(20)

118-127

岭回归估计均方误差的重要特性及其应用*

THE IMPORTANT PROPERTIES OF MEAN SQUARED ERROR OF RIDGE REGRESSION ESTIMATOR AND IT'S APPLICATIONS

何中市　　　何良材

He Zhongshi　　　He Liangcai

（重庆大学系统工程及应用数学系）

O212.1

摘　要　通过深入分析单参数岭估计 $\hat{\beta}(k)$ 的均方误差 MSE($\hat{\beta}(k)$)，得到了它的一个重要特性。并把较优均方误差的存在范围由 $(0, \sigma^2/\max_i \alpha_i^2]$ 扩大到 $(0, 2\sigma^2/\max_i \alpha_i^2]$；同时，还得到了较优均方误差的另一不可替代的存在范围 $(0, 2\sigma^2 \min_i \lambda_i/(\max_i \lambda_i \alpha_i^2 - \sigma^2))$；以及最优均方误差的存在范围 $[\sigma^2/\max_i \alpha_i^2, \sigma^2/\min_i \alpha_i^2]$。并结合实际问题，给出了岭参数 k 值选取的算法实例。

关键词　岭回归估计；均方误差；特 性；岭参数；最优均方误差　，单参数

中国图书资料分类法分类号　O212.1

ABSTRACT　An important property of mean squared error of ridge regression estimator is obtained with an analysis of MSE ($\hat{\beta}(k)$) to rhe ridge regression estimator $\hat{\beta}(k)$ of parameter vector, and the range of a better MSE is expanded from $(0, \sigma^2/\max_i \alpha_i^2]$ to $(0, 2\sigma^2/\max_i \alpha_i^2]$. Meanwhile, another range of smaller mean square error that can't be replaced is proposed and the interval of the minimum MSE value of k $[\sigma^2/\max \alpha_i^2, \sigma^2 \min_i \alpha_i^2]$ is acquired. Finally, a practical example is presented with algorithm for choosing the ridge parameter k value.

KEY WORDS　ridge regression estimator; mean sguare error; property; ridge parameter; MSE optimal.

0　引　　言

二年的事实证明，岭回归分析，在理论上，它是对回归分析的丰富和发展，在实际应用中，它无疑是重要的一种有偏估计[1]。对于岭估计，根本问题在于岭参数（单参数 k 或参数矩阵 K）的选取。在单参数岭估计中，统计工作者们根据均方误差（Mean Square Error，简记为 *MSE*）准则，提出了10余 k 值选取方法。

均方误差准则：使 $\hat{\beta}$ 的均方误差

$$\mathrm{MSE}(\hat{\beta}) \triangleq E\|\hat{\beta} - \beta\|^2$$

* 收文日期　1991—10—15

达到最小。其中 $\hat{\beta}$ 为 β 的一个估计。

由于岭估计 $\hat{\beta}(k)$ 的均方误差 $MSE(\hat{\beta}(k))$ 是 α,σ^2,k 的复杂函数，所以，现有的10余种 k 值选取方法则是侧面考虑了 $MSE(\hat{\beta}(k))$ 的某此特性或从某种角度的合理性而言提出来的。事实上，目前，尚没有能使人们信服的 k 值选取方法。为此，k 值选取对于统计工作者仍然具有很大的吸引力。

鉴于此，本文作者在深入分析 $MSE(\hat{\beta}(k))$ 的函数关系基础上，提出了均方误差的重要特性，并将较优均方误差的存在范围由 $(0,\sigma^2/\max_i \alpha_i^2]$[2,3]扩大了一倍，达到 $(0,2\sigma^2/\max_i \alpha_i^2]$，同时还给出了较优均方误差另一不可替代的存在范围：$(0,2\sigma^2\min_i \lambda_i/(\max_i \lambda_i\alpha_i^2-\sigma^2))$，并指出了最优均方误差能也只可能在 $[\sigma^2/\max_i \alpha_i^2,\sigma^2/\min_i \alpha_i^2]$ 上达到。

