第 15 卷第 1 期
2003 年 3 月

江苏工业学院学报
JOURNAL OF JIANGSU POLYTECHNIC UNIVERSITY

Vol. 15 No. 1
Mar. 2003

文章编号：1005－8893（2003）01－0039－04

岭参数的又一确定方法*

汪明瑾
（江苏工业学院 信息科学系，江苏 常州 213016）

摘要：文献［1］给出了一种从小于最优岭参数 k_0 的初值出发逐步改进岭参数的方法。这种方法改进了 Hoerl 和 Kennard 的结果。本文给出了另外一种从大于最优岭参数 k_0 的初值出发逐步改进岭参数的方法。在实际应用中，这 2 种方法互为补充。

关键词：线性模型；岭估计；最小二乘估计；均方误差

中图分类号：O 212.4　　　　文献标识码：A

考虑 Gauss－Markov 模型

$$\boldsymbol{Y}=\boldsymbol{X}\beta+\varepsilon,\ \mathrm{E}(\varepsilon)=0,\ \mathrm{Var}(\varepsilon)=\sigma^2\boldsymbol{I}_n \tag{1}$$

其中 $\boldsymbol{Y}$ 为 $n\times 1$ 观察向量，$\boldsymbol{X}$ 为 $n\times m$ 列满秩设计矩阵，β 为 $m\times 1$ 未知参数向量。关于参数的估计问题一般可用 β 的最小二乘估计（LSE）$\hat{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ 去估计 β。这种估计具有较好的性质，其中最主要的一点是 $\boldsymbol{p}'\hat{\beta}$ 是 $\boldsymbol{p}'\beta$ 的最佳线性无偏估计（BLUE）。若进一步假定 $\varepsilon\sim N_n(0,\ \sigma^2\boldsymbol{I}_n)$，则 $\boldsymbol{p}'\hat{\beta}$ 还是 $\boldsymbol{p}'\beta$ 的一致最小方差无偏估计（UMVUE）。因此 LSE 获得了广泛的应用。但若设计矩阵 $\boldsymbol{X}$ 有多重复共线性关系（即 $\boldsymbol{X}'\boldsymbol{X}$ 呈现病态）时，LSE 的性能可能变得很坏。针对这种情况众多的统计学家研究了最小二乘估计的改进，提出了许多改进 LSE 的新的估计方法。一种考虑是从减小 β 的 LSE 的均方误差 MSE（$\hat{\beta}$）出发的，有岭估计，广义岭估计，压缩估计等[1]，其中 $\mathrm{MSE}(\hat{\beta})=\mathrm{E}(\hat{\beta}-\beta)'(\hat{\beta}-\beta)$。岭估计（Ridge estimate）是由 Hoerl 和 Kennard 于 1970 年提出[2,3]。自 1970 年以来，这种估计的研究和应用得到广泛的重视，成为目前最有影响的有偏估计之一。对于线性模型（1），β 的岭估计定义为

$$\hat{\beta}(k)=(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2}$$

这里 $k\geqslant 0$ 称为岭参数或偏参数。如果 k 取与试验数据 $\boldsymbol{Y}$ 无关的常数，则 $\hat{\beta}(k)$ 为线性估计，不然的话，就是非线性估计。取不同的 k，得到不同的岭估计。所以式（2）就定义了一个很大的估计类。特别，取 $k=0$，$\hat{\beta}(0)=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ 就是 β 的 LS 估计。我们研究 k 为常数的情况。

引进线性模型（1）的典则形式。设 $\boldsymbol{p}_1,\ \boldsymbol{p}_2,\ \cdots,\ \boldsymbol{p}_m$ 为对应于特征根 $\lambda_1,\ \cdots,\ \lambda_m$ 的标准正交化特征向量。记 $\boldsymbol{P}=(\boldsymbol{p}_1,\ \boldsymbol{p}_2,\ \cdots,\ \boldsymbol{p}_m)$，则 $\boldsymbol{P}'\boldsymbol{P}=\boldsymbol{I}$ 且有 $\boldsymbol{P}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{P}=\Lambda=\mathrm{diag}(\lambda_1,\ \cdots,\ \lambda_m)$。

