应用数学和力学编委会编
重庆出版社出版

多元广义岭估计及K值选取的Q(c)准则*

陈世基　　曾志斌

（福建师范大学数学系，1991年10月18日收到）

摘　要

当自变量间存在复共线性时，最小二乘估计就表现出不稳定并可能导致错误的结果。本文采用广义岭估计$\beta(K)$来估计多元线性模型的回归系数$\beta=\mathrm{vec}(B)$，通过岭参数K值的选取，可使广义岭估计的均方误差MSE小于最小二乘估计的MSE。指出了广义岭估计中根据MSE准则选取K值存在的主要缺陷，采用了一种选取K值的新准则$Q(c)$，它包含MSE准则和最小二乘LS准则作为特例，从理论上证明和讨论了$Q(c)$准则的优良性，阐明了c值的统计含义，并给出了确定c值的方法。

关键词　最小二乘估计　广义岭估计　均方误差

一、引　言

我们考虑q个因变量的n个观测值Y，假设它满足标准的多元线性回归模型

$$Y=XB+\varepsilon \tag{1.1}$$

其中　X为列满秩的$n\times p$的已知矩阵，B为待估的$p\times q$的未知系数矩阵，记$\varepsilon=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_q)$，通常的假设是

$$E(\varepsilon_j)=0,\ \mathrm{cov}(\varepsilon_j,\varepsilon_l)=v_{jl}I_n \qquad (j,l=1,2,\cdots,q)$$

这里，$V=(v_{jl})>0$表示$q\times q$的正定矩阵。

当$q=1$时，模型(1.1)为一元线性回归模型

$$y=X\beta_I+e,\ E(e)=0,\ \mathrm{cov}(e)=\sigma^2I_n \tag{1.2}$$

这里y为$n\times1$的观测向量，β_I为$p\times1$的未知参数向量，e为$n\times1$的随机误差向量，$\sigma^2>0$为误差方差。

容易得到，模型(1.1)的典则形式为

$$Y=ZA+\varepsilon \tag{1.3}$$

这里　$Z=XQ$，$A=Q'B$，Q为$p\times p$的正交矩阵，它使得

$$Q'X'XQ=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p),$$

---

* 福建省自然科学基金资助项目。
林宗池推荐。

$\lambda_i(i=1,2,\cdots,p)$为矩阵$X'X$的特征根，并称$\Lambda$为典则参数矩阵.

模型(1.1)和(1.3)可分别化成如下的一元线性回归模型

$$\left.\begin{aligned}&\mathrm{vec}(Y)=(I_q\otimes X)\mathrm{vec}(B)+\mathrm{vec}(\varepsilon),\\&E(\mathrm{vec}(\varepsilon))=0\\&\mathrm{cov}(\mathrm{vec}(\varepsilon))=V\otimes I_n\end{aligned}\right\}\tag{1.4}$$

$$\left.\begin{aligned}&\mathrm{vec}(Y)=(I_q\otimes Z)\mathrm{vec}(A)+\mathrm{vec}(\varepsilon)\\&E(\mathrm{vec}(\varepsilon))=0\\&\mathrm{cov}(\mathrm{vec}(\varepsilon))=V\otimes I_n\end{aligned}\right\}\tag{1.5}$$

令$\beta=\mathrm{vec}(B)$，$\alpha=\mathrm{vec}(A)$，则$\alpha=(I_q\otimes Q')\beta$，称$\alpha$为典则参数向量.

很容易地可以得到，B和A的LS估计分别为

$$B^*=(X'X)^{-1}X'Y\tag{1.6}$$

和
$$A^*=(Z'Z)^{-1}Z'Y=\Lambda^{-1}Z'Y\tag{1.7}$$

β和α相应的LS估计分别为

$$\beta^*=[I_q\otimes(X'X)^{-1}X']\mathrm{vec}(Y)\tag{1.8}$$

和
$$\alpha^*=(I_q\otimes\Lambda^{-1}Z')\mathrm{vec}(Y)\tag{1.9}$$

且$\beta^*=(I_q\otimes Q')\alpha^*$，$E(\beta^*)=\beta$，$\mathrm{cov}(\beta^*)=V\otimes(X'X)^{-1}$.

通过直接计算，β^*的均方误差为

$$\mathrm{MSE}(\beta^*)=(\mathrm{tr}V)\sum_{i=1}^{p}\lambda_i^{-1}\tag{1.10}$$

这里$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$为$X'X$的特征根，因为$X'X>0$，所以$\lambda_i>0(i=1,2,\cdots,p)$.

从(1.10)式可知，若$X'X$至少有一特征根很小，即接近于零，则$\mathrm{MSE}(\beta^*)$就很大，此时虽然LS估计β^*的方差在线性无偏估计类中最小，但其值很大，使得估计量表现出相当的不稳定. [1]提出，$X'X$有多少个特征根接近于零，回归自变量之间就存在多少个近似的线性关系，此时回归自变量之间存在复共线性，并称设计矩阵X呈"病态". 为克服LS估计存在的缺陷，本文将讨论模型(1.1)回归系数的广义岭估计，得到广义岭估计的主要性质，并应用选取K值的$Q(c)$准则对广义岭估计作了进一步的探讨.

