文章编号: 1672 - 058X(2009)06 - 0517 - 05

有约束的多元线性回归模型的 Minimax估计

王理峰[1],朱道元[2*]
(1. 南京铁道职业技术学院 社科部, 南京 210015; 2. 东南大学 数学系,南京 210096)

摘　要:在损失函数 $\mathrm{tr}(\hat{B}-B)'A(\hat{B}-B)$下给出 B在线性估计类中的 Minimax估计,研究了其性质. 在一些特殊的情形下,该估计包括了多元功效岭回归估计 (Ridge regression estimation),多元 Stein估计等.

关键词:多元线性回归模型;多元线性 Minimax估计;多元岭回归估计
中图分类号: O212. 1　　　　**文献标志码**: A

为了以下计算方便,先介绍几个符号:符号 $a\wedge b=\min(a,b)$, $a\vee b=\max(a,b)$, $x_+=\max(x,0)$.

若 A为 $n\times n$阶对称矩阵,有谱分解 $A=PDP'$,其中 P为 $n\times n$阶正交阵,D为对角阵,其对角元记为 d_i, $i=1,2,\cdots,n$. 定义:$D_+=\mathrm{diag}((d_1\vee 0),(d_2\vee 0),\cdots,(d_i\vee 0),\cdots,(d_n\vee 0))$. 考虑多元线性回归模型:

$$\begin{cases} Y=XB+E \\ \mathrm{vec}(E)\sim(0,W\otimes I) \end{cases} \tag{1}$$

其中 Y为 $n\times q$阶观测矩阵, X为 $n\times p$设计矩阵, $rk(X)=p$. $E=(e_1,e_2,\cdots,e_q)$为 $n\times q$阶误差矩阵,$W=(w_{ij})$为已知的 q阶非零非负定阵. B为 $p\times q$阶未知参数矩阵,满足 $\mathrm{tr}(B'X'FXB)\le\rho$,其中 F为 $n\times n$阶非负定阵. 记 $\Theta=\{B\mid\mathrm{tr}(B'X'FXB)\le\rho\}$. 式 (1)的最小二乘估计为 $\hat{B}_{LS}=(X'X)^{-1}X'Y$,此处采用的损失函数为:$L(\hat{B},B,A)=\mathrm{tr}(\hat{B}-B)'A(\hat{B}-B)$,$A$为 $p\times p$阶正定阵,其相应的风险函数为 $EL(\hat{B},B,A)$.

定义 1　B^*为约束条件下 B的线性 Minimax估计,若 $\inf\limits_{\hat{B}\in l}\sup\limits_{B\in\Theta}EL(\hat{B},B,A)=\sup\limits_{B\in\Theta}EL(B^*,B,A)$,其中 l为 B的线性估计类,记 $l=\{\hat{B}:\hat{B}=CY,\ C为\ p\times n阶矩阵\}$.

线性回归的许多问题,包括 Minimax估计问题,在典则形式下很容易理解.

对 X进行奇异值分解: $X=U\tilde{D}V'$. U,V分别为 $n\times n$, $p\times p$阶正交阵,$\tilde{D}$为 $n\times p$阶矩阵. $\tilde{D}$的 (i,i)元为 d_i $(d_1\ge d_2\ge\cdots\ge d_p>0)$,其余位置为 0,则 $X'X=V\tilde{D}'\tilde{D}V'\overset{\Delta}{=}VD^2V'$.其中 $D=\mathrm{diag}(d_1,d_2,\cdots,d_p)$, $d_i=\lambda_i^{-\frac{1}{2}}$, λ_i为 $X'X$的非零特征根,其中 $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_p>0$.

令 $Z=U'Y$, $\tilde{D}=U'XV$, $R=V'B$, $\varepsilon=U'E$, $u=XB$, $\eta=EZ=U'u=(\eta_1,\eta_2,\cdots,\eta_q)$,其中,$\eta_i=(\eta_{1i},\eta_{2i},\cdots,\eta_{pi},0,\cdots,0)'$,则模型 (1)可化为:

$$\begin{cases} Z=\tilde{D}R+\varepsilon \\ \mathrm{vec}(\varepsilon)\sim(0,W\otimes I_n) \end{cases} \tag{2}$$