令 $\boldsymbol{Z}=\boldsymbol{X}\boldsymbol{P}$，$\alpha=\boldsymbol{P}'\beta=(\alpha_1,\ \cdots,\ \alpha_m)'$ 则模型（1）可改写为

$$\boldsymbol{Y}=\boldsymbol{X}\beta+\varepsilon=\boldsymbol{X}\boldsymbol{P}\boldsymbol{P}'\beta+\varepsilon=\boldsymbol{Z}\alpha+\varepsilon \tag{3}$$

称 $\boldsymbol{Y}=\boldsymbol{Z}\alpha+\varepsilon$ 为线性模型（1）的典则形式。其中 $\boldsymbol{Z}'\boldsymbol{Z}=\Lambda$。

从（3）式导出的 α 的 LS 估计可表为 $\hat{\alpha}=\Lambda^{-1}\boldsymbol{Z}'\boldsymbol{Y}$，相应的岭估计分别为 $\hat{\alpha}(k)=(\Lambda+k\boldsymbol{I})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$ 和 $\hat{\beta}(k)=\boldsymbol{P}\hat{\alpha}(k)$。

因为均方误差在估计和参数的正交变换下保持不变，所以典则参数 α 和原参数 β 的 LS 估计（或岭估计）有相同的均方误差。所以我们只对线性模

* 收稿日期：2002－07－06
基金项目：江苏省教育厅自然科学研究基金资助（02KJD110002）；江苏工业学院科技基金资助
作者简介：汪明瑾（1961－），男，安徽歙县人，硕士，副教授，主要研究方向为概率统计。

型的典则形式进行讨论。

引进岭估计的目的是减少均方误差，统计学家已经证明了[4]存在 $k>0$，使得

$$\mathrm{MSE}(\hat{\beta}(k))<\mathrm{MSE}(\hat{\beta})\Leftrightarrow\mathrm{MSE}(\hat{\alpha}(k))<\mathrm{MSE}(\hat{\alpha})$$

因此在均方误差意义下，岭估计可改进 LS 估计。但是 k 的最优值不但依赖于模型未知参数 α、σ^2 而且这种依赖关系没有显示表示，这使得 k 值的确定变得十分困难。对于这个应用上非常重要的问题统计学家们作了很多工作，提出了多种选 k 的方法。常用的方法由 Hoerl 和 Kennard 在 1970 年提出。他们取 $k=\sigma^2/\max\alpha_i^2$。当 α、σ^2 为已知时，这种 k 值可使 $H(k)=\mathrm{MSE}(\hat{\alpha}(k))$ 具有比 $H(0)$ 较小的值。事实上 σ^2 和 α_i 皆未知，这时用 LSE σ^2，α_i 代替未知参数 α_i、σ^2。文献 [1] 给出了一种从小于最优岭参数 k_0 的初值出发逐步改进岭参数的方法。这种方法改进了 Hoerl 和 Kennard 的结果。本文给出了另外一种从大于最优岭参数的初值出发逐步改进岭参数的方法。在实际应用中，这 2 种方法互为补充。

由文献 [4]，我们有

引理 1：记 $H(k)=\mathrm{MSE}(\hat{\alpha}(k))$，则有

$$H(k)=\sum_{i=1}^{m}\frac{1}{(\lambda_i+k)^2}(\sigma^2\lambda_i+k^2\alpha_i^2)$$

$$H'(k)=\sum_{i=1}^{m}\frac{2\lambda_i}{(\lambda_i+k)^3}(k\alpha_i^2-\sigma^2)\qquad k\geqslant 0$$

由引理 1 知 $H(k)$ 为光滑函数。又因为 $H'(0)<0$，$H'(+\infty)>0$，故使 $H(k)$ 达到最小的 k 必存在。记 $k_0=\inf\{k: H'(k)=0\}$，显然 $H(k_0)<H(0)$，$\hat{\alpha}(k_0)$ 改进了 LSE $\hat{\alpha}(0)$。下面的引理给出了 k_0 的范围。

引理 2：$\dfrac{\sigma^2}{\alpha_{(m)}^2}\leqslant k_0\leqslant\dfrac{\sigma^2}{\alpha_{(1)}^2}$。其中 $\alpha_{(1)}^2\leqslant\cdots\leqslant\alpha_{(m)}^2$ 是 $\alpha_1^2\cdots\alpha_m^2$ 按升序排列。且记 $\alpha_{(i)}=\sqrt{\alpha_{(i)}^2}$。