1 岭估计的均方误差

对线性模型

$$Y = X\beta + \varepsilon \quad , \quad E\varepsilon = 0, COV(\varepsilon) = \sigma^2 I \tag{1}$$

其典则形式为

$$Y = Z\alpha + \varepsilon \tag{2}$$

其中：

$Z=XP'$，$\alpha=P\beta$，$P'=(\varphi_1,\cdots,\varphi_p)$，$\varphi_1,\cdots,\varphi_p$ 为 $X'X$ 相应于其特征根 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p>0$ 的规范化正交特征向量。

$\hat{\alpha}=\Lambda^{-1}Z'Y$， $\hat{\alpha}(k)=(\Lambda+kI)^{-1}Z'Y$

分别为(2)式中 α 的最小二乘估计(LSE)和岭估计(RE)。其中 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p)$

$\hat{\beta}=P'\hat{\alpha}$， $\hat{\beta}(k)=P'\hat{\alpha}(k)$

分别为(1)式中 β 的 LSE 和 RE。则 $\hat{\beta}$、$\hat{\beta}(k)$ 的均方误差分别为：

$$MSE(\hat{\beta}) \triangleq E\|\hat{\beta}-\beta\|^2 = \sigma^2\sum_{i=1}^{p} 1/\lambda_i \tag{3}$$

$$\begin{aligned} MSE(\hat{\beta}(k)) &\triangleq E\|\hat{\beta}(k)-\beta\|^2 = E\|\hat{\alpha}(k)-\alpha\|^2 \\ &= \alpha^T[\Lambda(\Lambda+kI)^{-1}-I]^2\alpha + \sigma^2\mathrm{tr}[Z\Lambda+kI)^{-2}Z] \\ &= \sum_{i=1}^{p} k^2\alpha_i^2/(\lambda_i+k)^2 + \sigma^2\sum_{i=1}^{p}\lambda_i/(\lambda_i+k)^2 \\ &= \sum_{i=1}^{p}(\lambda_i\sigma^2+k^2\alpha_i^2)/(\lambda_i+k)^2 \end{aligned}$$

为了方便，记

$$H(k)=\mathrm{MSE}(\hat{\beta}(\mathrm{k}))=\sum_{i=1}^{p}(\lambda_i\sigma^2+\mathrm{k}^2\alpha_i^2)/(\lambda_i+\mathrm{k})^2 \tag{4}$$

则
$$H(0)=\mathrm{MSE}(\hat{\beta}(0))=MSE(\hat{\beta})=\sigma^2\sum_{i=1}^{p}1/\lambda_i \tag{5}$$

对于均方误差H(k)有如下性质：

C. M. Theobald 提出[1]：

引理1 当$0<\mathrm{k}<2\sigma^2/\alpha'\alpha$时，有

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

其给出的证明过程甚繁。

A. E. Hoerl 和 *R. W. Kennard* 提出[2]

引理2 当$0<\mathrm{k}<\sigma^2/\max\limits_i\alpha_i^2$时，有

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

同时，本文作者之一的何良材利用分析方法简要证明了此性质[3]。

对于以上性质，在本文作者给出以下重要性质及其应用之后，引理1，引理2仅是其一种特殊情形或其简单推论。

2 均方误差的重要特性

引理3 对于$\hat{\beta}(\mathrm{k})$的均方误差H(k)有

1) $0\leqslant\mathrm{k}\leqslant\sigma^2/\max\limits_i\alpha_i^2$时，H(k)严格单调减

2) $\mathrm{k}>\sigma^2/\min\limits_i\alpha_i^2$时，H(k)严格单调增

3) $\mathrm{H}(0)=\sigma^2\sum_{i=1}^{p}1/\lambda_i$，　$\mathrm{H}(+\infty)=\sum_{i=1}^{p}\alpha_i^2$

