

岭回归估计的向量参数方法

朱尚伟

(山西财经大学应用数学学院, 太原, 030006)

李景华*

(山西财经大学统计学院, 太原, 030006)

摘要: 本文提出岭回归估计的向量参数方法, 选择均方误差函数的负梯度方向作为参数向量方向, 根据均方误差与拟合误差的预期约束条件选择确定参数向量模长. 文中获得了两个单调性结论, 向量参数岭回归估计的均方误差是参数向量模长的单调减函数, 而拟合误差是参数向量模长的单调增函数. 基于两类误差的单调性结论, 本文创建了关于两类误差的预期约束条件, 预期条件约束下的向量参数岭回归方法有望成为兼备均方误差次优与拟合误差适度的双赢估计. 文章最后是一个应用实例.

关键词: 方向; 模长; 均方误差; 拟合误差; 期望约束

中图分类号: F224.0

英文引用格式: ZHU S W, LI J H. Vector parameter for ridge regression [J]. Chinese J Appl Probab Statist, 2018, 34(5): 501-514. (in Chinese)

§1. 引言

岭回归估计出现于上个世纪六十年代, 由于比最小二乘估计具有更小的均方误差, 而在多重共线性模型的估计中具有重要的意义. 文献中岭回归参数可见两种介入方式, 其一是单一参数的纯量矩阵形式 (本文称之为纯量参数岭回归), 其二是数组参数的对角矩阵形式 (本文称之为数组参数岭回归, 部分文献中亦名广义岭回归). 关于岭回归参数的确定方法, 文献中可见的有:

1) 岭迹分析方法^[1,2]. 岭迹分析方法只适用于纯量参数岭回归估计, 其唯一的岭回归参数通过对岭迹曲线的平缓性判断做出选择, 在岭回归实践中最为常见. 但是, 如何描述岭迹曲线的平缓, 所描述的平缓是否一定存在, 岭迹函数性态与均方误差关系等问题却一直未见有逻辑论证.

2) 方差界值方法^[3]. 方差界值方法通过先验取定的方差界值选择确定岭回归参数, 文献中可见其在纯量参数岭回归估计中的讨论与应用. 但是作为有偏估计, 方差只是均方误差的一部分, 仅仅参照方差界值决定岭回归估计其片面性明显受到质疑.

3) 极值模拟方法^[4,5]. 极值模拟方法通过模拟均方误差函数的极小值选择确定岭回归参数, 文献中可见其在数组参数岭回归估计中的讨论与应用. 均方误差的单目标最优优化受

*通讯作者, E-mail: ljh323@126.com.

本文 2017 年 2 月 23 日收到, 2017 年 7 月 21 日收到修改稿.

到质疑. 多重共线性问题使得最小二乘估计是携同其拟合误差的最优性被舍弃的, 作为后继的岭回归估计是否一定要谋求另一个极端均方误差的最优性呢? 并且作为数组参数方法, 极值模拟过程的高维度决策也是不可避免的.

多重共线性模型的最小二乘估计所遭遇的是大均方误差问题, 以上岭回归参数的决策方法都是在方差或均方误差目标下寻求改进. 但是, 回归分析的研究基础是统计观测, 估计方法所面对的应该是两类误差. 其一是估计量的均方误差, 以一个样本的估计所得作为对真实规律的认知, 估计量必须有足够优秀的均方误差, 它为估计结果能够用以分析与预测提供正当性理由. 其二是估计结果与所使用样本的拟合误差, 样本数据对研究者而言是需要尊重的历史真实, 估计结果与样本数据的高度拟合, 为研究结论能够以实证冠名提供正当性理由. 没有拟合程度保证就不是实证研究, 也就没有回归分析. 如同经典线性回归框架下, Gauss 与 Markov 为具备最小拟合误差的最小二乘估计奠定均方误差性质, 岭回归方法在解决均方误差问题时, 也有必要为岭回归估计保证拟合优度.

本文提出岭回归估计的向量参数方法. 选取均方误差函数的负梯度方向为参数向量的方向, 参照关于均方误差与拟合误差的预期条件选择确定参数向量的模长. 文中所创建的误差预期条件表达了研究者关于均方误差的改进程度预期以及关于拟合误差的损失程度预期, 预期条件约束下的向量参数岭回归估计有望同时具备令人满意的均方误差与可以接受的拟合误差.

本文以 I_m 表示 m 阶单位矩阵, 或者简记为 I ; 以 $\text{diag}(c_1, c_2, \dots, c_m)$ 表示以 c_1, c_2, \dots, c_m 为对角元的 m 阶对角矩阵, 或者简记为 $\text{diag}(c_i)$; 以 1_m 记 m 维分量 1 向量 $(1, 1, \dots, 1)^T$; 以 $\|w\| = \sqrt{\sum_i w_i^2}$ 表示向量 w 的模长. 随机变量 ξ 的方差记为 $\text{Var } \xi$; 随机向量 U 的协方差矩阵记为 $\text{Cov}(U)$. 估计量 $\hat{\alpha}$ 的均方误差记为 $\text{MSE } \hat{\alpha}$, 其中 $\hat{\alpha}$ 为纯量或者向量. 所有符号均在第一次出现时给予说明.

本文中关于线性回归模型的全部经典假定都成立. 并且由于本文涉及多重共线性问题, 模型包含至少两个或者两个以上解释变量.

§2. 向量参数岭回归估计

2.1 线性回归模型与最小二乘估计

考虑线性回归模型:

$$Y = \beta_0 1_n + X\beta + U, \quad (1)$$

其中 $Y = (Y_1, Y_2, \dots, Y_n)^T$ 为被解释变量的观测向量; $X_{n \times k} = (X_1, X_2, \dots, X_k)$ 为解释变量数据矩阵, $X_i = (X_{i1}, X_{i2}, \dots, X_{in})^T$ 为解释变量 X_i 的数据向量; β_0 为模型截距项参数, $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ 为模型斜率项参数向量; $U = (U_1, U_2, \dots, U_n)^T$ 为随机扰动向量, n 为

样本容量.