证明见文献 [1]。

文献 [1] 的思路是将 k_0 的初值取为它所在区间的左端点，然后逐步改进岭参数。而本文的思路是将 k_0 的初值取为它所在区间的右端点，然后逐步改进岭参数。在实际应用中究竟那种方法更好，需要通过计算所对应的均方误差来回答。

我们将 $\lambda_1,\cdots,\lambda_m$ 中与 $\alpha_{(i)}$ 相对应的值记为 $\lambda_{(i)}$，$i=1,\cdots,m$。

定理 1：$1\leqslant r\leqslant m-1$，若 $\alpha_{(r)}<\alpha_{(r+1)}$ 且

$$\sum_{i=1}^{r}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}\left(\frac{\alpha_{(i)}^2}{\alpha_{(r)}^2}-1\right)+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}\left(\frac{\alpha_{(i)}^2}{\alpha_{(r+1)}^2}-1\right)\geqslant 0 \qquad (4)$$

则有 $k_0<\dfrac{\sigma^2}{\alpha_{(r)}^2}$，使得 $\mathrm{MSE}(\hat{\alpha}(k_0))<\mathrm{MSE}\left(\hat{\alpha}\left(\dfrac{\sigma^2}{\alpha_{(r)}^2}\right)\right)$，其中 $k_0=\max\left(k'_0,\ \dfrac{\sigma^2}{\alpha_{(r+1)}^2}\right)$，$k'_0=$

$$\frac{\displaystyle\sum_{i=1}^{r}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}}{\displaystyle\sum_{i=1}^{r}\frac{\lambda_{(i)}\alpha_{(i)}^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\alpha_{(i)}^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}}$$

证明：当 $\dfrac{\sigma^2}{\alpha_{(r+1)}^2}\leqslant k\leqslant\dfrac{\sigma^2}{\alpha_{(r)}^2}$ 时

$$\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}$$

且 $k\alpha_{(i)}^2-\sigma^2\leqslant 0$，$i=1,\cdots,r$

$k\alpha_{(i)}^2-\sigma^2\geqslant 0$，$i=r+1,\cdots,m$

故

$$\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)$$

$$i=1,\cdots,r \qquad (5)$$

$$\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)$$

$$i=r+1,\cdots,m \qquad (6)$$

对（5）式，（6）式分别求和得

$$\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2) \qquad (7)$$

$$\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant$$

$$\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2) \quad (8)$$

（7）式，（8）式相加得

$$\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant H'(k)\leqslant \sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2) \quad (9)$$

令$\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)=0$

得$k'_0=$

$$\frac{\sum_{i=1}^{r}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}}{\sum_{i=1}^{r}\frac{\lambda_{(i)}\alpha_{(i)}^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\alpha_{(i)}^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}}$$

由（4）式知：$\frac{\sigma^2}{\alpha_{(r)}^2}\geqslant k'_0$，若 $k'_0>\frac{\sigma^2}{\alpha_{(r+1)}^2}$，则 $k_0=k'_0$，此时有：当 $k_0<k\leqslant\frac{\sigma^2}{\alpha_{(r)}^2}$时，$H'(k)>0$，即在$\left[k_0,\ \frac{\sigma^2}{\alpha_{(r)}^2}\right]$上 $H(k)$ 是严格单调增函数，故 $H(k_0)<H\left(\frac{\sigma^2}{\alpha_{(r)}^2}\right)$，即

$$\mathrm{MSE}(\hat{\alpha}(k_0))<\mathrm{MSE}\left(\hat{\alpha}\left(\frac{\sigma^2}{\alpha_{(r)}^2}\right)\right)$$

若 $k'_0\leqslant\frac{\sigma^2}{\alpha_{(r+1)}^2}$，则 $k_0=\frac{\sigma^2}{\alpha_{(r+1)}^2}$，此时对所有的 $\frac{\sigma^2}{\alpha_{(r+1)}^2}\leqslant k\leqslant\frac{\sigma^2}{\alpha_{(r)}^2}$，有 $H'(k)>0$，即在$\left[\frac{\sigma^2}{\alpha_{(r+1)}^2},\ \frac{\sigma^2}{\alpha_{(r)}^2}\right]$上 $H(k)$ 是严格单调增函数，故 $H(k_0)<H\left(\frac{\sigma^2}{\alpha_{(r)}^2}\right)$，即

$$\mathrm{MSE}(\hat{\alpha}(k_0))<\mathrm{MSE}\left(\hat{\alpha}\left(\frac{\sigma^2}{\alpha_{(r)}^2}\right)\right)$$