式 (2)的最小二乘估计为 $\hat{R}_{LS}=(\tilde{D}'\tilde{D})^{-1}\tilde{D}'Z$. 记 $Z=(z_{ij})=(z_1,z_2,\cdots,z_q)=(\tilde{Z}'\ *)'$,其中 $\tilde{Z}=(\tilde{z}_1,\tilde{z}_2,\cdots,\tilde{z}_q)$, $\tilde{z}_i=(z_{1i},z_{2i},\cdots,z_{pi})'$, $\tilde{Z}$为 Z上面 $p\times q$阶子块组成的矩阵. 记 $\varepsilon=(\varepsilon_{ij})=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_q)=(\varepsilon'\ *)'$ 其中 ε

收稿日期: 2009 - 08 - 30;**修回日期**: 2009 - 09 - 20.
作者简介:王理峰 (1981 -),女,河南平顶山人,硕士,从事多元统计分析及其应用研究.
*通讯作者:朱道元 (1947 -),男,东南大学数学系教授,从事多元统计分析与数学建模研究. e - mail: zhudy@seu. edu. cn.

$=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_q)$, $\varepsilon_i=(\varepsilon_{1i},\varepsilon_{2i},\cdots,\varepsilon_{pi})'$, ε为 ε上面 $p\times q$阶子块组成的矩阵. 记 $R=(r_{ij})=(r_1,r_2,\cdots r_q)$,则

$$\begin{cases}\tilde{Z}=DR+\varepsilon\\ \mathrm{vec}(\varepsilon)\sim(0,W\otimes I_p)\end{cases}\tag{3}$$

要估计 R只需考虑模型 (3)即可,而模型中 D为对角阵,它使求解方便、简洁. 将模型 (3)拉直得:

$$\begin{cases}\mathrm{vec}\tilde{Z}=(I\otimes D)\,\mathrm{vec}R+\mathrm{vec}\varepsilon\\ \mathrm{vec}(\varepsilon)\sim(0,W\otimes I_p)\end{cases}\tag{4}$$

为使 $L(\hat{B},B,A)$、$tr(B'X'FXB)$做相应的变换能化为简洁形式,规定 A, F满足以下条件:

$A=V'\tilde{A}V$ 记 $\tilde{A}=\mathrm{diag}(a_1,a_2,\cdots,a_p)$, $a_i>0$; $F=U'\tilde{F}U$, 记 $\tilde{F}=\mathrm{diag}(\tilde{f}_1,\tilde{f}_2,\cdots,\tilde{f}_p,\cdots,\tilde{f}_n)$, $\tilde{f}_i\geq 0$。记 $\tilde{F}=\mathrm{diag}(\tilde{f}_1,\tilde{f}_2,\cdots,\tilde{f}_p)$.

参数约束空间 $\Theta=\{B\mid \mathrm{tr}(B'X'FXB)\leq\rho\}=\{R\mid \mathrm{tr}(R'\tilde{D}'\tilde{F}\tilde{D}R)\leq\rho\}=\{R\mid \mathrm{tr}(R'D'\tilde{F}DR\leq\rho\}=\{R\mid \sum_{i=1}^{p}\sum_{j=1}^{q}\tilde{f}_i\lambda_i r_{ij}^2\leq\rho\}\triangleq\Theta$.

损失函数 $L(\hat{B},B,A)=\mathrm{tr}(\hat{B}-B)'A(\hat{B}-B)=\mathrm{tr}(\hat{R}-R)'\tilde{A}(\hat{R}-R)=\sum_{j=1}^{q}(\hat{r}_j-r_j)'\tilde{A}(\hat{r}_j-r_j)=\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(\hat{r}_{ij}-r_{ij})^2\triangleq L(\hat{R},R,\tilde{A})$.

1 多元线性 Minimax估计的求解

定理 1 将模型 (3)写成元素形式为: $z_{ij}=d_i r_{ij}+\varepsilon_{ij}$; $E\varepsilon_{ij}=0$, $\mathrm{cov}(\varepsilon_{ij},\varepsilon_{kl})=w_{jl}\delta_{ki}$; $i=1,2,\cdots,p$; $j=1,2,\cdots,q$