记 $\bar{X}_i = n^{-1} \sum_m X_{im}$ 为 X_i 的样本数据均值, $x_i = X_i - \bar{X}_i 1_n$ 为 X_i 的中心化数据向量, $\tilde{x}_i = x_i / \|x_i\|$ 为 X_i 的标准化数据向量, $i = 1, 2, \dots, k$. 并记中心化数据阵 $x_{n \times k} = (x_1, x_2, \dots, x_k)$, 标准化数据阵 $\tilde{x}_{n \times k} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k)$, $M = \text{diag}(\|x_i\|)$. 可以验证, 模型 (1) 的最小二乘估计如下*:

$$\hat{\beta} = M^{-1}(\tilde{x}^T \tilde{x})^{-1} \tilde{x}^T Y, \quad (2)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 - \dots - \hat{\beta}_k \bar{X}_k. \quad (3)$$

2.2 向量参数岭回归估计

我们称

$$\check{\beta}(z) = (\check{\beta}_j(z_j))_{k \times 1}, \quad (4)$$

$$\check{\beta}_0(z) = \bar{Y} - \check{\beta}_1(z_1) \bar{X}_1 - \check{\beta}_2(z_2) \bar{X}_2 - \dots - \check{\beta}_k(z_k) \bar{X}_k \quad (5)$$

为 (1) 的向量参数岭回归估计, 其中 $\check{\beta}(z)$ 的分量

$$\check{\beta}_j(z_j) = e_j^T M^{-1}(\tilde{x}^T \tilde{x} + z_j I)^{-1} \tilde{x}^T Y, \quad (6)$$

e_j 为 k 维单位向量, $j = 1, 2, \dots, k$; 岭回归参数向量 $z = (z_1, z_2, \dots, z_k)^T > 0$ 待定.

有必要指出, 向量参数岭回归估计式中 $\check{\beta}(z)$ 的每一个分量 $\check{\beta}_j(z_j)$, 有只属于自己的岭回归参数 z_j . 这一点与纯量参数岭回归估计不同, 纯量参数岭回归估计公式

$$\tilde{\beta}(\lambda) = M^{-1}(\tilde{x}^T \tilde{x} + \lambda I)^{-1} \tilde{x}^T Y \quad (7)$$

中 $\tilde{\beta}(\lambda)$ 的全体分量共同拥有同一个参数 λ . 这一点也与数组参数岭回归估计不同, 数组参数岭回归估计公式

$$\dot{\beta}(z_1, z_2, \dots, z_k) = M^{-1}[\tilde{x}^T \tilde{x} + \text{diag}(z_1, z_2, \dots, z_k)]^{-1} \tilde{x}^T Y \quad (8)$$

中 $\dot{\beta}(z_1, z_2, \dots, z_k)$ 的全体分量共同拥有同一个参数小组 z_1, z_2, \dots, z_k .

三种形式的岭回归公式显示:

$$\check{\beta}(z)|_{z=(0,0,\dots,0)^T} = \dot{\beta}(0,0,\dots,0) = \tilde{\beta}(0) = \hat{\beta}.$$

它们在各自的参数零点处都等同于最小二乘估计.

*文献中通常可见的公式是 $\hat{b} = (\tilde{x}^T \tilde{x})^{-1} \tilde{x}^T Y$, 其中 \hat{b} 是 $M\beta$ 的估计, 二者的关系是 $\hat{\beta} = M^{-1}\hat{b}$. 采用形式 (2) 便于本文所进行的讨论.

§3. 向量参数岭回归估计的均方误差

3.1 均方误差

首先证明以下引理.

引理 1 记 $\lambda_1, \lambda_2, \dots, \lambda_k$ 为正定矩阵 $\tilde{x}^\top \tilde{x}$ 的特征值, q_1, q_2, \dots, q_k 为相应的单位正交特征向量组. 记 $M^{-1}q_i q_i^\top M^{-1} = H(i) = (h_{ms}(i))_{k \times k}$, $i = 1, 2, \dots, k$, 并且 $Q = (q_1, q_2, \dots, q_k)$, 则

$$\text{Cov}[\tilde{\beta}(\lambda)] = \sigma^2 \sum_i \frac{\lambda_i}{(\lambda_i + \lambda)^2} H(i), \quad (9)$$

$$\mathbb{E}\tilde{\beta}(\lambda) - \beta = -\lambda M^{-1}Q \text{diag}\left(\frac{1}{\lambda_i + \lambda}\right)Q^\top M\beta. \quad (10)$$

证明: 记

$$D = \begin{pmatrix} 1 & & \\ & M & \end{pmatrix}, \quad G = \begin{pmatrix} 1 & \bar{X}_1 & \dots & \bar{X}_k \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

容易验证

$$(1_n, X) = (1_n, x)G = (1_n, \tilde{x})DG. \quad (11)$$

在 (7) 式中代入 (1) 与 (11) 有

$$\begin{aligned} \tilde{\beta}(\lambda) &= M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \left[(1_n, X) \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix} + U \right] \\ &= M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \left[(1_n, \tilde{x})DG \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix} + U \right] \\ &= M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \tilde{x} M\beta + M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top U. \end{aligned}$$

于是

$$\mathbb{E}\tilde{\beta}(\lambda) = M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \tilde{x} M\beta. \quad (12)$$

