

Minimax线性估计的影响分析*

田 保 光

(山西师范大学数学系, 临汾)

§ 1 引言

考虑线性回归模型

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

这里 X 是 $n \times q$ 列满秩阵, β 是 $q \times 1$ 未知参数, Y 是 $n \times 1$ 观察向量, e 是 $n \times 1$ 随机误差向量. 估计 β 通常采用最小二乘估计 $\hat{\beta}$ (简记为LS估计). 但当设计矩阵 X 有复共线性存在时, $\hat{\beta}$ 的性能很不好. 针对这个问题, 近二十年来, 特别是近十年来, 许多统计学者相继提出了一些新的方法, 企图改进LS估计. 例如, Stein在1960年提出的压缩估计; Massy于1965年提出的主成分估计; Hoerl和Kennard于1970年提出的岭估计等. 这些估计在均方误差的意义下优于LS估计. Kuks和Olman于1972年证明了 β 的Minimax线性估计在均方误差的意义下也优于LS估计, Theobald于1974年证明了在相当广泛的一类损失函数下, Minimax线性估计优于LS估计. 因此, Minimax线性估计是值得感兴趣的. 本文主要研究一组或多组试验数据对Minimax线性估计及拟合值的影响问题, 提出了Minimax线性估计中COOK统计量, W—K统计量和广义方差比. 用这些统计量度量了试验数据对Minimax线性估计和拟合值的影响大小, 得到了一些有实用价值的结果.

§ 2 Minimax线性估计的定义与广义方差比

定义 设 $h > 0$, A 是 $q \times q$ 正定矩阵, a 是 $q \times 1$ 向量

$$S = \{\beta, \sigma: \beta' A \beta / \sigma^2 \leq h\} \quad (2)$$

如果 $\tilde{c}$ 满足 $\min_{C \in R^q, \beta, \sigma \in S} E[(\tilde{c}'Y - a'\beta)^2] = \max_{\beta, \sigma \in S} E[(\tilde{c}'Y - a'\beta)^2]$

我们称 $\tilde{c}'Y$ 是 $a'\beta$ 基于 (2) 的Minimax线性估计. 这里 R^q 是 q 维欧氏空间.

关于Minimax线性估计的具体表达式我们有下面的定理1.

定理1 基于 (2) 的 $a'\beta$ 的Minimax线性估计 $\tilde{c}'Y$ 满足 $\tilde{c}'Y = a'\hat{\beta}_{A,h}$

其中 $\hat{\beta}_{A,h} = [X'X + (1/h)A]^{-1}X'Y$. Kuks和Olman于1972年证明了对任意 $a \in R^q$,

$$E[(a'\hat{\beta}_{A,h} - a'\beta)^2] \leq E_{\beta, \sigma}[(a'\hat{\beta} - a'\beta)^2]$$

* 1987年4月27日收到.

当 β, σ 满足: $\beta' A \beta / \sigma^2 \leq 2h$ 时 (3)

Theobald 1974 年证明了更一般的结果, 当 β, σ 满足 (3) 式时, 对任一非负定阵 $M_{q \times q}$ 有

$$E[(\hat{\beta}_{A,h} - \beta)' M (\hat{\beta}_{A,h} - \beta)] \leq E_{\beta, \sigma}[(\hat{\beta} - \beta)' M (\hat{\beta} - \beta)] \quad (4)$$

这表明当参数满足一定的条件时, 在均方误差的意义下 β 的 Minimax 线性估计 $\hat{\beta}_{A,h}$ 确是优于 LS 估计 $\hat{\beta}$.

为了度量试验数据对 $\hat{\beta}_{A,h}$ 的影响大小, 我们用下列统计量——广义方差比:

$$e(\hat{\beta}_{A,h}) = \frac{|\text{COV}(\hat{\beta}_{A,h})|}{|\text{COV}(\hat{\beta}_{A,h}^{(i)})|} \quad (5)$$

其中 $\hat{\beta}_{A,h}^{(i)} = [X_{(i)}' X_{(i)} + (1/h)A]^{-1} X_{(i)}' Y_{(i)}$ 是剔除第 i 组试验数据 (x_i', y_i) 后得到的 Minimax 线性估计. $X_{(i)}$ 表示剔除第 i 组试验数据后得到的矩阵, $Y_{(i)}$ 是剔除第 i 个分量后的向量.