其中 $\delta_{ki}=\begin{cases}1 & k=i\\ 0 & k\neq i\end{cases}$, $\Theta_o=\{R\mid \sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\tilde{f}_i r_{ij}^2\leq\rho\}$为约束空间, $L(\hat{R},R,\tilde{A})=\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(\hat{r}_{ij}-r_{ij})^2$为损失函数,其中 $a_i>0$, $\tilde{f}_i\geq 0$. 记 c_i为 C的第 i个对角元. r_{ij}的线性 Minimax估计为: $\hat{r}_{M,ij}=c_i^* z_{ij}$, $i,j=1,2,\cdots,p$. 若 $\tilde{f}_i>0$, $c_i^*=d_i^{-1}\left(1-h\left(\frac{\tilde{f}_i\lambda_i}{a_i}\right)^{\frac{1}{2}}\right)_+$, 其中 h由 $\sum_{i=1}^{p}\lambda_i\tilde{f}_i(\sum_{j=1}^{q}r_{ij}^2)^*=\rho$决定, $(\sum_{j=1}^{q}r_{ij}^2)^*=\sum_{j=1}^{q}w_{jj}d_i^{-2}\left(\frac{1}{h}\left(\frac{a_i}{\tilde{f}_i\lambda_i}\right)^{\frac{1}{2}}-1\right)_+$; 若 $\tilde{f}_i=0$, Θ_0对第 i个坐标方向无约束, $c_i=\frac{1}{d_i}$.

令 ζ表示所有 $p\times p$阶矩阵组成的类, $\tilde{Z}=(\tilde{z}_1,\tilde{z}_2,\cdots,\tilde{z}_q)$, $\tilde{z}_i=(z_{1i},z_{2i},\cdots,z_{pi})'$,则模型 (3)的 Minimax风险为:

$$\inf_{C\in\zeta}\sup_{R\in\Theta_O}E\,\mathrm{tr}(C\tilde{Z}-R)'\tilde{A}(C\tilde{Z}-R)=\inf_{c_i}\sup_{R\in\Theta_O}E\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_i z_{ij}-r_{ij})^2=$$

$$\sup_{R\in\Theta_O}\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_i^* z_{ij}-r_{ij})^2=\sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i d_i^{-2}(1-h(\frac{\tilde{f}_i\lambda_i}{a_i})^{\frac{1}{2}})_+$$

证明 (1)首先,利用 Speckman的讨论来证明使极大风险极小化的矩阵 C是对角阵,记 C的 (i,j)元为 c_{ij}.

$$J(C)=\sup_{R\in\Theta_O}E\,\mathrm{tr}(C\tilde{Z}-R)'\tilde{A}(C\tilde{Z}-R)=$$

$$\sup_{R\in\Theta_O}\mathrm{vec}R'[I\otimes(CD-I)'\tilde{A}(CD-I)]\mathrm{vec}R+\mathrm{tr}(W\otimes\tilde{A}CC')=$$

$$\sup_{R\in\Theta_O}\sum_{j=1}^{q}\{r_j'(CD-I_p)'\tilde{A}(CD-I_p)r_j+w_{jj}\mathrm{tr}(\tilde{A}CC')\}$$

记 $t_i=\frac{(c_{ii}d_i-1)^2 a_i}{\lambda_i\tilde{f}_i}$, $t_k=\max_{1\leq i\leq p}t_i$.

$$J(\mathrm{diag}C) = \sup_{R\in\Theta_0}\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_{ii}d_i-1)^2 r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i c_{ii}^2 =$$

$$\sup_{R\in\Theta_0}\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\tilde{f}_i t_i r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i c_{ii}^2 \le t_k\rho + \sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i c_{ii}^2$$

取 R满足: $i\neq k$时, $r_{ij}=0$. 则 $J(\mathrm{diag}C)\ge\sup_{R\in\Theta_0}\sum_{j=1}^{q}\lambda_k\tilde{f}_k t_k r_{kj}^2+\sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i c_{ii}^2 = t_k\rho+\sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i c_{ii}^2$. 所以

$$J_O(C)\ \hat{=}\max_{1\le i\le p}\frac{\rho}{\lambda_i\tilde{f}_i}(c_{ii}d_i-1)^2a_i+\sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_i c_{ii}^2 = J(\mathrm{diag}C)$$

$$J(C) = \sup_{R\in\Theta_O}\sum_{j=1}^{q}\{r_j'(CD-I_p)\tilde{A}(CD-I_p)r_j + w_{jj}tr(\tilde{A}CC')\}\ge$$

$$\sup_{R\in\Theta_O}\sum_{i=1}^{p}\sum_{j=1}^{q}r_{ij}^2e_i'(CD-I_p)\tilde{A}(CD-I_p)e_i+\sum_{j=1}^{q}w_{jj}\sum_{i=1}^{p}a_i\sum_{l=1}^{p}c_{il}^2\ge$$