二、广义岭估计及其性质

因为对一切的β都有$E\|\beta^*\|^2>\|\beta\|^2$，所以从平均的角度来看LS估计$\beta^*$总是偏长，因此有必要对$\beta^*$进行压缩.

模型(1.1)回归系数B的广义岭估计定义为

$$B^*=(X'X+QKQ')^{-1}X'Y\tag{2.1}$$

其中 Q为正交矩阵，使得$Q'X'XQ=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p)$，$K=\mathrm{diag}(k_1,k_2,\cdots,k_p)$ $(k_i\geqslant 0,\ i=1,2,\cdots,p)$，称$k_i(i=1,2,\cdots,p)$为广义岭参数. 当$K=kI_p$时，则(2.1)式化成$B^*(k)=(X'X+kI_p)^{-1}X'Y$，即为$B$的狭义岭估计；当$K=0$时，(2.1)即为$B$的LS估计；若$q=1$，则(2.1)即为模型(1.2)回归系数$\beta_I$的广义岭估计.

记$\beta^*(K)=\mathrm{vec}(B^*(K))$. 为了叙述上的方便，也把$\beta^*(K)$称为$\beta=\mathrm{vec}(B)$的广义岭估计. 从(2.1)式立即得到

$$\beta^*(K)=[I_q\otimes(X'X+QKQ')^{-1}](I_q\otimes X'X)\beta^* \tag{2.1$'$}$$

定理2.1 当$K>0$时，$\beta^*(K)$ 是β的压缩型有偏估计。

证明 因为 $\beta^*(K)=[I_q\otimes(X'X+QKQ')^{-1}](I_q\otimes X'X)\beta^*$

$$=(I_q\otimes Q)[I_q\otimes(\Lambda+K)^{-1}](I_q\otimes\Lambda)(I_q\otimes Q')\beta^*$$

所以 $$\|\beta^*(K)\|^2=\|[I_q\otimes(\Lambda+K)^{-1}\Lambda](I_q\otimes Q')\beta^*\|^2$$
$$<\|(I_q\otimes Q)\beta^*\|^2=\|\beta^*\|^2$$

可见，$\beta^*(K)$是LS估计β^*向原点压缩而得到的。

又 $$E(\beta^*(K))=[I_q\otimes(X'X+QKQ')^{-1}](I_q\otimes X'X)\beta\neq\beta,$$

否则必得到

$$I_q\otimes X'X=I_q\otimes(X'X+QKQ')$$

这说明$K=0$，与$K>0$矛盾，因此$\beta^*(K)$是β 的压缩型有偏估计。

根据(2.1)$'$式，α的广义岭估计为

$$\alpha^*(K)=[I_q\otimes(\Lambda+K)^{-1}\Lambda]\alpha^* \tag{2.2}$$

且 $$\beta^*(K)=(I_q\otimes Q)\alpha^*(K) \tag{2.3}$$

定理2.2 $\beta^*(K)$的均方误差为

$$\text{MSE}(\beta^*(K))=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\lambda_i v_{jj}+k_i^2\alpha_{ij}^2}{(\lambda_i+k_i)^2} \tag{2.4}$$

这里α_{ij}为$A=Q'B$ 的元素，v_{jj}为V主对角线上的元素。

证明 根据(2.3)式，$\beta^*(K)$的均方误差为

$$\text{MSE}(\beta^*(K))=\text{MSE}(\alpha^*(K))=\text{tr}\ \text{cov}(\alpha^*(K))+\|E\alpha^*(K)-\alpha\|^2$$

因为$E\alpha^*(K)=[I_q\otimes(\Lambda+K)^{-1}\Lambda]\alpha$

$$\text{cov}(\alpha^*(K))=[I_q\otimes(\Lambda+K)^{-1}\Lambda](V\otimes\Lambda^{-1})[I_q\otimes\Lambda(\Lambda+K)^{-1}]$$
$$=V\otimes(\Lambda+K)^{-1}\Lambda(\Lambda+K)^{-1}$$

所以$\text{tr}\ \text{cov}(\alpha^*(K))=\text{tr}[V\otimes(\Lambda+K)^{-1}\Lambda(\Lambda+K)^{-1}]$

$$=\left(\sum_{i=1}^{q}v_{jj}\right)\left[\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k_i)^2}\right]$$

$$\|E\alpha^*(K)-\alpha\|^2=\alpha'[I_q\otimes(\Lambda+K)^{-2}K^2]\alpha$$
$$=\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\frac{k_i\alpha_{ij}}{\lambda_i+k_i}\right)^2$$