计算

$$\begin{aligned} \text{Cov}[\tilde{\beta}(\lambda)] &= \mathbb{E}[\tilde{\beta}(\lambda) - \mathbb{E}\tilde{\beta}(\lambda)][\tilde{\beta}(\lambda) - \mathbb{E}\tilde{\beta}(\lambda)]^\top \\ &= M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \mathbb{E}UU^\top \tilde{x}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} M^{-1} \\ &= \sigma^2 M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \tilde{x}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} M^{-1} \end{aligned}$$

$$\begin{aligned}
&= \sigma^2 M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1}(\tilde{x}^\top \tilde{x} + \lambda I - \lambda I)(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} M^{-1} \\
&= \sigma^2 M^{-1}[(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} - \lambda(\tilde{x}^\top \tilde{x} + \lambda I)^{-2}] M^{-1}.
\end{aligned}$$

注意到 Q 正交, $Q^\top(\tilde{x}^\top \tilde{x})Q = \text{diag}(\lambda_i)$, 可有

$$(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} = Q \text{diag}\left(\frac{1}{\lambda_i + \lambda}\right) Q^\top, \quad (\tilde{x}^\top \tilde{x} + \lambda I)^{-2} = Q \text{diag}\left[\frac{1}{(\lambda_i + \lambda)^2}\right] Q^\top. \quad (13)$$

从而

$$\text{Cov}[\tilde{\beta}(\lambda)] = \sigma^2 M^{-1} Q \text{diag}\left[\frac{\lambda_i}{(\lambda_i + \lambda)^2}\right] Q^\top M^{-1}.$$

协方差矩阵即为

$$\text{Cov}[\tilde{\beta}(\lambda)] = \sigma^2 M^{-1} \left[\sum_i \frac{\lambda_i}{(\lambda_i + \lambda)^2} q_i(q_i)^\top \right] M^{-1} = \sigma^2 \sum_i \frac{\lambda_i}{(\lambda_i + \lambda)^2} H(i).$$

由 (12) 以及 (13), 偏倚向量为

$$\begin{aligned}
E\tilde{\beta}(\lambda) - \beta &= [M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} \tilde{x}^\top \tilde{x} M - I] \beta \\
&= [M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1}(\tilde{x}^\top \tilde{x} + \lambda I - \lambda I) M - I] \beta \\
&= -\lambda M^{-1}(\tilde{x}^\top \tilde{x} + \lambda I)^{-1} M \beta \\
&= -\lambda M^{-1} Q \text{diag}\left(\frac{1}{\lambda_i + \lambda}\right) Q^\top M \beta. \quad \square
\end{aligned}$$

籍此引理可获取 $\check{\beta}_j(z_j)$ 的均方误差函数.

定理 2 记 $M^{-1}Q = A = (a_{ms})_{k \times k}$, $Q^\top M \beta = \gamma = (\gamma_m)_{k \times 1}$, 则 $\check{\beta}_j(z_j)$ 有均方误差:

$$\text{MSE} \check{\beta}_j(z_j) = \sigma^2 \sum_i \frac{\lambda_i h_{jj}(i)}{(\lambda_i + z_j)^2} + z_j^2 \sum_{s,m} \frac{a_{js} a_{jm} \gamma_s \gamma_m}{(\lambda_s + z_j)(\lambda_m + z_j)}, \quad j = 1, 2, \dots, k. \quad (14)$$

证明: 根据定义 (6) 与 (7) 可知

$$\check{\beta}_j(\cdot) = \tilde{\beta}_j(\cdot), \quad j = 1, 2, \dots, k.$$

由 (9) 可有

$$\text{Var} \check{\beta}_j(z_j) = \sigma^2 \sum_i \frac{\lambda_i h_{jj}(i)}{(\lambda_i + z_j)^2}, \quad j = 1, 2, \dots, k.$$

再由 (10) 可有

$$\begin{aligned}
&E[E\check{\beta}_j(z_j) - \beta_j]^2 \\
&= z_j^2 e_j^\top M^{-1} Q \text{diag}\left(\frac{1}{\lambda_i + z_j}\right) Q^\top M \beta \beta^\top M Q \text{diag}\left(\frac{1}{\lambda_i + z_j}\right) Q^\top M^{-1} e_j \\
&= z_j^2 (a_{j1}, a_{j2}, \dots, a_{jk}) \left[\text{diag}\left(\frac{1}{\lambda_i + z_j}\right) \gamma \right] \left[\text{diag}\left(\frac{1}{\lambda_i + z_j}\right) \gamma \right]^\top (a_{j1}, a_{j2}, \dots, a_{jk})^\top \\
&= z_j^2 \sum_{s,m} \frac{a_{js} a_{jm} \gamma_s \gamma_m}{(\lambda_s + z_j)(\lambda_m + z_j)}.
\end{aligned}$$

两项相加, 即得均方误差式 (14). \square

3.2 均方误差性质

定理 3 存在 $\delta_j > 0$ 使得 $\text{MSE} \check{\beta}_j(z_j)$ 在 $z_j \in [0, \delta_j]$ 上单调减少.

证明: 记 $a = \min\{\lambda_i\}$, 由 $\tilde{x}^\top \tilde{x}$ 正定, 知 $a > 0$, $\text{MSE} \check{\beta}_j(z_j)$ 在区间 $(-a, +\infty)$ 内连续可微,

$$\begin{aligned} \frac{d \text{MSE} \check{\beta}_j(z_j)}{dz_j} = & -2\sigma^2 \sum_i \frac{\lambda_i h_{jj}(i)}{(\lambda_i + z_j)^3} + 2z_j \sum_{s,m} \frac{a_{js} a_{jm} \gamma_s \gamma_m}{(\lambda_s + z_j)(\lambda_m + z_j)} \\ & - z_j^2 \sum_{s,m} \frac{a_{js} a_{jm} \gamma_s \gamma_m (2z_j + \lambda_s + \lambda_m)}{(\lambda_s + z_j)^2 (\lambda_m + z_j)^2}, \end{aligned}$$

