# 病态线性回归模型系数的0－c型岭估计

张建军

（基础部）

【摘要】针对病态线性模型病态的实质，提出回归系数的0－c型岭估计。首先研究它的均方误差的最优化问题，给出了参数的最优值或最优值的一个上界和下界，证明可以选择参数，使它在均方误差的意义下优于LS估计。其次，研究它的偏差，证明存在0－c型岭估计优于LS估计，且比岭估计$\hat{\beta}(k)$具有较小偏差。最后证明它的可容许性，并讨论在实用中它的岭参数选择方法的优点。

关键词：0－c型岭估计；岭估计；平均平方误差；偏差；可容许性

## 1 引言

考虑线性回归模型：

$$Y = X_{\beta} + e, E(e) = 0, \mathrm{COV}(e) = \sigma^2 I_n。\tag{1}$$

其中$Y$为$n\times 1$随机观测向量，$X$为$n\times p$的已知设计矩阵，均已中心标准化，且$R(X)=p\leqslant n$，$\beta$为$p\times 1$未知参数向量，$\sigma^2>0$为未知参数，$e$为$n\times 1$随机误差向量，$I_n$为$n$阶单位矩阵。

为研究有偏估计，在模型中引进典则变量$Z=X\phi$及典则参数$\alpha=\phi'\beta$，模型(1)变为如下典则形式：

$$Y = Z\alpha + e, E(e) = 0, \mathrm{COV}(e) = \sigma^2 I_n。\tag{2}$$

其中$\phi=(\varphi_1,\varphi_2,\cdots,\varphi_p)$，$\varphi_1,\varphi_2,\cdots,\varphi_p$为$S=X'X$的对应于特征根$\lambda_1\geqslant\lambda_2\geqslant\lambda_p>0$的标准正交化特征向量，记$\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p)$，此时$Z'Z=\phi'S\phi=\Lambda$，原始参数$\beta$的LS估计$\hat{\beta}=\phi\hat{\alpha}$。

均方误差在估计和参数的正交变换下保持不变。典则参数$\alpha$的任一估计$\tilde{\alpha}$及相应原始参数$\beta$的估计$\tilde{\beta}=\phi\tilde{\alpha}$有相同均方误差，从而可以将(1)式中参数$\beta$的有偏估计的研究转化到对其典则参数$\alpha$的有偏估计的研究。

当$X$的列向量存在近似线性关系时，$S$接近奇异，模型(1)存在复共线性，称为病态线性回归模型。此时尽管$\hat{\beta}$与$n$组数据拟合得最好，但$\mathrm{MSE}(\hat{\beta})=\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i}$却很大，LS估计的性能变坏。

当线性回归模型(1)为病态时，为减小估计的均方误差，Hoerl和Kennard于1970年提出了著名的有偏估计——岭估计(Ridge Estimate)。对模型(1)，回归系数$\beta$的岭估计定义为

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y。\tag{3}$$

本文1995年12月13日收到.

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.c

其中 $k>0$ 为偏参数，与数据 $Y$ 无关。

与 LS 估计 $\hat{\beta}$ 相比，岭估计是把 $X'X$ 换成 $X'X+kI$ 得到的。当 $X$ 呈病态时，$X'X$ 的特征根至少有一个非常接近于 0，而 $X'X+kI$ 的特征根 $\lambda_1+k,\cdots,\lambda_p+k$ 接近于 0 的程度就会得到改善，从而“打破”原设计阵的复共线性，牺牲无偏性，使 $\hat{\beta}(k)$ 比 $\hat{\beta}$ 具有较小的均方误差，他们证明了，存在 $k>0$，使 $\mathrm{MSE}(\hat{\beta}(k))<\mathrm{MSE}(\hat{\beta})$。

