应用数学和力学，第11卷第11期（1990年11月）
Applied Mathematics and Mechanics

应用数学和力学编委会编
重庆出版社出版

最小二乘估计精度及其界限*

杨 虎

（重庆交通学院，1988年12月19日收到）

摘 要

Puntanen[1]提出用均方误差来度量最小二乘估计的精度，以后Styan[2]，Rao[3]等相继讨论了这种精度及其界限．本文考虑采用广义方差，从而引进了一种新的最小二乘估计精度的度量并讨论了它的界．

关键词 精度 相对效率 均方误差 广义方差 矩阵微商 最佳线性无偏估计

一、引 言

我们考虑Gauss-Markoff模型：

$$Y=X\beta+\varepsilon,\ E(\varepsilon)=0,\ \mathrm{Cov}(\varepsilon)=\Sigma \tag{1.1}$$

其中Y和ε为n维向量，β为p维向量，X为$n\times p$列满秩矩阵，Σ为ε的协方差矩阵，$\Sigma>0$（正定阵）．在实际中，β的最佳线性无偏估计(the best linear unbased estimator：简记为BLUE)为$\beta^*=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$，然而一个实际的问题是$\Sigma$往往有误差，这时$\beta^*$将不再是最佳的．为解决这个问题，我们可假定$\Sigma=I$(单位矩阵)，从而得到(1.1)的最小二乘估计(the least squares estimator：简记为LSE)$\hat\beta$，我们只需考虑$\hat\beta$与β^*究竟有多大的偏离，或用$\hat\beta$作为β的估计究竟有多大的误差．

需要说明的是假设$\Sigma=I$并不妨碍问题的一般性，假如我们实际采用正定矩阵$\Sigma_0\neq\Sigma$，考虑如下的模型：

$$\tilde Y=\tilde X\beta+\tilde\varepsilon,\ E(\tilde\varepsilon)=0,\ \mathrm{Cov}(\tilde\varepsilon)=\tilde\Sigma \tag{1.2}$$

其中 $\tilde Y=\Sigma_0^{-\frac12}Y,\tilde X=\Sigma_0^{-\frac12}X,\tilde\varepsilon=\Sigma_0^{-\frac12}\varepsilon,\quad \tilde\Sigma=\Sigma_0^{-\frac12}\Sigma\Sigma_0^{-\frac12}$，则考虑$\hat\beta$与$\beta^*$的差异就相当于对(1.2)假设了$\tilde\Sigma=I$．

刻画$\hat\beta$与β^*的差异有两种方法，其一是$\hat\beta$相对于β^*的效率．这方面的文献可参阅Bloomfield，Watson[4]，Knott[5]，王松桂[6-7]及本文作者[8]的工作．另一种方法就是$\hat\beta$相对于β^*的精度，这方面的工作除了在摘要中提到的三篇外，还可参见文献[9—10]．

文献[1—3]讨论并提出了如下的精度：

$$MSE(X\hat\beta)-MSE(X\beta^*), \tag{1.3}$$

* 王志忠推荐．

$MSE(\cdot)$表示均方误差(the mean squared error)。由(1.3)可得

$$\begin{aligned}MSE(X\hat{\beta})-MSE(X\beta^*)&=\mathrm{tr}[\mathrm{Cov}(X\hat{\beta})]-\mathrm{tr}[\mathrm{Cov}(X\beta^*)]\\&=\mathrm{tr}[X\mathrm{Cov}(\hat{\beta})X']-\mathrm{tr}[X\mathrm{Cov}(\beta^*)X']\\&=\mathrm{tr}[X'X\mathrm{Cov}(\hat{\beta})]-\mathrm{tr}[X'X\mathrm{Cov}(\beta^*)]\\&\triangleq\mathrm{tr}[h\mathrm{Cov}(\hat{\beta})]-\mathrm{tr}[h\mathrm{Cov}(\beta^*)]\end{aligned}\tag{1.4}$$

在这篇文章中，我们将考虑的类似的精度度量形式为：

$$\det[h\mathrm{Cov}(\hat{\beta})]-\det[h\mathrm{Cov}(\beta^*)]\tag{1.5}$$

这里$\det(\cdot)$表示行列式，$\det[\mathrm{Cov}(\cdot)]$表示广义方差，$h=X'X$（见(1.4)式的定义）。下一节，我们将给出(1.5)的上界。

