^2}{\sigma^2}\leqslant\dfrac{2k+1+\lambda_p}{2k^2+2k+\frac{1}{4}}$ 时, $MSE(\tilde{\alpha}(k))\leqslant MSE(\hat{\alpha}_R(k))$.

相对效率是比较两个估计量优劣的另一主要标准. C. R. Rao[9] 讨论了无偏估计的四种不同的效, 对于有偏估计, 沿用[10]中的相对效率概念:

设 α_1 为 α 的线性有偏估计, α_2 为 α 的另一线性估计, 称 $\rho(\alpha_1|\alpha_2)=1-\dfrac{MSE(\alpha_1)}{MSE(\alpha_2)}$ 为 α_1 相

对于 α_2 的效率.

显然, $MSE(\alpha_1)\leqslant MSE(\alpha_2)$时, $0\leqslant\rho(X_1|X_2)\leqslant1$, 且它越接近于1, 表明 α_1 改进 α_2 的程度越高.

定理3.4 i) 若$(1+\alpha'\alpha/\sigma^2)\lambda_p\leqslant1$, 则

$$\rho(\tilde{\alpha}(k)|\hat{\alpha})\geqslant\frac{1-\lambda_p-\lambda_p\alpha'\alpha/\sigma^2}{1+(p-1)\lambda_p},$$

ii) 在椭球 $\alpha'C_1\alpha\leqslant\sigma^2$ 内, 若 $0<k<2\sigma^2/\alpha'\alpha$, 则

$$\rho(\tilde{\alpha}(k)|\hat{\alpha})\geqslant\rho(\hat{\alpha}_R(k)|\hat{\alpha})$$

其中 C_1 如定理3.2中所给, 以下皆同.

证: i) 记$(A\Lambda^{-1})_{p-1}$为 $A\Lambda^{-1}$ 的前 $p-1$ 阶对角阵, 又注意到 $\frac{x+a}{x+b}$, 在 $a<b$ 时是 X 的增函数, 由定义2.1中 A 的性质, 立得

$$\rho(\tilde{\alpha}(k)|\hat{\alpha})\geqslant1-\frac{\sigma^2\mathrm{tr}(A\Lambda^{-1})_{p-1}+\sigma^2+\alpha'\alpha}{\sigma^2\mathrm{tr}(A\Lambda^{-1})_{p-1}+\sigma^2/\lambda_p}\geqslant1-\frac{\sigma^2(p-1)+\sigma^2+\alpha'\alpha}{\sigma^2(p-1)+\sigma^2/\lambda_p}$$
$$=\frac{1-\lambda_p-\lambda_p\alpha'\alpha/\sigma^2}{1+(p-1)\lambda_p}.$$

ii) 是显然的, 因为这时有 $MSE(\tilde{\alpha}(k))\leqslant MSE(\hat{\alpha}_R(k))\leqslant MSE(\hat{\alpha})$. [证毕].

上面定理的证明实际上证实了, 广义压缩 OLS 类[7]中的估计, 只要其对应的对角压缩阵 A 满足 $A\Lambda^{-1}<I$, 则都有i)成立, 即都可较好地改进 OLS. 由于结论依赖于待估参数, 下面的结果是它的某种补偿.

定理3.5 若存在正整数 l, 使得 $\lambda_{p-l}\approx\cdots\approx\lambda_p\approx0$, 则在椭球 $\alpha'C\alpha\leqslant\sigma^2$ 内, 有

$$\rho(\tilde{\alpha}(k)|\hat{\alpha})\geqslant\frac{l(1-\lambda_p^2)-\frac{k^2+2k}{k+1}\lambda_p}{l+1+(p-l-1)\lambda_p}\approx\frac{l}{l+1}\Big(1-\frac{p-l-1}{l+1}\lambda_p\Big).$$

证: 首先证明在椭球 $\alpha'C\alpha\leqslant\sigma^2$ 内, 有 $\alpha'(I-A)^2\alpha\leqslant\frac{1-(1+k)^2\lambda_p^2}{\lambda_p}\sigma^2$. 事实上, 注意到

$$(I-A^2)^{-1}\Lambda=\mathrm{diag}\Big(\frac{\lambda_1(\lambda_1+k)^2}{k(2\lambda_1+k)},\ \cdots,\ \frac{\lambda_r(\lambda_r+k)^2}{k(2\lambda_r+k)},\ \frac{(1+k)^2\lambda_{r+1}}{(1+k)^2-\lambda_{r+1}},\ \cdots,\ \frac{(1+k)^2\lambda_p}{(1+k)^2-\lambda_p^2}\Big)$$
$$\geqslant\frac{(1+k)^2\lambda_p}{(1+k)^2-\lambda_p}I$$