且在零点处有导数值:

$$\left. \frac{d \text{MSE} \check{\beta}_j(z_j)}{dz_j} \right|_{z_j=0} = -2\sigma^2 \sum_i \frac{h_{jj}(i)}{\lambda_i^2}. \quad (15)$$

根据 $H(i)$ 定义, 对于一切 $i, j = 1, 2, \dots, k$,

$$h_{jj}(i) = \frac{q_{ji}^2}{\|x_j\|^2} \geq 0.$$

又由于 Q 为正交特征向量矩阵, 对一切 $j = 1, 2, \dots, k$, 可有 $\sum_i h_{jj}(i) > 0$. 于是,

$$\left. \frac{d \text{MSE} \check{\beta}_j(z_j)}{dz_j} \right|_{z_j=0} < 0.$$

而又 $\text{MSE} \check{\beta}_j(z_j)$ 连续可微, 存在 $\delta_j > 0$ 使得

$$\frac{d \text{MSE} \check{\beta}_j(z_j)}{dz_j} < 0, \quad z_j \in [0, \delta_j],$$

$\text{MSE} \check{\beta}_j(z_j)$ 在 $[0, \delta_j]$ 上单调减少. \square

记向量参数岭回归估计的均方误差函数为

$$\begin{aligned} F(z) = \text{MSE} \check{\beta}(z) &= \sum_j \text{MSE} \check{\beta}_j(z_j) \\ &= \sum_j \sum_i \frac{\sigma^2 \lambda_i h_{jj}(i)}{(\lambda_i + z_j)^2} + \sum_j \sum_{s,m} \frac{z_j^2 a_{js} a_{jm} \gamma_s \gamma_m}{(\lambda_s + z_j)(\lambda_m + z_j)}, \quad z \in \prod [0, \delta_j]. \end{aligned} \quad (16)$$

式 (15) 以及定理 3 表明 $F(z)$ 在零点处有梯度向量

$$\nabla F(0) = -2\sigma^2 \left[\sum_i \frac{h_{11}(i)}{\lambda_i^2}, \sum_i \frac{h_{22}(i)}{\lambda_i^2}, \dots, \sum_i \frac{h_{kk}(i)}{\lambda_i^2} \right] < 0. \quad (17)$$

由此即可有以下推论:

推论 4 $F(z)$ 在每一个正方向 $\rho^\top = (\rho_1, \rho_2, \dots, \rho_k) > 0$ 上单调减少; 特别地 $F(z)$ 在对角方向 $\nu = 1_k / \|1_k\|$ 上单调减少.

记 $\mu = -\nabla F(0) / \|\nabla F(0)\| > 0$ 为 $F(z)$ 在零点的负梯度方向, 则 μ 方向向量参数岭回归估计有以下性质.

定理 5 对于每一个正方向 $\rho^\top = (\rho_1, \rho_2, \dots, \rho_k) > 0$ ($\rho \neq \mu$), 存在 $\theta_\rho > 0$ 使得对一切 $\lambda \in (0, \theta_\rho]$, 成立 $F(\lambda\mu) < F(\lambda\rho)$.

证明: 对于每一个正方向 $\rho^\top = (\rho_1, \rho_2, \dots, \rho_k) > 0$ ($\rho \neq \mu$), 考虑 $g(\lambda) = F(\lambda\mu) - F(\lambda\rho)$. $g(\lambda)$ 连续可微且

$$g'(\lambda) = [\nabla F(\lambda\mu)]^\top \mu - [\nabla F(\lambda\rho)]^\top \rho.$$

由 $\rho \neq \mu$, $\mu^\top \rho < 1 = \mu^\top \mu$, 可有

$$g'(0) = -\|\nabla F(0)\|(\mu^\top \mu - \mu^\top \rho) < 0.$$

由于 $g'(\lambda)$ 连续, 存在 $\theta_\rho > 0$, 使 $\rho_j \theta_\rho < \delta_j$, $j = 1, 2, \dots, k$, 有

$$g'(\lambda) < 0, \quad \lambda \in [0, \theta_\rho].$$

$g(\lambda)$ 在 $[0, \theta_\rho]$ 上单调减少. 注意到 $g(0) = 0$, 即有

$$g(\lambda) = F(\lambda\mu) - F(\lambda\rho) < 0, \quad \lambda \in (0, \theta_\rho]. \quad \square$$

推论 6 在对角方向 ν 上, 存在 $\theta_\nu > 0$ 使 $F(\lambda\mu) < F(\lambda\nu)$, $\lambda \in (0, \theta_\nu]$.

纯量参数岭回归估计是向量参数岭回归估计取对角方向的特例:

$$\check{\beta}(z)|_{z=(\lambda, \lambda, \dots, \lambda)^\top} = \tilde{\beta}(\lambda), \quad \text{MSE } \check{\beta}(z)|_{z=(\lambda, \lambda, \dots, \lambda)^\top} = \text{MSE } \tilde{\beta}(\lambda).$$

本节推论 4 即是纯量参数岭回归估计的原理, 推论 6 则表明 μ 方向向量参数岭回归估计与纯量参数岭回归估计相比可以获得更大的均方误差降幅.

向量参数岭回归估计在形式上与数组参数岭回归估计一样, 都包含 k 个岭回归参数. 但是梯度向量表达式 (17) 表明, 参数向量方向 μ 完全由模型样本先验确定, 只有参数向量模长需要选择决定, 这就使得 μ 方向向量参数岭回归估计本质上是一个 1 维决策过程. 而数组参数岭回归估计则不然, 它的 k 个岭回归参数是在一个优化过程中被同时确定的, 从而是一个 k 维决策过程. 两者相比, μ 方向向量参数岭回归估计明显具有低维度决策优势.