证明：由(4)式可得

$$\begin{aligned}\frac{\mathrm{d}H(k)}{\mathrm{d}k}&=\frac{d}{dk}\sum_{i=1}^{p}(\lambda_i\sigma^2+k^2\alpha_i^2)/(\lambda_i+k)^2\\&=\sum_{i=1}^{p}[2\lambda_i/(\lambda_i+k)^3]\cdot(\alpha_i^2k-\sigma^2)\\&=\begin{cases}<0,\text{若 }\alpha_i^2k-\sigma^2<0,i=1,\cdots,p\\>0,\text{若 }\alpha_i^2k-\sigma^2>0,i=1,\cdots,p\end{cases}\end{aligned}$$

1) 当$0<\mathrm{k}<\sigma^2/\max\limits_i\alpha_i^2$时，有$\alpha_i^2\mathrm{k}-\sigma^2<0$，$\mathrm{i}=1,\cdots,\mathrm{p}$，

即
$$\frac{dH(k)}{dk}<0$$

所以 H(k)随 k 严格单调递减

2） 当 $k>\sigma^2/\min\limits_i\alpha_i^2$ 时，有 $\alpha_i^2k-\sigma^2>0, i=1,\cdots,p$

即 $$\frac{dH(k)}{dk}>0$$

所以 H(k)随 k 严格单调递增。

3） 由(4)式

a） 令 k=0 得 $H(0)=MSE(\hat{\beta}(0))=\sigma^2\sum\limits_{i=1}^{p}1/\lambda_i$

b） $H(+\infty)=\lim\limits_{k\to+\infty}\sum\limits_{i=1}^{p}\frac{\lambda_i\sigma^2+k^2\alpha_i^2}{(\lambda_i+k)^2}=\sum\limits_{i=1}^{p}\alpha_i^2$ 证毕

由此可以推出，关于最优均方误差有以下结论

推论1 设 $H(k^*)=\min\limits_{k\geqslant 0}H(k)$，则

$$k^*\in[\sigma^2/\max_i\alpha_i^2,\sigma^2/\min_i\alpha_i^2]$$

证明：由引理3有

1） 当 $0\leqslant k<\sigma^2/\max\limits_i\alpha_i^2$ 时，$H(k)>H(\sigma^2/\max\limits_i\alpha_i^2)$成立。

2） 当 $k>\sigma^2/\min\limits_i\alpha_i^2$ 时，$H(k)>H(\sigma^2/\min\limits_i\alpha_i^2)$成立。由此可得：

$$H(k^*)=\min_{k\geqslant 0}H(k)=\min_{\frac{\sigma^2}{\max\limits_i\alpha_i^2}\leqslant k\leqslant\frac{\sigma^2}{\min\limits_i\alpha_i^2}}H(k)$$ 证毕

还可以看出，若 $\alpha_i^2=a_i^2$时，$i=1,\cdots,p$，则

$$k^*=\sigma^2/\max_i\alpha_i^2=\sigma^2/\min_i\alpha_i^2$$

即 k^* 可在$[\sigma^2/\max\limits_i\alpha_i^2,\sigma^2/\min\limits_i\alpha_i^2]$的两个端点处取得。

定理1 设 $\alpha\neq 0$，令 $k_p=\sigma^2/\max\limits_i\alpha_i^2$，$\forall\ d\in(0,1]$，$k_l=k_p(1-d)$，$k_u=k_p(1+d)$，则

$$H(k_l)>H(k_u)$$

证明：由(4)式可得

$$H(k_l)=\sum_{i=1}^{p}[(\lambda_i\sigma^2+k_l^2\alpha_i^2)/(\lambda_i+k_l)^2]$$

$$H(k_u)=\sum_{i=1}^{p}[(\lambda_i\sigma^2+k_u^2\alpha_i^2)/(\lambda_i+k_u)^2]$$

令
$$A_i = (\lambda_i\sigma^2 + k_l^2\alpha_i^2)(\lambda_i + k_u)^2 - (\lambda_i\sigma^2 + k_u^2\alpha_i^2)(\lambda_i + k_l)^2$$

则
$$H(k_l) - H(k_u) = \sum_{i=1}^{p} A_i / [(\lambda_i + k_l)^2(\lambda_i + k_u)^2] \tag{6}$$

