第 25 卷 第 1 期　　应用数学学报　　Vol. 25 No. 1
2002 年 1 月　　ACTA MATHEMATICAE APPLICATAE SINICA　　Jan., 2002

多元测量误差模型的稳健 GM- 估计量*

马江洪
(西安交通大学理学院信息与系统科学研究所, 西安 710049)
(长安大学应用数学教研室, 西安 710064)

张文修
(西安交通大学理学院信息与系统科学研究所, 西安 710049)

摘　要　由于 EV (Errors-in-Variables) 模型 (也称测量误差模型) 的最大似然估计由正交回归给出, 而正交回归对污染数据是敏感的, 所以, 需要采用稳健的统计方法来估计模型参数. 本文在多元 EV 模型中引入稳健 GM- 估计量, 把一元正态 EV 模型的若干结果推广到多元情形, 所得的稳健性结果不仅更具一般性, 而且还修正了文献中对一元情形给出的一个错误结果.

关键词　EV 模型, GM- 估计, 稳健性

1　引言

考虑通常的线性 EV(Errors-in-Variables) 模型 (也称 ME(Measurement Error) 模型或测量误差模型)

$$y=\alpha+x^T\beta,\qquad y\in R,\ \ x\in R^p,\ \ \beta\in R^p, \tag{1.1}$$

$$Y=y+\varepsilon,\qquad X=x+\delta, \tag{1.2}$$

其中真实变量 x 和 y 不能被直接观测, 而只能观测到 X 和 Y, 随机变量 δ,ε 和 x 是相互独立的, $E\varepsilon=0$, $E\delta=0$. 这样, 观测到的 x 和 y 都带有误差.

设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 是取自上述 EV 模型总体 (X,Y) 的简单随机样本, 相应的 x 值记为 $x_1,\cdots,x_n$. 在统计学中, 若诸 x_i 是 IID 随机变量且与误差 δ 独立, 则称 (1.1) 是一个结构模型; 若 x_i 独立但不同分布 (均值可能不同) 而同方差, 则称 (1.1) 是一个超结构 (ultrastructural) 模型; 若 x_i 为确定型变量, 则称 (1.1) 是一个函数模型. 本文考虑结构关系模型.

关于 EV 模型, 较详细的讨论可在 [1] 和 [2] 中找到. 这种模型的许多重要应用可参见 [2] 和 [3] 等. 现在, 我们先说明 EV 模型的正交回归方法. 与传统线性回归中所用的纵向距离不同, 正交回归用的是点 (X_i,Y_i) 到回归超平面 $f(X)=\alpha+X^T\beta$ 的垂直距离, 即点 (X_i,Y_i) 到超平面上点 $(X,\alpha+X^T\beta)$ 的距离 $d_i(X)$ 的最小值 $d_i^{\perp}$.

一般地, $d_i^2=d_i^2(X)=(X-X_i)^T(X-X_i)+(\alpha+X^T\beta-Y_i)^2$, $\forall X\in R^p$. 由条件

本文 1999 年 10 月 20 日收到. 2000 年 8 月 7 日收到修改稿.
* 国家自然科学基金 (40101021 号) 资助项目.

$\frac{\partial}{\partial X}d_i^2(X)=0$, 得 $(I+\beta\beta^T)X_i^0=X_i+(Y_i-\alpha)\beta$, 从而

$$\begin{aligned}X_i^0 =&(I+\beta\beta^T)^{-1}\left[X_i+(Y_i-\alpha)\beta\right]=\left(I-k^{-2}(\beta)\beta\beta^T\right)\left[X_i+(Y_i-\alpha)\beta\right]\\ =&X_i-k^{-2}(\beta)\left[X_i^T\beta+(Y_i-\alpha)\right]\beta,\end{aligned}$$

其中 $k^2(\beta)\equiv 1+\beta^T\beta$. 于是 $d_i^{\perp}=d_i(X_i^0)=\left|Y_i-(\alpha+X_i^T\beta)\right|/k(\beta)$. 这样, 正交 L_p- 线性回归的目标函数为

$$L_p(\alpha,\beta)=\sum_{i=1}^n(d_i^{\perp})^p=\sum_{i=1}^n\left|Y_i-(\alpha+X_i^T\beta)\right|^p/k^p(\beta). \tag{1.3}$$

令 $\theta=\begin{pmatrix}\beta\\ \alpha\end{pmatrix}$, $Z_i=\begin{pmatrix}X_i\\ 1\end{pmatrix}$, $A=(Z_1,Z_2,\cdots,Z_n)^T$, $Y=(X_1,X_2,\cdots,X_n)^T$, 则 (1.3) 式还可表为

$$L_p(\alpha,\beta)=\|A\theta-Y\|_p^p/\left(1+\|\beta\|_2^2\right)^{p/2}, \tag{1.4}$$

其中 $\|\cdot\|_p$ 是向量的 p- 范数.

但是, 这种正交 L_p- 线性回归是不稳健的. 为此, 人们已考虑用稳健 EV 模型来代替它. 有关 EV 模型稳健性的早期讨论可参见 [4] 和 [5]. 1989 年, Zamar[6] 把经典 M-估计引入到 EV 模型. 1992 年, Cheng et al.[7] 研究了一元正态 EV 模型的稳健 GM-估计以及 Hampel 最优性问题. 最近, 国内也有不少讨论 EV 模型的文献, 如 [8] 和 [9] 等, 但基于稳健性考虑的尚不多见.

本文在多元正态 EV 回归中引入 GM- 估计量, 研究相应的有界影响估计的稳健性, 将 [7] 的一元模型推广到多元情形. 不仅所用方法的思想简捷、结果更具一般性, 而且还对 [7] 的一个错误结果进行了修正.

2 稳健正交回归估计量

容易证明, 模型 (1.1), (1.2) 的最大似然等价于极小化加权和

$$\min_{\alpha,\beta}\sum_{i=1}^n w_i(\beta)(Y_i-\alpha-X_i^T\beta)^2=\min_{\alpha,\beta}\sum_{i=1}^n\left(\frac{Y_i-\alpha-X_i^T\beta}{k(\beta)}\right)^2, \tag{2.1}$$

这恰好对应极小化 $L_2(\alpha,\beta)$ 的正交 L_2 回归估计量. 与普通 LS 回归估计的稳健化相类似, 自然引入 Huber 的经典 M- 估计量

$$\min_{\alpha,\beta}\sum_{i=1}^n\rho\Big(\frac{Y_i-\alpha-X_i^T\beta}{k(\beta)}\Big)=\min_{\alpha,\beta}\int\rho\Big(\frac{Y-\alpha-X^T\beta}{k(\beta)}\Big)\,\mathrm{d}F_n(X,Y), \tag{2.2}$$

其中 F_n 是样本 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 的经验分布函数. 适当选择 ρ- 函数, 就可得到 α 和 β 的各种稳健估计量, 写成泛函形式即为 $\widehat{\alpha}_n=S(F_n)$, $\widehat{\beta}_n=T(F_n)$, 其中 $S(F)$ 和 $T(F)$ 是下式的解

$$\min_{S,T}\int\rho\Big(\frac{Y-S(F)-X^TT(F)}{k(T(F))}\Big)\,\mathrm{d}F(X,Y). \tag{2.3}$$