于是得到(2.4)式。

定理2.3 若 $0<k_i<k_i^*(i=1,2,\cdots,p)$，则

$$\text{MSE}(\beta^*(K))<\text{MSE}(\beta^*)$$

这里 $$k_i^*=(\text{tr}V)\Big/\sum_{j=1}^{q}\alpha_{ij}^2\qquad(i=1,2,\cdots,p) \tag{2.5}$$

证明 由定理2.2有

$$\frac{\partial\text{MSE}(\beta^*(K))}{\partial k_i}=-2\lambda_i(\lambda_i+k_i)^{-3}\left(\text{tr}V-k_i\sum_{j=1}^{q}\alpha_{ij}^2\right)$$

令 $k_i^* = (\mathrm{tr}V)\Big/\sum_{j=1}^{q} \alpha_{ij}^2 \quad (i=1,2,\cdots,p)$，得

$$\frac{\partial \mathrm{MSE}(\beta^*(K))}{\partial k_i} \begin{cases} >0 & (\text{当}\ k_i > k_i^*) \\ =0 & (\text{当}\ k_i = k_i^*) \\ <0 & (\text{当}\ k_i < k_i^*) \end{cases}$$

所以当 $0<k_i<k_i^*$，必有

$$\mathrm{MSE}(\beta^*(K)) < \mathrm{MSE}(\beta^*(0)) = \mathrm{MSE}(\beta^*).$$

记$K^*=\mathrm{diag}(k_1^*, k_2^*, \cdots, k_p^*)$，则从定理2.3的证明过程可知，$\mathrm{MSE}(\beta^*(K))$在$K=K^*$达到最小值，其最小值为

$$\min_{K\geqslant 0} \mathrm{MSE}(\beta^*(K)) = \mathrm{MSE}(\beta^*(K^*)) = (\mathrm{tr}V)\sum_{i=1}^{p} \frac{\alpha_{i\cdot}^2}{\lambda_i \alpha_{i\cdot}^2 + \mathrm{tr}V}.$$

其中 $\alpha_{i\cdot}^2 = \sum_{j=1}^{q} \alpha_{ij}^2 \quad (i=1,2,\cdots,p)$.

定理2.4 $\beta^*(K)$是β的可容许估计.

证明 令$C = I_q \otimes (X'X+QKQ')^{-1}X'X$,

$$W = (I_q\otimes X')(V^{-1}\otimes I_n)(I_q\otimes X) = V^{-1}\otimes X'X,$$

则有$$\begin{aligned} CW^{-1} - CW^{-1}C' &= [I_q\otimes(X'X+QKQ')^{-1}X'X][V\otimes(X'X)^{-1}] \\ &\quad - [I_q\otimes(X'X+QKQ')^{-1}X'X][V\otimes(X'X)^{-1}][I_q\otimes X'X(X'X+QKQ')^{-1}] \\ &= V\otimes(X'X+QKQ')^{-1} - V\otimes(X'X+QKQ')^{-1}X'X(X'X+QKQ')^{-1} \\ &= V\otimes(X'X+QKQ')^{-1}QKQ'(X'X+QKQ')^{-1} \end{aligned}$$

因为$V>0$，$K\geqslant 0$，且Q为正交矩阵，使得

$$Q'X'XQ = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p),$$

所以 $$(X'X+QKQ')^{-1}QKQ'(X'X+QKQ')^{-1} \geqslant 0$$

于是 $$CW^{-1} - CW^{-1}C' = V\otimes(X'X+QKQ')^{-1}QKQ'(X'X+QKQ')^{-1} \geqslant 0.$$

根据[2]的推论1.7.2，得知$\beta^*(K) = C\beta^*$为β的可容许估计.

定理2.3说明了$\beta^*(K)$比β^*有较小的均方误差. 但(2.5)式与未知参数α和V有关，所以我们必须通过观测到的数据来确定它们，这样就导致非线性估计.

因为$V^* = (n-p)^{-1}Y'[I - X(X'X)^{-1}X']Y$为$V$的无偏估计，$\alpha^* = (I_q\otimes \Lambda^{-1}Z')\mathrm{vec}(Y)$为$\alpha = \mathrm{vec}(A)$的LS估计，所以一种自然的作法是在(2.5)式中以V^*及α_{ij}^*代替V及α_{ij}得到k_i^*的估计.

$$\hat{k}_i^* = \mathrm{tr}(V^*)\Big/\sum_{j=1}^{q} \alpha_{ij}^{*2} \qquad (i=1,2,\cdots,p) \tag{2.6}$$

也可采用Hemmerle-Brantle方法来估计k_i^*，即

$$\hat{k}_i^* = \frac{\mathrm{tr}(V^*)}{\sum_{j=1}^{q} \alpha_{ij}^{*2} - \lambda_i^{-1}\mathrm{tr}(V^*)} \qquad (i=1,2,\cdots,p) \tag{2.7}$$

当$\sum_{j=1}^{q} \alpha_{ij}^{*2} - \lambda_i^{-1}\mathrm{tr}(V^*) \leqslant 0$时，取$\hat{k}_i^* = +\infty (i=1,2,\cdots,p)$.