岭估计被提出以来，统计学家们对它作了大量工作，它已成为目前最有影响的一种有偏估计，对岭估计的进一步研究已成为一个应用前景广阔的热门课题。

本文正是这种研究的一个重要方面。基于模型病态的根本原因，考虑到致 $\mathrm{MSE}(\hat{\beta}(k))$ 比 $\mathrm{MSE}(\hat{\beta})$ 为小的主要原因在于在那些 $X'X$ 的非常接近于 0 的特征根（不妨设为 $\lambda_{r+1}\geqslant\lambda_{r+2}\geqslant\cdots\geqslant\lambda_p>0$）上各自加上 $k$。而其他特征根各自加上 $k$ 后对均方 MS 误差减小的贡献不大，且 $\lambda_{r+1},\cdots,\lambda_p$ 各自加上 $k$ 后对均方误差减小的贡献也不同。基于此，本文对模型(2)中参数 $\alpha$ 提出 0－c 型岭估计。

定义 1.1 模型(2)中参数 $\alpha$ 的 0－c 型岭估计定义为

$$\hat{\alpha}_{0-c}(k) = (\Lambda + kI_{0,c})^{-1}Z'Y。\tag{4}$$

其中 $k>0$，$I_{0,c}=\mathrm{diag}(0,0,\cdots,0,c_{r+1},c_{r+2},\cdots,c_p)$，且 $c_i>0$，$i=r+1,\cdots,p$。它们均与数据 $Y$ 无关。

相应定义 $\hat{\beta}_{0-c}(k)=\phi\hat{\alpha}_{0-c}(k)$ 为 $\beta$ 的 0－c 型岭估计。

如令 $c_{r+1}=c_{r+2}=\cdots=c_p=1$，就是文献[5]中的特殊的 0－c 型岭估计——0－k 型岭估计 $\hat{\beta}_{0-k}$。

本文首先研究 $\mathrm{MSE}(\hat{\beta}_{0-c}(k))$ 关于参数的最优化问题，给出了参数的最优值或最优值的一个上界和下界，证明可以选择参数，使 $\hat{\beta}_{0-c}(k)$ 在均方误差的意义下优于 LS 估计 $\hat{\beta}$。其次研究偏差的性质，证明可以选择参数，使 $\hat{\beta}_{0-c}(k)$ 优于 $\hat{\beta}$ 且比 $\hat{\beta}(k)$ 具有较小偏差。最后证明它的可容许性，讨论在实际应用中它的岭参数选择法的优点。

## 2 MSE($\hat{\beta}_{0-c}(k)$)的最优化

$\hat{\beta}_{0-c}(k)$ 写出来就是 $(X'X+k\phi I_{0,c}\phi')^{-1}X'Y$，它具有以下的基本性质。

引理 $\hat{\beta}_{0-c}(k)$ 具有以下基本性质：

(1) $\hat{\beta}_{0-c}(k)=B_{k,c}\hat{\beta}$，这里 $B_{k,c}=(X'X+k\phi I_{0,c}\phi)^{-1}X'X$。这表明 $0-c$ 型岭估计是 LS 估计 $\hat{\beta}$ 的一个线性变换。

(2) $E(\hat{\beta}_{0-c}(k))=B_{k,c}\beta$，只要 $B_{k,c}\neq I$，等价地 $k\neq 0$，$0-c$ 型岭估计就是有偏估计。

(3) 对任意 $k>0$，$c_i>0$，$i=r+1,\cdots,p$，$\|\beta\|>0$，总有

$$\|\hat{\beta}_{0-c}(k)\| < \|\hat{\beta}\|$$

即 0－c 型岭估计是 LS 估计向原点的一种压缩。

(4) $$\|E(\hat{\beta}_{0-c}(k))-\beta\|^2 = k^2\sum_{i=r+1}^{p}\frac{C_i^2}{(\lambda_i+kc_i)^2}d_i^2。\tag{5}$$

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

(5)
$$\mathrm{trCOV}(\hat{\beta}_{0-c}(k)) = \sigma^2\sum_{i=1}^{r}\frac{1}{\lambda_i} + \sigma^2\sum_{i=r+1}^{p}\frac{\lambda_i}{(\lambda_i + kc_i)^2}。\tag{6}$$

证明：只证(3)～(5)：

证(3) 从

$$\begin{aligned}\hat{\alpha}_{0-c}(k) &= (\Lambda + kI_{0,c})^{-1}Z'Y\\ &= (\Lambda + kI_{0,c})^{-1}\Lambda\hat{\alpha}\\ &= [\Lambda^{-1}(\Lambda + kI_{0,c})]^{-1}\hat{\alpha},\end{aligned}$$

