

文章编号: 1000-5013(2011)05-0584-04

复共线性下的主成分全最小二乘估计

刘文丽, 吕书龙, 梁飞豹

(福州大学 数学与计算机科学学院, 福建 福州 350108)

摘要: 针对最小二乘法在参数估计中的局限性, 在多维解释变量存在复共线性时, 提出主成分全最小二乘估计, 避免奇异矩阵求逆的问题. 经多组大量测试, 计算得到的回归系数的平均绝对偏差均较小, 且表现稳定, 其效果明显地优于最小二乘估计和全最小二乘估计.
关键词: 最小二乘估计; 全最小二乘估计; 主成分; 复共线性
中图分类号: O 212 **文献标志码:** A

1 基本知识

在参数估计中, 最常见且广泛应用的是最小二乘法. 值得注意的是, 在 Gauss-Markov 模型假设中, 最小二乘估计表现出众多优良的性质. 模型中只有响应变量受随机因素的影响, 设为随机变量, 而所有解释变量均为可精确控制的普通变量, 且忽略其测量误差. 显然, 这在很多实际应用中是很难满足的. 近年来, 同时考虑解释变量和响应变量误差的 EV(errors-in-variables)模型和全最小二乘方法(也称正交最小二乘)引起了众多学者的研究兴趣, 并在统计分析、线性和非线性回归^[1-4], 信号处理^[5-6], 系统辨识^[7-8]和网络构造^[9]中有了广泛的应用, 成为了数理统计、数值代数和决策控制等方向的热点之一.

考虑一般的线性回归模型, 即

$$\left. \begin{aligned} \hat{\mathbf{y}} &= \beta_0 + \boldsymbol{\beta}^T \hat{\mathbf{x}}, \\ \mathbf{y} &= \hat{\mathbf{y}} + \boldsymbol{\varepsilon}, \\ \mathbf{x} &= \hat{\mathbf{x}} + \boldsymbol{\eta}. \end{aligned} \right\} \tag{1}$$

式(1)中: $\hat{\mathbf{y}}, \hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_p)^T$ 是满足线性关系, 但不可精确观测的普通变量; $\mathbf{y}, \mathbf{x} = (x_1, x_2, \dots, x_p)^T$ 是相应观测的随机变量; $\boldsymbol{\varepsilon}$ 是响应变量 \mathbf{y} 与 $\hat{\mathbf{y}}$ 间的误差; $\boldsymbol{\eta}$ 是解释变量 x_1, x_2, \dots, x_p 与 $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_p$ 间的 p 维随机误差; β_0 和 $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ 是未知参数.

假设解释变量 x_1, x_2, \dots, x_p 和响应变量 \mathbf{y} 的 n 组观测值为 $(x_{i,1}, \dots, x_{i,p}; y_i), i = 1, 2, \dots, n$. 于是, 模型(1)可表示为

$$\left. \begin{aligned} \hat{\mathbf{y}} &= \beta_0 \mathbf{I}_n + \hat{\mathbf{X}} \boldsymbol{\beta}, \\ \mathbf{y} &= \hat{\mathbf{y}} + \boldsymbol{\varepsilon}, \\ \mathbf{X} &= \hat{\mathbf{X}} + \boldsymbol{\eta}. \end{aligned} \right\} \tag{2}$$

式(2)中: \mathbf{I}_n 是所有分量皆为 1 的 $n \times 1$ 向量; $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$; $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,p} \\ x_{21} & x_{22} & \cdots & x_{2,p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n,p} \end{bmatrix}$; $\hat{\mathbf{y}}, \boldsymbol{\varepsilon}$ 均为 $n \times 1$ 向量; $\hat{\mathbf{X}}, \boldsymbol{\eta}$ 均为 $n \times p$ 矩阵.

当随机矩阵 $\boldsymbol{\eta}$ 为零矩阵时, 模型(2)退化为一般的 Gauss-Markov 模型.

全最小二乘法就是要寻找参数 $\beta_0, \beta_1, \dots, \beta_p$, 使上述 n 个数据点到回归超平面 $\hat{\mathbf{y}} = \beta_0 + \boldsymbol{\beta}^T \hat{\mathbf{x}}$ 的垂直距离之和

$$Q_{\text{TSL}}(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2}{1 + \sum_{j=1}^p \beta_j^2} \tag{3}$$

达到最小.