事实上，$\sum_{j=1}^{q} \alpha_{ij}^2$的无偏估计为

$$\sum_{j=1}^{q} \hat{\alpha}_{ij}^2 = \sum_{j=1}^{q} \alpha_{ij}^{*2} - \lambda_i^{-1}\mathrm{tr}(V^*)$$

这是因为

$$E(\alpha_{ij}^{*2}) = \mathrm{Var}(\alpha_{ij}^*) + (E\alpha_{ij}^*)^2 = \lambda_i^{-1} v_{jj} + \alpha_{ij}^2,$$

$$E(v_{jj}^*) = v_{jj} \qquad (i=1,2,\cdots,p, j=1,2,\cdots,q)$$

所以 $E(\alpha_{ij}^{*2} - \lambda_i^{-1} v_{jj}^*) = \alpha_{ij}^2,$

因此 $E\left(\sum_{j=1}^{q} \alpha_{ij}^{*2} - \lambda_i^{-1}\mathrm{tr}(V^*)\right) = \sum_{j=1}^{q} \alpha_{pj}^2 \quad (i=1,2,\cdots,p)$

最后，在(2.5)式中分别以V和$\sum_{j=1}^{q} \alpha_{ij}^2$的无偏估计$V^*$和$\sum_{j=1}^{q} \hat{\alpha}_{ij}^2$代替$V$和$\sum_{j=1}^{q} \alpha_{ij}^2$就得到 (2.7)式.

三、选取K值的Q(c)准则

1. 广义岭估计均方残差的性质和根据MSE准则选取K值存在的主要缺陷

比较LS估计β^*和广义岭估计$\beta^*(K)(K>0)$，我们可以发现它们有两个主要区别

第一，β^*不含未知参数，它可直接由数据估计出来，但$\beta^*(K)$的表达式(2.1)中含有参数K，它需要通过数据来估计.

第二，两者的优化准则不同，β^*是基于LS准则（即，最小二乘准则）得到的最佳线性无偏估计，而$\beta^*(K)$中选取K值通常总是基于MSE的减少(MSE准则是要使$\mathrm{MSE}(\beta^*(K))$达到最小)．$\beta^*(K)$是有偏估计，它在减少MSE的同时，势必要增大对模型拟合的均方残差.

定理3.1 对$\beta^*(K)$的均方残差

$$\mathrm{MSR}(\beta^*(K)) = E\|\mathrm{vec}(Y) - (I_q \otimes X)\beta^*(K)\|^2$$

必成立着

$$\mathrm{MSR}(\beta^*(K)) = \mathrm{MSR}(\beta^*) + \Delta\mathrm{MSR}(\beta^*(K))$$

这里 $\mathrm{MSR}(\beta^*) = E\|\mathrm{vec}(Y) - (I_q \otimes X)\beta^*\|^2 = (n-p)\mathrm{tr}V$ 为LS估计β^*的均方残差，$\Delta\mathrm{MSR}(\beta^*(K)) = E\|(I_q \otimes X)\beta^*(K) - (I_q \otimes X)\beta^*\|^2$为$\beta^*(K)$对$\beta^*$的修正所引起的MSR增量，且

$$\Delta\mathrm{MSR}(\beta^*(K)) = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(v_{jj} + \lambda_i \alpha_{ij}^2) k_i^2}{(\lambda_i + k_i)^2} \tag{3.1}$$

证明 $\mathrm{MSR}(\beta^*(K)) = E\|\mathrm{vec}(Y) - (I_q \otimes X)\beta^*(K)\|^2$

$$=E\|\mathrm{vec}(Y)-(I_q\otimes X)\beta^*+(I_q\otimes X)\beta^*-(I_q\otimes X)\beta^*(K)\|^2$$
$$=E\|\mathrm{vec}(Y)-(I_q\otimes X)\beta^*\|^2+E\|(I_q\otimes X)\beta^*-(I_q\otimes X)\beta^*(K)\|^2$$
$$=\mathrm{MSR}(\beta^*)+\Delta\mathrm{MSR}(\beta^*(K)),$$

这里　$\min\limits_{K\geqslant 0}\mathrm{MSR}(\beta^*(K))=\mathrm{MSR}(\beta^*(0))=\mathrm{MSR}(\beta^*)=(n-p)\mathrm{tr}V.$