二、主 要 结 果

设$\varphi(C)$为$n\times p$矩阵C的一个实值函数，矩阵

$$\frac{\partial\varphi(C)}{\partial C}\triangleq\begin{pmatrix}\frac{\partial\varphi(C)}{\partial c_{11}} & \frac{\partial\varphi(C)}{\partial c_{12}}\cdots & \frac{\partial\varphi(C)}{\partial c_{1p}}\\ \frac{\partial\varphi(C)}{\partial c_{21}} & \frac{\partial\varphi(C)}{\partial c_{22}}\cdots & \frac{\partial\varphi(C)}{\partial c_{2p}}\\ \cdots\cdots & & \\ \frac{\partial\varphi(C)}{\partial c_{n1}} & \frac{\partial\varphi(C)}{\partial c_{n2}}\cdots & \frac{\partial\varphi(C)}{\partial c_{np}}\end{pmatrix}.\tag{2.1}$$

定义为$\varphi(C)$对C的矩阵微商，并记为$\left(\frac{\partial\varphi(C)}{\partial c_{ij}}\right)_{n\times p}$，这里$(c_{ij})_{n\times p}=C$。

引理 设B对称，则

(1) $\frac{\partial[\det(C'BC)]^\delta}{\partial C}=2\delta[\det(C'BC)]^\delta BC(C'BC)^{-1}$,

(2) $\frac{\partial\mathrm{tr}[(C'C-I)B]}{\partial C}=2CB$.

其中 $\mathrm{tr}(\cdot)$表示矩阵的迹，δ为任意实数。

证明 $\frac{\partial[\det(C'BC)]^\delta}{\partial C}$

$$=\left(\frac{\partial[\det(C'BC)]^\delta}{\partial c_{ij}}\right)_{n\times p}$$

$$=\left(\delta(\det(C'BC)]^{\delta-1}\frac{\partial\det(C'BC)}{\partial c_{ij}}\right)_{n\times p}$$

$$=\delta[\det(C'BC)]^{\delta-1}\frac{\partial\det(C'BC)}{\partial C},$$

我们用D^*记矩阵$D=(d_{ij})_{n\times p}$的伴随矩阵，即$D^*=(D_{ji})_{p\times n}$，D_{ji}为d_{ij}的代数余子式，从而

$$\frac{\partial[\det(C'BC)]^\delta}{\partial C}=\delta[\det(C'BC)]^{\delta-1}\sum_{ij}(C'BC)^{*\prime}_{ij}\frac{\partial(C'BC)_{ij}}{\partial C}\tag{2.2}$$

这里$(C'BC)_{ij}$为$C'BC$的第i,j元，D'表示D的转置，因为

$$\frac{\partial(C'BC)}{\partial c_{ij}}=E'_{ij}BC+C'BE_{ij} \tag{2.3}$$

这里E_{ij}表示仅第i,j元为1，其余都为零的矩阵，由矩阵微商的转换定理，得

$$\sum_{ij}(C'BC)^{*}_{ij}\frac{\partial(C'BC)_{ij}}{\partial C}$$

$$=\sum_{ij}\det(C'BC)(C'BC)^{-1}_{ij}(BCE'_{ij}+BCE_{ij})$$

$$=\det(C'BC)BC\sum_{ij}(C'BC)^{-1}_{ij}(E'_{ij}+E_{ij})$$

$$=2\det(C'BC)BC(C'BC)^{-1}$$

故(1)得证。而又因为

$$\frac{\partial\operatorname{tr}[(C'C-I)B]}{\partial C}=\sum_{i}\frac{\partial[(C'C-I)B)_{ii}}{\partial C}$$

且
$$\frac{\partial[(C'C-I)B]}{\partial c_{ij}}=E'_{ij}CB+C'E_{ij}B.$$

同样由转换定理得知

$$\frac{\partial[\operatorname{tr}((C'C-I)B)]}{\partial C}=\sum_{i}(CBE_{ii}+CBE_{ii})=2CB \tag{2.4}$$

从而引理得证。

定理 设$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$为$\Sigma$的$n$个顺序特征根，$\operatorname{rank}X=p$且$n\geqslant 2p$，则

$$\det[h\operatorname{Cov}(\hat\beta)]-\det[h\operatorname{Cov}(\beta^*)]$$
$$\leqslant K\left[\prod_{i=1}^{p}\frac{\lambda_i+\lambda_{n-i+1}}{2}-\prod_{i=1}^{p}\left(\frac{\lambda_i^{-1}+\lambda_{n-i+1}^{-1}}{2}\right)^{-1}\right]\triangleq KS \tag{2.5}$$