(令 $\vartheta_i=\frac{1}{\lambda_i\tilde{f}_i}[a_i(c_{ii}d_i-1)^2+\sum_{j\neq i}a_ic_{ij}^2d_j^2]\ \vartheta_m=\max_{1\le i\le p}\vartheta_i$. 取 R满足: $i\neq m$时, $r_{ij}=0$)

$$\sup_{R\in\Theta_O}\sum_{j=1}^{q}\{\lambda_m\tilde{f}_m\vartheta_m r_{mj}^2+w_{jj}\sum_{i=1}^{p}a_i\sum_{l=1}^{p}c_{il}^2\}=$$

$$\max_{1\le i\le p}\frac{\rho}{\lambda_i\tilde{f}_i}[a_i(c_{ii}d_i-1)^2+\sum_{j\neq i}a_ic_{ij}^2d_j^2]+\sum_{j=1}^{q}w_{jj}\sum_{i=1}^{p}a_i\sum_{l=1}^{p}c_{il}^2\ge J_O(C)$$

当且仅当 C为对角阵时等号成立. 因此使极大风险极小化的矩阵 C是对角阵.

(2) $E\sum_{j=1}(c_iz_{ij}-r_{ij})^2=\sum_{j=1}\{\mathrm{var}(c_iz_{ij}-r_{ij})+[E(c_iz_{ij}-r_{ij})]^2\}=\sum_{j=1}^{q}c_i^2w_{jj}+\sum_{j=1}^{q}r_{ij}^2(d_ic_i-1)^2$

在 $c_i=\dfrac{d_i(\sum_{j=1}^{q}r_{ij}^2)}{\sum_{j=1}^{q}(w_{jj}+d_i^2r_{ij}^2)}$, $i=1,2,\cdots,p$处取得最小值.

记 $\omega\ \hat{=}\sum_{j=1}^{q}w_{jj}$, $s_i\ \hat{=}\sum_{j=1}^{q}r_{ij}^2$, $\xi_i\ \hat{=}\sum_{j=1}^{q}(w_{jj}+d_i^2r_{ij}^2)=\omega+d_i^2s_i$, $\sup_{R\in\Theta_0}\inf_{c_i}\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_iz_{ij}-r_{ij})^2=\sup_{R\in\Theta_0}\sum_{i=1}^{p}\frac{a_i\omega s_i}{\xi_i}\ \hat{=}\ \nu^2$.

显然 s_i越大, $\sum_{i=1}^{p}\frac{a_i\omega s_i}{\xi_i}$越大, ν^2 在约束条件边界上达到. 利用 Lagrange乘子法可求出 ν^2.

对于 $h>0$, $i=1,2,\cdots,p$, $G(s_i,h)=\sum_{i=1}^{p}\frac{\omega a_is_i}{\xi_i}-h^2\sum_{i=1}^{p}\tilde{f}_i\lambda_is_i$,则 $\frac{\partial G(s_i,h)}{\partial s_i}=\frac{a_i\omega}{\xi_i}-\frac{d_i^2a_i\omega s_i}{\xi_i^2}-h^2\tilde{f}_i\lambda_i$.

令 $\frac{\partial G(s_i,h)}{\partial s_i}=0$,得 $(s_i)^{**}=\omega d_i^{-2}\left(\frac{1}{h}\left(\frac{a_i}{\tilde{f}_i\lambda_i}\right)^{\frac{1}{2}}-1\right)$. 而 $\frac{\partial^2G(s_i,h)}{\partial^2s_i}=-\frac{2a_id_i^2\omega^2}{\xi_i^3}$,故当 $s_i\ge(s_i)^{**}$时, $\frac{\partial^2G(s_i,h)}{\partial^2s_i}<0$, 所以 $s_i\ge(s_i)^{**}$时, $\frac{\partial G(s_i,h)}{\partial s_i}\le0$, 可知 $G(s_i,h)$在此范围内为减函数. 故 $s_i^*=\omega d_i^{-2}(\frac{1}{h}(\frac{a_i}{\tilde{f}_i\lambda_i})^{\frac{1}{2}}-1)_+$时,得 $\nu^2=\sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_id_i^{-2}(1-h(\frac{\tilde{f}_i\lambda_i}{a_i})^{\frac{1}{2}})_+$,其中 h由 $\sum_{i=1}^{p}\lambda_i\tilde{f}_is_i^*=\rho$决定.