若假定 ρ 有导数 $\psi = \rho'$ 且容许求导和积分交换次序的话，则 (2.3) 式可变成

$$\begin{cases} \displaystyle\int \psi\Big(\frac{Y - S(F) - X^T T(F)}{k\big(T(F)\big)}\Big)\,\mathrm{d}F(X,Y) = 0, \\ \displaystyle\int \psi\Big(\frac{Y - S(F) - X^T T(F)}{k\big(T(F)\big)}\Big)\big[X + YT(F) + X\big(T(F)\big)^T T(F) \\ \qquad -\big(X^T T(F)\big)T(F)\big]\,\mathrm{d}F(X,Y) = \mathbf{0}. \end{cases} \tag{2.4}$$

为进一步简化方程组，可将数据作中心化处理，即令 $\widetilde{X} = X - EX$, $\widetilde{Y} = Y - EY$. 为简单计，$\widetilde{X}$ 和 $\widetilde{Y}$ 仍用 X 和 Y 表示，$(\widetilde{X},\widetilde{Y})$ 的分布函数仍用 F 来表示，也就是说，以下我们总假定：$EX = \mathbf{0}$, $EY = \mathbf{0}$. 这样，(2.4) 式可简化为

$$\int \psi\big[r(T(F))\big]\,w(T(F))\,\mathrm{d}F(X,Y) = \mathbf{0}, \tag{2.5}$$

其中

$$r(\beta) = (Y - X^T\beta)/k(\beta), \tag{2.6}$$

$$w(\beta) = \big[X + \beta Y + X\beta^T\beta - (X^T\beta)\beta\big]/k(\beta), \tag{2.7}$$

除假定所有数据都经过中心化处理外，本文还假定所有估计量泛函均是 Fisher 相合的，即 $T(F_\beta) = \beta$, 这里 F_β 为 (X,Y) 的模型分布. 由 [10] 可得 β 的估计量泛函 $T(F)$ 在 F_β 处的影响函数为

$$IF(X,Y;T,F_\beta) = M^{-1}(\psi,F_\beta)\,\psi\big(r(\beta)\big)\,w(\beta), \tag{2.8}$$

其中

$$M(\psi,F_\beta) = -\int \frac{\partial}{\partial\xi}\big[\psi\big(r(\xi)\big)w(\xi)\big]_{\xi=\beta}\,\mathrm{d}F_\beta(X,Y).$$

注意从 (2.8) 可以看出，对一般的 ψ, IF 无界. 因此，EV 模型的经典 M- 估计不是有界影响估计，单个异常点对估计的影响可任意大. 为构造有界影响估计，[7] 在一元情形引入了 GM- 估计. 类似地，这里我们给出多元 GM- 估计.

用下式代替 (2.5) 式

$$\int \eta\big[w(T(F)),\, r(T(F))\big]\,w(T(F))\,\mathrm{d}F(X,Y) = \mathbf{0}, \tag{2.9}$$

则称方程 (2.9) 的解 $T(F)$ 为参数向量 β 的 GM- 估计泛函，GM- 估计量 $\widehat{\beta}_n$ 为方程

$$\sum_{i=1}^{n} \eta\big[w_i(\widehat{\beta}_n),\, r_i(\widehat{\beta}_n)\big]\,w_i(\widehat{\beta}_n) = \mathbf{0} \tag{2.10}$$

的解，其中

$$r_i(\widehat{\beta}_n) = (Y_i - X_i^T\widehat{\beta}_n)/k(\widehat{\beta}_n),$$
$$w_i(\widehat{\beta}_n) = \big[X_i + \widehat{\beta}_n Y_i + X_i\widehat{\beta}_n^T\widehat{\beta}_n - (X_i^T\widehat{\beta}_n)\widehat{\beta}_n\big]/k(\widehat{\beta}_n),$$

$\eta: R^p \times R \to R$ 满足正则条件

(i) 对所有 $w\in R^p$, $\eta(w,\cdot)$ 在 $R-C(w;\eta)$ 上连续, 这里 $C(w;\eta)$ 为有限集, 在其中每个点处, $\eta(w,\cdot)$ 有有限的左右极限;

(ii) 对所有 $w\in R^p, \eta(w,\cdot)$ 为奇函数且 $\eta(w,r)\geq 0$, $r\in R^+$;

(iii) 对所有 $w\in R^p$, 使 $\eta(w,\cdot)$ 连续但 $\partial\eta(w,r)/\partial r$ 无定义或不连续的所有点形成的集合 $D(w;\eta)$ 有限;

(iv) 以下两矩阵在 F_β 处均存在且非奇异

$$M(T(F))\equiv M(\eta,F)\equiv-\int\frac{\partial}{\partial t}\big[\eta\big(w(t),r(t)\big)w(t)\big]_{t=T(F)}\,\mathrm{d}F(X,Y),\tag{2.11}$$

$$Q(T(F))\equiv Q(\eta,F)\equiv\int\eta^2\big[w(T(F)),\,r(T(F))\big]\,w(T(F))w^T(T(F))\,\mathrm{d}F(X,Y).\tag{2.12}$$

需要指出的是, EV 模型的 M 和 Q 矩阵均与向量 β 有关, 而普通回归模型的 M 和 Q 矩阵均与向量 β 无关. 另外, EV 模型的 M 矩阵也未必是对称阵. 注意到 EV 模型和普通模型的这些重要差异对正确建立多元 EV 模型的稳健估计理论是有益的.

不难看出, T 在 F 处的影响函数和渐近协差阵分别为

$$IF(X,Y;T,F)=M^{-1}(\eta,F)\eta\big[w(T(F)),r(T(F))\big]w(T(F)),\tag{2.13}$$

$$\begin{aligned}V(T(F))=&V(T,F)=\int IF(X,Y;T,F)\big[IF(X,Y;T,F)\big]^T\,\mathrm{d}F(X,Y)\\=&M^{-1}(\eta,F)\,Q(\eta,F)\,M^{-T}(\eta,F).\end{aligned}\tag{2.14}$$

定理 2.1 假定 EV 模型 (见 (1.1) 和 (1.2)) 中涉及的随机变量均正态分布, 则对 (2.6) 和 (2.7) 定义的 w 和 r, 我们有

$$\begin{pmatrix}w\\r\end{pmatrix}\sim N_{p+1}(\mathbf{0},\widetilde{\Sigma}),\qquad\widetilde{\Sigma}=\begin{pmatrix}\Sigma_w&\Sigma_{wr}\\\Sigma_{wr}^T&\sigma_r^2\end{pmatrix},$$

其中

$$\Sigma_w=k^2(\beta)(\Sigma_x+\Sigma_\delta)+\frac{1}{k^2(\beta)}(\beta^T\Sigma_\delta\beta+\sigma_\varepsilon^2)\beta\beta^T-(\beta\beta^T\Sigma_\delta+\Sigma_\delta\beta\beta^T),\tag{2.15}$$

$$\Sigma_{wr}=\frac{1}{k^2(\beta)}(\beta^T\Sigma_\delta\beta+\sigma_\varepsilon^2)\beta-\Sigma_\delta\beta,\tag{2.16}$$

$$\sigma_r^2=\frac{1}{k^2(\beta)}(\beta^T\Sigma_\delta\beta+\sigma_\varepsilon^2).\tag{2.17}$$