根据[2]的定理2.3.1，有

$$\Delta\mathrm{MSR}(\beta^*(K))=E\|(I_q\otimes X)\beta^*(K)-(I_q\otimes X)\beta^*\|^2$$
$$=E\|(I_q\otimes Z)\alpha^*(K)-(I_q\otimes Z)\alpha^*\|^2$$
$$=E\|[(I_q\otimes Z)(I_q\otimes(\Lambda+K)^{-1}\Lambda)-(I_q\otimes Z)]\alpha^*\|^2$$
$$=E\|(I_q\otimes Z)[I_q\otimes(\Lambda+K)^{-1}K]\alpha^*\|^2$$
$$=\mathrm{tr}\{[I_q\otimes(\Lambda+K)^{-2}K^2\Lambda]\mathrm{cov}(\alpha^*)\}+\alpha'[I_q\otimes(\Lambda+K)^{-2}K^2\Lambda]\alpha$$
$$=\mathrm{tr}\{V\otimes(\Lambda+K)^{-2}K^2\}+\alpha'[I_q\otimes(\Lambda+K)^{-2}K^2\Lambda]\alpha$$
$$=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{k_i^2 v_{jj}}{(\lambda_i+k_i)^2}+\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\lambda_i k_i^2\alpha_{ij}^2}{(\lambda_i+k_i)^2}$$
$$=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{(v_{jj}-\lambda_i\alpha_{ij}^2)k_i^2}{(\lambda_i+k_i)^2}.$$

这样，要讨论$\beta^*(K)$的MSR，只需讨论$\Delta\mathrm{MSR}(\beta^*(K))$即可.

当$k=0$时，

$$\Delta\mathrm{MSR}(\beta^*(K))=0,$$

当$k_i\to+\infty(i=1,2,\cdots,p)$时，

$$\Delta\mathrm{MSR}(\beta^*(K))\to p\mathrm{tr}V+\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i\alpha_{ij}^2$$

当$K>0$时，

$$\frac{\partial[\Delta\mathrm{MSR}(\beta^*(K)]}{\partial k_i}=\frac{2\lambda_i(v_{jj}+\lambda_i\alpha_{ij}^2)k_i}{(\lambda_i+k_i)^3}>0\quad(i=1,2,\cdots,p)$$

这说明$\Delta\mathrm{MSR}(\beta^*(K))$关于$k_i$单调递增$(i=1,2,\cdots,p)$.

对模型(1.4)中$\beta=\mathrm{vec}(B)$的广义岭估计类

$$\mathscr{G}=\{\beta^*(K):K=\mathrm{diag}(k_1,k_2,\cdots,k_p)\geqslant 0\}$$

关于均方误差和均方残差有如下的主要结果：

$$\max_{K>0}\mathrm{MSE}(\beta^*(K))=\max\left\{(\mathrm{tr}V)\sum_{i=1}^{p}\lambda_i^{-1},\ \sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_{ij}^2\right\}$$

$$\min_{K\geqslant 0}\mathrm{MSE}(\beta^*(K))=(\mathrm{tr}V)\sum_{i=1}^{p}\frac{\alpha_{i\cdot}^2}{\lambda_i\alpha_{i\cdot}^2+\mathrm{tr}V},$$

$$\max_{K\geqslant 0}\Delta\mathrm{MSR}(\beta^*(K))=p\mathrm{tr}V+\sum_{i=1}^{p}\lambda_i\alpha_{i\cdot}^2,$$

$\mathrm{MSR}(\beta^*(K))$和$\Delta\mathrm{MSR}(\beta^*(K))$随$k_i$变化的情况大致如表1.

表 1

k_i	0	$0<k_i<(\mathrm{tr}V)/a_{i\cdot}^2$	$(\mathrm{tr}V)/a_{i\cdot}^2$	$(\mathrm{tr}V)/a_{i\cdot}^2<k_i<+\infty$	$+\infty$
$\mathrm{MSE}(\beta^*(K))$	$(\mathrm{tr}V)\sum_{i=1}^{p}\lambda_i^{-1}$	↓	minimum	↑	$\sum_{i=1}^{p}\sum_{j=1}^{q}a_{ij}^2$
$\Delta\mathrm{MSR}(\beta^*(K))$	0	↑	$(\mathrm{tr}V)^2\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{v_{jj}+\lambda_i a_{ij}^2}{(\lambda_i a_{i\cdot}^2+\mathrm{tr}V)^2}$	↑	$p\mathrm{tr}V+\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_i a_{ij}^2$

综上可见，广义岭估计$\beta^*(K)$在改进LS估计使得MSE减小的同时却引起 ΔMSR 的增大，导致模型的拟合变坏．MSE的最大改进可由$(\mathrm{tr}V)\sum\limits_{i=1}^{p}\lambda_i^{-1}$降低到$(\mathrm{tr}V)\cdot\sum\limits_{i=1}^{p}\frac{1}{\lambda_i+(\mathrm{tr}V)/a_{i\cdot}^2}$，但与此同时，$\Delta$MSR却由零增大到$(\mathrm{tr}V)^2\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{q}\frac{v_{jj}+\lambda_i a_{ij}^2}{(\lambda_i a_{i\cdot}^2+\mathrm{tr}V)^2}$．

根据MSE准则选取K值有如下的两点缺陷：

第一，当模型(1.1)的设计矩阵呈“病态”时，才有必要对LS估计作修正，因此K值的选取应依赖于设计矩阵的“病态”程度．但MSE准则对各种的“病态”情况均同等看待，这不够恰当．