其中 $K=2C_2(1+\sqrt{1+pC_1C_2S})^{-1}$，并且

$$C_1=\prod_{i=1}^{p}\frac{2\lambda_i^{-1}\lambda_{n-i+1}}{\lambda_i+\lambda_{n-i+1}},\quad C_2=\left(\frac{1}{2}+\frac{1}{2}\prod_{i=1}^{p}\frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i+\lambda_{n-i+1})^2}\right)^{-p}.$$

推论 1 当$\Sigma=I$时，$S=0$且

$$\det[h\operatorname{Cov}(\hat\beta)]=\det[h\operatorname{Cov}(\beta^*)]=1 \tag{2.6}$$

推论 2 当 $C_2\leqslant 1+\frac{1}{4}pC_1S$ 时，有

$$\det[h\operatorname{Cov}(\hat\beta)]-\det[h\operatorname{Cov}(\beta^*)]\leqslant S. \tag{2.7}$$

三、定理的证明

设Σ的谱分解为$Q'\Lambda Q$，Q为正交矩阵，$\Lambda=\operatorname{diag}(\lambda_1, \lambda_2, \cdots\lambda_n)$，记$V=QX(X'X)^{-\frac{1}{2}}$，则有$V'V=I$，且

$$\begin{cases}\det[h\operatorname{Cov}(\hat\beta)]=\det(V'\Lambda V),\\ \det[h\operatorname{Cov}(\beta^*)]=\det(V'\Lambda^{-1}V)^{-1}\end{cases} \tag{3.1}$$

因此只需考虑如下的极大化问题

$$\begin{cases} \max\{\det[h\mathrm{Cov}(\hat{\beta})]-\det[h\mathrm{Cov}(\beta^*)]\} \\ V'V=I \end{cases}$$

记 $H_1=\det[h\mathrm{Cov}(\hat{\beta})]$, $H_2=\det[h\mathrm{Cov}(\beta^*)]$, 此问题即:

$$\begin{cases} \max(H_1-H_2) \\ V'V=I \end{cases} \tag{3.2}$$

考虑如下的 Lagrange 表达式

$$H_1-H_2-\mathrm{tr}[(V'V-I)A],$$

将它对V求微商并令其为零，由引理，我们可以得到

$$H_1\Lambda V(V'\Lambda V)^{-1}+H_2\Lambda^{-1}V(V'\Lambda^{-1}V)^{-1}=VA \tag{3.3}$$

从而$A=(H_1+H_2)I$，我们从(3.3)立即得知

$$H_1V\Lambda^2V(V'\Lambda V)^{-1}+H_2(V'\Lambda^{-1}V)^{-1}=V'\Lambda V(H_1+H_2) \tag{3.4}$$

(3.4)式两端同时右乘$V'\Lambda V$，得

$$(V'\Lambda^{-1}V)^{-1}V'\Lambda V=\frac{H_1+H_2}{H_2}(V'\Lambda V)^2-\frac{H_1}{H_2}V\Lambda^2V \tag{3.5}$$

故$(V'\Lambda^{-1}V)^{-1}$与$V'\Lambda V$可交换次序，从而可同时对角化，设Z为正交阵，使

$$(V'\Lambda^{-1}V)^{-1}=ZFZ',\ V'\Lambda V=ZEZ' \tag{3.6}$$

这里 $F=\mathrm{diag}(f_1,f_2,\cdots,f_p)$, $E=\mathrm{diag}(e_1,e_2,\cdots,e_p)$, 由(3.3)式，

$$H_1\Lambda VZE^{-1}Z'+H_2\Lambda^{-1}VZFZ'=V(H_1+H_2)$$

记 $VZ=B=(b_{ij})_{n\times p}$, 则对$\forall j$, $0\leqslant j\leqslant p$

$$H_1\lambda_ib_{ij}e_j^{-1}+H_2\lambda_i^{-1}b_{ij}f_j=b_{ij}(H_1+H_2),\ i=1,2,\cdots,n \tag{3.7}$$

从而，对$\forall j$, (3.7)仅有两个b_{ij}不为零，设b_{rj}与b_{sj}不为零，则方程

$$H_1e_j^{-1}\lambda^2-(H_1+H_2)\lambda+H_2f_j=0 \tag{3.8}$$

有两个根λ_r和λ_s，由根与系数的关系，不难得到

$$\begin{cases} f_j=\dfrac{H_1+H_2}{H_2}\dfrac{\lambda_r\lambda_s}{\lambda_r+\lambda_s} \\ e_j=\dfrac{H_1}{H_1+H_2}(\lambda_r+\lambda_s) \end{cases} \tag{3.9}$$