证 由 $w(\beta)$ 和 $r(\beta)$ 的定义式 (2.6) 和 (2.7), 可得

$$\begin{pmatrix}w\\r\end{pmatrix}=\frac{1}{k(\beta)}\begin{pmatrix}k^2(\beta)I-\beta\beta^T&\beta\\-\beta^T&1\end{pmatrix}\begin{pmatrix}X\\Y\end{pmatrix}\equiv A\begin{pmatrix}X\\Y\end{pmatrix},\tag{2.18}$$

其中 I 表示单位阵. 将 EV 模型写成矩阵形式

$$\begin{pmatrix}X\\Y\end{pmatrix}=\begin{pmatrix}I&I&\mathbf{0}\\\beta^T&\mathbf{0}^T&1\end{pmatrix}\begin{pmatrix}x\\\delta\\\varepsilon\end{pmatrix}\equiv B\begin{pmatrix}x\\\delta\\\varepsilon\end{pmatrix},\qquad\begin{pmatrix}x\\\delta\\\varepsilon\end{pmatrix}\sim N_{2p+1}(\mathbf{0},\Sigma_e),$$

其中 **0** 表示零向量，$\Sigma_e = \mathrm{diag}\,(\Sigma_x, \Sigma_\delta, \sigma_\varepsilon^2)$，于是

$$\begin{aligned}
&\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+1}(\mathbf{0}, \Sigma), \\
&\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} = B\Sigma_e B^T = \begin{pmatrix} \Sigma_x + \Sigma_\delta & \Sigma_x\beta \\ \beta^T\Sigma_x & \beta^T\Sigma_x\beta + \sigma_\varepsilon^2 \end{pmatrix}.
\end{aligned} \tag{2.19}$$

从而

$$\begin{aligned}
\begin{pmatrix} w \\ r \end{pmatrix} &\sim N_{p+1}(\mathbf{0}, \widetilde{\Sigma}), \\
\widetilde{\Sigma} &= \begin{pmatrix} \Sigma_w & \Sigma_{wr} \\ \Sigma_{wr}^T & \sigma_r^2 \end{pmatrix} = A\Sigma A^T \\
&= \frac{1}{k^2(\beta)} \begin{pmatrix} k^2(\beta)I - \beta\beta^T & \beta \\ -\beta^T & 1 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} \begin{pmatrix} k^2(\beta)I - \beta\beta^T & -\beta \\ \beta^T & 1 \end{pmatrix} \\
&= \frac{1}{k^2(\beta)} \begin{pmatrix} k^2(\beta)\Sigma_{11} - \beta\beta^T\Sigma_{11} + \beta\Sigma_{12}^T & k^2(\beta)\Sigma_{12} - \beta\beta^T\Sigma_{12} + \beta\Sigma_{22} \\ -\beta^T\Sigma_{11} + \Sigma_{12}^T & -\beta^T\Sigma_{12} + \Sigma_{22} \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} k^2(\beta)I - \beta\beta^T & -\beta \\ \beta^T & 1 \end{pmatrix}, \\
k^2(\beta)\Sigma_w &= \big(k^2(\beta)I - \beta\beta^T\big)\Sigma_{11}\big(k^2(\beta)I - \beta\beta^T\big) + \beta\Sigma_{12}^T\big(k^2(\beta)I - \beta\beta^T\big) \\
&\quad + k^2(\beta)\Sigma_{12}\beta^T - \beta\beta^T\Sigma_{12}\beta + \beta\Sigma_{22}\beta^T \\
&= \big(k^2(\beta)I - \beta\beta^T\big)(\Sigma_x + \Sigma_\delta)\big(k^2(\beta)I - \beta\beta^T\big) + \beta(\beta^T\Sigma_x)\big(k^2(\beta)I - \beta\beta^T\big) \\
&\quad + \big(k^2(\beta)I - \beta\beta^T\big)(\Sigma_x\beta)\beta^T + \beta(\beta^T\Sigma_x\beta + \sigma_\varepsilon^2)\beta^T \\
&= k^4(\beta)(\Sigma_x + \Sigma_\delta) + (\beta^T\Sigma_\delta\beta + \sigma_\varepsilon^2)\beta\beta^T - k^2(\beta)(\beta\beta^T\Sigma_\delta + \Sigma_\delta\beta\beta^T), \\
k^2(\beta)\Sigma_{wr} &= \big(k^2(\beta)\Sigma_{11} - \beta\beta^T\Sigma_{11} + \beta\Sigma_{12}^T\big)(-\beta) + k^2(\beta)\Sigma_{12} - \beta\beta^T\Sigma_{12} + \beta\Sigma_{22} \\
&= \big(k^2(\beta)I - \beta\beta^T\big)(\Sigma_x + \Sigma_\delta)(-\beta) - \beta(\beta^T\Sigma_x)\beta + \big(k^2(\beta)I - \beta\beta^T\big)\Sigma_x\beta \\
&\quad + \beta(\beta^T\Sigma_x\beta + \sigma_\varepsilon^2) \\
&= (\beta^T\Sigma_\delta\beta + \sigma_\varepsilon^2)\beta - k^2(\beta)\Sigma_\delta\beta, \\
k^2(\beta)\sigma_r^2 &= (-\beta^T\Sigma_{11} + \Sigma_{12}^T)(-\beta) - \beta^T\Sigma_{12} + \Sigma_{22} = \beta^T\Sigma_\delta\beta + \sigma_\varepsilon^2.
\end{aligned}$$

为讨论方便，我们作以下假设条件

(A.1) $\Sigma_\delta = \sigma_\delta^2 I$, $\sigma_\delta^2 = \sigma_\varepsilon^2 \equiv \sigma^2$.

(A.2) 方程 (2.9) 定义的 GM- 估计泛函 $T(F)$ 唯一.

(A.3) 对所有 $w \in R^p$, 有 $E_r\eta(w, r) = 0$.

由 (2.16) 式立即有

推论 1 在条件 (A.1) 之下，w 和 r 相互独立且均服从正态分布.

推论 2 在条件 (A.1)-(A.3) 之下，$T(F)$ 具有 Fisher 相合性.

证 由推论 1 知 w 和 r 相互独立，于是，w 和 r 的联合分布 G_β 可表示为 $G_\beta(w, r) = G_1(w)G_2(r)$，其中 G_1 和 G_2 分别为 w 和 r 的分布函数. 因此，

$$\begin{aligned}
&\int \eta\big(w(\beta), r(\beta)\big)\, w(\beta)\, \mathrm{d}F_\beta(X, Y) = \int \eta(w, r) w\, \mathrm{d}G_\beta(w, r) \\
= &\int \Big[\int \eta(w, r)\, \mathrm{d}G_2(r)\Big]\, w\, \mathrm{d}G_1(w) = \int \big[E_r\eta(w, r)\big]\, w\, \mathrm{d}G_1(w) = \mathbf{0}.
\end{aligned}$$

由方程 (2.9) 解的唯一性，即得 $T(F_\beta)=\beta$.