证毕

如果在定理 1 中，k_0 取为 $k'_0>\frac{\sigma^2}{\alpha_{(r+1)}^2}$时，按定理 2 它还可以被改进。

定理 2：接定理 1，若$\frac{\sigma^2}{\alpha_{(r+1)}^2}<k_0<\frac{\sigma^2}{\alpha_{(r)}^2}$，则存在 $k_1<k_0$，使得

$$\mathrm{MSE}(\hat{\alpha}(k_1))<\mathrm{MSE}(\hat{\alpha}(k_0))$$

其中 $k_1=\max\left(k'_1,\ \frac{\sigma^2}{\alpha_{(r+1)}^2}\right)$

$k'_1=$

$$\frac{\sum_{i=1}^{r}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+k_0)^3}}{\sum_{i=1}^{r}\frac{\lambda_{(i)}\alpha_{(i)}^2}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\alpha_{(i)}^2}{(\lambda_{(i)}+k_0)^3}}$$

证明：首先有

$$\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k_0\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}(k_0\alpha_{(i)}^2-\sigma^2)>\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k_0\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r)}^2)^3}(k_0\alpha_{(i)}^2-\sigma^2) \quad (10)$$

当$\frac{\sigma^2}{\alpha_{(r+1)}^2}\leqslant k\leqslant k_0$ 时，

$\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}$ 且 $k\alpha_{(i)}^2-\sigma^2\leqslant0$，$i=1, \cdots, r$；$k\alpha_{(i)}^2-\sigma^2\geqslant0$，$i=r+1, \cdots, m$，故

$$\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}(k\alpha_{(i)}^2-\sigma^2)$$

$$i=1, \cdots, r \quad (11)$$

$$\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)$$

$$i=r+1, \cdots, m \quad (12)$$

对（11）式，（12）式分别求和再相加得

$$\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}(k\alpha_{(i)}^2-\sigma^2)\leqslant H'(k)\leqslant\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}(k\alpha_{(i)}^2-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha_{(r+1)}^2)^3}(k\alpha_{(i)}^2-\sigma^2)$$

令$\sum_{i=1}^{r}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha^2_{(r+1)})^3}(k\alpha^2_{(i)}-\sigma^2)+\sum_{i=r+1}^{m}\frac{2\lambda_{(i)}}{(\lambda_{(i)}+k_0)^3}(k\alpha^2_{(i)}-\sigma^2)=0$

得$k'_1=$

$$\frac{\sum_{i=1}^{r}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha^2_{(r+1)})^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+k_0)^3}}{\sum_{i=1}^{r}\frac{\lambda_{(i)}\alpha^2_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha^2_{(r+1)})^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\alpha^2_{(i)}}{(\lambda_{(i)}+k_0)^3}}$$

由（10）式知 $k'_1<k'_0$。若 $k'_1>\frac{\sigma^2}{\alpha^2_{(r+1)}}$，则 $k_1=k'_1$，当 $k_1<k<k_0$ 时，$H'(k)>0$，即在 $[k_1,\ k_0]$ 上 $H(k)$ 是严格单调增函数，故 $H(k_1)<H(k_0)$。即

$\mathrm{MSE}(\hat{\alpha}(k_1))<\mathrm{MSE}(\hat{\alpha}(k_0))$

若 $k'_1\leqslant\frac{\sigma^2}{\alpha^2_{(r+1)}}$，则 $k_1=\frac{\sigma^2}{\alpha^2_{(r+1)}}$，此时对所有的$\frac{\sigma^2}{\alpha^2_{(r+1)}}\leqslant k\leqslant k_0$，有 $H'(k)>0$，故