很明显，当$\lambda_r=\lambda_s$时，$H_1-H_2=0$，故H_1-H_2的最大值不会在这种情况下达到，我们仅考虑(3.9)在$\lambda_r\neq\lambda_s$时的情形，有

$$H_1-H_2=\left(\frac{2H_1}{H_1+H_2}\right)^p\prod_{i=1}^{p}\frac{\lambda_{r_i}+\lambda_{s_i}}{2}-\left(\frac{H_1+H_2}{2H_2}\right)^p\prod_{i=1}^{p}\left(\frac{\lambda_{r_i}^{-1}+\lambda_{s_i}^{-1}}{2}\right)^{-1} \tag{3.10}$$

这里r_i, s_i($i=1,2,\cdots,p$)为1到n中的自然数，上式经过整理可得

$$H_1-H_2=BC+\left[\left(\frac{2H_1}{H_1+H_2}\right)^p-\left(\frac{H_1+H_2}{2H_2}\right)^p\right]\prod_{i=1}^{p}\frac{2\lambda_{r_i}\lambda_{s_i}}{\lambda_{r_i}+\lambda_{s_i}},$$

这里

$$\begin{cases} C=\prod_{i=1}^{p}\frac{\lambda_{r_i}+\lambda_{s_i}}{2}-\prod_{i=1}^{p}\left(\frac{\lambda_{r_i}^{-1}+\lambda_{s_i}^{-1}}{2}\right)^{-1} \\ B=\left(\frac{2H_1}{H_1+H_2}\right)^p \end{cases} \tag{3.11}$$

类似于[4～6]中的讨论，我们有$C\leqslant S$，且根据[4]

$$H_1H_2^{-1}\leqslant\prod_{i=1}^{p}\frac{(\lambda_i+\lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}} \tag{3.12}$$

容易证明$B\leqslant C_2$，此外

$$\left(\frac{2H_1}{H_1+H_2}\right)^p-\left(\frac{H_1+H_2}{2H_2}\right)^p=-\frac{(H_1-H_2)^2}{2H_2(H_1+H_2)}$$

$$\cdot\sum_{j=0}^{p-1}\left(\frac{2H_1}{H_1+H_2}\right)^{p-j-1}\left(\frac{H_1+H_2}{2H_2}\right)^j \tag{3.13}$$

考虑到$H_1>H_2$，得

$$-\left[\left(\frac{2H_1}{H_1+H_2}\right)^p-\left(\frac{H_1+H_2}{2H_2}\right)^p\right]\prod_{i=1}^{p}\frac{2\lambda_{r_i}\lambda_{s_i}}{\lambda_{r_i}+\lambda_{s_i}}\geqslant\frac{p(H_1-H_2)^2}{2H_2(H_1+H_2)}\prod_{i=1}^{p}\frac{2\lambda_{r_i}\lambda_{s_i}}{\lambda_{r_i}+\lambda_{s_i}} \tag{3.14}$$

又

$$\frac{2^p}{H_2(H_1+H_2)}\prod_{i=1}^{p}\frac{\lambda_{r_i}\lambda_{s_i}}{\lambda_{r_i}+\lambda_{s_i}}\geqslant\frac{2^p}{2\prod_{i=1}^{p}\lambda_i^2}\prod_{i=1}^{p}\frac{\lambda_i\lambda_{n-i+1}}{\lambda_i+\lambda_{n-i+1}}$$

$$=2^{p-1}\prod_{i=1}^{p}\frac{\lambda_i^{-1}\lambda_{n-i+1}}{\lambda_i+\lambda_{n-i+1}}=C_1/2, \tag{3.15}$$

综合(3.11)，(3.14)及(3.15)，我们从(3.10)式得到

$$H_1-H_2\leqslant C_2S-\frac{p}{4}(H_1-H_2)^2C_1 \tag{3.16}$$

记$t=H_1-H_2$，得到不等式

$$\frac{p}{4}C_1t^2+t-C_2S\leqslant 0 \tag{3.17}$$

因为上式左端关于t的二次多项式的判别式$1+pC_1C_2S>0$，故由不等式(3.17)可得

$$t\leqslant\frac{2\sqrt{1+pC_1C_2S}-2}{pC_1}=\frac{2C_2S}{\sqrt{1+pC_1C_2S}+1}=KS$$

定理(2.5)式得证。

四、若 干 注 记

1．考虑矩阵二次型$X'AX$。我们若假定$X'X=I$，定理实际上给出了

$$0\leqslant\det(X'AX)-\det(X'A^{-1}X)^{-1}\leqslant KS \tag{4.1}$$

这里 K、S见上节定理，A对称。关于两个二次型$X'AX$，$(X'A^{-1}X)^{-1}$在某种度量下的商目前已有很多结果，其中以[4]，[8]中用行列式和迹度量的两类广义Kantorovich不等式应用最直接且研究也较深入，现已有很多很好的结果，[11～12]从代数的角度对这一类问题作了总结，尽管其中有些推广实用价值不大，但仍不失为一项有意义的工作，从代数的角

度讲，(4.1)也可作类似的推广.