得
$$\|\hat{\alpha}_{0c}(k)\| < \|\hat{\alpha}\| \quad 即\ \|\hat{\beta}_{0-c}(k)\| < \|\hat{\beta}\|。$$

证(4) 从
$$\begin{aligned}&\|E(\hat{\alpha}_{0-c}(k)) - \alpha\|^2\\ &= \|(\Lambda + kI_{0,c})^{-1}(\Lambda - \Lambda - kI_{0,c})\alpha\|^2\\ &= \|[k(\Lambda + kI_{0,c})^{-1}I_{0,c}]\alpha\|^2\\ &= k^2\sum_{i=r+1}^{p}\frac{C_i^2}{(\lambda_2 + kc_i)^2}\alpha_i^2,\end{aligned}$$

及
$$\begin{aligned}\|E(\hat{\beta}_{0-c}(k)) - \beta\|^2 &= \|\phi(E(\hat{\alpha}_{0-c}(k)) - \alpha)\|^2\\ &= (E(\hat{\alpha}_{0-c}(k)) - \alpha)'\phi'\phi(E(\hat{\alpha}_{0-c}(k)) - \alpha)\\ &= \|E(\hat{\alpha}_{0-c}(k)) - \alpha\|^2,\end{aligned}$$

即得到证明。

证(5) 事实上，

$$\begin{aligned}\mathrm{COV}(\hat{\alpha}_{0-c}(k)) &= \mathrm{COV}[(\Lambda + kI_{0,c})^{-1}\Lambda\hat{\alpha}]\\ &= [(\Lambda + kI_{0,c})^{-1}\Lambda]\mathrm{COV}(\hat{\alpha})[(\Lambda + kI_{0,c})^{-1}\Lambda]'\\ &= [(\Lambda + kI_{0,c})^{-1}\Lambda]\mathrm{COV}(\phi'\hat{\beta})[\Lambda(\Lambda + kI_{0,c})^{-1}]\\ &= [(\Lambda + kI_{0,c})^{-1}\Lambda]\cdot\sigma^2\phi'(X'X)^{-1}\phi\cdot[\Lambda(\Lambda + kI_{0,c})^{-1}]\\ &= \sigma^2[(\Lambda + kI_{0,c})^{-1}]\cdot\Lambda^{-1}\cdot[\Lambda(\Lambda + kI_{0,c})^{-1}]\\ &= \sigma^2(\Lambda + kI_{0,c})^{-1}\Lambda(\Lambda + kI_{0,c})^{-1},\end{aligned}$$

故
$$\begin{aligned}&\mathrm{trCOV}(\hat{\beta}_{0-c}(k)) = \mathrm{trCOV}(\phi\hat{\alpha}_{0-c}(k))\\ &= \mathrm{tr}[\phi\cdot\mathrm{COV}(\hat{\alpha}_{0-c}(k))\phi']\\ &= \mathrm{tr}[\mathrm{COV}(\hat{\alpha}_{0-c}(k))\phi'\phi]\\ &= \sigma^2\sum_{i=1}^{r}\frac{1}{\lambda_i} + \sigma^2\sum_{i=r+1}^{p}\frac{\lambda_i}{(\lambda_i + kc_i)^2}。\end{aligned}$$

引理得证。

现在给出本节的主要结论：

定理 2.1 对给定的 $c_i > 0, i = r+1, \cdots, p$，存在相应 $k > 0$，使得 $\mathrm{MSE}(\hat{\beta}_{0-c}(k)) < \mathrm{MSE}(\hat{\beta})$，即存在 $k > 0$，使得在均方误差意义下，$0-c$ 型岭估计优于 LS 估计。

证明：令

$$g_{c,1}(k) \triangleq \mathrm{trCOV}(\hat{\beta}_{0-c}(k)),\ g_{c,2}(k) \triangleq \|E(\hat{\beta}_{0-c}(k)) - \beta\|^2,$$

