第16卷第2期　　数学研究与评论　　Vol. 16 No. 2
1996年5月　　JOURNAL OF MATHEMATICAL RESEARCH AND EXPOSITION　　May 1996

多元线性回归置信域的局部影响*

岳 荣 先

(南京气象学院，南京 210044)

摘　要　运用 Cook(1986)的局部影响法评价多元线性回归模型的微小扰动对回归系数置信域的影响，扰动方式包括协方差阵扰动，自变量扰动和因变量扰动.

关键词　模型扰动，诊断，局部影响.

分类号　AMS(1991) 62H12/CCL O212.1

考虑 p 个因变量 $Y_1,\cdots,Y_p$ 关于 $m-1$ 个自变量 $X_1,\cdots,X_{m-1}$ 的线性回归模型[1]

$$\begin{cases}Y = XB+\varepsilon,\\ \varepsilon \sim N_{np}(0,V\otimes I_n),\end{cases}\tag{1}$$

其中 $\underset{n\times p}{Y}=(y_1,\cdots,y_n)'$ 为因变量的 n 次观测数据阵，$\underset{n\times m}{X}=(x_1,\cdots,x_n)'$ 为自变量的设计阵，$\underset{m\times p}{B}=(\beta_0,\cdots,\beta_{m-1})'$ 为回归系数阵，$\underset{n\times p}{\varepsilon}=(\varepsilon_1,\cdots,\varepsilon_n)'$ 为随机误差阵. B 与 V 的极大似然估计(MLE)分别为

$$\hat{B}=(X'X)^{-1}X'Y\sim N_{mp}(B,V\otimes(X'X)^{-1}),$$

$$\hat{V}=\frac{1}{n}(Y-X\hat{B})'(Y-X\hat{B})=\frac{1}{n}Y'(I_n-H)Y,$$

其中 $H=X(X'X)^{-1}X'=(h_{ij})_{n\times n}$. 在现有文献中，对模型(1)的统计诊断工作主要是考虑删除一组或多组数据对回归系数最小二乘估计的影响[2],[3]. 本文采用局部影响法[4]研究模型(1)的微小扰动对 B 其及子集的置信域的影响，扰动包括协方差阵扰动，自变量扰动及因变量扰动. 本文假定协方差阵 V 为多余参数，用 $\hat{V}=\frac{1}{n}Y'(I_n-H)Y$ 代替. 记 $\mathrm{vec}(\hat{B})$ 的协方差阵估计量为

$$\Sigma(\hat{B})=\hat{V}\otimes U^{-1}(\hat{B}),$$

当 $\hat{B}=(X'X)^{-1}X'Y$ 时，$U(\hat{B})=X'X$. B 的渐近置信椭球为

$$C(\hat{B})=\{B:\mathrm{tr}[\hat{V}^{-1}(\hat{B}-B)'U(\hat{B})(\hat{B}-B)]\leqslant\chi^2(mp;1-\alpha)\}.\tag{2}$$

假设(1)受微小扰动，$\omega=(\omega_1,\cdots,\omega_q)'$ 为描述扰动因素的向量，$\hat{B}(\omega)$ 表示扰动后 B 的 MLE，设存在 ω_0 使 $\hat{B}(\omega_0)=\hat{B}$. 记 $\mathrm{vec}(\hat{B}(\omega))$ 的协方差阵估计量为

$$\Sigma(\hat{B}(\omega))=\hat{V}\otimes U^{-1}(\hat{B}(\omega)),$$

则在扰动模型中，B 的渐近置信椭球为

$$C(\hat{B}(\omega))=\{B:\mathrm{tr}[\hat{V}^{-1}(\hat{B}(\omega)-B)'U(\hat{B}(\omega))(\hat{B}(\omega)-B)]\geqslant\chi^2(mp;1-\alpha)\}.\tag{3}$$

* 1993年6月1日收到. 94年7月4日收到修改稿.

两种置信椭球(2),(3)的体积分别为

$$\mathrm{vol}[C(\hat{B})] \propto [\chi^2(mp;1-\alpha)]^{mp/2}|\hat{V}^{-1}\otimes U(\hat{B})|^{-1/2},$$

$$\mathrm{vol}[C(\hat{B}(\omega))] \propto [\chi^2(mp;1-\alpha)]^{mp/2}|\hat{V}^{-1}\otimes U(\hat{B}(\omega))|^{-1/2},$$