将 s_i^*代入(*)得 $c_i^*=d_i^{-1}\left(1-h\left(\frac{\tilde{f}_i\lambda_i}{a_i}\right)^{\frac{1}{2}}\right)_+$,则:

$$\sup_{R\in\Theta_O}E\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_iz_{ij}-r_{ij})^2\ge\inf_{c_i}\sup_{R\in\Theta_O}E\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_iz_{ij}-r_{ij})^2\ge\sup_{R\in\Theta_O}\inf_{c_i}E\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_iz_{ij}-r_{ij})^2=\nu^2$$

当 $c_i=c_i^*$, $i=1,2,\cdots,p$时上面不等式仍成立. 而

$$\sup_{R\in\Theta_O}E\sum_{i=1}^{p}\sum_{j=1}^{q}a_i(c_i^*z_{ij}-r_{ij})^2=\sup_{R\in\Theta_O}\sum_{i=1}^{p}\sum_{j=1}^{q}r_{ij}^2((h^2\tilde{f}_i\lambda_i)\wedge a_i)+\sum_{i=1}^{p}\sum_{j=1}^{q}w_{jj}a_id_i^{-2}\left(1-h\left(\frac{\tilde{f}_i\lambda_i}{a_i}\right)^{\frac{1}{2}}\right)_+^2\le$$

$$\rho h^2 + \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj} a_i d_i^{-2}\left(1 - h\left(\frac{\tilde f_i\lambda_i}{a_i}\right)^{-\frac{1}{2}}\right)_+^2 = h^2\sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}\tilde f_i\lambda_i d_i^{-2}\left(\frac{1}{h}\left(\frac{a_i}{\tilde f_i\lambda_i}\right)^{-\frac{1}{2}} - 1\right)_+ +$$

$$\sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj} a_i d_i^{-2}\left(1 - h\left(\frac{\tilde f_i\lambda_i}{a_i}\right)^{-\frac{1}{2}}\right)_+^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj} a_i d_i^{-2}\left(1 - h\left(\frac{\tilde f_i\lambda_i}{a_i}\right)^{-\frac{1}{2}}\right)_+ = \nu^2$$

由上面的不等式及证明中的 (1)部分可知:

$$\sup_{R\in\Theta_O} E\sum_{i=1}^{p}\sum_{j=1}^{q} a_i (c_i^* z_{ij} - r_{ij})^2 = \nu^2 \le \sup_{R\in\Theta_O} E\sum_{i=1}^{p}\sum_{j=1}^{q} a_i (c_i z_{ij} - r_{ij})^2 \le \sup_{R\in\Theta_O} \mathrm{Etr}(C\tilde Z - R)'\tilde A(C\tilde Z - R)$$

由定义 1知 $\hat R_M = C^* Z$为 R的 Minimax估计. 其中 $C^* = \mathrm{diag}(c_1^*, c_2^*, \cdots, c_p^*)$.

定理 2 对于多元线性回归模型 (1), $B\in\Theta = \{B \mid \mathrm{tr}(B'X'FXB)\le\rho\}$, A, F满足条件 1,则 B的线性 Minimax估计为 $\hat B_M = (I - h(X'FX)^{-\frac{1}{2}}A^{-\frac{1}{2}})_+\hat B_{LS}$. 其中 h满足: $\mathrm{tr}W\cdot\mathrm{tr}\{(X'X)^{-1}(h^{-1}A^{\frac{1}{2}}(X'FX)^{-\frac{1}{2}} - X'FX)_+\}$ $= \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}\lambda_i^{-1}(h^{-1}(a_i\tilde f_i\lambda_i)^{\frac{1}{2}} - \lambda_i\tilde f_i)_+ = \rho$

线性 Minimax估计的风险为: $\inf_{S\in\tau}\sup_{B\in\Theta} E\{tr(SY - B)'A(SY - B)\} = \sup_{B\in\Theta} E\{\mathrm{tr}(\hat B_M - B)'A(\hat B_M - B)\} = \mathrm{tr}W\cdot\mathrm{tr}$ $\{(X'X)^{-1}(A - hA^{\frac{1}{2}}(X'FX)^{\frac{1}{2}})_+\}$.

其中 τ表示任意的 $p\times n$阶矩阵组成的类.