定义 1 若 $\hat{\beta}_{\text{TSL}}^0, \hat{\boldsymbol{\beta}}_{\text{TSL}}$ 满足

$$Q_{\text{TSL}}(\hat{\beta}_{\text{TSL}}^0, \hat{\boldsymbol{\beta}}_{\text{TSL}}) = \min_{\beta_0, \boldsymbol{\beta}} Q_{\text{TSL}}(\beta_0, \boldsymbol{\beta}), \tag{4}$$

则称 $\hat{\beta}_{\text{TSL}}^0, \hat{\boldsymbol{\beta}}_{\text{TSL}}$ 为参数 $\beta_0, \boldsymbol{\beta}$ 的全最小二乘估计.

将式(3)对 β_0 求导, 令其为 0 并求解, 易得如下引理.

引理 1 $\hat{\beta}_{\text{TSL}}^0 = \bar{\mathbf{y}} - \hat{\boldsymbol{\beta}}_{\text{TSL}}^T \bar{\mathbf{x}}$. 其中: $\bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n y_i; \bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)^T; \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{i,j}, j=1, 2, \dots, p$.

该引理说明全最小二乘回归平面经过数据中心点 $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p; \bar{y})$. 假设数据点已进行了中心化, 即 $\bar{y}=0, \bar{x}_j=0, j=1, 2, \dots, p$, 模型(2)的全最小二乘问题可简化为

$$\left. \begin{aligned} \hat{\mathbf{y}} &= \hat{\mathbf{X}}\boldsymbol{\beta}, \\ \mathbf{y} &= \hat{\mathbf{y}} + \boldsymbol{\varepsilon}, \\ \mathbf{X} &= \hat{\mathbf{X}} + \boldsymbol{\eta}. \end{aligned} \right\} \tag{5}$$

该模型通常称为 EV 模型.

定义 2 全最小二乘法就是优化 $\min_{\boldsymbol{\beta}, \boldsymbol{\varepsilon}, \boldsymbol{\eta}} \|\boldsymbol{\beta}, \boldsymbol{\varepsilon}\|_F$, 满足 $(\mathbf{X} - \boldsymbol{\eta})\boldsymbol{\beta} = \mathbf{y} - \boldsymbol{\varepsilon}$. 这里, $\|\cdot\|_F$ 为 Frobenius 范数. 该优化问题中 $\boldsymbol{\beta}$ 的解, 称为全最小二乘估计 $\hat{\boldsymbol{\beta}}_{\text{TSL}}$.

残差 $[\boldsymbol{\eta}, \boldsymbol{\varepsilon}] = [\hat{\mathbf{X}} - \mathbf{X}, \hat{\mathbf{y}} - \mathbf{y}] = [\hat{\mathbf{X}}, \hat{\mathbf{y}}] - [\mathbf{X}, \mathbf{y}]$, 而 $[\hat{\mathbf{X}}, \hat{\mathbf{y}}]$ 在回归平面 $\hat{\mathbf{y}} = \beta_0 + \hat{\boldsymbol{\beta}}^T \hat{\mathbf{x}}$ 上, 可见 $[\boldsymbol{\varepsilon}, \boldsymbol{\eta}]$ 垂直于该平面时, 其相应的 Frobenius 范数达到最小, 则关于全最小二乘的定义 1 和定义 2 是一致的.

引理 2^[1-2] 当 $\sigma'_p > \sigma_{p+1}$ 时, 在模型(5)中, 回归系数 $\boldsymbol{\beta}$ 的全最小二乘解可表示为

$$\hat{\boldsymbol{\beta}}_{\text{TSL}} = (\mathbf{X}^T \mathbf{X} - \sigma_{p+1}^2 \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}, \tag{6}$$

式(6)中: σ'_p, σ_{p+1} 分别是 $\mathbf{X}, [\mathbf{X}^T \mathbf{y}]$ 的最小奇异值.

明显地, 当 $\sigma_{p+1}^2 = 0$, 即 \mathbf{X} 的 p 个列向量与 \mathbf{y} 之间有着完全的线性关系时, 全最小二乘估计 $\hat{\boldsymbol{\beta}}_{\text{TSL}}$ 为一般的最小二乘估计 $\hat{\boldsymbol{\beta}}_{\text{LS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$.