令 $r(\omega)=\dfrac{2}{p}\ln\{\mathrm{vol}[C(\hat{B})]/\mathrm{vol}[C(\hat{B}(\omega))]\}$，则有

$$r(\omega)=\ln|U(\hat{B}(\omega))|-\ln|U(\hat{B})|. \tag{4}$$

$r(\omega)$ 反映模型扰动对回归系数置信域的影响. 我们感兴趣的是 $r(\omega)$ 在 ω_0 处的最大局部变化方向即梯度方向 $d_{\max}$. 易知

$$d_{\max} \,/\!/\, \dot{r}=\frac{\partial r(\omega)}{\partial\omega}\Big|_{\omega_0}.$$

1 协方差阵扰动

由(1)知 $y_i\sim N_p(B'x_i,V)$，假设 V 被扰动成为 $\omega_i^{-1}V\ (0<\omega_i\leqslant 1)$，则相应的扰动模型为

$$\begin{cases}Y=XB+\varepsilon,\\ \varepsilon\sim N_{np}(0,V\otimes\Omega^{-1}),\end{cases} \tag{5}$$

其中 $\Omega=\mathrm{diag}(\omega_1,\cdots,\omega_n)$，记 $\omega=(\omega_1,\cdots,\omega_n)'$，则 $\omega=1$ 对应于无扰动模型.

定理 1 在模型(5)情形下，$r(\omega)$ 的梯度方向为

$$d_{\max}^{(1)} \,/\!/\, \dot{r}=(h_{11},\cdots,h_{nn})',$$

其中 h_{ii} 是 $H=X(X'X)^{-1}X'$ 的主对角元，$i=1,\cdots,n$.

证明 首先可以求得(5)中 B 的 MLE 为

$$\hat{B}(\omega)=(X'\Omega X)^{-1}X'\Omega Y,$$

而且 $\hat{B}(\omega)\sim N_{mp}(B,V\otimes(X'\Omega X)^{-1})$. 因此 $U(\hat{B}(\omega))=X'\Omega X$，(4)式成为

$$r(\omega)=\ln|X'\Omega X|-\ln|X'X|,$$

运用矩阵微商公式，并记 $\underset{n\times n}{L_i}=\mathrm{diag}(0,\cdots,0,\underset{(i)}{1},0,\cdots,0)$，得

$$\frac{\partial r(\omega)}{\partial\omega_i}\Big|_{\omega=1}=\mathrm{tr}\{(X'X)^{-1}\frac{\partial\Omega}{\partial\omega_i}\Big|_{\omega=1}X\}=\mathrm{tr}\{(X'X)^{-1}X'L_iX\}$$
$$=\mathrm{tr}\{(X'X)^{-1}x_ix_i'\}=x_i'(X'X)^{-1}x_i=h_{ii},$$

所以

$$\dot{r}=\frac{\partial r(\omega)}{\partial\omega}\Big|_{\omega=1}=(h_{11},\cdots,h_{nn})'.$$

可见，对于协方差阵扰动，$d_{\max}$ 的 n 个分量恰是 n 组数据的杠杆值，因此 $r(\omega)$ 的局部强影响点与高杠杆点是一致的.

2 自变量扰动

设一个自变量(如 X_j)被扰动为

$$X_{j\omega}=X_j+s_j\omega,$$

其中 $\omega=(\omega_1,\cdots,\omega_n)'$，$s_j$ 为尺度因子，$\omega=0$ 表示无扰动. 扰动模型可表示为

$$\begin{cases}Y=X_\omega B+\varepsilon, \quad X_\omega=X+s_j\omega\delta_j',\\ \varepsilon\sim N_{np}(0,V\otimes I_n),\end{cases} \tag{6}$$

其中 $\underset{m\times 1}{\delta_j} = (0,\cdots,0,\underset{(j)}{1},0,\cdots,0)'$.

定理 2 在模型(6)情形下，$r(\omega)$ 的梯度方向为

$$d_{\max}^{(2)} \mathbin{/\!/} \dot{r} = 2s_j X(X'X)^{-1}\delta_j.$$

证明 由(6)易得 B 的 MLE 为

$$\hat{B}(\omega) = (X_\omega' X_\omega)^{-1} X_\omega' Y \sim N_{mp}(B, V \otimes (X_\omega' X_\omega)^{-1}),$$

则(4)式成为

$$r(\omega) = \ln|X_\omega' X_\omega| - \ln|X'X|,$$

其中 $X_\omega = X + s_j \omega \delta_j'$，对 ω 求导即可得所证.