定理 2 $e(\hat{\beta}_{A,h}) = (1 - h_{ii})^{-1} (1 - h_{ii}^{(A)})^2$

这里 $h_{ii} = x_i' (X' X)^{-1} x_i$, $h_{ii}^{(A)} = x_i' (X' X + (1/h)A)^{-1} x_i$

证明 为方便记 $k = 1/h$ (本文以下都记 $k = 1/h$)

$$\begin{aligned} e(\hat{\beta}_{A,h}) &= \frac{|(X' X + kA)^{-1} X' X (X' X + kA)^{-1}|}{|(X_{(i)}' X_{(i)} + kA)^{-1} X_{(i)}' X_{(i)} (X_{(i)}' X_{(i)} + kA)^{-1}|} \\ &= (1 - h_{ii})^{-1} \frac{|X' X + kA|^{-2}}{|X_{(i)}' X_{(i)} + kA|^{-2}} \\ &= (1 - h_{ii})^{-1} \frac{|X' X + kA|^{-2}}{|(X' X + kA)^{-1} + (X' X + kA)^{-1} x_i x_i' (X' X + kA)^{-1} / (1 - h_{ii}^{(A)})|^2|} \\ &= (1 - h_{ii})^{-1} \frac{1}{|I + x_i x_i' (X' X + kA)^{-1} / (1 - h_{ii}^{(A)})|^2} = (1 - h_{ii})^{-1} (1 - h_{ii}^{(A)})^2 \end{aligned}$$

此定理表明, $e(\hat{\beta}_{A,h})$ 仅依赖于 h_{ii} , $h_{ii}^{(A)}$, 而 $0 < h_{ii}^{(A)} \leq h_{ii} \leq 1$. 因此, 当 h_{ii} 很小时 $e(\hat{\beta}_{A,h})$ 接近 1, 即 $\hat{\beta}_{A,h}$ 与 $\hat{\beta}_{A,h}^{(i)}$ 很接近, 也就是说第 i 组试验数据 (x_i', y_i) 对 $\hat{\beta}_{A,h}$ 的影响小. 反之, 当 $e(\hat{\beta}_{A,h})$ 很小时, 一般说来 $h_{ii}^{(A)}$ 较大, 这时第 i 组试验数据对 $\hat{\beta}_{A,h}$ 的影响大. 从此定理中可简便地计算出一组试验数据对 $\hat{\beta}_{A,h}$ 的影响大小, 因此, 可决定在实际工作中应重视那些试验数据.

其次, 我们考虑一次剔除多组试验数据的情况.

设 $I = \{i_1, \dots, i_m\}$ 表示数据的指标集. (I) 表示剔除指标属于 I 的数据. 如 $X_{(I)}$ 表示 X 中去掉第 $i_1 \dots i_m$ 行后得到的矩阵, $Y_{(I)}$ 表示去掉第 $i_1 \dots i_m$ 个分量得到的 $n - m$ 维向量.

类似剔除一组试验数据时的情况, 我们称

$$e(\hat{\beta}_{A,h}^{(I)}) = \frac{|\text{COV}(\hat{\beta}_{A,h})|}{|\text{COV}(\hat{\beta}_{A,h}^{(I)})|} \quad (6)$$

为剔除多组试验数据时的广义方差比. 其中 $\hat{\beta}_{A,h}^{(I)} = [X_{(I)}' X_{(I)} + kA]^{-1} X_{(I)}' Y_{(I)}$

定理 3 $e(\hat{\beta}_{A,h}^{(I)}) = |I - V_I|^{-1} |I - V_I^{(A)}|^2$ (7)

这里 $X_I = (x_{i_1} \dots x_{i_m})'$, $V_I = X_I (X' X)^{-1} X_I'$, $V_I^{(A)} = X_I (X' X + kA)^{-1} X_I'$

证明 仿定理 2 的证明即得.