注 此推论不仅推广了 [7] 的讨论结果，而且这里所用的方法较其一元情形的旋转主成分方法更为简便.

3 多元 EV 模型的有界影响估计

首先，我们给出几个敏感度定义.

定义 1 估计量泛函 $T(F)$ 在模型分布 F_β 处的 (非标准化) 过失误差敏感度定义为

$$\gamma_u^*(\eta)=\sup_{X,Y}\left\|IF(X,Y;T,F_\beta)\right\|=\sup_{w,r}|\eta(w,r)|\cdot\|M^{-1}(\beta)w\|.$$

定义 2 若 $V(T,F_\beta)$ 存在，则自标准化敏感度定义为

$$\gamma_s^*(\eta)=\begin{cases}\sup\limits_{X,Y}\left[(IF(X,Y;T,F_\beta))^T\big(V(T,F_\beta)\big)^{-1}IF(X,Y;T,F_\beta)\right]^{1/2}, & \\ & \text{若}V(T,F_\beta)\text{非奇异},\\ \infty, & \text{否则}.\end{cases}$$

定义 3 称估计量 T 是 B_u- 稳健 (或 B_s- 稳健) 的，如果 $\gamma_u^*(\eta)$(或 $\gamma_s^*(\eta)$) 有限；称 T 是最 B_u(或 B_s)- 稳健的，如果 T 具有最小的 $\gamma_u^*(\eta)$(或 $\gamma_s^*(\eta)$).

引理 3.1 设条件 (A.1) 成立，则

$$\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)=\Sigma_x\Sigma_w^{-1}\Big[\frac{1}{\sigma^2}rw+ww^T\Sigma_w^{-1}\beta-\beta\Big],\tag{3.1}$$

其中 f_β 是 F_β 的密度函数.

证 由推论 1, 在正态 EV 模型假定下，w 和 r 的联合密度为

$$g_\beta(w,r)=g_1(w)\cdot g_2(r)=(2\pi)^{-(p+1)/2}|\Sigma_w|^{-1/2}\sigma_r^{-1}\exp\Big[-\frac{1}{2}\Big(w^T\Sigma_w^{-1}w+\frac{r^2}{\sigma^2}\Big)\Big].$$

由 (2.18) 式可知，(X,Y) 的联合密度为

$$f_\beta(X,Y)=\frac{1}{|A^{-1}|}g_\beta\big(A(X^T,Y^T)^T\big)=g_\beta(w,r).$$

于是

$$\begin{aligned}\mathrm{d}\ln f_\beta(X,Y)=&\mathrm{d}\Big[-\frac{p+1}{2}\ln(2\pi)-\frac{1}{2}\ln|\Sigma_w|-\ln\sigma_r-\frac{1}{2}\Big(w^T\Sigma_w^{-1}w+\frac{r^2}{\sigma^2}\Big)\Big]\\=&-\frac{1}{2}\mathrm{d}\big(\ln|\Sigma_w|\big)-\frac{1}{2}\mathrm{d}\big[\mathrm{tr}\,(\Sigma_w^{-1}ww^T)\big]-\frac{1}{2\sigma^2}\mathrm{d}(r^2)\\=&-\frac{1}{2}\mathrm{tr}\,(\Sigma_w^{-1}\mathrm{d}\Sigma_w)-\frac{1}{2}\mathrm{tr}\,\big[(\mathrm{d}\Sigma_w^{-1})ww^T+\Sigma_w^{-1}\mathrm{d}(ww^T)\big]-\frac{r}{\sigma^2}\mathrm{d}r\\=&-\frac{1}{2}\mathrm{tr}(\Sigma_w^{-1}\mathrm{d}\Sigma_w)+\frac{1}{2}\mathrm{tr}\,\big[\Sigma_w^{-1}(\mathrm{d}\Sigma_w)\Sigma_w^{-1}ww^T\big]\\&-\frac{1}{2}\mathrm{tr}\,\big[\Sigma_w^{-1}((\mathrm{d}w)w^T+w(\mathrm{d}w)^T)\big]-\frac{r}{\sigma^2}\,\mathrm{d}r,\end{aligned}$$

其中用到了公式 $\mathrm{d}\ln|D| = \mathrm{tr}(D^{-1}\mathrm{d}D)$, $\mathrm{d}D^{-1} = -D^{-1}(\mathrm{d}D)D^{-1}$ 以及在引理条件下 $\sigma_r^2 = \sigma_\varepsilon^2 = \sigma^2$ (见 (2.17)), σ_r^2 与 β 无关. 由 (2.15) 可得

$$\begin{aligned}
\Sigma_w =& k^2(\beta)(\Sigma_x + \sigma^2 I) - \sigma^2\beta\beta^T,\\
\mathrm{d}\Sigma_w =& 2\beta^T(\mathrm{d}\beta)(\Sigma_x + \sigma^2 I) - \sigma^2\big[(\mathrm{d}\beta)\beta^T + \beta(\mathrm{d}\beta)^T\big]\\
=& \frac{2}{k^2(\beta)}\beta^T(\mathrm{d}\beta)(\Sigma_w + \sigma^2\beta\beta^T) - \sigma^2\big[(\mathrm{d}\beta)\beta^T + \beta(\mathrm{d}\beta)^T\big].
\end{aligned}$$

注意到 (2.18) 式，则有

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A^{-1}\begin{pmatrix} w \\ r \end{pmatrix} = \frac{1}{k(\beta)}\begin{pmatrix} I & -\beta \\ \beta^T & 1 \end{pmatrix}\begin{pmatrix} w \\ r \end{pmatrix} = \frac{1}{k(\beta)}\begin{pmatrix} w - \beta r \\ \beta^T w + r \end{pmatrix}.$$

这样

$$\begin{aligned}
\mathrm{d}w =& \mathrm{d}\Big[\frac{1}{k(\beta)}(X + Y\beta + X\beta^T\beta - \beta X^T\beta)\Big]\\
=& \mathrm{d}\Big[\frac{1}{k(\beta)}\Big](X + Y\beta + X\beta^T\beta - \beta X^T\beta) + \frac{1}{k(\beta)}\mathrm{d}(X + Y\beta + X\beta^T\beta - \beta X^T\beta)\\
=& -\frac{1}{k^3(\beta)}(X + Y\beta + X\beta^T\beta - \beta X^T\beta)\beta^T\mathrm{d}\beta\\
&+ \frac{1}{k(\beta)}(Y\mathrm{d}\beta + 2X\beta^T d\beta - (X^T\beta)\mathrm{d}\beta - \beta X^T\mathrm{d}\beta)\\
=& -\frac{1}{k^3(\beta)}\big[(Y - X^T\beta)(k^2(\beta)I - \beta\beta^T) + k^2(\beta)(X\beta^T - \beta X^T)\big]\mathrm{d}\beta\\
=& \frac{1}{k^2(\beta)}\big[r(k^2(\beta)I - \beta\beta^T) + (w\beta^T - \beta w^T)\big]\mathrm{d}\beta,\\
\mathrm{d}r =& \mathrm{d}\Big[\frac{1}{k(\beta)}(Y - X^T\beta)\Big] = \mathrm{d}\Big[\frac{1}{k(\beta)}\Big](Y - X^T\beta) + \frac{1}{k(\beta)}\mathrm{d}(Y - X^T\beta)\\
=& -\frac{1}{k^3(\beta)}\big[(Y - X^T\beta)\beta^T + k^2(\beta)X^T\big]\mathrm{d}\beta = -\frac{1}{k^2(\beta)}w^T\mathrm{d}\beta.
\end{aligned}$$