3 因变量的扰动

设因变量的 n 次观测有扰动

$$y_{i\omega} = y_i + s_i\omega_i, i = 1,\cdots,n,$$

其中 $\omega_i = (\omega_{i1},\cdots,\omega_{ip})'$ 为扰动向量，$\omega = (\omega_1,\cdots,\omega_n)'$ 为扰动矩阵，$S = \mathrm{diag}(s_1,\cdots,s_n)$ 为尺度矩阵，$\omega = 0$ 表示无扰动. 扰动模型可表为

$$\begin{cases} Y_\omega = XB + \varepsilon, \quad Y_\omega = Y + S\omega, \\ \varepsilon \sim N_{np}(0, V \otimes I_n). \end{cases} \tag{7}$$

此时 B 的 MLE 为

$$\hat{B}(\omega) = (X'X)^{-1}X'Y_\omega = \hat{B} + (X'X)^{-1}X'S\omega,$$

由于 $\Sigma(\hat{B}(\omega)) = \Sigma(\hat{B})$，所以 $r(\omega) \equiv 0$. 这说明上述因变量的扰动对参数的置信域没有影响.

4 对子集参数置信域的局部影响

考虑模型扰动对 B 的第 l 行 $\underset{p\times 1}{\beta_l}$ 的置信域的影响. 因 $\beta_l = B'\delta_l$，在无扰动模型中，β_l 的 MLE 为 $\hat{\beta}_l = \hat{B}'\delta_l \sim N_p(\beta_l, a_l V)$，其中 $a_l = \delta_l'(X'X)^{-1}\delta_l$，$\beta_l$ 的渐近置信椭球为

$$C(\hat{\beta}_l) = \{\beta_l : (\beta_l - \hat{\beta}_l)' a_l^{-1} \hat{V}^{-1} (\beta_l - \bar{\beta}_l) \leqslant \chi^2(p;1-\alpha)\}.$$

在 ω 扰动模型中，

$$\hat{\beta}_l(\omega) = \hat{B}'(\omega)\delta_l \sim N_p(\beta_l, a_l(\omega)V),$$

β_l 的渐近置信椭球为

$$C(\hat{\beta}_l(\omega)) = \{\beta_l : (\beta_l - \hat{\beta}_l(\omega))' a_l^{-1}(\omega) \hat{V}^{-1} (\beta_l - \bar{\beta}_l(\omega)) \leqslant \chi^2(p;1-\alpha)\}.$$

则有

$$\mathrm{vol}[C(\hat{\beta}_l)] \propto [\chi^2(p;1-\alpha)]^{p/2} |a_l \hat{V}|^{-1/2},$$

$$\mathrm{vol}[C(\hat{\beta}_l(\omega))] \propto [\chi^2(p;1-\alpha)]^{p/2} |a_l(\omega) \hat{V}|^{-1/2},$$

$$r(\omega,\beta_l) = \frac{2}{p}\ln\{\mathrm{vol}[C(\hat{\beta}_l)]/\mathrm{vol}[C(\hat{\beta}_l(\omega))]\} = \ln a_l(\omega) - \ln a_l.$$

对于上述定义的各种扰动，可分别求出 $r(\omega,\beta_l)$ 在 $\omega = \omega_0$ 处变化最敏感的方向 $d_{\max}(\hat{\beta}_l)$. 例如，在协方差阵扰动模型(5)中，

$$a_l(\omega) = \delta_l'(X'\Omega X)^{-1}\delta_l, \omega_0 = 1,$$

$$d_{\max}^{(1)}(\hat{\beta}_l) \mathbin{/\!/} a_l^{-1}(v_{1l}^2,\cdots,v_{nl}^2)',$$

其中 $v_{il}=x_i'(X'X)^{-1}\delta_l, i=1,\cdots,n$. 在自变量扰动模型(6)中,

$$a_l(\omega)=\delta_l'(X_\omega'X_\omega)^{-1}\delta_l, X_\omega=X+s_j\omega\delta_j',$$

则有

$$d_{\max}^{(2)}(\hat{\beta}_l) \,/\!/\, [s_j a_l^{-1}\delta_l'(X'X)^{-1}\delta_j]X(X'X)^{-1}\delta_l.$$

参 考 文 献

[1] 张尧庭、方开泰，多元统计分析引论，科学出版社，1982.

[2] 韦博成、鲁国斌、史建清，统计诊断引论，东南大学出版社，1991.

[3] E. B. Bruce and F. L. Robert, *Genreal classes of influence measures for multivariate regression*, JASA, 87(1992), 184—191.

[4] R. D. Cook, *Assessment of local inflence*, JRSS, B48(1984), 133—169.

Local Influence for Confidence Regions of Multivariate Regression Coefficients

Yue Rongxian

(Nanjing Institute of Meteorology, 210044)

Abstract

Local influence of small perturbations on confidence regions for multivariate regression coefficients are discussed using the method of Cook. The diagnostics for influence are based on local modifications of the covariance, explanatory variables, and responses.

Keywords model perturbations, diagnostics, local influence.