§4. 向量参数岭回归估计的拟合误差

作为非最小二乘估计, 向量参数岭回归估计必然伴随更大的拟合误差. 本文获得了向量参数岭回归估计拟合误差的以下性质.

定理 7 对任何方向 $\rho^\top = (\rho_1, \rho_2, \dots, \rho_k) > 0$ 以及 $\check{e}(\lambda\rho) = Y - \check{Y}(\lambda\rho)$, 存在 $\vartheta_\rho > 0$ 使得 $f(\lambda) = [\check{e}(\lambda\rho)]^\top \check{e}(\lambda\rho)$ 在 $[0, \vartheta_\rho]$ 上单调增加, 其中 $\check{Y}(\lambda\rho) = \check{\beta}_0(\lambda\rho)1_n + X\check{\beta}(\lambda\rho)$.

证明: 记 $U(\lambda) = (\tilde{x}^\top \tilde{x} + \lambda I)^{-1}$ 以及 $V(\lambda) = \sum_{j=1}^k e_j e_j^\top U(\lambda \rho_j)$. 重写

$$\check{\beta}(\lambda\rho) = \sum_{j=1}^k e_j e_j^\top M^{-1} U(\lambda \rho_j) \tilde{x}^\top Y.$$

由于 $e_j e_j^\top M^{-1} = M^{-1} e_j e_j^\top$, 又可有

$$\check{\beta}(\lambda\rho) = M^{-1} V(\lambda) \tilde{x}^\top Y.$$

记 $y = Y - \bar{Y}1_n$, 并注意到 $xM^{-1} = \tilde{x}$ 以及 $\tilde{x}^\top Y = \tilde{x}^\top y$, 知

$$\begin{aligned} \check{e}(\lambda\rho) &= Y - [1_n \check{\beta}_0(\lambda\rho) + X\check{\beta}(\lambda\rho)] \\ &= Y - [\bar{Y}1_n - \check{\beta}_1(\lambda\rho)\bar{X}_1 1_n - \check{\beta}_2(\lambda\rho)\bar{X}_2 1_n - \dots - \check{\beta}_k(\lambda\rho)\bar{X}_k 1_n + X\check{\beta}(\lambda\rho)] \\ &= y - x\check{\beta}(\lambda\rho) \\ &= y - \tilde{x}V(\lambda)\tilde{x}^\top y. \end{aligned}$$

于是

$$\begin{aligned} f(\lambda) &= y^\top y - 2y^\top \tilde{x}V(\lambda)^\top \tilde{x}^\top y + y^\top \tilde{x}V(\lambda)^\top \tilde{x}^\top \tilde{x}V(\lambda)\tilde{x}^\top y, \\ f'(\lambda) &= -2y^\top \tilde{x}V'(\lambda)\tilde{x}^\top y + 2y^\top \tilde{x}V'(\lambda)^\top \tilde{x}^\top \tilde{x}V(\lambda)\tilde{x}^\top y = 2y^\top \tilde{x}V'(\lambda)^\top [\tilde{x}^\top \tilde{x}V(\lambda) - I]\tilde{x}^\top y. \end{aligned}$$

注意到 $V(0) = U(0) = (\tilde{x}^\top \tilde{x})^{-1}$ 有 $f'(0) = 0$.

进一步还可有

$$f''(\lambda) = 2y^\top \tilde{x}V''(\lambda)^\top [\tilde{x}^\top \tilde{x}V(\lambda) - I]\tilde{x}^\top y + 2y^\top \tilde{x}V'(\lambda)^\top \tilde{x}^\top \tilde{x}V'(\lambda)\tilde{x}^\top y$$

以及

$$f''(0) = 2y^\top \tilde{x}V'(0)^\top \tilde{x}^\top \tilde{x}V'(0)\tilde{x}^\top y = 2(\tilde{x}^\top y)^\top [V'(0)^\top \tilde{x}^\top \tilde{x}V'(0)](\tilde{x}^\top y).$$

根据预解算子理论, 可以验证

$$V'(\lambda) = -\sum_{j=1}^k \rho_j e_j e_j^\top U(\lambda \rho_j)^2, \quad V'(0) = -\text{diag}(\rho_j), \quad U(0)^2 = -\text{diag}(\rho_j)(\tilde{x}^\top \tilde{x})^{-2}.$$

记 $B_{n \times k} = \tilde{x}V'(0)$. 则 $r(B) = r(\tilde{x}) = k$ 且 $B^T B$ 正定. 于是,

$$f''(0) = 2(\tilde{x}^T y)^T B^T B(\tilde{x}^T y) > 0.$$

由于 $f''(\lambda)$ 连续, 知存在 $\vartheta_\rho > 0$ 使得

$$f''(\lambda) > 0, \quad \lambda \in [-\vartheta_\rho, \vartheta_\rho].$$

由此,

$$f'(\lambda) = f'(0) + \int_0^\lambda f''(s)ds > 0 = \int_0^\lambda f''(s)ds > 0, \quad \lambda \in [0, \vartheta_\rho].$$

$f(\lambda)$ 在 $[0, \vartheta_\rho]$ 上单调增加. \square

这里顺便指出, 无论是纯量参数岭回归估计, 还是数组参数岭回归估计, 文献中都未见它们拟合误差性质的讨论.

§5. 向量参数岭回归估计实施

5.1 两个预期约束条件

之前讨论已经获知:

- 1) 向量参数岭回归估计的均方误差在 μ 方向单调减少, 并可获得最大降幅 (见定理 5);
- 2) 向量参数岭回归估计的拟合误差在 μ 方向单调增加 (见定理 7);
- 3) 作为参数向量零点的最小二乘估计具有最大均方误差与最小拟合误差.