则得. 据此有

$$\rho(\tilde{\alpha}(k)|\hat{\alpha})=1-\frac{\sigma^2\mathrm{tr}A^2\Lambda^{-1}+\alpha'(I-A)^2\alpha}{\sigma^2\mathrm{tr}\Lambda}$$
$$\geqslant1-\frac{\mathrm{tr}(A^2\Lambda^{-1})_{p-l-1}+(l+1)\lambda_p+\lambda_p^{-1}-(1+k)^{-2}\lambda_p}{\mathrm{tr}(A^2\Lambda^{-1})_{p-l-1}+(l+1)\lambda_p^{-1}}$$
$$\geqslant\frac{\Big(l+\frac{k^2+2k}{(1+k)^2}\lambda_p\Big)+\lambda_p^{-1}}{(p-l-1)+(l+1)\lambda_p^{-1}}\approx\frac{l}{l+1}\Big(1-\frac{p-l-1}{l+1}\lambda_p\Big).$$

[证毕].

§4. 对 OLS、ORR 抗干扰性、稳健性的改进

对模型(2.1), $OLS\hat{\alpha}$, $ORR\hat{\alpha}_R(k)$, 岭型组合主成分估计 $\tilde{\alpha}(k)$分别是方程组 $\Lambda\alpha=Z'Y$、$(\Lambda+kI)\alpha=XY$、$\Lambda\alpha=Z'Y$ 的解. 本质上涉及的都是线性方程组 $GX=d$ 的解的问题.

定义 4.1 若线性方程组 $GX=d$, 有 G^{-1} 存在, 则称 $K(G)=\|G\|\cdot\|G^{-1}\|$ 为其系数矩阵 G 的条件数, 其中 $\|G\|=\sqrt{\lambda_{max}(G'G)}$ 为谱范数, $\lambda_{max}(G'G)$ 为 $G'G$ 的最大特征根.

我们知道[10], 若 $GX=d$ 中 G 的条件数很大, 即使 G 和 d 的扰动很小, 也可能引起解 X 产生很大的偏离. 也就是说, $K(G)$ 愈小, 解 X 的抗干扰性能愈强. 因此要提高模型(2.1)中待估参数 α 的线性估计的抗干扰性能的一个可行方法是, 降低其对应线性方程组系数矩阵的条件数, 它也称为相应估计 $\tilde{\alpha}$ 的条件数, 并记为 $K(\tilde{\alpha})$.

对模型(2.1), 当 $k\geqslant 0$ 时, 总有

$$K(\tilde{\alpha}(k))=\frac{\lambda_1+k}{1+k}<\frac{\lambda_1+k}{\lambda_p+k}=K(\hat{\alpha}_R(k))\leqslant\frac{\lambda_1}{\lambda_p}=K(\hat{\alpha}).$$

因此, $\tilde{\alpha}(R)$ 提高了 OLS 和 ORR 的抗干扰性能, 即模型的扰动对它的相对影响均比对 $\tilde{\alpha}_R(k)$、$\hat{\alpha}$ 的相对影响小.

利用影响函数[11] 对估计量作影响分析还可证得, 使用 $\tilde{\alpha}(k)$ 时, 模型的扰动对其绝对影响也比 $\hat{\alpha}_R(k)$ 和 $\hat{\alpha}$ 小. 参用[11]中的有关定理和 Tayler 展开式, 易算得 n 充分大时, 删除样本点 (X_i', Y_i) 后, 它们的样本影响函数分别为,

$$SIF(i,\ \tilde{\alpha}(k))\approx -A\Lambda^{-1}Q'X_iY_i,$$
$$SIF(i,\ \tilde{\alpha}_R(k))\approx -(\Lambda+hI)^{-1}Q'X_iY_i,$$
$$SIF(i,\ \hat{\alpha})\approx -\Lambda^{-1}Q'X_iY_i.$$

它们是向量, 若考虑 Cook 距离 $d(i,\ \tilde{\alpha})=\frac{1}{n-1}\|SIF(i,\ \tilde{\alpha})\|$. 易见 $d(i,\ \tilde{\alpha}(k))\leqslant d(i,\hat{\alpha}_R(k))\leqslant d(i,\ \hat{\alpha})$, 即 $\tilde{\alpha}(k)$ 更具稳健性.