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cr

则
$$\begin{aligned}\mathrm{MSE}(\hat{\beta}_{0-c}(k)) &= \mathrm{trCOV}(\hat{\beta}_{0-c}(k)) + \| E(\hat{\beta}_{0-c}(k))\beta \|^2 \\ &= g_{c,1}(k) + g_{c,2}(k) \\ &= \sigma^2 \sum_{i=1}^{r} \frac{1}{\lambda_i} + \sigma^2 \sum_{i=r+1}^{p} \frac{\lambda_i}{(\lambda_i + kc_i)^2} + k^2 \sum_{i=r+1}^{p} \frac{C_i^2}{(\lambda_2 + kc_i)^2}\alpha_i^2 \\ &\triangleq g_c(k)。\end{aligned}$$

由
$$\frac{\partial g_{c,1}(k)}{\partial k} = -2\sigma^2 \sum_{i=r+1}^{p} \frac{\lambda_i c_i}{(\lambda_i + kc_i)^3}, \frac{\partial g_{c,2}(k)}{\partial k} = 2k \sum_{i=r+1}^{p} \frac{\lambda_i c_i^2 \alpha_i^2}{(\lambda_i + kc_i)^3},$$

知
$$\frac{\partial g_{c,1}(0)}{\partial k} < 0, \frac{\partial g_{c,2}(0)}{\partial k} = 0,$$

从而$\frac{\partial g_c(0)}{\partial k} < 0$，由$\frac{\partial g_c(k)}{\partial k}$在$k \geqslant 0$上的连续性，可知当$k > 0$且充分小时，$\frac{\partial g_c(k)}{\partial k} = \frac{\partial g_{c,1}(k)}{\partial k} + \frac{\partial g_{c,2}(k)}{\partial k} < 0$，，说明$g_c(k) = \mathrm{MSE}(\hat{\beta}_{0-c}(k))$在$k > 0$充分小时，随$k$的增大而减小。故对给定$c_i > 0, i = r+1, \cdots, p$，存在相应$k > 0$，致$g_c(k) < g_c(0)$，即$\mathrm{MSE}(\hat{\beta}_{0-c}(k)) < \mathrm{MSE}(\hat{\beta})$，结论得证。

与岭估计一样，定理中的$k$值将与未知参数$\beta$、$\sigma^2$有关，关于$k$的最优值$k^*$有以下结论。

定理 2.2 对给定的$c_i > 0, i = r+1, \cdots, p$，存在相应$k^*$，满足$\frac{\sigma^2}{\max\limits_{i \geqslant r+1}\{c_i\alpha_i^2\}} \leqslant k^* \leqslant \frac{\sigma^2}{\min\limits_{i \geqslant r+1}\{c_i\alpha_i^2\}}$在$k = k^*$处，$\mathrm{MSE}(\hat{\beta}_{0-c}(k))$达到最小值。

证明：注意到$\frac{\partial^2 g_c(k)}{\partial k^2}$在$k \geqslant 0$上连续，可知$g_c(k) = \mathrm{MSE}(\hat{\beta}_{0-c}(k))$的最小值点必存在，且满足

$$\frac{\partial g_c(k)}{\partial k} |_{k=k^*} = 0。$$

整理后知$k^*$满足下式

$$k^* = \sigma^2 \sum_{i=r+1}^{p} \frac{\lambda_i c_i}{(\lambda_i + k^* c_i)^3} / [\sum_{i=r+1}^{p} \frac{\lambda_i c_i}{(\lambda_i + k^* c_i)^3} c_i \alpha_i^2]。$$

从上式不能得到$k^*$的显式表达，但注意到右端分母是$c_i\alpha_i^2 (i = r+1, \cdots, p)$的一个加权平均，有

$$\frac{\sigma^2}{\max\limits_{i \geqslant r+1}\{c_i\alpha_i^2\}} \leqslant k^* \leqslant \frac{\sigma^2}{\min\limits_{i \geqslant r+1}\{c_i\alpha_i^2\}},$$

结论得证。

由定理 2.2 还可得到

$$k^* \geqslant \frac{\sigma^2}{\max\limits_{i \geqslant r+1}\{c_i\alpha_i^2\}} \geqslant \frac{\sigma^2}{\max\limits_{i \geqslant r+1}\{c_i\} \cdot \max\limits_{i \geqslant r+1}\{\alpha_i^2\}} > \frac{\sigma^2}{\max\limits_{i \geqslant r+1}\{c_i\} \| \beta \|^2} > 0。$$