此定理给出了度量一组试验数据对 Minimax 线性估计 $\hat{\beta}_{A,h}$ 影响的一个标准. 当 $e(\hat{\beta}_{A,h}^{(I)})$ 接近 1 时, 指标属于 I 的这组试验数据对 $\hat{\beta}_{A,h}$ 的影响不大. 当 $e(\hat{\beta}_{A,h}^{(I)})$ 很小时, 这组试验数据

对 $\hat{\beta}_{A,h}$ 的影响大,可能是影响子集。(7)式的计算也不复杂,因此,这是一个可实用的度量。

§3 COOK 统计量

度量一组试验数据对 $\hat{\beta}_{A,h}$ 的影响大小,最直观的方法是考查 $\hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)}$ 。但 $\hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)}$ 是一向量,有其应用上的不便之处。因此,可用它的某种数量化来度量试验数据对 $\hat{\beta}_{A,h}$ 的影响。我们称下列模:

$$D(M, C) = \frac{1}{C} (\hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)})' M (\hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)}) \quad (8)$$

为 Minimax 线性估计的 COOK 统计量: 其中 $c > 0$ 是常数, M 是 $q \times q$ 正定阵。

$D(M, C)$ 度量了一组试验数据对 $\hat{\beta}_{A,h}$ 影响的大小。由于

$$\begin{aligned} \hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)} &= (X'X + kA)^{-1} X'Y - (X'_{(i)} X_{(i)} + kA)^{-1} X'_{(i)} Y_{(i)} \\ &= (X'X + kA)^{-1} X'Y - [(X'X + kA)^{-1} + \frac{(X'X + kA)^{-1} x_i x_i' (X'X + kA)^{-1}}{1 - h_{ii}^{(A)}}] (X'Y - x_i y_i) \\ &= (X'X + kA)^{-1} x_i [y_i - \frac{x_i' \hat{\beta}_{A,h}}{1 - h_{ii}^{(A)}} + \frac{y_i h_{ii}^{(A)}}{1 - h_{ii}^{(A)}}] \\ &= \frac{(X'X + kA)^{-1} x_i (y_i - x_i' \hat{\beta}_{A,h})}{1 - h_{ii}^{(A)}} = \frac{(X'X + kA)^{-1} x_i e_i^{(A)}}{1 - h_{ii}^{(A)}} \end{aligned} \quad (9)$$

其中 $e_i^{(A)} = y_i - x_i' \hat{\beta}_{A,h}$ 是第 i 个残差。

我们可得:

$$\text{定理 4} \quad D(M, C) = \frac{(e_i^{(A)})^2}{c(1 - h_{ii}^{(A)})^2} x_i' (X'X + kA)^{-1} M (X'X + kA)^{-1} x_i$$

特别, 当 $M = X'X + kA$, $c = q\hat{\sigma}^2$ 时

$$D((X'X + kA), q\hat{\sigma}^2) = \frac{h_{ii}^{(A)} (t_i^{(A)})^2}{q(1 - h_{ii}^{(A)})}$$

这里 $t_i^{(A)} = \frac{e_i^{(A)}}{\hat{\sigma}^2(1 - h_{ii}^{(A)})}$ 是学生化残差 $\hat{\sigma}^2 = \frac{1}{n - q} (Y - X\hat{\beta}_{A,h})' (Y - X\hat{\beta}_{A,h})$

由定理 4 可知 $D(X'X + kA, q\hat{\sigma}^2)$ 是 $h_{ii}^{(A)}$ 的单增函数, 且和第 i 个学生化残差 $t_i^{(A)}$ 成正比。因此, 当 $h_{ii}^{(A)}$ 和 $t_i^{(A)}$ 较大时, $D(X'X + kA, q\hat{\sigma}^2)$ 也较大, 这说明第 i 个试验点 (x_i', y_i) 的影响大。当学生化残差 $t_i^{(A)}$ 和 $h_{ii}^{(A)}$ 较小时, $D(X'X + kA, q\hat{\sigma}^2)$ 也较小, 此时 (x_i', y_i) 的影响小。因此, 使用此度量可方便地确定一组试验数据对 $\hat{\beta}_{A,h}$ 的影响大小, 可找出强影响点和平常点。

§4 W—K 统计量

前两节用广方差比和 COOK 统计量度量了试验数据对估计 $\hat{\beta}_{A,h}$ 的影响。本节从拟合值的观点出发, 用 W—K 统计量研究试验数据对拟合值的影响问题。

$$\begin{aligned} \text{由于 } \text{Var}(x_j' \hat{\beta}_{A,h}) &= x_j' [(X'X + kA)^{-1} X'X (X'X + kA)^{-1}] x_j \sigma^2 \\ &= [x_j' (X'X + kA)^{-1} x_j - x_j' k (X'X + kA)^{-1} A (X'X + kA)^{-1} x_j] \sigma^2 \\ &= (h_{jj}^{(A)} - kW_{jj}) \sigma^2 \end{aligned} \quad (11)$$