证明 由第一节知 $Z = U'Y, \tilde D = U'XV, R = V'B, \tilde A = V'AV, \tilde F = U'FU$. 则

$$\tilde A^{\frac{1}{2}} = V'A^{\frac{1}{2}}V,\quad (\tilde D'\tilde F\tilde D)^{\frac{1}{2}} = V'(X'FX)^{\frac{1}{2}}V$$

由定理 1知 $\hat r_{M,ij} = c_i^* z_{ij} = d_i^{-1}\left(1 - h\left(\frac{\tilde f_i\lambda_i}{a_i}\right)^{-\frac{1}{2}}\right)_+ z_{ij}$, $i = 1, 2, \cdots, p$, $j = 1, 2, \cdots, q$ 写成矩阵形式为:

$$\hat R_M = (I - h(\tilde D'\tilde F\tilde D)^{-\frac{1}{2}}\tilde A^{-\frac{1}{2}})_+(DD)^{-1}D\tilde Z = (I - h(\tilde D'\tilde F\tilde D)^{-\frac{1}{2}}\tilde A^{-\frac{1}{2}})_+(\tilde D'\tilde D)^{-1}\tilde D'Z =$$

$$V'(I - h(X'FX)^{-\frac{1}{2}}A^{-\frac{1}{2}})_+(X'X)^{-1}X'Y = V'\hat B_M$$

所以 $\hat B_M = (I - h(X'FX)^{-\frac{1}{2}}A^{-\frac{1}{2}})_+\hat B_{LS}$, h满足

$$\rho = \sum_{i=1}^{p}\lambda_i\tilde f_i\left(\sum_{j=1}^{q} r_{ij}^2\right)^* = \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}\lambda_i^{-1}(h^{-1}(a_i\tilde f_i\lambda_i)^{\frac{1}{2}} - \lambda_i\tilde f_i)_+ =$$

$$\mathrm{tr}W\cdot tr\{(D'D)^{-1}(h^{-1}\tilde A^{\frac{1}{2}}(\tilde D'\tilde F\tilde D)^{\frac{1}{2}} - \tilde D'\tilde F\tilde D)_+\} =$$

$$\mathrm{tr}W\cdot tr\{(X'X)^{-1}(h^{-1}A^{\frac{1}{2}}(X'FX)^{\frac{1}{2}} - X'FX)_+\}$$

下面证明 $\hat B_M$ 为 B的线性 Minimax估计:

$\hat R_M$ 为 R的线性 Minimax估计,由定义 1知 $\sup_{R\in\Theta_O}\mathrm{Etr}(\hat R_M - R)'\tilde A(\hat R_M - R)\le\sup_{R\in\Theta_O}\mathrm{Etr}(\hat R - R)'\tilde A(\hat R - R)$. $\hat R$为 R的任意估计. 而 $\mathrm{Etr}(\hat B - B)'A(\hat B - B) = \mathrm{Etr}(\hat R - R)'\tilde A(\hat R - R)$, $\hat B$为 B的任意估计. $\mathrm{Etr}(\hat B_M - B)'A(\hat B_M - B) = \mathrm{Etr}(\hat R_M - R)'\tilde A(\hat R_M - R)$,所以 $\sup_{B\in\Theta}\mathrm{Etr}(\hat B_M - B)'\tilde A(\hat B_M - B)\le\sup_{B\in\Theta}\mathrm{Etr}(\hat B - B)'\tilde A(\hat B - B)$.

由定义 1知 $\hat B_M$ 为 B的线性 Minimax估计. 由定理 1知 $\hat R_M$ 的 Minimax风险为:

$$\inf_{C\in\sigma}\sup_{R\in\Theta_O}\mathrm{Etr}(C\tilde Z - R)'\tilde A(C\tilde Z - R) = \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj} a_i d_i^{-2}\left(1 - h\left(\frac{\tilde f_i\lambda_i}{a_i}\right)^{-\frac{1}{2}}\right)_+ =$$

$$\mathrm{tr}W\cdot\mathrm{tr}\{(D'D)^{-1}(\tilde A - h\tilde A^{\frac{1}{2}}(\tilde D'\tilde F\tilde D)^{\frac{1}{2}})_+\} =$$

$$\mathrm{tr}W\cdot tr\{(X'X)^{-1}(A - hA^{\frac{1}{2}}(X'FX)^{\frac{1}{2}})_+\}$$

由 $\sup_{B\in\Theta}\mathrm{Etr}(\hat B_M - B)'A(\hat B_M - B) = \sup_{R\in\Theta_O}\mathrm{Etr}(\hat R_M - R)'\tilde A(\hat R_M - R)$,所以线性 Minimax估计的风险为:

$$\inf_{S\in\tau}\sup_{B\in\Theta}\mathrm{Etr}(SY - B)'A(SY - B) = \sup_{B\in\Theta}\mathrm{Etr}(\hat B_M - B)'A(\hat B_M - B) =$$

$\mathrm{tr}W \cdot \mathrm{tr}\{(X'X)^{-1}(A - hA^{\frac{1}{2}}(X'FX)^{\frac{1}{2}})_+\}$.