定理 2.1 及 2.2 为$k$值的选取提供了理论依据和参考。

以下考虑当$k > 0$给定时，$\mathrm{MSE}(\hat{\beta}_{0-c}(k))$关于$c_i > 0, i = r+1, \cdots, p$的最小化问题，下面给出$c_i$的最优值$c_i^*, i = r+1, \cdots, p$及$\mathrm{MSE}(\hat{\beta}_{0-c}(k))$的最小值。

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cn

定理2.3　当 $k>0$ 给定时，当 $c_i=c_i^*=\sigma^2/(k\alpha_i^2)$，$i=r+1,\cdots,p$ 时，$\mathrm{MSE}(\hat{\beta}_{0-c}(k))$ 对变参数 $c_i$，$i=r+1,\cdots,p$ 达到最小值 $\mathrm{MSE}(\hat{\beta})-\sigma^4\sum_{i=r+1}^{p}\frac{1}{\lambda_i\alpha_i^2+\lambda_i\sigma^2}$。

证明：

$$\mathrm{MSE}(\hat{\beta}_{0-c}(k)) = g_c(k)$$

$$= \sigma^2\sum_{i=1}^{r}\frac{1}{\lambda_i}+\sigma^2\sum_{i=r+1}^{p}\frac{\lambda_i}{(\lambda_i+kc_i)^2}+k^2\sum_{i=r+1}^{p}\frac{c_i^2}{(\lambda_i+kc_i)^2}d_i^2。$$

当 $c_i=0$，$i=r+1,\cdots,p$ 时，$\mathrm{MSE}(\hat{\beta}_{0-c}(k))=\mathrm{MSE}(\hat{\beta})$；

当 $c_i\to\infty$，$i=r+1,\cdots,p$ 时，$\mathrm{MSE}(\hat{\beta}_{0-c}(k))\to\sigma^2\sum_{i=1}^{r}\frac{1}{\lambda_i}+\sum_{i=r+1}^{p}\alpha_i^2$。

考虑
$$\frac{\partial[\mathrm{MSE}(\hat{\beta}_{0-c}(k))]}{\partial c_i}=\frac{\partial g_c(k)}{\partial c_i}$$

$$=\frac{2\lambda_i k}{(\lambda_i+kc_i)^3}(kc_i\alpha_i^2-\sigma^2),$$

$$\begin{cases} >0,\ 当\ c_i>\sigma^2/(k\alpha_i^2); \\ =0,\ 当\ c_i=\sigma^2/(k\alpha_i^2); \qquad i=r+1,\cdots,p \\ <0,\ 当\ c_i<\sigma^2/(k\alpha_i^2), \end{cases}$$

即 $k>0$ 给定时，$\mathrm{MSE}(\hat{\beta}_{0-c}(k))$ 在 $c_i=c_i^*=\frac{\sigma^2}{k\alpha_i^2}$，$i=r+1,\cdots,p$ 时达到最小，容易验证此时

$$\mathrm{MSE}(\hat{\beta}_{0-c^*}(k))-\mathrm{MSE}(\hat{\beta})=-\sigma^4\sum_{i=r+1}^{p}\frac{1}{\lambda_i\alpha_i^2+\lambda_i\sigma^2}。$$

其中 $c^*$ 表 $c_i$ 取值 $c_i^*$，$i=r+1,\cdots,p$，结论得证。

$c_i^*$，$i=r+1,\cdots,p$ 等与未知参数 $\sigma^2$ 与 $\alpha_i(p\geqslant i\geqslant r+1)$ 有关，由定理2.3有以下推论。