由
$$k_l = k_p(1-d), \qquad k_u = k_p(1+d)$$

有
$$\begin{aligned} A_i &= \lambda_i\sigma^2[(\lambda_i + k_u)^2 - (\lambda_i + k_l)^2] + \alpha_i^2[(k_l\lambda_i + k_lk_u)2 - (k_u\lambda_i + k_uk_l)^2] \\ &= \lambda_i\sigma^2[2(\lambda_i + k_p)\cdot 2dk_p] + \alpha_i^2[(2\lambda_ik_p + 2(1-d^2)k_p)(-2dk_p\lambda_i)] \\ &= 4\lambda_ik_p\sigma^2(\lambda_i + k_p)d + 4\lambda_i\alpha_i^2[(-k_p^2)(\lambda_i + k_p)d + k_p^3d^3] \\ &= 4\lambda_ik_p(\lambda_i + k_p)(\sigma^2 - k_p\alpha_i^2)d + 4\lambda_ik_p^3\alpha_i^2d^3 \end{aligned}$$

∵
$$\lambda_i > 0, k_p > 0, \sigma^2 - k_p\alpha_i^2 \geqslant 0, i = 1,\cdots,p, d > 0$$

∴
$$A_i \geqslant 0,且 \sum_{i=1}^{p} A_i > 0$$

否则，由 $\sum_{i=1}^{p} Ai = 0$ 得 $Ai = 0$，这与 $\alpha \neq 0$ 矛盾。

由(6)式得

$$H(k_l) - H(k_u) > 0$$

即
$$H(k_l) > H(k_u)$$
证毕

3 应　　用

利用以上重要性质，可以将较优均方误差的存在范围扩大一倍，即有

定理2　当 $0 < k \leqslant 2\sigma^2/\max_i \alpha_i^2$ 时，有 $H(k) < H(0)$

即
$$\mathrm{MSE}(\hat{\beta}(k)) < MSH(\hat{\beta}) \tag{7}$$

证明：记 $k_p = \sigma^2/\max_i \alpha_i^2$，则

(1)　当 $0 < k \leqslant k_p$ 时，由引理3知(7)式成立。

(2)　当 $k_p<k\leqslant 2k_p$ 时，有

$$0<k-k_p\leqslant k_p$$

令　$d=(k-k_p)/k_p$，显然 $d\in(0,1]$

由定理1得到

$$\begin{aligned}H(k)&=H(k_p+(k-k_p))=H(k_p+dk_p)\\&=H(k_p)<H(k_p)=H(k_p-dk_p)\\&=H(k_p-(k-k_p))=H(2k_p-k)\end{aligned}$$

即
$$H(k)<H(2k_p-k) \tag{8}$$

而由 $k_p<k\leqslant 2k_p$　知　$0\leqslant 2k_p-k<k_p$

再结合(1)式知 $H(2k_p-k)\leqslant H(0)$

结合(8)式亦得证(7)式成立。　证毕

由定理1，因 $\max\limits_i \alpha_i^2\leqslant\sum\limits_{i=1}^{p}\alpha_i^2=\alpha'\alpha$，故

$$2\sigma^2/\max_i\alpha_i^2\geqslant 2\sigma^2/\alpha'\alpha，从而有：$$

推论2　当 $0<k\leqslant 2\sigma^2/\alpha'\alpha$ 时，有 $H(k)<H(0)$

定理3　若 $0<k<2\sigma^2\min\limits_i\lambda_i/(\max\limits_i\lambda_i\alpha_i^2-\sigma^2)$，则 $H(k)<H(0)$，即有：

$$(\hat{\beta}) \tag{9}$$

证明：由 $0<k<2\sigma^2\cdot\min\limits_i\lambda_i/(\max\limits_i\lambda_i\alpha_i^2-\sigma^2)$　得