§5. Pitman 准则下的优良性

Pitman 准则是近年来人们经常用于比较估计量优良性的另一重要标准. 以下设模型中误差向量 $\varepsilon\sim N(0,\ \sigma^2I_n)$.

定义 5.1 设 $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ 是参数 α 的两个估计, $L(\cdot,\ \alpha)$ 为损失函数, 定义 $\tilde{\alpha}_1$ 相对于 $\tilde{\alpha}_2$ 的 Pitman 度量(PMC)值为

$$PMC_L(\tilde{\alpha}_1,\ \tilde{\alpha}_2,\ \alpha)=P_r(L(\tilde{\alpha}_1,\ \alpha)\leqslant L(\tilde{\alpha}_2,\ \alpha)).$$

若 $PMC_L(\tilde{\alpha}_1,\ \tilde{\alpha}_2;\ \alpha)\geqslant 0.5$, 则称 $\tilde{\alpha}_1$ 在损失函数 $L(\cdot,\ \alpha)$ 下, PMC 优于 $\tilde{\alpha}_2$.

以下 $PMC(\tilde{\alpha}_1,\ \tilde{\alpha}_2;\ \alpha)$ 和 $PMC_F(\tilde{\alpha}_1,\ \tilde{\alpha}_2;\ \alpha)$ 分别记二次损失和 Fisher 损失下的 PMC 值. Fisher 损失是指, 估计量 $\tilde{\alpha}$ 满足 $\tilde{\alpha}=Gb$, $b_{p\times 1}$ 是样本统计量, $\mathrm{Cov}(b)=\Sigma>0$ 存在时, 将损失取为 $L(\tilde{\alpha},\ \alpha)=(\tilde{\alpha}-\alpha)'\Sigma^{-1}(\tilde{\alpha}-\alpha)$.

定理 5.1 设 $p>2$, 则 $k\geqslant 0$ 时, 有

i) $$\left(\frac{\alpha}{\sigma}\right)'\begin{pmatrix}(2I+k\Lambda_1^{-1})^{-2}\Lambda_1 & 0\\ 0 & (I+\Lambda_2^{-1}+k\Lambda_2^{-1})^{-2}\Lambda_2\end{pmatrix}\left(\frac{\alpha}{\sigma}\right)$$
$$\leqslant\frac{2k(p-2)(2\lambda_1+k)\lambda_p^2}{2[(1+k)^2-\lambda_p^2](\lambda_1+k)^2-2k\lambda_p^2(2\lambda_1+k)}$$

时, $\tilde{\alpha}(k)$ 在 PMC_F 下优于 $\hat{\alpha}$.

ii) i)中不等式左边 $\geqslant\frac{2p\lambda_1^2[(1+k)^2-\lambda_p^2]}{\lambda_1^2\lambda_p^2+2(1+k)^2(\lambda_1+k)^2-3(1+k)^2\lambda_1^2}$ 时, $\tilde{\alpha}(k)$ 在 PMC_F 下不可

能优于 $\hat{\alpha}$.

证：利用[13]中的结果，这时 $\Sigma=\sigma^2\Lambda^{-1}$, $C_0=\sigma^2\Lambda(I-\Lambda^2)>0$, $G\to(I+\Lambda)^{-1}$. 令 $\alpha^*\triangleq\frac{1}{\sigma}\Lambda^{\frac{1}{2}}(I-G)\alpha$, $\hat{\alpha}^*=\frac{1}{\sigma}\Lambda^{\frac{1}{2}}(\hat{\alpha}-\alpha)\sim N_p(0,\ I)$. 故 $U^*=(\hat{\alpha}^*+\alpha^*)'(\hat{\alpha}^*+\alpha^*)\sim\chi_p^2(\lambda)$, $\lambda=\frac{1}{2\sigma^2}\alpha'(I-G)'\Lambda(I-G)\alpha$. 记 $h_1=\frac{(1+k)^2}{\lambda_p^2}-1$, $d_p=1-\frac{\lambda_1^2}{(\lambda_1+k)^2}$, 则有

$$PMC_F(\tilde{\alpha}(k),\ \hat{\alpha};\alpha)=P_r(U\geqslant\alpha'G\Lambda G\alpha)\geqslant P_r\left(U^*\geqslant\frac{2\lambda h}{d_p}\right)$$