籍此我们提出两个预期约束条件, 以选择决定 μ 方向岭回归估计的参数向量模长.

关于均方误差我们提出以下的降幅预期条件:

$$\begin{aligned} r_1 &= \left| \left[-2\sigma^2 \sum_j \sum_i \frac{\lambda \mu_j h_{jj}(i)}{\lambda_i^2} \right] / \left[\sigma^2 \sum_j \sum_i \frac{h_{jj}(i)}{\lambda_i} \right] \right| \\ &= \left| \left[-2 \sum_j \sum_i \frac{\lambda \mu_j h_{jj}(i)}{\lambda_i^2} \right] / \left[\sum_j \sum_i \frac{h_{jj}(i)}{\lambda_i} \right] \right| \geq \varepsilon_1, \end{aligned}$$

其中第一个分式的分母是最小二乘估计的均方误差 (见 (9)), 分子是由微分替代的 μ 方向岭回归估计的均方误差增量 (见 (15)), r_1 成为向量参数岭回归估计的均方误差降幅. 不等式右手的 ε_1 则是由建模者认同的降幅预期.

由于岭回归方法是针对多重共线性模型的估计, 共线性问题的严重程度成为均方误差降幅预期的首要提示. 共线性问题越严重, 建模者会对均方误差降幅有越高预期. 基于最小二乘估计的诸多指标能够明确昭示共线性问题的严重性, 诸如方差扩大化因子, 相关系数, 标准差估值等, 都可以成为均方误差降幅预期的重要参照指标.

关于拟合误差我们提出以下的增幅预期条件:

$$r_2 = \left| \sqrt{\frac{\tilde{e}^T \tilde{e}(\lambda \mu)}{n}} / |Y| - \sqrt{\frac{\hat{e}^T \hat{e}}{n}} / |Y| \right| \leq \varepsilon_2,$$

其中不等式左边是 μ 方向岭回归估计的拟合误差率与最小二乘估计拟合误差率之差, r_2 成为向量参数岭回归估计的拟合误差率增幅. 不等式右边的 ε_2 则是由建模者认同的增幅预期.

拟合误差均值与因变量样本均值的比率具有清晰的拟合误差率意义 (见文献 [6]), 1%–3% 左右的误差率水平是常识可以接受的. 考虑到最小二乘估计的最佳拟合效果, 建模者可以参照其拟合优度设置这里的增幅预期. 最小二乘估计的拟合优度越高, 建模者的拟合误差率增幅预期可以越为宽松.

岭回归估计是多重共线性条件下关于最小二乘估计的改良方法, “改良” 性质使得最小二乘估计客观上成为岭回归估计的参照标准. 以上的预期条件构造充分体现了这一点, ε_1 , ε_2 的赋值也因此而带有了某种先验性质. 向量参数岭回归过程中, 均方误差预期与拟合误差预期应该是相互参照折衷选定的. 建模者可以先行确定具有明确预期的拟合误差率增幅区间, 随后再折衷选择自己的均方误差降幅预期. 本文关于误差函数的单调性结论, 为这种均衡抉择提供了可行性保障.

5.2 设置 λ

多重共线性问题使得矩阵 $\tilde{x}^T \tilde{x}$ 存在接近于零的最小特征值, 而这直接导致了最小二乘估计的大均方误差 (见文献 [7]). 各类岭回归方法也因此最小二乘公式中以矩阵 $\tilde{x}^T \tilde{x} + z_j I$, $\tilde{x}^T \tilde{x} + \lambda I$ 或者 $\tilde{x}^T \tilde{x} + \text{diag}(z_1, z_2, \dots, z_k)$ 的逆运算取代矩阵 $\tilde{x}^T \tilde{x}$ 的逆运算 (参见 (6), (7), (8)). 岭回归参数的出现使得逆运算矩阵特征值获得增量, 大均方误差问题因而有望得以改善.

由于本文两个误差函数的单调性结论是在参数向量模长零点右边邻域的性质 (定理 5, 定理 7), 岭回归计算的模长起点不能离开零点太远. 并且还由于不同的模型岭迹函数性态的差异, 其计算设置步长也不宜太大. 我们建议参照最小特征值 $\min\{\lambda_i\}$ 为 λ 赋值. 为了争取两类误差的满意结果, 应用实践中 λ 的赋值常常需要根据具体模型灵活变化与调整.

5.3 操作建议

- 1) 计算 $\|x_i\|$, λ_i , q_i , $H(i)$, $i = 1, 2, \dots, k$;
- 2) 计算

$$\frac{1}{2\sigma^2} \frac{\partial F}{\partial z_j} \Big|_{z_j=0} = - \sum_i \frac{h_{jj}(i)}{\lambda_i^2}, \quad j = 1, 2, \dots, k,$$

获取向量 μ ;

3) 试探性设置系列 λ 数值, 计算相应的 $\check{\beta}_j(\lambda\mu_j)$, $\check{\beta}_0(\lambda\mu)$;

4) 计算均方误差降幅:

$$r_1 = \left| \left[2 \sum_j \sum_i \frac{\lambda\mu_j h_{jj}(i)}{\lambda_i^2} \right] / \left[\sum_j \sum_i \frac{h_{jj}(i)}{\lambda_i} \right] \right|;$$

5) 计算拟合误差增幅:

$$r_2 = \left| \sqrt{\frac{\check{e}^\tau \check{e}(\lambda\mu)}{n}} / |\overline{Y}| - \sqrt{\frac{\widehat{e}^\tau \widehat{e}}{n}} / |\overline{Y}| \right|;$$

6) 根据可以接受的预期约束条件最终选择确定 λ .

§6. 一个应用实例

本文提供一个应用研究示例. 为了提高拟合程度, 模型采用中间变量形式进行了数据转换. 原始数据源自《中国金融年鉴》, 2010 经济与社会发展统计公报以及美国劳工统计网站. 全部计算通过 E-views 完成.