于是

$$\begin{aligned}
&\mathrm{tr}\big[\Sigma_w^{-1}\mathrm{d}\Sigma_w\big]\\
=& \mathrm{tr}\Big[\frac{2}{k^2(\beta)}\Sigma_w^{-1}\beta^T(\mathrm{d}\beta)(\Sigma_w + \sigma^2\beta\beta^T) - \sigma^2\Sigma_w^{-1}\big((\mathrm{d}\beta)\beta^T + \beta(\mathrm{d}\beta)^T\big)\Big]\\
=& \frac{2}{k^2(\beta)}\beta^T\mathrm{d}\beta + \frac{2\sigma^2}{k^2(\beta)}\beta^T\mathrm{d}\beta(\beta^T\Sigma_w^{-1}\beta) - 2\sigma^2\beta^T\Sigma_w^{-1}\mathrm{d}\beta\\
=& \frac{2}{k^2(\beta)}\beta^T\Sigma_w^{-1}\big[\Sigma_w + \sigma^2\beta\beta^T - \sigma^2 k^2(\beta)I\big]\mathrm{d}\beta = 2\beta^T\Sigma_w^{-1}\Sigma_x\mathrm{d}\beta,\\
&\mathrm{tr}\big[\Sigma_w^{-1}(\mathrm{d}\Sigma_w)\Sigma_w^{-1}ww^T\big]\\
=& \mathrm{tr}\Big\{\Sigma_w^{-1}\Big[\frac{2}{k^2(\beta)}\beta^T(\mathrm{d}\beta)(\Sigma_w + \sigma^2\beta\beta^T) - \sigma^2\big((\mathrm{d}\beta)\beta^T + \beta(\mathrm{d}\beta)^T\big)\Big]\Sigma_w^{-1}ww^T\Big\}\\
=& \frac{2}{k^2(\beta)}w^T\Sigma_w^{-1}w\beta^T\mathrm{d}\beta + \frac{2\sigma^2}{k^2(\beta)}w^T\Sigma_w^{-1}\beta w^T\Sigma_w^{-1}\beta\beta^T\mathrm{d}\beta - 2\sigma^2 w^T\Sigma_w^{-1}\beta w^T\Sigma_w^{-1}\mathrm{d}\beta
\end{aligned}$$

$$=\frac{2}{k^2(\beta)}w^T\Sigma_w^{-1}\left[w\beta^T+\sigma^2\beta w^T\Sigma_w^{-1}\beta\beta^T-\sigma^2k^2(\beta)\beta w^T\Sigma_w^{-1}\right]\mathrm{d}\beta$$
$$=\frac{2}{k^2(\beta)}w^T\Sigma_w^{-1}\left[w\beta^T+\sigma^2\beta w^T\Sigma_w^{-1}\left(\beta\beta^T-k^2(\beta)I\right)\right]\mathrm{d}\beta$$
$$=\frac{2}{k^2(\beta)}w^T\Sigma_w^{-1}\left[w\beta^T+\beta w^T\Sigma_w^{-1}\left(k^2(\beta)\Sigma_x-\Sigma_w\right)\right]\mathrm{d}\beta$$
$$=2w^T\Sigma_w^{-1}\left[\beta w^T\Sigma_w^{-1}\Sigma_x+\frac{1}{k^2(\beta)}(w\beta^T-\beta w^T)\right]\mathrm{d}\beta,$$
$$\operatorname{tr}\left[\Sigma_w^{-1}\left((\mathrm{d}w)w^T+w(\mathrm{d}w)^T\right)\right]$$
$$=2w^T\Sigma_w^{-1}(\mathrm{d}w)=\frac{2}{k^2(\beta)}w^T\Sigma_w^{-1}\left[r(k^2(\beta)I-\beta\beta^T)+(w\beta^T-\beta w^T)\right]\mathrm{d}\beta.$$

故

$$\begin{aligned}\mathrm{d}\ln f_\beta(X,Y)=&-\beta^T\Sigma_w^{-1}\Sigma_x\mathrm{d}\beta+w^T\Sigma_w^{-1}\left[\beta w^T\Sigma_w^{-1}\Sigma_x+\frac{1}{k^2(\beta)}(w\beta^T-\beta w^T)\right]\mathrm{d}\beta\\&-w^T\Sigma_w^{-1}\left[r\left(I-\frac{1}{k^2(\beta)}\beta\beta^T\right)+\frac{1}{k^2(\beta)}(w\beta^T-\beta w^T)\right]\mathrm{d}\beta+\frac{r}{\sigma^2k^2(\beta)}w^T\mathrm{d}\beta\\=&\frac{r}{\sigma^2}w^T\Sigma_w^{-1}\left[\frac{1}{k^2(\beta)}\Sigma_w-\sigma^2I+\frac{\sigma^2}{k^2(\beta)}\beta\beta^T\right]\mathrm{d}\beta-\beta^T\Sigma_w^{-1}\Sigma_x\mathrm{d}\beta\\&+\beta^T\Sigma_w^{-1}ww^T\Sigma_w^{-1}\Sigma_x\mathrm{d}\beta\\=&\left[\frac{r}{\sigma^2}w^T\Sigma_w^{-1}\Sigma_x-\beta^T\Sigma_w^{-1}(\Sigma_w-ww^T)\Sigma_w^{-1}\Sigma_x\right]\mathrm{d}\beta.\end{aligned}$$

由公式 $\mathrm{d}f(x)=\left(\frac{\partial f}{\partial x}\right)^T\mathrm{d}x$, 即证 (3.1) 式.

特别地, 当 $p=1$ 时, 在引理 3.1 的条件下, $\Sigma_w=\sigma_w^2=k^2(\beta)\sigma_x^2+\sigma^2$. 由公式 (3.1) 得

$$\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)=\left(\frac{\sigma_x^2}{\sigma^2\sigma_w^2}\right)rw+\left(\frac{\sigma_x^2}{\sigma_w^4}\beta\right)w^2-\frac{\sigma_x^2}{\sigma_w^2}\beta. \tag{3.2}$$

可见, [7] 引理 4.1 的结论 (4.2) 式

$$\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)=\alpha(\beta)rw$$

并不成立 (其中 $\alpha(\beta)$ 即为本文的 $\sigma_x^2/(\sigma^2\sigma_w^2)$), 而此式是该文获得并引用的重要结论之一. 这里, 我们不仅给出了多元情形的一般结论, 而且对 [7] 的一元结果作了修正.