因此，$\frac{2h_1\lambda}{d_p}\leqslant\lambda+p-2$ 时，$P_r(U>\alpha'G\Lambda G\alpha)\geqslant0.5$, 从而 i)成立。

另记 $h_p=\left(1+\frac{k}{\lambda_1}\right)^2-1$, $d_1=1-\frac{\lambda_p^2}{(1+k)^2}$, 则

$$P_r(U>\alpha'G\Lambda G\alpha)\leqslant P_r(U^*\geqslant2h_p\lambda/d_1),$$

故 $\frac{2\lambda h_p}{d_1}\geqslant\lambda+p$ 时，$P_r\left(U^*\geqslant\frac{2\lambda h_p}{d_1}\right)\leqslant0.5$, 即 ii)成立。 [证毕].

按相同方式剖分下，记 $\alpha=\begin{pmatrix}\alpha^{(1)}\\ \alpha^{(2)}\end{pmatrix}\begin{matrix}r\times1\\ (p-r)\times1\end{matrix}$, $\tilde{\alpha}(k)=\begin{pmatrix}\hat{\alpha}^{(1)}(k)\\ \tilde{\alpha}^{(2)}(k)\end{pmatrix}$, $\hat{\alpha}=\begin{pmatrix}\hat{\alpha}^{(1)}\\ \hat{\alpha}^{(2)}\end{pmatrix}$, $\hat{\alpha}_R(k)=\begin{pmatrix}\hat{\alpha}_R^{(1)}(k)\\ \hat{\alpha}_R^{(2)}(k)\end{pmatrix}$, 又 $Z=\begin{pmatrix}Z_1 & Z_2\\ n\times r & n\times(p\cdot r)\end{pmatrix}$, 将定理 5.1 中所涉及的向量和矩阵换成其第二块，由类似方法立得，

定理 5.2 记 $p>r+2$, 则对 $k\in[0,\ 1-\lambda_{r+1})$, 若 $\left(\frac{\alpha^{(2)}}{\sigma}\right)'[(I-G^{(2)})]^2\Lambda_2\left(\frac{\alpha^{(2)}}{\sigma}\right)\leqslant\frac{2(p-r-2)d_p^{(2)}}{2h_{r+1}^{(2)}-d_p^{(2)}}$, 则在 PMC_F 下，$\tilde{\alpha}(k)$优于 ORR. 其中，$G^{(2)}=(1+k)(\Lambda_2+kI)\Lambda_2^{-1}[\Lambda_2+(2k+1)I]^{-1}$, $d_p^{(2)}=\min\limits_{r+1\leqslant i\leqslant p}\left\{\frac{\lambda_i^2}{(\lambda_i+k)^2}-\frac{\lambda_i^2}{(1+k)^2}\right\}$, $h_{r+1}^{(2)}=\max\limits_{r+1\leqslant i\leqslant p}\left\{\frac{\begin{Bmatrix}(1-\lambda_i)(1+\lambda_i+2k)\\ \cdot(\lambda_i^2+k\lambda_i-k^2-k)\lambda_i^2\end{Bmatrix}}{(\lambda_i+k)^3(1+k)^3}\right\}$.