这里 $W_{jj} = x_j'(X'X + kA)^{-1}A(X'X + kA)^{-1}x_j$

$$\text{我们称 } W - K = \frac{x_j'(\hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)})}{(\hat{\sigma}\sqrt{h_{ii}^{(A)}} - kW_{jj})} \quad (12)$$

为Minimax线性估计中的 $W-K$ 统计量. 它度量了第 i 组试验数据对 x_j 处拟合值的影响.

$$\text{定理 5} \quad W - K = t_i^{(A)} \left(\frac{h_{ii}}{1 - h_{ii}^{(A)}} \right)^{1/2} \rho(x_j'\hat{\beta}_{A,h}, x_i'\hat{\beta}) \quad (13)$$

其中 ρ 是简单相关系数.

证明 由(9)式知 $\hat{\beta}_{A,h} - \hat{\beta}_{A,h}^{(i)} = (X'X + kA)^{-1}x_i e_i^{(A)} / (1 - h_{ii}^{(A)})$

再利用(11)式, 我们有

$$\begin{aligned} W - K &= \frac{h_{ij}^{(A)} e_i^{(A)}}{(1 - h_{ii}^{(A)}) \hat{\sigma} \sqrt{h_{jj}^{(A)}} - kW_{jj}} \\ &= \frac{e_i^{(A)}}{\hat{\sigma} \sqrt{1 - h_{ii}^{(A)}}} \cdot \frac{\sqrt{h_{ii}^{(A)}}}{\sqrt{1 - h_{ii}^{(A)}}} \cdot \frac{h_{ij}^{(A)}}{\sqrt{h_{ii}^{(A)}} \sqrt{h_{jj}^{(A)}} - kW_{jj}} \end{aligned}$$

又因 $\text{Var}(x_i'\hat{\beta}) = h_{ii}\sigma^2$

$$\text{COV}(x_j'\hat{\beta}_{A,h}, x_i'\hat{\beta}) = x_j'(X'X + kA)^{-1}X'X(X'X)^{-1}x_i\sigma^2 = x_j'(X'X + kA)^{-1}x_i\sigma^2 = h_{ij}^{(A)}\sigma^2$$

$$\text{故 } \rho(x_j'\hat{\beta}_{A,h}, x_i'\hat{\beta}) = \frac{h_{ij}^{(A)}}{\sqrt{h_{ii}^{(A)}} \sqrt{h_{jj}^{(A)}} - kW_{jj}}$$

$$\text{从而 } W - K = t_i^{(A)} \left(\frac{h_{ii}}{1 - h_{ii}^{(A)}} \right)^{1/2} \rho(x_j'\hat{\beta}_{A,h}, x_i'\hat{\beta})$$

从此定理可看出, 第 i 组试验数据对 x_j 处拟合值的影响依赖于学生化残差 $t_i^{(A)}$, h_{ii} , $h_{ii}^{(A)}$ 及 $x_j'\hat{\beta}_{A,h}$ 与 $x_i'\hat{\beta}$ 的相互系数. 影响的大小与这些因子成正比. 影响的上界不超过 $t_i^{(A)} \left(\frac{h_{ii}}{1 - h_{ii}^{(A)}} \right)^{1/2}$.

参 考 文 献

- [1] Lawrence peele and Thomas P. Ryan. "Minimax Linear Regression Estimators with Application to Ridge Regression". Technometrics Vol. 24, No. 2 May (1982) 157—159.
- [2] Casella, G. (1980) "Minimax Ridge Regression Estimation", Annals of statistics, 8, 1036—1056.
- [3] 吴诗泳, 于义良. "W—K统计量与相关系数" (待发表).
- [4] R. R. Hocking "Developments in linear Regression Methodology: 1959—1982" Technometrics Vol. 25, No. 3 (1983).

The Influence Analysis of Minimax Linear Estimator

Tian Baoguang

(Shan xi normal university)

In this paper, we propose the ratio of the generalized variances, cook statistic and $W-K$ statistic as tools to measure the influence of data on the Minimax linear estimator and the fitted value. Their properties were discussed and some practical results were reached. The relation between $W-K$ statistic and correlation coefficient was built.