第二，当$K=\mathrm{diag}(k_1,k_2,\cdots,k_p)$时，根据$\mathrm{MSR}(\beta^*(K))$和$\Delta\mathrm{MSR}(\beta^*(K))$的性质，$\beta^*(K)$在使MSE减小的同时，必使MSR增大．MSE准则要求$\mathrm{MSE}(\beta^*(K))$达到最小，但此时$\mathrm{MSR}(\beta^*(K))$较大，使得$\beta^*(K)$对模型(1.4)的拟合变坏．

2. Q(c)准则及其优良性

基于以上的分析，我们引入如下的准则．

$Q(c)$准则 在广义岭估计类$\mathscr{G}$中，选取K值使得

$$Q(c)=c\mathrm{MSE}(\beta^*(K))+(1-c)\Delta\mathrm{MSR}(\beta^*(K)) \tag{3.2}$$

为最小，其中$c\in[0,1]$为给定的常数．

对于不同的c，(3.2)给出了不同的准则，所以$Q(c)$准则实际上是一族准则．LS准则和MSE准则分别为$c=0$和$c=1$时$Q(c)$准则的特例．

由(2.4)、(3.1)二式得

$$Q(c)=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\{[c+(1-c)\lambda_i]a_{ij}^2+(1-c)v_{jj}\}k_i^2}{(\lambda_i+k_i)^2}+c\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\lambda_i v_{jj}}{(\lambda_i+k_i)^2}$$

$$\frac{\partial Q(c)}{\partial k_i}=\sum_{j=1}^{q}\frac{2\lambda_i\{[c+(1-c)\lambda_j]a_{ij}^2+(1-c)v_{jj}\}k_i-2c\lambda_i v_{jj}}{(\lambda_i+k_i)^2}.$$

令$\frac{\partial Q(c)}{\partial k_i}=0$，则得$Q(c)$准则下的唯一$K$值为

$$k_i(Q(c))=\frac{c\mathrm{tr}V}{[c+(1-c)\lambda_i]a_{i\cdot}^2+(1-c)\mathrm{tr}V}\quad(c\in[0,1]\ i=1,2,\cdots,p).$$

又 $$\frac{dk_i(Q(c))}{dc}=\frac{(\lambda_i\alpha_{i.}^2+\mathrm{tr}V)\mathrm{tr}V}{\{[c+(1-c)\lambda_i]\alpha_{i.}^2+(1-c)\mathrm{tr}V\}^2}>0,$$

$$(c\in[0,1],\ i=1,2,\cdots,p)$$

所以 $k_i(Q(c))$ 关于 c 严格单调上升．当 $c=0$ 时，$k_i(Q(c))=0$；当 $c=1$ 时，$k_i(Q(c))=(\mathrm{tr}V)/\alpha_{i.}^2=k_i^*$.

引理3.1 对线性回归模型(1.1)或(1.4)，$\forall c\in[0,1]$，由 $Q(c)$ 准则确定的 K 满足

$$0\leqslant k_i\leqslant(\mathrm{tr}V)/\alpha_{i.}^2.\qquad(i=1,2,\cdots,p)$$

定理3.2 对 $K=\mathrm{diag}(k_1,k_2,\cdots,k_p)$，令

$$\mathscr{G}=\{\beta^*(K)=[I_q\otimes(X'X+QKQ')^{-1}](I_q\otimes X'X)\beta^*:K\geqslant0\}$$

$$\mathscr{G}_1=\{\beta^*(K):0\leqslant k_i\leqslant(\mathrm{tr}V)/\alpha_{i.}^2,\ i=1,\ 2,\ \cdots,\ p\}$$

$$\mathscr{G}_{Q(c)}=\{\beta^*(K):K\geqslant0,\ K\text{依}Q(c)\text{准则选取}\}$$

则 (1) $\mathscr{G}_{Q(c)}\subset\mathscr{G}_1$,

(2) $\forall\beta^*(K)\in\mathscr{G}$，必存在 $\beta^*(K_0)\in\mathscr{G}_{Q(c)}$ 使得

$$\mathrm{MSR}(\beta^*(K_0))\leqslant\mathrm{MSE}(\beta^*(K))\tag{3.3}$$

$$\Delta\mathrm{MSR}(\beta^*(K_0)\leqslant\Delta\mathrm{MSR}(\beta^*(K))\tag{3.4}$$

证明 (1)由引理3.1即得

(2) 记 $K^*=\mathrm{diag}(k_1^*,k_2^*,\cdots,k_p^*)$，$k_i^*=(\mathrm{tr}V)/\alpha_{i.}^2$．$(i=1,2,\cdots,p)$，则 $\beta^*(K^*)\in\mathscr{G}_{Q(c)}$，且 $\mathrm{MSE}(\beta^*(K^*))=\min\limits_{\beta^*(K)\in\mathscr{G}}\mathrm{MSE}(\beta^*(K))$

令 $$\mathscr{G}^*=\{\beta(K)\in\mathscr{G}:0<\Delta\mathrm{MSR}(\beta^*(K))<\Delta\mathrm{MSR}(\beta^*(K^*))\}$$

对 $\forall\beta^*(K)\in\mathscr{G}$.