另一方面, (3.2) 式亦可直接进行验证. 为此, 可直接计算 $\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)$. 根据 (2.19) 式及定理 2.1 的证明可知

$$f_\beta(X,Y)=(2\pi)^{-(p+1)/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(X\quad Y)\Sigma^{-1}\begin{pmatrix}X\\Y\end{pmatrix}\right],$$

而

$$|\Sigma|=\begin{vmatrix}\sigma_x^2+\sigma^2 & \sigma_x^2\beta\\ \sigma_x^2\beta & \sigma_x^2\beta^2+\sigma^2\end{vmatrix}=\sigma^2\left[\sigma_x^2k^2(\beta)+\sigma^2\right],$$
$$\Sigma^{-1}=\frac{1}{|\Sigma|}\begin{pmatrix}\sigma_x^2\beta^2+\sigma^2 & -\sigma_x^2\beta\\ -\sigma_x^2\beta & \sigma_x^2+\sigma^2\end{pmatrix},$$

于是

$$\begin{aligned}
\frac{\partial}{\partial\beta}\ln f_\beta(X,Y) &= \frac{\partial}{\partial\beta}\left[-\frac{p+1}{2}\ln(2\pi)-\frac{1}{2}\ln|\Sigma|-\frac{1}{2}(X\ \ Y)\Sigma^{-1}\begin{pmatrix}X\\Y\end{pmatrix}\right]\\
&=-\frac{1}{2|\Sigma|}\frac{\partial|\Sigma|}{\partial\beta}-\frac{1}{2}\frac{\partial}{\partial\beta}\Big[\frac{1}{|\Sigma|}\big((\sigma_x^2\beta^2+\sigma^2)X^2-2\sigma_x^2\beta XY+(\sigma_x^2+\sigma^2)Y^2\big)\Big]\\
&=-\frac{1}{|\Sigma|}\sigma^2\sigma_x^2\beta+\frac{1}{2|\Sigma|^2}\frac{\partial|\Sigma|}{\partial\beta}\left[\sigma_x^2(Y-X\beta)^2+\sigma^2(X^2+Y^2)\right]\\
&\quad-\frac{1}{2|\Sigma|}\big[2\sigma_x^2(Y-X\beta)(-X)\big],
\end{aligned}$$

将 $\begin{pmatrix}X\\Y\end{pmatrix}=\frac{1}{k(\beta)}\begin{pmatrix}w-\beta r\\ \beta w+r\end{pmatrix}$ 代入上式右端，整理可得

$$\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)=-\frac{1}{|\Sigma|}\sigma^2\sigma_x^2\beta+\frac{1}{|\Sigma|^2}\sigma^4\sigma_x^2\beta w^2+\frac{1}{|\Sigma|}\sigma_x^2 rw,$$

再利用 $|\Sigma|=\sigma^2\left(k^2(\beta)\sigma_x^2+\sigma^2\right)=\sigma^2\sigma_w^2$，即有 (3.2) 式.

在模型分布 F_β 处对 (2.9) 式两边关于 β 求偏导，则有

$$\begin{aligned}
&\int\frac{\partial}{\partial\beta}\left[\eta\left(w(\beta),r(\beta)\right)w(\beta)\right]\mathrm{d}F_\beta(X,Y)\\
&+\int\eta\big(w(\beta),r(\beta)\big)w(\beta)\Big[\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)\Big]^T\mathrm{d}F_\beta(X,Y)=\mathbf{0}.
\end{aligned}$$

根据 (2.11) 和 (3.1) 式，可推出

$$\begin{aligned}
M(\beta)&=\int\eta\big(w(\beta),r(\beta)\big)w(\beta)\Big[\frac{\partial}{\partial\beta}\ln f_\beta(X,Y)\Big]^T\mathrm{d}F_\beta(X,Y)\\
&=\int\eta(w,r)w\Big[\frac{r}{\sigma^2}w^T+\beta^T\Sigma_w^{-1}ww^T-\beta^T\Big]\Sigma_w^{-1}\Sigma_x\,\mathrm{d}G_\beta(w,r)\\
&=\frac{1}{\sigma^2}\int\eta(w,r)rww^T\mathrm{d}G_\beta(w,r)\Sigma_w^{-1}\Sigma_x.
\end{aligned}\tag{3.3}$$

下面，我们给出 $\gamma_u^*(\eta)$ 和 $\gamma_s^*(\eta)$ 的下界以及相应于最 B_u- 稳健和最 B_s- 稳健估计量的 η- 函数.

定理 3.1 在假设 (A.1)–(A.3) 之下，总有

$$\gamma_u^*(\eta)\ge \mathrm{d}_u(\beta)\equiv\sqrt{\frac{\pi}{2}}\,\frac{p\sigma}{E\|\Sigma_x\Sigma_w^{-1}w\|},\tag{3.4}$$

并且若 $E(\Sigma_x\Sigma_w^{-1}w)(\Sigma_x\Sigma_w^{-1}w)^T\|\Sigma_x\Sigma_w^{-1}w\|^{-1}$ 为数量阵 (即数乘单位阵) 时，$\eta(w,r)=\mathrm{sgn}\,(r)\|\Sigma_x\Sigma_w^{-1}w\|^{-1}$ 达到此下界.

证　由 (3.3) 式，有

$$I=\frac{1}{\sigma^2}\int\eta(w,r)rM^{-1}(\beta)ww^T\mathrm{d}G_\beta(w,r)\Sigma_w^{-1}\Sigma_x.$$

取迹得

$$\begin{aligned}p=&\frac{1}{\sigma^2}\int r[\Sigma_x\Sigma_w^{-1}w]^T[\eta(w,r)M^{-1}(\beta)w]\,\mathrm{d}G_\beta(w,r)\\ \le&\frac{1}{\sigma^2}\int |r|\cdot\|\Sigma_x\Sigma_w^{-1}w\|\cdot\|\eta(w,r)M^{-1}(\beta)w\|\,\mathrm{d}G_\beta(w,r)\\ \le&\frac{1}{\sigma^2}\gamma_u^*(\eta)E|r|E\|\Sigma_x\Sigma_w^{-1}w\|=\sqrt{\frac{\pi}{2}}\frac{1}{\sigma}\gamma_u^*(\eta)E\|\Sigma_x\Sigma_w^{-1}w\|,\end{aligned}$$

故 (3.4) 式成立. 若 $E(\Sigma_x\Sigma_w^{-1}w)(\Sigma_x\Sigma_w^{-1}w)^T\|\Sigma_x\Sigma_w^{-1}w\|^{-1}=aI_p,\quad a\in R$, 则取迹可得 $a=p^{-1}E\|\Sigma_x\Sigma_w^{-1}w\|$, 取 $\eta(w,r)=\mathrm{sgn}(r)\|\Sigma_x\Sigma_w^{-1}w\|^{-1}$, 则由 (3.3) 式知

$$\begin{aligned}M(\beta)=&\frac{1}{\sigma^2}\int |r|ww^T\|\Sigma_x\Sigma_w^{-1}w\|^{-1}\,\mathrm{d}G_\beta(w,r)\Sigma_w^{-1}\Sigma_x\\ =&\frac{1}{\sigma^2}E|r|(\Sigma_x\Sigma_w^{-1})^{-1}\int(\Sigma_x\Sigma_w^{-1}w)(\Sigma_x\Sigma_w^{-1}w)^T\|\Sigma_x\Sigma_w^{-1}w\|^{-1}\,\mathrm{d}G_\beta(w,r)\\ =&\sqrt{\frac{2}{\pi}}\frac{1}{\sigma}(\Sigma_x\Sigma_w^{-1})^{-1}\frac{1}{p}E\|\Sigma_x\Sigma_w^{-1}w\|I_p,\end{aligned}$$

故

$$\gamma_u^*(\eta)=\sup_{w,r}|\eta(w,r)|\cdot\|M^{-1}(\beta)w\|=\sqrt{\frac{\pi}{2}}\frac{p\sigma}{E\|\Sigma_x\Sigma_w^{-1}w\|}.$$

定理 3.2 总有 $\gamma_s^*(\eta)\ge\sqrt{p}$. 若 $Eww^T\|w\|^{-1}$ 为数量阵, 则 $\eta(w,r)=\mathrm{sgn}\,(r)\|w\|^{-1}$ 达到此下界.