当设计阵 $X = I$ (此时 $n = p$), $A = I$时,有如下定理:

定理 3 对于多元标准线性模型: $\begin{cases} Y = B + E \\ \mathrm{vec}(E) \sim (0, W \otimes I) \end{cases}$, $B \in \Theta = \{B \mid \mathrm{tr}(B'FB) \le \rho\}$,损失函数为 $L(\hat{B}, B, A) = \mathrm{tr}(\hat{B} - B)'(\hat{B} - B)$. 对于任意的满足 $T \ge 0$和 $(I - T) \ge 0$的 n阶方阵 T,若 $F = (I - T)^2$、$\rho = \mathrm{tr}W\{\mathrm{tr}T - \mathrm{tr}T^2\}$,则 TY为 B的线性 Minimax估计.

证明 对 T进行谱分解 $T = P\Gamma P'$, $\Gamma = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_n)$, $1 \le \gamma_i \le 1$, $i = 1$, $i = 2, \cdots, p$ 则:

$$F = (I - T)^2 = P(1 - \Gamma)^2 P'. \text{所以 } \tilde{f}_i = (1 - \gamma_i)^2, \rho = \sum_{i=1}^{n}\sum_{j=1}^{q} w_{jj}(\gamma_i - \gamma_i^2) = \sum_{i=1}^{n}\sum_{j=1}^{q} w_{jj}(\tilde{f}_i^{\frac{1}{2}} - \tilde{f}_i)$$

由定理 2知 $h = 1$, $SY = (1 - F^{\frac{1}{2}})_+ \hat{B}_{LS} = TY$为 B的线性 Minimax估计,极大化风险为: $\mathrm{tr}W \cdot \mathrm{tr}T$.

这个定理说明:对于特征根在 $[0, 1]$之间的任意 n阶对称阵 T, $\hat{B} = TY$使 $\max_{\hat{B} \in \zeta}\{E\mathrm{tr}(\hat{B} - B)'(\hat{B} - B): \mathrm{tr}(B'(I - T)^2 B) \le \mathrm{tr}W(\mathrm{tr}T - \mathrm{tr}T^2)\}$达到最小. 其中 ζ为 B的所有估计组成的类,又因为 $E\mathrm{tr}(TY - B)'(TY - B) = \mathrm{tr}\{B'(T - I)^2 B\} + \mathrm{tr}W \cdot \mathrm{tr}T^2$,所以 Θ 可以写成 $\Theta = \{B: E\mathrm{tr}(TY - B)'(TY - B) \le \mathrm{tr}W \cdot \mathrm{tr}T\}$.

2 多元线性 Minimax估计的性质及特例

记 $\hat{\beta}_M = \mathrm{vec}(\hat{B}_M)$, $\hat{\beta}_{LS} = \mathrm{vec}(\hat{B}_{LS})$. 则 $\hat{\beta}_M = (I \otimes (I - h(X'FX)^{\frac{1}{2}}A^{-\frac{1}{2}})_+)\hat{\beta}_{LS}$. 容易得到如下几条性质:

性质 1 $\hat{\beta}_M = (I \otimes (I - h(X'FX)^{\frac{1}{2}}A^{-\frac{1}{2}})_+)\hat{\beta}_{LS}$为 β的压缩有偏估计,即 $\| E\hat{\beta}_M \| < \| \beta \|$, $E\hat{\beta}_M \ne \beta$.

性质 2 若 $W > 0$, $\hat{\beta}_M$ 是 β的可容许估计.