推论　对给定的 $k>0$，当 $c_i=c_i^*$，$i=r+1,\cdots,p$ 时，0－c型岭估计在均方误差的意义下优于0－k型岭估计，即

$$\mathrm{MSE}(\hat{\beta}_{0-c^*}(k))<\mathrm{MSE}(\hat{\beta}_{0-k})。$$

其中 $c^*$ 表示 $c_i$ 取值 $C_i^*$，$i=r+1,\cdots,p$。

它表明，可以选择 $\hat{\beta}_{0-c^*}(k)$，在 $\hat{\beta}_{0-k}$ 的基础上进一步改进LS估计。

## 3　0－c型岭估计对 $\beta$ 的偏差

考察 $\hat{\beta}_{0-c}(k)$ 对 $\beta$ 的偏差，有如下结论。

定理3.1　对给定的 $k>0$，$\|E(\hat{\beta}_{0-c}(k))-\beta\|^2$ 对每个 $c_i(i=r+1,\cdots,p)$ 为递增的，它具有一个上界 $\sum_{i=r+1}^{p}\alpha_i^2$，这是它在 $c_i\to+\infty$ 时极限。

证明：由§2中引理，

$$\|E(\hat{\beta}_{0-c}(k))-\beta\|^2$$

$$=k^2\sum_{i=r+1}^{p}\frac{c_i^2\alpha_i^2}{(\lambda_i+kc_i)^2}\triangleq g_{c,2}(k),$$

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cn

从而$\frac{\partial g_{c,2}(k)}{\partial c_i}=2\lambda_i k^2\alpha_i^2 c_i/(\lambda_i+kc_i)^3>0, i=r+1,\cdots,p$，且对给定的 $k>0$，如果 $c_i=0, i=r+1,\cdots,p$，则

$$g_{c,2}(k) = k^2\sum_{i=r+1}^{p}\frac{0^2\alpha_i^2}{(\lambda_i+k\cdot 0)^2} = 0 = \|E(\hat{\beta})-\beta\|^2,$$

这是偏差平方的最小值。若 $c_i\to+\infty, i=r+1,\cdots p$，则

$$g_{c,2}(k)\to k^2\sum_{i=r+1}^{p}\frac{1}{k^2}\alpha_i^2 = \sum_{i=r+1}^{p}\alpha_i^2,$$

结论得证。

由此可见，0－c 型岭估计是以牺牲无偏性为代价来改进 $\hat{\beta}$的，若模型的复共线性越严重，则 $r$ 越小，一般说来，$\hat{\beta}_{0-c}(k)$所损失的无偏性就越多。在实用中，当 $k$ 选定后，选取 $c_i$ 时应兼顾使 $\hat{\beta}_{0-c}(k)$的方差及偏差均较小。

特别地，有如下推论：

推论　对给定的 $k>0$，若 $0<c_i<1, i=r+1,\cdots,p$，则

$$\|E(\hat{\beta}_{0-c}(k))-\beta\| \leqslant \|E(\hat{\beta}_{0-k})-\beta\| < \|E(\hat{\beta}(k))-\beta\|。$$

证明：由
$$\|E(\hat{\beta}_{0-c}(k))-\beta\|^2 = k^2\sum_{i=r+1}^{p}\frac{\alpha_i^2}{(\lambda_i+kc_i)^2}c_i^2,$$

得
$$\|E(\hat{\beta}_{0-k})-\beta\|^2 = k^2\sum_{i=r+1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^2},$$

与
$$\|E(\hat{\beta}(k))-\beta\|^2 = k^2\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^2}$$

比较，再由定理 3.1，结论得证。

由定理 2.1 及以上推论，可得以下重要结论。

定理 3.2　存在 $0<c_i<1, i=r+1,\cdots,p$ 及 $k>0$，使得在均方误差的意义下，$\hat{\beta}_{0-c}(k)$优于 $\hat{\beta}$，且 $\hat{\beta}_{0-c}(k)$比 $\hat{\beta}(k)$具有较小偏差。

## 4　$\hat{\beta}_{0-c}(k)$的可容许性及岭参数的选择

为证明 $\hat{\beta}_{0-c}(k)$的可容许性，先给出如下引理。

引理 4.1　对模型(1)式，若 $R(X_{n\times p})=p$，则

$$A\hat{\beta}\sim C_{\beta}\Leftrightarrow A(X'X)^{-1}A'\leqslant A(X'X)^{-1}C'。$$

由上引理，即得如下结果：

定理 4.1　在 $\beta$ 的线性估计类中，$\hat{\beta}_{0-c}(k)$是 $\beta$ 的可容许估计，即 $\hat{\beta}_{0-c}(k)\sim\beta$。