2 主成分全最小二乘估计

设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ 为 $\mathbf{X}^T \mathbf{X}$ 的特征根, $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_p$ 为相对应的标准正交化特征向量, 记 $\boldsymbol{\Phi} = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_p)$, 它为正交阵. 又令 $\mathbf{Z} = \mathbf{X}\boldsymbol{\Phi} = (\mathbf{X}\boldsymbol{\varphi}_1, \mathbf{X}\boldsymbol{\varphi}_2, \dots, \mathbf{X}\boldsymbol{\varphi}_p)$, 则有

$$\boldsymbol{\Phi}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\Phi} = \mathbf{Z}^T \mathbf{Z} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p).$$

此时, $\boldsymbol{\beta}$ 的全最小二乘估计 $\hat{\boldsymbol{\beta}}_{\text{TSL}}$ 也可表示为

$$\hat{\boldsymbol{\beta}}_{\text{TSL}} = \boldsymbol{\Phi}(\mathbf{Z}^T \mathbf{Z} - \sigma_{p+1}^2 \mathbf{I})^{-1} \mathbf{Z}^T \mathbf{y}. \tag{7}$$

当 \mathbf{X} 存在复共线关系时, 有一些 $\mathbf{X}^T \mathbf{X}$ 的特征根很小. 此时, 矩阵 $\mathbf{Z}^T \mathbf{Z}$ 接近奇异, 则 $\mathbf{Z}^T \mathbf{Z} - \sigma_{p+1}^2 \mathbf{I}$ 更接近奇异矩阵. 明显地, 全最小二乘解 $\hat{\boldsymbol{\beta}}_{\text{TSL}}$ 变坏.

借助挑选主成分的思想, 提出主成分全最小二乘法. 不妨假设 $\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_p \approx 0$, 此时后面的 $p-r$ 个主成分 $(\mathbf{X}\boldsymbol{\varphi}_i, i=r+1, r+2, \dots, p)$ 取值变动就很小. 于是用前面 r 个主成分 $(\mathbf{X}\boldsymbol{\varphi}_i, i=1, 2, \dots, r)$ 与响应变量 \mathbf{p} , 应用全最小二乘法进行回归, 再变回到原来的解释变量, 就得到了主成分全最小二乘估计.

定义 3 当 $\lambda_r > \sigma_{r+1}$ 时, 称 $\hat{\boldsymbol{\beta}}^* = \boldsymbol{\Phi}_1(\mathbf{Z}_1^T \mathbf{Z}_1 - \sigma_{r+1}^2 \mathbf{I})^{-1} \mathbf{Z}_1^T \mathbf{y}$ 为模型(5)中 $\boldsymbol{\beta}$ 的主成分全最小二乘估计. 其中: $\boldsymbol{\Phi}_1 = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_r); \mathbf{Z}_1 = \mathbf{X}\boldsymbol{\Phi}_1; \sigma_{r+1}$ 是矩阵 $[\mathbf{Z}_1, \mathbf{y}]$ 的最小奇异值. 明显地, 当 $r=p$ 时, 主成分全最小二乘估计 $\hat{\boldsymbol{\beta}}^*$ 退化为全最小二乘估计 $\hat{\boldsymbol{\beta}}_{\text{TSL}}$.

另外, 下面给出的性质说明当 Gauss-Markov 模型中参数满足一定条件时, 在均方误差准则下, 主成分全最小二乘估计 $\hat{\boldsymbol{\beta}}^*$ 优于最小二乘估计.

性质 1 在模型(5)中,若误差向量 $\boldsymbol{\varepsilon}, \boldsymbol{\eta}$ 满足 $E(\boldsymbol{\varepsilon})=0, D(\boldsymbol{\varepsilon})=\sigma^2 \boldsymbol{I}, \boldsymbol{\eta}=0$, 即在一般的 Gauss-Markov 模型下,当满足

$$\sum_{i=1}^r (\frac{\sigma_{r+1}^2}{\lambda_i - \sigma_{r+1}^2} \boldsymbol{\varphi}_i^T \boldsymbol{\beta})^2 + \sum_{i=r+1}^p (\boldsymbol{\varphi}_i^T \boldsymbol{\beta})^2 \leq \sigma^2 [\sum_{i=1}^r \frac{\sigma_{r+1}^2 (\sigma_{r+1}^2 - 2\lambda_i)}{\lambda_i (\lambda_i - \sigma_{r+1}^2)^2} + \sum_{i=r+1}^p \frac{1}{\lambda_i}], \tag{8}$$

则有 $MSE(\hat{\boldsymbol{\beta}}^*) \leq MSE(\hat{\boldsymbol{\beta}}_{LS})$.