(i) 若 $\beta^*(K)\in\mathscr{G}-\mathscr{G}^*$，当 $\Delta\mathrm{MSR}(\beta^*(K^*))\leqslant\Delta\mathrm{MSR}(\beta^*(K))$ 时，取 $K_0=K^*$，则

$$\mathrm{MSE}(\beta^*(K_0))=\mathrm{MSE}(\beta^*(K^*))\leqslant\mathrm{MSE}(\beta^*(K))$$

$$\Delta\mathrm{MSR}(\beta^*(K_0))=\Delta\mathrm{MSR}(\beta^*(K^*))\leqslant\Delta\mathrm{MSR}(\beta^*(K)).$$

当 $\Delta\mathrm{MSR}(\beta^*(K))=0$ 时，取 $K_0=0$，由 $\Delta\mathrm{MSR}(\beta^*(K))=0$ 得 $K=0$，于是

$$\mathrm{MSE}(\beta^*(K_0))=\mathrm{MSE}(\beta^*(K))$$

$$\Delta\mathrm{MSR}(\beta^*(K_0))=\Delta\mathrm{MSR}(\beta^*(K)).$$

(ii) 若 $\beta^*(K)\in\mathscr{G}^*$，因为

$$\Delta\mathrm{MSR}(\beta^*(K))=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{(v_{jj}+\lambda_i\alpha_{ij}^2)k_i^2}{(\lambda_i+k_i)^2}$$

对 k_i 连续，

$$k_i(Q(c))=\frac{c\mathrm{tr}V}{[c+(1-c)\lambda_i]\alpha_{i.}^2+(1-c)\mathrm{tr}V}$$

关于 c 于[0,1]上连续 $(i=1,2,\cdots,p)$.

所以 $\Delta\mathrm{MSR}[\beta^*(K(Q(c)))]$ 关于 c 于[0,1]上连续.

又 $\Delta\mathrm{MSR}[\beta^*(K(Q(0)))]=0$，$\Delta\mathrm{MSR}[\beta^*(K(Q(1)))]=\Delta\mathrm{MSR}[\beta^*(K)]$.

根据连续函数的介值定理，必存在 $c_0\in(0,1)$ 使得

$$\Delta\mathrm{MSR}[\beta^*(K(Q(c_0)))]=\Delta\mathrm{MSR}[\beta^*(K)]$$

即 $K_0=K(Q(c_0))$ 满足(3.4)式.

由 $\beta^*(K_0)$ 的含义，有

$$c_0\mathrm{MSE}[\beta^*(K_0)]+(1-c_0)\Delta\mathrm{MSR}[\beta^*(K_0)]$$
$$\leqslant c_0\mathrm{MSE}[\beta^*(K)]+(1-c_0)\Delta\mathrm{MSR}[\beta^*(K)]$$

考虑到 $\Delta\mathrm{MSR}[\beta^*(K_0)]=\Delta\mathrm{MSR}[\beta^*(K)]$

由此有 $\mathrm{MSE}(\beta^*(K_0))\leqslant\mathrm{MSE}(\beta^*(K))$

因此(3.3)式成立。

依定理3.2，$\mathscr{G}-\mathscr{G}_{Q(c)}$中的岭估计是不足取的，因为在$\mathscr{G}_{Q(c)}$中有MSE和ΔMSR 都比它小的$\beta^*(K)$存在。

虽然当模型(1.1)的设计矩阵呈现“病态”时，$\beta=\mathrm{vec}(B)$的LS估计将有较大的MSE，但它毕竟是β的最佳线性无偏估计，并没有完全失去它的优良性。根据“病态”的不同程度，LS估计还有不同程度的参考价值。$Q(c)$准则通过c值在[0,1]中的选取在一定程度上考虑了“病态”严重程度的因素。利用了这种参考价值，在这种意义下，对多元广义岭估计来说，$Q(c)$准则优于MSE准则。

3. 关于参数 c 及 K 的确定

（1） 因为$Q(c)$准则中c值的大小，反映了$\beta^*(K)$的MSE和$\beta^*(K)$对模型拟合好坏这两者的重视程度，所以在实用中，参数c可根据设计矩阵的“病态”严重程度来确定。例如[1]中指出方阵的条件数$\mathrm{cond}(X'X)=\frac{\lambda_{\max}}{\lambda_{\min}}$，直观上度量了$X'X$的特征根散布程度，所以可用来判断复共线性是否存在及复共线性的严重程度。在实际应用中，若$0<\mathrm{cond}(X'X)<100$，则认为没有复共线性；若$100\leqslant\mathrm{cond}(X'X)\leqslant1000$，则认为存在中等程度或较强的复共线性；若$\mathrm{cond}(X'X)>1000$，则认为存在严重的复共线性。于是可取$c=1-(\mathrm{cond}(X'X))^{-1}=(\lambda_{\max}-\lambda_{\min})/\lambda_{\max}$。