证明：注意到

$$\hat{\beta}_{0-c}(k) = (X'X+k\phi I_{0,c}\phi')^{-1}X'X\hat{\beta}。$$

故此处
$$A = (X'X+k\phi I_{0,c}\phi')^{-1}X'X,\ C = I。$$

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

只须证
$$A(X'X)^{-1}A' \leqslant A(X'X)^{-1}C',$$

即
$$\begin{aligned}&(X'X + k\phi I_{0,c}\phi')^{-1}X'X \cdot (X'X)^{-1}(X'X)(X'X + k\phi I_{0,c}\phi')^{-1}\\ &\leqslant (X'X + k\phi I_{0,c}\phi')^{-1}X'X \cdot (X'X)^{-1} \cdot I'\\ &\Leftrightarrow (X'X + k\phi I_{0,c}\phi')^{-1}[(X'X + k\phi I_{0,c}\phi') - X'X](X'X + k\phi I_{0,c}\phi')^{-1} \geqslant 0\\ &\Leftrightarrow (X'X + k\phi I_{0,c}\phi')^{-1}(k\phi I_{0,c}\phi')(X'X + k\phi I_{0,c}\phi')^{-1} \geqslant 0\\ &\Leftrightarrow (\phi'X'X + k\phi'\phi I_{0,c}\phi')^{-1}(kI_{0,c})(X'X\phi + k\phi I_{0,c}\phi'\phi)^{-1} \geqslant 0\\ &\Leftrightarrow [(\phi'X'X + kI_{0,c}\phi')^{-1}\sqrt{k}\,\mathrm{diag}(0,\cdots,0,\sqrt{c_{r+1}},\cdots,\sqrt{c_p})]\\ &\cdot [(\phi'X'X + kI_{0,c}\phi')^{-1}\sqrt{k}\,\mathrm{diag}(0,\cdots,0,\sqrt{c_{r+1}},\cdots,\sqrt{c_p})]' \geqslant 0,\end{aligned}$$

最后一式显然,故结论成立。

由此可见,在 $\beta$ 的线性估计类中,一致地改进 $\hat{\beta}_{0-c}(k)$是不可能的。

以下我们讨论 0－c 型岭估计在岭参数选择方法上的特点及优点。

如前所述,$\hat{\beta}_{0-c}(k)$的参数 $k$ 的最优值依赖未知参数$\beta$、$\sigma^2$,且无显式表示。$k$ 的选取与$c_i$,$i=r+1,\cdots,p$ 等的选取有一定关系,我们在 $c_i$ 等已经选定(如可选 $c_i=1$, $i=r+1,\cdots,p$)的情况下,简要讨论如何通过试验数据选取 $k$。

(1) 岭迹法

考虑$(\hat{\alpha}_{0-c}(k))_i = \dfrac{1}{\lambda_i}(Z'Y)_i, i = 1,2,\cdots,r$;$(\hat{\alpha}_{0-c}(k))_i = \dfrac{1}{\lambda_i kc_i}(Z'Y)_i, i = r+1,\cdots,p$。

这里,我们只须描述出$(\alpha_{0-c}(k))_i (i\geqslant r+1)$等对 $k$ 变化的曲线,选取 $k$ 值使同一个图上的这 $p-r$ 条曲线都大体稳定,且兼顾回归系数没有不合理的符号,残差平方和上升不太多等。如图 1,可选取 $k=k^*$ 即可。

与岭估计岭迹法相比,这种方法有两个优点:

① 它的岭迹图易于分析,$(\alpha_{0-c}(k))i(i\geqslant r+1)$等的曲线都具有单调性,比 $\hat{\beta}_i(k)(i\geqslant 1)$等的比较杂乱的曲线易于分析,这是一个显著的优点。

② 实用上,对于大型病态线性回归模型,其自变量个数较多($p$ 较大),而自变量之间的复共线关系仅一两个,这时 $p-r$ 很小(根据度量复共线性严重程度的特征分析法),这种情形下,它的岭迹图分析起来较岭估计 $\hat{\beta}(k)$的岭迹图分析简捷得多,这是它在实用上重要的优越性。