证 由 $\gamma_s^*(\eta)$ 的定义及 (2.13) 和 (2.14) 式, 可得

$$\begin{aligned}\gamma_s^*(\eta)=&\sup_{X,Y}\left[\left(IF(X,Y;T,F_\beta)\right)^T\left(V(T,F_\beta)\right)^{-1}IF(X,Y;T,F_\beta)\right]^{1/2}\\ =&\sup_{w,r}\left[\eta^2(w,r)w^TQ^{-1}(\beta)w\right]^{1/2}.\end{aligned}$$

因为 $Q(\beta)=\int\eta^2(w,r)ww^T\,\mathrm{d}G_\beta(w,r)$, 所以

$$\begin{aligned}p=&\mathrm{tr}(I_p)=\mathrm{tr}\int\eta^2(w,r)ww^TQ^{-1}(\beta)\,\mathrm{d}G_\beta(w,r)\\ =&\int\eta^2(w,r)w^TQ^{-1}(\beta)w\,\mathrm{d}G_\beta(w,r)\le\left[\gamma_s^*(\eta)\right]^2.\end{aligned}$$

若 $Eww^T\|w\|^{-2}=aI_p$, 则 $a=p^{-1}$, 取 $\eta(w,r)=\mathrm{sgn}(r)\|w\|^{-1}$, 则

$$Q(\beta)=\int ww^T\|w\|^{-2}\,\mathrm{d}G_\beta(w,r)=\frac{1}{p}I_p,$$

于是, $\gamma_s^*(\eta)=\sup\limits_{w,r}\left[\eta^2(w,r)w^TQ^{-1}(\beta)w\right]^{1/2}=\sqrt{p}$.

以上我们分别给出了使 γ_u^* 和 γ_s^* 达到最小的估计量. 进一步地, 我们还可以考虑在 Hampel 意义下的最优性问题.

4　最优稳健的估计量

Hampel 最早对一元 M- 估计量提出了这样一个最优性问题，就是在影响函数有界的条件下，寻求使渐近方差极小化的估计量. 后来这个思想被扩展到了多元 M- 估计量的情形 [10]. 但存在性问题在一般情形下并未得到解决. 对普通线性模型，讨论 Hampel 最优性问题的有 [11] 和 [12]. 对 EV 模型，[7] 在一元正态情形下研究了这个问题，虽然其结论的推证用到了该文的 (4.2) 式，但所幸的是其结果的正确性并未受到影响. 下面，我们来解决多元 EV 模型的 Hampel 最优性问题.

定理 4.1　设条件 (A.1)-(A.3) 成立，常数 $c>0$ 足够大，那么，在满足 $\gamma_u^*\le c$ 的所有 η- 函数中使相应估计量 T 的渐近协差阵的迹 $\mathrm{tr}\,V(T)$ 达到最小的 η- 函数为

$$\eta_c(w,r)=\frac{1}{\|ADw\|}\psi_c\{\|ADw\|\},\tag{4.1}$$

其中 $\psi_c(u)\equiv\mathrm{sgn}(u)\min(|u|,c)$ 是 Huber 函数，$D\equiv D(\beta)\equiv\sigma^{-2}\Sigma_x\Sigma_w^{-1}$，$A$ 是 $p\times p$ 阶非奇异方阵，满足方程

$$E\Big\{\Big[2\Phi\Big(\frac{c}{\|ADw\|}\Big)-1\Big]Dw(Dw)^T\Big\}=A^{-1}.\tag{4.2}$$

证　记似然得分向量为

$$s\equiv s(w,r)\equiv\frac{\partial}{\partial\beta}g_\beta(w,r),\qquad t\equiv t(w)\equiv\Sigma_x\Sigma_w^{-1}[ww^T\Sigma_w^{-1}\beta-\beta],$$

则由引理 3.1, 有 $s-t=\Sigma_x\Sigma_w^{-1}\sigma^{-2}rw=rDw$. 对满足假设的任何 η- 函数，其相应的影响函数简记为 $L\equiv L(w,r)$, 则 $L\equiv L(w,r)=M^{-1}(\eta,G_\beta)\eta(w,r)w$ (根据 (2.13) 式). 注意到 t 与 r 无关，由定理 2.1 的两个推论可推出

$$\int L(w,r)t(w)^T\mathrm{d}G_\beta(w,r)=\int t(w)L(w,r)^T\,\mathrm{d}G_\beta(w,r)=0.$$

再根据公式 (见 (3.3))$\int L(w,r)s(w,r)^T\,\mathrm{d}G_\beta(w,r)=I_p$, 就有

$$\begin{aligned}&\int\big[L(w,r)-A(s(w,r)-t(w))\big]\big[L(w,r)-A(s(w,r)-t(w))\big]^T\,\mathrm{d}G_\beta(w,r)\\
=&\int LL^T\,\mathrm{d}G_\beta-A-A^T+A\Big[\int ss^T\,\mathrm{d}G_\beta-\int st^T\,\mathrm{d}G_\beta-\int ts^T\,\mathrm{d}G_\beta+\int tt^T\,\mathrm{d}G_\beta\Big]A^T\\
=&\int LL^T\mathrm{d}G_\beta-A-A^T+A\Big[J(\beta)-\int tt^T\,\mathrm{d}G_\beta\Big]A^T,\end{aligned}\tag{4.3}$$

其中 $J(\beta)\equiv\int ss^T\,\mathrm{d}G_\beta$ 为 Fisher 信息阵, 此处利用了 $\int st^T\,\mathrm{d}G_\beta=\int ts^T\,\mathrm{d}G_\beta=\int t[rDw+t]^T\,\mathrm{d}G_\beta=\int tt^T\,\mathrm{d}G_\beta$ (因为 $r\sim N(0,\sigma_r^2)$ 且 r 与 w 相互独立), (4.3) 两边取迹得

$$\begin{aligned}&\int\|L(w,r)-rADw\|^2\,\mathrm{d}G_\beta(w,r)\\
=&\int\|L(w,r)\|^2\mathrm{d}G_\beta(w,r)-2\mathrm{tr}A+\mathrm{tr}[J(\beta)A^TA]-\int\|At(w)\|^2\,\mathrm{d}G_\beta(w,r),\end{aligned}$$