Y = 人民币实际有效汇率指数;

$X_1 = \text{PROD}^{0.01} + 35$, 其中 PROD 为中国相对劳动生产率;

$X_2 = (\text{TOT} - 0.96)^4 + 35$, 其中 TOT 为贸易条件;

$X_3 = (\text{OPEN} - 0.2)^2 + 35$, 其中 OPEN 贸易开放度;

$X_4 = (G - 0.1)^2 + 30$, 其中 G 为政府支出比率;

$X_5 = \text{RM2}$, 其中 RM2 为实际货币供应量.

模型最小二乘估计如下:

$$\hat{Y} = -217021.7 + 2341.6X_1 + 1205.6X_2 + 532.63X_3 + 2400.9X_4 - 14.12X_5.$$

$$(737.82) \quad (363.79) \quad (248.78) \quad (402.41) \quad (7.95)$$

$$(3.17) \quad (3.31) \quad (2.14) \quad (5.97) \quad (-1.78)$$

$$R^2 = 0.9182, \quad \text{SE} = 7.44, \quad \text{DW} = 2.07, \quad F = 44.92, \quad n = 26.$$

回归报告显示, 全部 $\hat{\beta}_i$ 有大标准差, X_1 与 X_5 线性相关系数高达 $r_{x_1x_5} = 0.87$. 模型存在多重共线性问题. 本文以此例演示向量参数岭回归估计.

计算 $\tilde{x}^\tau \tilde{x}$ 矩阵特征值:

$$0.086413, \quad 0.403070, \quad 0.874906, \quad 0.977287, \quad 2.658323.$$

试探性设置 $\lambda = 0, 0.01, \dots, 0.09$, 计算 $\check{\beta}_j$ 以及相应的预期约束比率 $r(1)$ 与 $r(2)$, 列表如下:

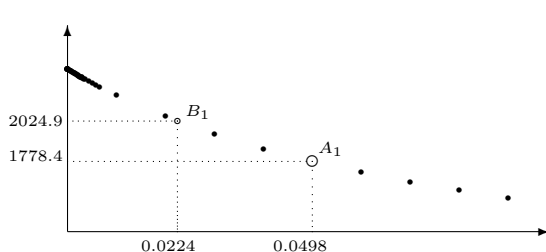
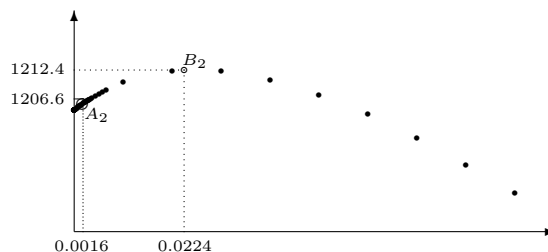
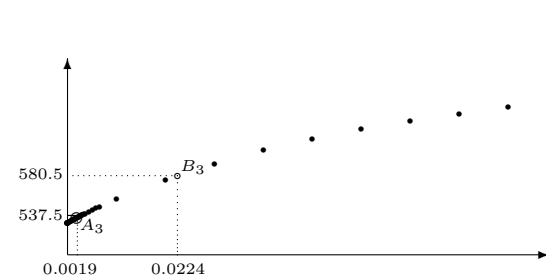
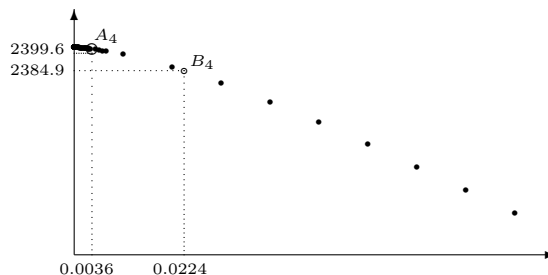
表 1 $\check{\beta}_j$ 以及相应的预期约束比率 $r(1)$ 与 $r(2)$

λ	$\check{\beta}_1$	$\check{\beta}_2$	$\check{\beta}_3$	$\check{\beta}_4$	$\check{\beta}_5$	$\check{\beta}_0$	r_1	r_2
0	2341.6	1205.6	532.63	2400.9	-14.12	-217021.7	0	0
0.01	2182.4	1205.85	533.62	2400.7	-14.12	-211329.3	0.1360	0.0009
0.02	2053.0	1206.0	534.60	2400.5	-14.12	-206710.5	0.2720	0.0029
0.03	1945.8	1206.2	535.58	2400.2	-14.12	-202888.5	0.4080	0.0054
0.04	1855.5	1206.4	536.55	2399.9	-14.12	-199673.8	0.5440	0.0080
0.05	1778.4	1206.6	537.51	2399.6	-14.12	-196932.6	0.6801	0.0106
0.06	1711.8	1206.8	538.46	2399.3	-14.12	-194567.3	0.8161	0.0132
0.07	1653.7	1207.0	539.40	2398.9	-14.12	-192505.5	0.9521	0.0155
0.08	1602.5	1207.2	540.34	2398.5	-14.12	-190692.1	1.0881	0.0178
0.09	1557.1	1207.3	541.27	2398.1	-14.12	-189084.4	1.2241	0.0198

列表显示, 参数向量模长为 $\lambda = 0.05$ 所决定的岭回归估计, 均方误差比最小二乘估计下降 68.01%, 拟合误差优度比最小二乘估计损失 1.06%. 如果这样的均方误差降幅与拟合优度损失符合研究者预期, 向量参数岭回归函数即为

$$\check{Y} = -196932.6 + 1778.4X_1 + 1206.6X_2 + 537.5X_3 + 2399.6X_4 - 14.12X_5.$$

作为比较, 这里给出本模型的主流岭回归估计 (MATLAB 与 SPSS 的岭回归程序). 它具有纯量岭回归参数, 并且按照岭迹分析方法确定岭回归参数. 借助 10×5 ($\lambda \times \mu$) 个试算岭回归参数, 可以获得 5 条岭迹曲线. 但是主流岭回归估计遭遇岭迹分析方法的困境, 5 条岭迹曲线并没有呈现共同平稳的趋势.