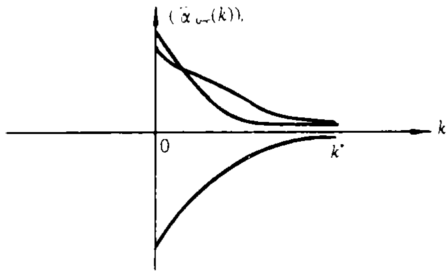


图 1　$k$ 变化曲线

(2) 直观选取法

以 $\alpha$ 及$\sigma^2$ 的 LS 估计 $\hat{\alpha}$及$\hat{\sigma}^2$ 分别代替 $\alpha$ 及$\sigma^2$,描出 MSE($\hat{\alpha}_{0-c}(k)$)的曲线,选取 $k^*$,使之在该点处取最小值。应该看到,由定理 2.2

$$\frac{\sigma^2}{\max\limits_{i\geqslant r+1}\{c_i\alpha_i^2\}} < k^* < \frac{\sigma^2}{\min\limits_{i\geqslant r+1}\{c_i\alpha_i^2\}}。$$

当 $p$ 较小,$k^*$ 的范围较为明确时,可以采用这种方法。

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cr

从 $k^*$ 的隐式表达式

$$k^* = \sigma^2 \sum_{i=r+1}^{p} \frac{\lambda_i c_i}{(\lambda_i + k^* c_i)^3} / [\sum_{i=r+1}^{p} \frac{\lambda_i c_i}{(\lambda_i + k^* c_i)^3} c_i \alpha_i^2],$$

可以采用数值法或图解法大致确定 $k^*$。

同岭估计一样,我们可以采用公式法(选取 $\hat{k} = \hat{\sigma}^2 \max_{i \geqslant r+1} \{c_i \hat{\alpha}_i^2\}$)、McDorald - Galarneau 法、方差扩大因子法、$c_p$ 准则法,双 $h$ 公式法等方法选取 $k$。

0 - c 型岭估计的得到,也可以象广义岭估计一样,采取迭代的方法得到。

前述方法,$k$ 的选取都依赖 $\alpha$ 及 $\sigma^2$。近年来一些学者提出了一些富有吸引力的选法(如[4]),它们不依赖 $\alpha$ 及 $\sigma^2$,且具实用价值。0 - c 型岭估计中 $k$ 值的选取,也可以尝试这些方法。

具体的例子,可以参见[5]中关于特殊的 0 - c 型岭估计——0 - k 型岭估计的实例。

致谢:作者衷心感谢湖南大学数学系朱秀娟教授及中南工业大学数学系关家骥教授在本文写作过程中给予的指导、建议和帮助。

## 参 考 文 献

1 陈希孺,王松桂 . 近代回归分析——原理、方法及应用 . 合肥:安徽教育出版社,1987

2 王松桂 . 线性模型参数估计的新进展 . 数学进展 .1985,14:193～204

3 Marquardt D W, Snee R D. Ridge Regression in Practice. The American Statistican, 29(1):3～19

4 Sun W S. A Selection Method for Ridge Regression Parameter. Technical Reports (1989～1990), Yunnan University, 1990

5 凌晨飞 . 病态线性回归模型的系数估计 . 湖南大学学报,1990,2

# The 0 - c Class of Ridge Estimators of the Coefficients in the Ill - Condition Linear Regression Model

Zhang Jianjun

**Abstract**

In the light of the essence of the ill - condition, in this paper we propose the 0 - c class of ridge estimators of the coefficients in the ill - condition linear regression model. First we study the optimization of its MSE and give the optimal value of the parameters or one of its upper and lower bounds. It is proued that the 0 - c class ridge estimator can be superior to the LSE by selecting its parameters under the mean square error criterion. Second we study its deviation from the regression parameters and prove there exists the 0 - c class ridge estimator that is superior to the LSE and has less deviation. At last we prove it to be admissible and discuss the superiority to RRE of the selection methods of its ridge prarameter in practice.

**Key words:** 0 - c class of ridge estimators; RRE; MSE; Deviation; Admissibility

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cr