$$k\cdot\max_i\lambda_i\alpha_i^2-k\sigma^2<2\sigma^2\min_i\lambda_i$$

即
$$k\lambda_i\alpha_i^2<k\sigma^2+2\sigma^2\lambda_i,\ i=1,\cdots,p \tag{10}$$

∴
$$H(k)=\sum_{i=1}^{p}(\lambda_i\sigma^2-k^2\alpha_i^2)/(\lambda_i+k)^2$$

$$= \sum_{i=1}^{p} (\lambda_i^2\sigma^2 + k \cdot \lambda_i k \alpha_i^2)/[\lambda_i(\lambda_i + k)^2]$$

$$< \sum_{i=1}^{p} [\lambda_i^2\sigma^2 + k \cdot (k\sigma^2 + 2\sigma^2\lambda_i)]/[\lambda_i(\lambda_i + k)^2]$$

$$= \sum_{i=1}^{p} \sigma^2/\lambda_i \cdot (\lambda_i^2 + k^2 + 2k\lambda_i)/(\lambda_i + k)^2$$

$$= \sum_{i=1}^{p} \sigma^2/\lambda_i = H(0)$$

$\therefore$ $H(k) < H(0)$ 成立 证毕

注：$2\sigma^2 \cdot \min\limits_i \lambda_i / (\max\limits_i \alpha_i^2 \lambda_i - \sigma^2) = \sigma^2 / \max\limits_i \alpha_i^2 \cdot \dfrac{\max\limits_i \alpha_i^2 \cdot \min\limits_i \lambda_i^2}{\max\limits_i \lambda_i \sigma_i^2 - \sigma^2}$

$= k_r \cdot l$ 可知 l 既可能小于 l，亦可能大于1

事实上，若 $\lambda_i \equiv \lambda_1$，$i = 1, \cdots, p$，则

$$l = \lambda_1 \cdot \max_i \alpha_i^2 / (\lambda_1 \max_i \alpha_i^2 - \sigma^2) > 1$$

所以，定理2和定理3给出的较优均方误差存在范围，各有裨益，不可替代。

以上关于均方误差的研究结果，有利于寻求 k 值选取新方法，有关此新方法及相关问题，作者在文献[5~7]中作了具体研究。

4 岭参数 k 值选取算法实例

例：发电量需求模型——影响发电量需求量的指标有：水泥产量、原煤产量、生铁产量、钢材产量、化肥产量、棉纱产量等6个指标。限于客观，现收集了历史上12年某地的以上各指标的产量与发电量数据（如表1，注：表1中的数据是在实际数量的基础上减去统计年内年平均值的整数部分）。求其发电量需求模型。

表 1 原始数据（万吨）

统计年序	水泥 (x_1)	原煤 (x_2)	生铁 (x_3)	钢材 (x_4)	化肥 (x_5)	棉纱 (x_6)	发电量 y 亿 kw·h
1	1.5971	0.5154	−0.2229	0.1068	0.0600	0.0349	1.2465
2	−1.5532	0.1462	0.2378	−0.0503	0.0301	0.0491	−1.2947
3	−1.3668	−0.8257	−0.7218	−0.0730	0.0204	0.0097	−1.1685
4	1.3752	−0.9875	0.3013	−0.1160	0.1113	−0.0050	0.9154
5	1.5231	0.6920	−0.1160	−0.5368	−0.0403	0.0055	1.1957
6	1.4297	−0.9559	0.0864	0.0266	−0.3414	0.1340	0.9752
7	−1.0084	0.7746	0.5468	−0.1935	−0.6574	−0.2769	−0.6896

续表1

8	0.7731	−0.9280	−0.2699	0.5393	0.5861	0.6728	0.5201
9	0.6825	−0.4064	−1.4320	0.2361	0.6409	0.1917	1.0349
10	−0.2687	0.6110	0.2572	−1.2368	−0.2268	−0.8399	−1.0154
11	0.7963	0.6267	0.7104	0.2450	−1.3466	−0.2333	−1.0243
12	−0.3112	0.7123	0.1721	−1.0037	−0.2258	−1.3301	1.1193

通过分析，我们建立发电量 y 与指标 $x_1, x_2, \cdots, x_6$ 间的线性回归模型如下

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + e$$

结合给出的数据可得如下模型

$$Y = X\beta + e$$

其中 Y 为12×1随机观测向量，X 为12×6阶矩阵，β 为6×1参数向量，e 为12×1随机误差向量

$$E(e) = 0, \qquad \mathrm{COV}(e) = \sigma^2 I$$

现对模型参数向量 β 进行估计。

4.1　用 Jacobi 法求设计阵 $X'X$ 的特征值和特征向量。如表2.