下面讨论二次损失，即 PMC 下估计量的优良性. 令 $\xi=\frac{\hat{\alpha}}{\sigma}=\begin{pmatrix}\xi^{(1)}\\ \xi^{(2)}\end{pmatrix}\begin{matrix}r\\ p-r\end{matrix}$, 则 $\xi\sim N_p\left(\frac{\alpha}{\sigma},\ \Lambda^{-1}\right)$, $\xi^{(1)}\sim N_r\left(\frac{\alpha^{(1)}}{\sigma},\ \Lambda_1^{-1}\right)$, $\xi^{(2)}\sim N_{p-r}\left(\frac{\alpha^{(2)}}{\sigma},\ \Lambda_2^{-1}\right)$; 记 $A^{(1)}=(\Lambda_1+kI)^{-1}\Lambda_1$, $A^{(2)}=(1+k)^{-1}\Lambda_2$; 又记 $W(k)=\|\tilde{\alpha}(k)-\alpha\|^2-\|\hat{\alpha}-\alpha\|^2=\sigma^2\left[\left\|A\xi-\frac{\alpha}{\sigma}\right\|^2-\left\|\xi-\frac{\alpha}{\sigma}\right\|^2\right]$, $W_R(k)=\|\tilde{\alpha}(k)-\alpha\|^2-\|\tilde{\alpha}_R(k)-\alpha\|^2=\sigma^2\left[\left\|A\xi-\frac{\alpha}{\sigma}\right\|^2-\left\|B_r\xi-\frac{\alpha}{\sigma}\right\|^2\right]$.

定理 5.3 当 $\left\|\frac{\alpha^{(1)}}{\sigma}\right\|\leqslant r_{\sqrt{0.5}}^{(1)}\frac{\lambda_1}{\lambda_r}$, $\left\|\frac{\alpha^{(2)}}{\sigma}\right\|\leqslant\frac{r_{\sqrt{0.5}}^{(2)}(1+k-\lambda_{r+1})(1+k+\lambda_r+1)}{2(1+k)(1+k-\lambda_p)}$ 时，在 PMC 下，$\tilde{\alpha}(k)$ 优于 $\hat{\alpha}$, 其中 $r_{\sqrt{0.5}}^{(i)}$ 满足 $P_r(\|\xi\|^{(i)}\leqslant r_{\sqrt{0.5}}^{(i)})=\sqrt{0.5}$, $i=1,\ 2$.

证：$PMC(\tilde{\alpha}(k),\ \hat{\alpha};\alpha)=P_r(W(k)\leqslant0)$. 而

$$W(k)\leqslant0\Leftarrow\begin{cases}\left\|A^{(1)}\xi^{(1)}-\frac{\alpha^{(1)}}{\sigma}\right\|^2\leqslant\left\|\xi^{(1)}-\frac{\alpha^{(1)}}{\sigma}\right\|^2\Leftrightarrow\sum\limits_{i=1}^{r}\left(\frac{\lambda_i\xi_i}{\lambda_i+k}-\frac{\alpha_i}{\sigma}\right)^2\leqslant\sum\limits_{i=1}^{r}\left(\xi_i-\frac{\alpha_i}{\sigma}\right)^2\triangleq(1)\\ \left\|A^{(2)}\xi^{(2)}-\frac{\alpha^{(2)}}{\sigma}\right\|^2\leqslant\left\|\xi^{(2)}-\frac{\alpha^{(2)}}{\sigma}\right\|^2\Leftrightarrow\sum\limits_{i=r+1}^{p}\left(\frac{\lambda_i\xi_i}{1+k}-\frac{\alpha_i}{\sigma}\right)^2\leqslant\sum\limits_{i=r+1}^{p}\left(\xi_i-\frac{\alpha_i}{\sigma}\right)^2\triangleq(2).\end{cases}$$

由[12], $\left\|\frac{\alpha^{(1)}}{\sigma}\right\|\leqslant r_{\sqrt{0.5}}^{(1)}\frac{\lambda_1}{\lambda_r}$ 时，$P_r((1)成立)\geqslant\sqrt{0.5}$. 又注意到

$$2(1+k)\sum_{i=r+1}^{p}(1+k-\lambda_i)\frac{\alpha_i}{\sigma}\xi_i\leqslant[(1+k)^2-\lambda_{r+1}^2]\sum_{i=r+1}^{p}\zeta_i^2,$$

由 Cauchy-Schwarz 不等式，可得$\|\xi^{(2)}\|\geqslant\dfrac{2(1+k)(1+k-\lambda_p)}{(1+k-\lambda_{r+1})(1+k+\lambda_{r+1})}\dfrac{\|\alpha^{(2)}\|}{\sigma}$时，(2)成立，故 $P((2)\text{成立})\geqslant\sqrt{0.5}$. 而 $P(W(k)\leqslant 0)\geqslant P((1)\text{成立})\cdot P((2)\text{成立})\geqslant 0.5$. [证毕].