证明 令 $\boldsymbol{\Phi}_1=(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \cdots, \boldsymbol{\varphi}_r), \boldsymbol{Z}_1=\boldsymbol{X}\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2=(\boldsymbol{\varphi}_{r+1}, \boldsymbol{\varphi}_{r+2}, \cdots, \boldsymbol{\varphi}_p), \boldsymbol{Z}_2=\boldsymbol{X}\boldsymbol{\Phi}_2, \boldsymbol{A}_1=\boldsymbol{Z}_1^T \boldsymbol{Z}_1=\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r), \boldsymbol{A}_2=\boldsymbol{Z}_2^T \boldsymbol{Z}_2=\text{diag}(\lambda_{r+1}, \lambda_{r+2}, \cdots, \lambda_p)$, 则 $\hat{\boldsymbol{\beta}}_{LS}, \hat{\boldsymbol{\beta}}^*$ 可分别表示为

$$\hat{\boldsymbol{\beta}}_{LS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{A}^{-1} \boldsymbol{Z}^T \boldsymbol{y} = \begin{bmatrix} \boldsymbol{\Phi}_1 \boldsymbol{A}_1^{-1} \boldsymbol{Z}_1^T \boldsymbol{y} \\ \boldsymbol{\Phi}_2 \boldsymbol{A}_2^{-1} \boldsymbol{Z}_2^T \boldsymbol{y} \end{bmatrix}, \tag{9}$$

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{\Phi}_1 (\boldsymbol{A}_1 - \sigma_{r+1}^2 \boldsymbol{I})^{-1} \boldsymbol{Z}_1^T \boldsymbol{y} = \boldsymbol{\Phi}_1 (\boldsymbol{A}_1 - \sigma_{r+1}^2 \boldsymbol{I})^{-1} \boldsymbol{A}_1 \boldsymbol{\Phi}_1^T \cdot \boldsymbol{\Phi}_1 \boldsymbol{A}_1^{-1} \boldsymbol{Z}_1^T \boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{\Phi}^T \hat{\boldsymbol{\beta}}_{LS}. \tag{10}$$

其中 $\boldsymbol{A}=\begin{bmatrix} (\boldsymbol{A}_1 - \sigma_{r+1}^2 \boldsymbol{I})^{-1} \boldsymbol{A}_1 & 0 \\ 0 & 0 \end{bmatrix}_{p \times p}$. 若误差向量 $\boldsymbol{\varepsilon}, \boldsymbol{\eta}$ 满足 $E(\boldsymbol{\varepsilon})=0, D(\boldsymbol{\varepsilon})=\sigma^2 \boldsymbol{I}, \boldsymbol{\eta}=0$, 即在一般的 Gauss-Markov 模型下,有

$$E(\hat{\boldsymbol{\beta}}^*) = \boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{\Phi}^T E(\hat{\boldsymbol{\beta}}_{LS}) = \boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{\Phi}^T \boldsymbol{\beta}, \tag{11}$$

$$\text{cov}(\hat{\boldsymbol{\beta}}^*) = \boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{\Phi}^T \text{cov}(\hat{\boldsymbol{\beta}}) \boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{\Phi}^T = \sigma^2 \boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{A}^{-1} \boldsymbol{A} \boldsymbol{\Phi}^T, \tag{12}$$

$$MSE(\hat{\boldsymbol{\beta}}_{LS}) = E \|\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}\|_2 = \text{tr cov}(\hat{\boldsymbol{\beta}}_{LS}) + \|E(\hat{\boldsymbol{\beta}}_{LS}) - \boldsymbol{\beta}\|_2 =$$

$$\sigma^2 \text{tr}((\boldsymbol{X}^T \boldsymbol{X})^{-1}) = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}, \tag{13}$$

$$MSE(\hat{\boldsymbol{\beta}}^*) = \text{tr cov}(\hat{\boldsymbol{\beta}}^*) + \|E(\hat{\boldsymbol{\beta}}^*) - \boldsymbol{\beta}\|_2 = \sigma^2 \text{tr}(\boldsymbol{\Phi} \boldsymbol{A} \boldsymbol{A}^{-1} \boldsymbol{A} \boldsymbol{\Phi}^T) + \boldsymbol{\beta}^T \boldsymbol{\Phi} (\boldsymbol{I} - \boldsymbol{A})^2 \boldsymbol{\Phi}^T \boldsymbol{\beta} =$$

$$\sigma^2 \sum_{i=1}^r \frac{\lambda_i}{(\lambda_i - \sigma_{r+1}^2)^2} + \sum_{i=1}^r (\frac{\sigma_{r+1}^2}{\lambda_i - \sigma_{r+1}^2} \boldsymbol{\varphi}_i^T \boldsymbol{\beta})^2 + \sum_{i=r+1}^p (\boldsymbol{\varphi}_i^T \boldsymbol{\beta})^2, \tag{14}$$