表 2　特征值及特征向量表

特征值 λ_i	26.9390	6.6340	2.4986	0.8804	0.5952	0.2666
特征向量 φ_i	−0.435	−0.274	0.403	−0.427	−0.355	−0.485
	0.457	−0.080	−0.430	0.202	−0.178	−0.646
	0.371	0.453	0.732	0.292	−0.119	−0.125
	−0.375	0.542	−0.305	0.157	−0.658	0.082
	−0.415	−0.380	0.118	0.809	0.074	−0.072
	−0.390	0.525	−0.087	−0.040	0.626	−0.376

有 $X'X = P'\Lambda P$，其中 $P'(\varphi_1, \varphi_2, \cdots, \varphi_6$ 为正交矩阵，

$$\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_6)$$

可得 $X'X$ 的条件数(*condition number*)为

$$c=\frac{\max\lambda_i}{\min\lambda_i}=\frac{26.939}{0.2666}=101>100$$

所以，我们认为设计矩阵呈较严重病态，亦即此模型具有较严重的共线性性、于是，最小二乘估计不受欢迎，为了得到较好的估计，我们采用岭回归估计对参数向量 β 进行估计。

4.2 求 β 的岭估计

1) 求 α、β 的最小二乘估计 $\hat{\alpha}$、$\hat{\beta}$

由
$$\hat{\alpha}=A^{-1}Z'Y=A^{-1}-PX'Y$$
$$\hat{\beta}=P'\hat{\alpha}$$

得 $\hat{\alpha}=(-0.9235,1.2047,0.9364,-0.3027,1.7048,-2.1049)'$

$\hat{\beta}=(1.2009,0.1107,0.9288,-0.6848,0.1095,2.9726)'$

2) 估计 $k_p=\sigma^2/\max\limits_i\alpha_i^2$

由于 σ^2、α_i^2 的真值不知，故此处分别用 $\hat{\sigma}^2$、$\hat{\alpha}_i^2$ 代替 σ^2、α_i^2，则得 k_p 的估计值为

$$\hat{k}_p=\hat{\sigma}^2/\max_i\hat{\alpha}_i^2=2.1079/(-2.1049)^2=0.4758$$

其中
$$\hat{\sigma}^2=(Y-X\hat{\beta})'(Y-X\hat{\beta}/(n-p)=2.1079$$

3)岭参数 k^* 的选取

基于推论1和定理2(注，我们也可以基于推论1和定理3，方法与此相同)我们可选取 k^* 满足：

$$\begin{cases}H(k^*)=\min\limits_k H(k)=\sum\limits_{i=1}^{6}\dfrac{\lambda_i\hat{\sigma}^2+k^2\hat{\alpha}_i^2}{(\lambda_i+k)^2}\\ \text{s.t. } \hat{k}_p\leqslant k\leqslant 2\hat{k}_p\end{cases}$$

对此非线性规划问题，用一般的最优化方法如牛顿法、对分法、……等即可求得其解。

此处，鉴于实际问题的满意解与理论上最优解的关系，本文将用等分法(8等分)求较优解——满意解、即计算出 k 取区间 $[\hat{k}_p,2\hat{k}_p]$ 的端点和其内的各分点时的岭估计 $\hat{\alpha}(k)$(或 $\hat{\beta}(k)$)和均方误差 $H(k)$ 如表3

其中
$$\hat{\alpha}(k)=(A+kI)^{-1}A\hat{\alpha},H(k)=\text{MSE}\hat{\beta}(k))=\sum_{i=1}^{6}\frac{\lambda_i\hat{\sigma}^2+k^2\hat{\alpha}_i^2}{(\lambda_i+k)^2}$$