类似可得：

定理 5.4 设 $k\in[0,1-\lambda_{r+1})$，则当$\left\|\dfrac{\alpha^{(2)}}{\sigma}\right\|\leqslant\dfrac{\delta_{0.5}^{(2)}(\lambda_p+k)(1+\lambda_{r+1}+k)(1-\lambda_{r+1})}{2(1+k)(\lambda_{r+1}+k)^2(1-\lambda_p)}$时，在 PMC 意义下，岭型组合主成分估计优于岭估计. 其中 $\delta_{0.5}^{(2)}$ 满足 $P(\|A_2\xi^{(2)}\|\geqslant\delta_{0.5}^{(2)})=0.5$.

§6. 一些说明

在实际应用中，用有偏估计来改进 OLS，一般应兼顾偏度、均方误差、抗干扰性等等. 在应用中，岭型组合主成分估计宜取 $k>0$ 尽可能小. 选取时可借鉴岭估计中的岭迹法、方差扩大因子法、以及兼顾均方误差和均方残差的 $Q(C)$ 准则[16] 方法等等.

将岭型组合主成分估计 $\tilde{\alpha}(k)$ 看成 k 的有理函数时，沿[15]的方式，易得 $\tilde{\alpha}(k)$ 与 $\hat{\alpha}_R(k)$ 相似或独具的一些代数性质，如与符号有关的性质、渐近性质和变化速度等.

参 考 文 献

[1] Hoerl, A. E & Kennard, R. N. Ridge regression biased estimation for non-orthogonal problems. *Technometrics* 17(1970), pp. 55—67

[2] Massy, W. F. Principal component regression in exploratory statistical research *J. A. S. A.* 60 (1965), pp. 234—246.

[3] Baye & Parker Combining ridge and principal component regression: a money demand illustration. *Comm. Statist. -Theory & Methods*, 13 [2], pp. 197—205. (1984)

[4] Masuo Nonura & Tsunehara Ohkubo. A note on combining ridge and principal component regression. *Comm. Statist.-Theo. Meth.* 14(2) (1985) pp. 2489—2493

[5] 贾忠贞 回归系数的组合主成分估计. 数理统计与应用概率 Vol. 2 no. 2 (1987), 153—158.

[6] 王松桂 线性模型的理论及其应用，安徽教育出版社，1987.

[7] 王力群 广义压缩最小二乘估计 应用概率统计. Vol 6, No. 3, 1990, 225—231

[8] 郑昌光 约束条件下的线性估计 应用概率统计 Vol 2, no. 1, 1986, 5—12.

[9] C. R. Roo The efficienty of least squares extension of the Kantororich inequality. *Linear Algebra Appl.* 70 (1985) 249—255.

[10] 杨 虎 参数主成分估计 高校应用数学学报，Vol 4, no. 1, 1989, 74—80.

[11] 韦博成等 统计诊断方法 东南大学出版社

[12] Alessandro Bortuzzi & Aebarto Gandocfi. Ridge regression versus OLS by Pitman's Closeness under quadratic and Fisher's loss. *Comm. Statist.—Theory Meth.*, 20(11), 1991, 3581—3590.

[13] Robert L. Mason et al. Comparison of linear estimators using Pitman's Measure of Clossness. *JASA*, Vol. 85, 1990, 579—581.

[14] 王启应 回归系数岭估计的相合性 数理统计与应用概率，Vol. 3, no. 1, 1988, 43—51.

[15] Gary, c. Mc Donald. Some algebraic properties of ridge Coefficients. *JBSS*(*B*), 42, no. 1, 1980, 31—34

[16] 鲁国斌 广义岭回归估计中关于 K 值选择的 $Q(C)$ 准则. 数理统计与应用概率，Vol. 4, no. 2, 1989, 159—169.

RIDGE COMBINED PRINCIPAL COMPONENTS ESTIMATOR

XU, WEN LI LIN, JU GAN
(*East China Normal University*)

In this paper, We propose a ridge combined principal Component estimator for the regression coefficients. And the admissibility, restricted admissibility and consistcncy of the estimator are discussed. We show the regions preferable to OLS, ORR, under mean square error and Pitman's clossness respectively, and also that it can improve the resistence and sensitivity to data.

Keyword: Admissibility, Relative efficiency, Numerical stability, Consistency, Influential function. Pitman's clossness.