Figure $\check{\beta}_1$ Figure $\check{\beta}_2$ Figure $\check{\beta}_3$ Figure $\check{\beta}_4$

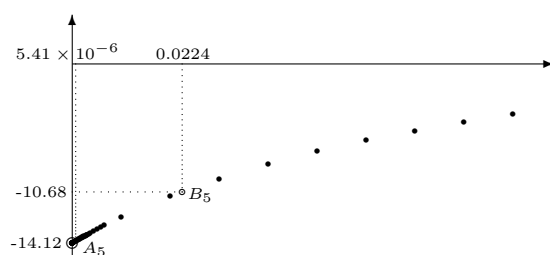
Figure β_5

图 1 5 条岭迹曲线

如果视主流岭回归估计为向量参数岭回归估计的特例, 补充计算对角方向模长 0.05 的参数设置值 $0.05 \times (1/\sqrt{5}) = 0.0224$, 可获得纯量参数岭回归函数:

$$\tilde{Y} = -207069.0 + 2024.9X_1 + 1212.4X_2 + 580.5X_3 + 2384.9X_4 - 10.68X_5.$$

主流岭回归估计的全部 $\tilde{\beta}_j$ (图中 B_j) 具有相同横坐标, 位于 5 条曲线的第 43 与 44 点之间. 而向量参数岭回归估计的全部 $\check{\beta}_j$ (图中 A_j) 按顺序为第 46 (A_1), 25 (A_2), 26 (A_3), 37 (A_4), 10 (A_5). 补充计算 $\tilde{r}_1 = 30.41\%$, 可见主流岭回归估计的均方误差改善程度明显劣于向量参数岭回归估计. 而又向量参数岭回归估计仅仅比最小二乘估计增加不超过 1.1% 的拟合误差, 权衡考虑的结果, 向量参数岭回归估计是更好的选择.

对比以上示例模型三种估计方法的三个样本回归函数, 可见两处显著差异. 一是 β_1 估值的显著差异, 它关系到有效汇率指数关于相对劳动生产率的边际关系, 经由最小二乘估计所获得的边际函数是向量参数岭回归估计的 1.3167 倍. 二是 β_5 估值的显著差异, 它关系到有效汇率指数关于实际货币供给量的边际系数, 经由主流岭回归估计所获得的边际系数是向量参数岭回归估计的 0.7559 倍. 基于向量参数岭回归方法在示例模型中所表现出的误差优势, 我们认为最小二乘估计严重高估了相对劳动生产率对有效汇率指数的影响强度, 而主流岭回归估计则严重低估了实际货币供给量对有效汇率指数的影响强度.

§7. 结 束 语

众所周知, 最小二乘估计具有与生俱来的最小拟合误差, 加之以 Gauss-Markov 条件后, 它同时又具备了线性无偏估计类中的最小均方误差. 但是多重共线性问题致使线性无偏估计类整体丧失了均方误差优势, 线性有偏的各类岭回归方法应运而生. 它们顺势在方差或均方误差目标下寻求改进, 却没有将拟合目标纳入旗下. 本文的向量参数岭回归方法较好地解决了两类误差的双重目标问题, 它基于对两类误差所获得的单调性结论, 通过预

期约束条件进行双误差目标下的均衡抉择. 向量参数岭回归估计可望是一种兼备均方误差次优同时拟合误差适度的双赢估计.

参 考 文 献

- [1] HOERL A E, KENNARD R W. Ridge regression: biased estimation for nonorthogonal problems [J]. *Technometrics*, 1970, **12**(1): 55–67.
- [2] HOERL A E, KENNARD R W. Ridge regression: applications to nonorthogonal problems [J]. *Technometrics*, 1970, **12**(1): 69–82.
- [3] MARQUARDT D W, SNEE R D. Ridge regression in practice [J]. *Amer Statist*, 1975, **29**(1): 3–20.
- [4] HOERL A E, KANNARD R W, BALDWIN K F. Ridge regression: some simulations [J]. *Comm Statist Theory Methods*, 1975, **4**(2): 105–123.
- [5] HEMMERLE W J, BRANTLE T F. Explicit and constrained generalized ridge estimation [J]. *Technometrics*, 1978, **20**(2): 109–120.
- [6] 李景华, 朱尚伟. 关于 R^2 的几点质疑 [J]. 数量经济技术经济研究, 2013, **30**(9): 152–160.
- [7] 陈希孺, 王松桂. 近代实用回归分析 [M]. 南宁: 广西人民出版社, 1984.

Vector Parameter for Ridge Regression

ZHU Shangwei

(School of Applied Mathematics, Shanxi University of Finance & Economics, Taiyuan, 030006, China)

LI Jinghua

(School of Statistics, Shanxi University of Finance & Economics, Taiyuan, 030006, China)

Abstract: In this paper, a vector parameter method for ridge regression is proposed. We choose the negative gradient of mean square error as vector direction and decide vector norm with the expectation constrains both of mean square error and of residual error. We come to conclusions that the mean square error is a decreasing function of vector norm while the residual error a increasing one. It is the monotonicity of the errors that leads to our expectation constrains. Since two conflict constrains are under consideration, our vector parameter ridge regression is expected to bear both satisfactory mean square error and acceptable residual error. Finally, a multi-collinearity model is given as an example.

Keywords: direction; norm; mean square error; residual error; expectation constraint

2010 Mathematics Subject Classification: 62J07