特例 1 当 $A = X'X$, $F = I$时, $\hat{B}_M = (1 - h)\hat{B}_{LS}$, $h = \dfrac{p\sum_{j=1}^{q} w_{jj}}{\rho + p\sum_{j=1}^{q} w_{jj}}$,此时 $\hat{B}_M$ 也为多元 Stein估计. 其 Minimax风险为: $\sup_{\mathrm{tr}B'X'XB \le \rho} E\mathrm{tr}(X\hat{B}_M - XB)'(X\hat{B}_M - XB) = \dfrac{p\rho\sum_{j=1}^{q} w_{jj}}{\rho + p\sum_{j=1}^{q} w_{jj}}$. 进一步,若 $\rho = p\rho^*$ 则 $\lim_{p\to\infty}\inf_{C}\sup_{\mathrm{tr}B'X'XB \le p\rho^*} p^{-1} E\mathrm{tr}(CY - B)'X'X(CY - B) =$

$$\lim_{p\to\infty}\sup_{\mathrm{tr}B'X'XB \le p\rho^*} p^{-1} E\mathrm{tr}(X\hat{B}_M - XB)'(X\hat{B}_M - XB) = \frac{\rho^* \sum_{j=1}^{q} w_{jj}}{\rho^* + \sum_{j=1}^{q} w_{jj}}.$$

当 $p \to \infty$时, C变为任意的可测矩阵,此时求得 Minimax估计称为渐近 Minimax估计.

特例 2 当 $A = X'X$, $F = U\mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_p^{-1}, f_{p+1}, \cdots, f_n)U'$, f_i为任意非负数, $i = p + 1, \cdots, n$ 则 $X'FX = I$,此种情况下 $\hat{B}_M = (I - h(X'X)^{-\frac{1}{2}})_+ \hat{B}_{LS}$, h满足 $\sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}\lambda_i^{-1}(h^{-1}\lambda_i^{\frac{1}{2}} - 1)_+ = \rho$,此时 $\hat{B}_M$ 也为多元岭估计,其 Minimax风险为: $\sup_{\mathrm{tr}B'B \le \rho} E\mathrm{tr}(X\hat{B}_M - XB)'(X\hat{B}_M - XB) = \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}(1 - h\lambda_i^{-\frac{1}{2}})_+$.

特例 3 当 $A = X'X$, $X'FX = (X'X)^{\delta}$ 时,此时 $\hat{B}_M = (I - h(X'X)^{\frac{\delta-1}{2}})_+ \hat{B}_{LS}$, h 满足 $\sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}\lambda_i^{\delta-1}(\lambda_i^{-\frac{\delta-1}{2}}h^{-1} - 1)_+ = \rho$,此时 $\hat{B}_M$ 也为多元功效岭估计,其 Minimax风险为: $\sup_{\mathrm{tr}B'(X'X)^{\delta}B \le \rho} E\mathrm{tr}(X\hat{B}_M - XB)'(X\hat{B}_M - XB) = \sum_{i=1}^{p}\sum_{j=1}^{q} w_{jj}(1 - h\lambda_i^{-\frac{\delta-1}{2}})_+$.

(下转第 525页)

Optimal sufficient conditions of a class of multiobjective programming about B-preinvex function

ZHANG Qi-mao

(College of Mathematics and Computer Science, Chongqing Normal University, Chongqing 400047, China)

Abstract: Under the assumption of B-preinvex function, this paper studies a class of multi-objective programming problems, and obtains Kuhn-Tunker sufficient optimality conditions

Key words: multiobject programming; optimal conditions; B-preinvex function; efficient solution; weakly efficient solution

责任编辑:李翠薇

(上接第 521页)

参考文献:

[1] BLAKER H. Minimax estimation in linear regression under restrictions[J]. Journal of Statistical Planning and Inference, 2000, 90: 35-55

[2] FRANK I E, FRIEDMAN J H. A Statistical view of some chemometrics regression tool(with discussion)[J]. Technometrics, 1993, 35: 109-148

[3] 朱道元. 多元统计分析及 SASS软件[M]. 上海:东南大学出版社, 1998

Minimax estimation in multivariate linear regression model under restrictions

WANG Li-feng[1] ZHU Dao-yuan[2]

(1. Department of Social Science. Nanjing Institute of Railway Technology, Nanjing 210015;

2. Department of Mathematics, Southeast University, Nanjing 210096, China)

Abstract: In this paper, a multivariate Minimax estimation in all linear estimators under the loss function $\mathrm{tr}(\hat{B} - B)' A (\hat{B} - B)$ is derived. We consider its properties. In special case, the multivariate Minimax estimation includes power Ridge regression, Stein's estimator and so on.

Key words: multivariate linear regression model; multivariate linear Minimax estimation; multivariate Ridge regression estimation

责任编辑:罗泽举