又 $\gamma_u^*(\eta) \le c$ 等价于 $\|L(w,r)\| \le c,\ \forall (w,r) \in R^{p+1}$, 这样在满足 $\gamma_u^*(\eta) \le c$ 的所有 η- 函数中, 相应于 $L(w,r) = h_c(rADw)$ 的 η- 函数使 $\operatorname{tr} V(T) = \operatorname{tr} \int L(w,r)L(w,r)^T \,\mathrm{d}G_\beta(w,r) = \int \|L(w,r)\|^2 \,\mathrm{d}G_\beta(w,r)$ 达到最小, 其中 $h_c(z) \equiv z \min\left(1, \frac{c}{\|z\|}\right)$ 是多元 Huber 函数 $(z \in R^p)$. 最后, 由 (3.3) 和 (4.2) 式不难验证

$$\begin{aligned} M(\beta) \equiv& M(\eta_c, G_\beta) = \int \eta_c(w,r) r w w^T D^T \,\mathrm{d}G_\beta(w,r) \\ &= \int r^2 \min\left(1, \frac{c}{|r| \cdot \|ADw\|}\right) w(Dw)^T \,\mathrm{d}G_\beta(w,r) = D^{-1}A^{-1}. \end{aligned}$$

于是

$$IF(w,r;\eta_c,G_\beta) = M^{-1}(\beta)\eta_c(w,r)w = r \min\left(1, \frac{c}{|r| \cdot \|ADw\|}\right) ADw = h_c(rADw).$$

记由 (4.1) 和 (4.2) 式定义的 η_c- 函数为 η_c^u, 根据定理 4.1 的证明立即有

推论 1 在定理 4.1 条件下, $\gamma_u^*(\eta_c^u) = c$.

注 1 定理 4.1 条件要求 c 足够大是为了保证方程 (4.2) 解矩阵 A 的存在性. 因为对足够大的 c, 可以证明 A 是存在的 (用类似于 [13] 或 [11] 的方法). 但是 c 究竟大到什么程度尚不清楚, 不过, c 有一个下界 (见定理 3.1).

注 2 正如 [7] 指出的那样, 普通线性模型和一元 EV 模型在 Hampel 意义下的最优估计量的主要差异在于, 前者与 β 无关, 而后者处处与 β 有关. 我们看到, 对多元 EV 模型, 情况亦如此. 不过, 估计量对未知参数的这种相依性在一般 M- 估计中也并非不常见 (见 [10]).

定义 4 给定一个估计量泛函类 J, 称 $T \in J$ 在 F 处是 J-(渐近地协差阵) 可容许的, 如果 $V(T,F)$ 存在且不存在泛函 $\widetilde{T} \in J$ 使 $V(\widetilde{T},F) \le V(T,F)$ 且 $V(\widetilde{T},F) \ne V(T,F)$, 此处不等号 " $\le$ " 意指协差阵的偏序 " $U \le V \Leftrightarrow V-U$ 半正定".

推论 2 设 $J_c^u \equiv \{T : IF(\cdot,\cdot;T,G_\beta)$ 和 $V(T,G_\beta)$ 都存在且 $\gamma_u^*(\eta) \le c\}$, 则在定理 4.1 的条件下, 最优 B_u- 估计量 T_c^u (相应于 η_c^u) 是 J_c^u- 可容许的.

证 假如存在 $\widetilde{T} \in J_c^u$ 使 $V(\widetilde{T},G_\beta) \le V(T_c^u,G_\beta)$ 且 $V(\widetilde{T},G_\beta) \ne V(T_c^u,G_\beta)$ 成立, 则 $V(T_c^u,G_\beta) - V(\widetilde{T},G_\beta)$ 半正定且非零, 于是 $\operatorname{tr}\left[V(T_c^u,G_\beta) - V(\widetilde{T},G_\beta)\right] = \operatorname{tr} V(T_c^u,G_\beta) - \operatorname{tr} V(\widetilde{T},G_\beta) > 0$. 这与定理 4.1 的结论矛盾.

对 $\gamma_s^*(\eta)$ 而言, 在普通线性模型中只能得到一个类似于上述推论 2 的较弱结果, 但是, 在 EV 模型中, 由于 M 阵和协差阵 V 对未知参数 β 的相依性, 却连这样的较弱结果也难以获得. 在这方面仍有许多问题值得进一步研究.

参 考 文 献

1 Kendall M, Stuart A. The Advanced Theory of Statistics 2, 4th ed. London: Charles Griffin, 1979

2 Fuller W A. Measurement Error Models. New York: Wiley, 1987

3 Cheng C L, Van Ness J W. Robust Calibration. *Technometrics*, 1997, 39: 401–411

4 Carroll R J, Gallo P P. Some Aspects of Robustness in the Functional Errors-in-variables Model. *Comm. Statist. Theory Methods*, 1982, 11: 2573–2585

5 Brown M L. Robust Line Estimation with Errors in both Variables. *J. Amer. Statist. Assoc.*, 1982, 77: 71–79

6 Zamar R H. Robust Estimation in the Errors-in-variables Model. *Biometrika*, 1989, 76: 149–160

7 Cheng C L, Van Ness J W. Generalized M-estimators for Errors-in-variables Regression. *Ann. Statist.*, 1992, 20: 385–397

8 崔恒健. EV 模型中广义最小一乘估计的渐近性质. 中国科学, 1997, 27(2): 119–131
(Cui Hengjian. Asymptotic Behavior of Generalized Least Absolute Estimators for the EV Model. *Science in China*, 1997, 27(2): 119–131)

9 高玉福, 梁华. 关于 EV 线性回归模型中的广义最小二乘估计. 数学的实践与认识, 1996, 26(4): 343–348
(Gao Yufu, Liang Hua. Generalized Least Squares Estimates on EV Linear Regression Models. *Mathematics in Practice and Theory*, 1996, 26(4): 343–348)

10 Hampel F R. et al. Robust Statistics. New York: Wiley, 1986

11 Krasker W S. Estimation in Linear Regression Models with Disparate Data Points. *Econometrica*, 1980, 46: 1333–1346

12 Ronchetti E, Rousseeuw P J. Change-of-variance Sensitivities in Regression. *Z. Wahrsch. Verw. Gebiete.*, 1985, 68: 503–519

13 马江洪. 拟合值影响有界的回归估计. 数理统计与应用概率, 1992, 7(1): 33–40
(Ma Jianghong. The Bounded Fitted-value Influence Regression Estimator. *Mathematical Statistics and Applied Probability*, 1992, 7(1): 33–40)

ROBUST GM-ESTIMATORS FOR MULTIVARIATE ERRORS-IN-VARIABLES MODELS

MA JIANGHONG

(*Institute for Information and System Science, Faculty of Science, Xi'an Jiaotong University, Xi'an* 710049)

(*Chang'an University, Xi'an* 710064)

ZHANG WENXIU

(*Institute for Information and System Science, Faculty of Science, Xi'an Jiaotong University, Xi'an* 710049)

Abstract Since the maximum likelihood estimates for errors-in-variables (EV) model are provided by orthogonal regression (OR), but OR is sensitive to contaminated data, so are robust methods needed to estimate model parameters. In this paper, we introduce robust GM-estimators for multivariate EV models, and generalize some results for univariate normal EV models to the multivariate case. The results we obtain are not only more general, but also can correct a result for the univariate case in the literature.

Key words Errors-in-Variables models, GM-estimators, robustness