$\mathrm{MSE}(\hat{\alpha}(k_1))<\mathrm{MSE}(\hat{\alpha}(k_0))$

证毕

仿定理 2 的证明，可得下述定理。

定理 3：接定理 2，若$\frac{\sigma^2}{\alpha^2_{(r+1)}}<k_1<k_0\leqslant\frac{\sigma^2}{\alpha^2_{(r)}}$，则存在 $k_2<k_1$，使得

$\mathrm{MSE}(\hat{\alpha}(k_2))<\mathrm{MSE}(\hat{\alpha}(k_1))$

其中 $k_2=\max\left(k'_2,\ \frac{\sigma^2}{\alpha^2_{(r+1)}}\right)$，$k'_2=$

$$\frac{\sum_{i=1}^{r}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+\sigma^2/\alpha^2_{(r+1)})^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\sigma^2}{(\lambda_{(i)}+k_1)^3}}{\sum_{i=1}^{r}\frac{\lambda_{(i)}\alpha^2_{(i)}}{(\lambda_{(i)}+\sigma^2/\alpha^2_{(r+1)})^3}+\sum_{i=r+1}^{m}\frac{\lambda_{(i)}\alpha^2_{(i)}}{(\lambda_{(i)}+k_1)^3}}$$

综上所述，若 $\alpha_{(i)}$（$i=1,\ 2,\ \cdots,\ m$）不全相等，则按定理 1，可以找到一个 k_0，$\frac{\sigma^2}{\alpha^2_{(r+1)}}\leqslant k_0\leqslant\frac{\sigma^2}{\alpha^2_{(r)}}$使 $\mathrm{MSE}(\hat{\alpha}(k_0))<\mathrm{MSE}\left(\hat{\alpha}\left(\frac{\sigma^2}{\alpha^2_{(r)}}\right)\right)$

若 $k_0>\frac{\sigma^2}{\alpha^2_{(r+1)}}$，则按定理 2，可以找到 k_1 满足 $\frac{\sigma^2}{\alpha^2_{(r+1)}}\leqslant k_1<k_0$，使

$\mathrm{MSE}(\hat{\alpha}(k_2))<\mathrm{MSE}(\hat{\alpha}(k_1))$

以此类推，只要 $k_2>\frac{\sigma^2}{\alpha^2_{(r+1)}}$，就可以按定理 3 的方法一直做下去，使得均方误差逐步减小。

参考文献：

[1] 汪明瑾，王静龙．岭回归中确定 K 值的一种方法［J］．应用概率统计，2001，17（1）：7－13.

[2] Hoerl A E，Kennard R W．Ridge Regression：Biased Estimation for Non－orthogonal Problems［J］．Technometrics，1970，12：55－88.

[3] Hoerl A E，Kennard R W．Ridge Regression：Application for Non－orthogonal Problems［J］．Technometrics，1970，12：69－72.

[4] 陈希孺，王松桂．近代回归分析［M］．合肥：安徽教育出版社，1987.

[5] 汪明瑾．Kantorovich 不等式的一种推广［J］．江苏石油化工学院学报，2002，14（1）：51－52.

Another Method to Determine the Ridge Parameter in Ridge Regression

WANG Ming－jin

(Department of Information Science，Jiangsu Polytechnic University，Changzhou 213016，China)

Abstract：In this paper，the author considered linear model $Y_{n\times 1}=X_{n\times m}\beta_{m\times 1}+\varepsilon_{n\times 1}$，$E(Y)=X\beta$，$\mathrm{Var}(Y)=\sigma^2 I_n$，$R(X)=m$．Its canonical model was $Y_{n\times 1}=Z_{n\times m}\alpha_{m\times 1}+\varepsilon_{n\times 1}$．Where $Z'Z=\Lambda=\mathrm{diag}(\lambda_1,\ \cdots,\ \lambda_m)$，$\lambda_1\geqslant 0,\ \cdots,\ \lambda_m\geqslant 0$ were the eigenvalues of $X'X$．The ridge estimator of α was $\hat{\alpha}(k)=(\Lambda+kI)^{-1}Z'Y$，and the ridge estimator of β was $P'\hat{\alpha}(k)$，where P was orthogonal matrix．So that $P'X'XP=\Lambda$．Paper (1) gave a new method to determine ridge parameter k．That method has improved the Hoerl－kennard formula．In this paper，another method to determine ridge parameter k in ridge regression was given．

Key words：linear model；ridge estimator；least square estimator；mean square error