又如，可根据最大熵原理来确定$K(Q(c))$，即假设c服从[0,1]上的均匀分布，这样一来，我们求得

$$k_i=\int_0^1 k_i(Q(c))dc$$
$$=\frac{\mathrm{tr}V}{\alpha_{i\cdot}^2-\lambda_i\alpha_{i\cdot}^2-\mathrm{tr}V}\left[1-\frac{\lambda_i\alpha_{i\cdot}^2+\mathrm{tr}V}{\alpha_{i\cdot}^2-\lambda_i\alpha_{i\cdot}^2-\mathrm{tr}V}\ln\frac{\alpha_{i\cdot}^2}{\lambda_i\alpha_{i\cdot}^2+\mathrm{tr}V}\right]\quad(i=1,2,\cdots,p)$$

（2） 在理论上得到k_i对未知参数的依赖关系的函数形式之后，为了实际中的应用，我们必须由模型的数据集将未知参数的函数$k_i(\alpha_{i\cdot}^2,V)$估计出来，简单又自然的作法是取$\hat{k}_i=k_i(\alpha_{i\cdot}^{*2},V^*)$，其中$\alpha_{i\cdot}^{*2}$和$V^*$分别为$\alpha_{i\cdot}^2$和$V$的无偏估计，即

$$\alpha_{i\cdot}^{*2}=\sum_{i=1}^{q}\alpha_{ij}^{*2}-\lambda_i^{-1}\mathrm{tr}V^*,\ (\alpha_{ij}^*\text{为}\alpha_{ij}\text{的LS估计})$$

$$V^*=(n-p)^{-1}Y'[I_n-Z(Z'Z)^{-1}Z']Y=(n-p)^{-1}Y'[I_n-Z\Lambda^{-1}Z']Y.$$

我们也可对α^*采用迭代法来求出k_i，这种方法是

$$\hat{k}_i^{(m)}=k_i((\alpha_{i\cdot}^{*(m)})^2,V^*),(\alpha_{i\cdot}^{*(m)})^2=\sum_{j=1}^{q}(\alpha_{ij}^{*(m)})^2.$$
$$(i=1,2,\cdots,p,\ m=1,2,\cdots)$$

其中 $\alpha^{*(m)}=[I_q\otimes(\Lambda+\hat{k}^{(m-1)})^{-1}Z']\mathrm{vec}(Y)\quad(m=1,2,\cdots)$,

$$\alpha^{*(0)}=(I_q\otimes\Lambda^{-1}Z')\mathrm{vec}(Y),$$
$$V^*=(n-p)^{-1}Y'[I_n-Z\Lambda^{-1}Z']Y$$

当然这还得考虑到迭代收敛性和迭代的技巧，至于确定参数c、K的其它方法及各种方法之间的比较，尚有待于进一步探讨。

参 考 文 献

[1] 陈希孺、王松桂，《近代回归分析——原理、方法及应用》，安徽教育出版社（1987）。

[2] 王松桂，《线性模型的理论及其应用》，安徽教育出版社（1987）。

[3] Brown, P. J. and J. V. Zidek, Adaptive multivariate ridge regression, *Ann. Statist.*, 8(1) (1980), 64—74.

[4] Haitovky, Yoel, On multivariate ridge regression, *Biometrika*, 74(3)(1987), 563—570.

[5] Hemmerle, W. J. and T. F. Brantle, Explicit and constrained generalized ridge estimation, *Technometrics*, 20 (1978), 109—120.

[6] Wichem, D. W. and G.A. Churchill, A comparison of ridge estimators, *Technometrics*, 20 (1978), 301—311.

[7] 鲁国斌，广义岭回归估计中关于K值选取的$Q(c)$准则，数理统计与应用概率，4(2)(1989), 159—169.

Generalized Multivariate Ridge Regression Estimate and Criteria Q(c) for Choosing Matrix K

Chen Shi-ji Zeng Zhi-bin

(Department of Mathematics, Fujian Normal University, Fuzhou)

Abstract

When multicollinearity is present in a set of the regression variables, the least square estimate of the regression coefficient tends to be unstable and it may lead to erroneous inference.

In this paper, generalized ridge estimate $\beta^*(K)$ of the regression coefficient $\beta=\mathrm{vec}(B)$ is considered in multivariate linear regression model. The MSE of the above estimate is less than the MSE of the least square estimate by choosing the ridge parameter matrix K. Moreover, it is pointed out that the criterion MSE for choosing matrix K of generalized ridge estimate has several weaknesses. In order to overcome these weaknesses, a new family of criteria $Q(c)$ is adpoted which includes the criterion MSE and criterion LS as its special case. The good properties of criteria $Q(c)$ are proved and discussed from theoretical point of view. The statistical meaning of the scale c is explained and the methods of determining c are also given.

Key words least square estimate, generalized ridge estimate, mean square error