表 3　岭估计 $\hat{\alpha}(k)$ 及相应的均方误差 $H(k)$

k	o	$\hat{k}_p$	$\frac{8}{9}\hat{k}_p$	$\frac{10}{8}\hat{k}_p$	$\frac{11}{8}\hat{k}_p$	$\frac{12}{8}\hat{k}_p$	$\frac{13}{8}\hat{k}_p$	$\frac{14}{8}\hat{k}_p$	$\frac{15}{8}\hat{k}_p$	$2\hat{k}_p$
$\hat{\alpha}(k)=$ $\hat{\alpha}_1(k)$	−0.9235	−0.9074	−0.9055	−0.9055	−0.9016	−0.8997	−0.8977	−0.8958	−0.8939	−0.8919
$\hat{\alpha}_2(k)$	1.2047	1.1241	1.1147	1.1056	1.0966	1.0877	1.0790	1.0704	1.0619	1.0536
$\hat{\alpha}_3(k)$	0.9364	0.7866	0.7712	0.7564	0.7422	0.7284	0.7152	0.7024	0.6901	0.6782
$\hat{\alpha}_4(k)$	−0.3027	−0.1966	−0.1883	−0.1807	−0.1737	−0.1672	−0.1612	−0.1557	−0.1504	−0.1459
$\hat{\alpha}_5(k)$	1.7049	0.9476	0.8978	0.8529	0.8124	0.7754	0.7418	0.7109	0.6824	0.6562
$\hat{\alpha}_6(k)$	−2.1049	−0.7561	−0.7001	−0.6518	−0.6097	−0.5727	−0.5399	−0.5107	−0.4845	−0.4609
H(k)	15.3211	6.5048	6.3735	6.2840	6.2244	6.1826	6.1527	6.1348	6.1319	6.1330

选取使 H(k)达到最小的 k 值作为岭参数 k*，即得：

$$k^* = \frac{15}{8}\hat{k}_p = \frac{15}{8} \times 0.4758 = 0.8921$$

4)求参数向量 β 的岭估计 $\hat{\beta}(k^*)$

$$\because \quad \hat{\alpha}(k^*) = (-0.8939, 1.0619, 0.6901, -0.1504, 0.6824, -0.4845)'$$

$$\therefore \quad \hat{\beta}(k^*) = P'\hat{\alpha}(k^*)$$

$$= (0.2683, -0.4860, 0.4558, 0.1451, 0.0097, 1.1290)'$$

此时均方误差 H(k)为6.1319，较最小二乘估计 $\hat{\beta}$ 的均方误差 H(0)=15.3211减少了1.5倍。所以，所求发电量需求模型为

$$y = 0.2683x_1 - 0.4860x_2 + 0.4558x_3 + 0.1451x_4 + 0.0097x_5 + 1.129x_6$$

参 考 文 献

1 王松桂. 线性模型参数估计的新进展. 数学进展. 1985. 14(3)：193～204

2 Hoerl A E, Kennard R W. Ridge regression：Biased estimation for nonorthogonal problems. Technometrics. 12(1)：55～67

3 何良材. 岭回归估计 $\hat{\beta}(k)$的一个特性及其应用. 重庆大学学报. 1990. 13(1)：127～133

4 Theobald C M. Generalizalizations of mean square error applied to ridge regression. J Royal Statist So. B. 1974, 36(1)：103～106

5 何良材、何中市. 广义岭回归 Q(c)准则下 k 值选取迭代算法的收敛性定理及极限. 数理统计与应用概率. 1992, 7(2)：197～206

6 何中市. 岭回归分析及岭回归技术. 1990. 重庆大学硕士学位论文. 26～42

7 He Liangcai(何良材), He Zhongshi(何中市). The convergence Thdorem and the limit kr iterative algorithm for choosing matrix K under criteria Q(c) in the generalized ridge regression in Wang Xuereng. The Fourth China-Japan Symposium on Statistics, Kunming：Yunnan University, 1991, 124-127.