2. 当$X'\Sigma Z=0$时. 这里Z满足$X'Z=0$且具有最大秩，则β^*是稳健的，这时$\hat{\beta}$仍然是β的BLUE. Zyskind给出了它的一个等价条件$P_X\Sigma=\Sigma P_X$（参见文献[6]），亦即

$$P_X\Sigma^2P_X=P_X\Sigma P_X(P_X\Sigma P_X) \tag{4.2}$$

[4]考虑了如下的精度

$$\mathrm{tr}(P_X\Sigma^2P_X)-\mathrm{tr}(P_X\Sigma P_X\Sigma P_X) \tag{4.3}$$

并给出了它的界. 由于$\det(P_X\Sigma^2P_X)=0$，我们对(4.2)稍作处理就会得到它的等价条件

$$X'\Sigma^2X=X'\Sigma P_X\Sigma X \tag{4.4}$$

当X列满秩时，考虑如下的精度

$$\det[X'\Sigma^2Xh)-\det(X'\Sigma X)^2 \tag{4.5}$$

它的上界也可类似地得出.

3. 用范数来度量$\hat{\beta}$的精度在[9～10]中给出了，这对于实际运用同样是有益的. 文中考虑了如下形式的精度

$$\|X'AX-(X'A^{-1}X)^{-1}\| \tag{4.6}$$

这里A正定，$\|\cdot\|$表示欧氏范数或谱范数，并且还给出了一些相应的Kantorovich型不等式.

4. 对(4.1)式的另一个可以考虑的形式为

$$\det(X'AX-(X'A^{-1}X)^{-1}) \tag{4.7}$$

其中 A正定. 它的界可以用类似于本文的方法得出. 为节约篇幅，这里从略.

参 考 文 献

[1] Puntanen, Simo, *Personal communication* (1982).

[2] Styan, G. P. H., On some inequalities associated with ordinary least squares and the Kantorovich inequality, *Festschrift for Eino Haikala on His Seventieth Birthday*, Univ. of Tampere (1983), 158—166.

[3] Rao, C. R., The inefficiency of least squares: Extensions of the Kantorovich inequality, *Linear Algebra and Its Applications*, 70 (1985), 249—255.

[4] Bloomfield, P. and G. S., Watson, The inefficiency of least squares, *Biometrika*, 62 (1975), 121—128.

[5] Knott, M., On the minimum efficiency of least square, *Biometrika*, 62 (1975), 129—132.

[6] 王松桂, 《线性模型的理论及其应用》, 安徽教育出版社 (1987).

[7] 王松桂, 广义相关系数与估计效率, 科学通报, 19 (1985), 1521—1524.

[8] 杨 虎, Kantorovich不等式的延拓与均方误差比效率, 应用数学, 4 (1988) 85—90.

[9] Wang, Song-gui (王松桂) and Yang, Hu (杨虎), Kantorovich-type inequlities and the measures of inefficiency of the GLSE, *Acta. Math. Appli. Sinica.*, 5, 4 (1989), 372—381.

[10] 杨虎、王松桂, 条件数、谱范数与估计精度, 应用概率统计 (即将发表).

[11] Khatri, C. and G., C. R. Rao, Some extension of the Kantorovich inequality and statistical applications, *J. Multi. Anal.*, 11 (1981), 498—505.

[12] Khatri, C. G. and C. R. Rao, Some generalizations of Kantorovich inequality, *Sankhya Ser. A*, 44 (1982), 91—102.

The Inefficiency of the Least Squares Estimator and Its Bound

Yang Hu

(*Chongqing Jiaotong Institute, Chongqing*)

Abstract

It was suggested by Pantanen[1] that the mean squared error may be used to measure the inefficiency of the least squares estimator. Styan[2] and Rao[3] et al. discussed this inefficiency and its bound later. In this paper we propose a new inefficiency of the least squares estimator with the measure of generalized variance and obtain its bound.

Key words inefficiency, relative efficiency, mean squared error, generalized variance, matrix derivative, best linear unbased estimator