$$MSE(\hat{\boldsymbol{\beta}}_{LS}) - MSE(\hat{\boldsymbol{\beta}}^*) =$$

表 1 模型的平均回归系数及平均绝对偏差

Tab. 1 Means of coefficients and absolute deviations

样本量	参数	最小二乘法		
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
50	\bar{K}	1.000 167	0.998 566	1.001 339
	MAD	0.008 020	0.124 320	0.124 352
200	\bar{K}	1.000 465	1.002 934	0.996 972
	MAD	0.004 242	0.057 869	0.057 909
500	\bar{K}	0.999 198	1.005 038	0.995 188
	MAD	0.002 676	0.046 112	0.045 806

样本量	参数	全最小二乘法		
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
50	\bar{K}	1.010 324	2.316 927	-0.320 87
	MAD	0.023 553	3.240 553	3.242 226
200	\bar{K}	0.999 962	0.774 187	1.225 752
	MAD	0.005 448	0.942 585	0.942 372
500	\bar{K}	0.999 025	1.068 177	0.932 111
	MAD	0.002 998	0.426 457	0.426 129

样本量	参数	主成分全最小二乘法		
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
50	\bar{K}	0.999 870	1.000 014	0.999 977
	MAD	0.008 083	0.001 256	0.001 273
200	\bar{K}	1.000 054	0.999 946	1.000 067
	MAD	0.004 228	0.000 610	0.000 664
500	\bar{K}	0.998 786	1.000 198	1.000 129
	MAD	0.002 692	0.000 469	0.000 423

3 数据模拟

构造了 $\hat{x}_1 \sim N(4, 1), \hat{x}_2 = \hat{x}_1 + e, e \sim N(0, 0.005^2)$, 误差 $\varepsilon, \eta_1, \eta_2$ 均服从 $N(0, 0.01^2)$ 分布, 且 $\beta_0 = \beta_1 = \beta_2 = 1$, 令

$$\left. \begin{aligned} \hat{\boldsymbol{y}} &= \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2, \\ \boldsymbol{y} &= \hat{\boldsymbol{y}} + \boldsymbol{\varepsilon}, \\ x_1 &= \hat{x}_1 + \eta_1, x_2 = \hat{x}_2 + \eta_2. \end{aligned} \right\} \tag{16}$$

在随机实验中,分别取样本量为 50, 200, 500, 并各做 1 000 次随机模拟. 最小二乘法、全最小二乘法和主成分全最小二乘法等 3 个模型的回归系数平均值(\bar{K})及绝对偏差平均值(MAD), 如表 1 所示. 表

1 中: $\bar{K} = \frac{1}{1\,000} \sum_{i=1}^{1\,000} \hat{\beta}_j^{(i)}; MAD = \frac{1}{1\,000} \times$

$\sum_{i=1}^{1\,000} |\hat{\beta}_j^{(i)} - \beta_j|$; $\hat{\beta}_j^{(i)}$ 为 β_j 的第 i 次估计值.

显然在构造的数据中, 解释变量 $\mathbf{x}_1, \mathbf{x}_2$ 间存在严重的复共线性. 在样本量为 50 的一次随机模拟中, $\mathbf{X}^T \mathbf{X}$ 的两个特征根分别为 92.466 650 和 0.005 615, 因此回归系数的最小二乘估计的偏差较大; 而此时 $[\mathbf{X}, \mathbf{y}]^T [\mathbf{X}, \mathbf{y}]$ 的最小特征根即式(6)中的 $\sigma_{p+1}^2 = 0.005\,250$. 于是, $0.005\,615 - 0.005\,250$ 的差值更接近于零, 因此全最小二乘估计在回归系数的平均值及绝对偏差平均值方面均表现得比较差.

4 结束语

经典的最小二乘估计只考虑了响应变量的误差, 而全最小二乘估计还考虑了解释变量的误差, 故全最小二乘法更符合实际情况, 有着更广泛的应用. 在解释变量为多维且解释变量间存在严重复共线性时, 最小二乘估计和全最小二乘估计都会涉及奇异阵求逆的问题. 提出的主成分全最小二乘估计, 避免了奇异阵求逆的问题, 并在一定情况下证明了它优于最小二乘估计. 另外, 在一般的 EV 模型中, 全最小二乘估计和主成分全最小二乘估计的性质还值得进一步研究.

参考文献:

[1] VAN HUFFEL S, ZHA H. The total least squares problem[M]. Amsterdam: Frontiers in Applied Mathematics, 1993.

[2] 张洪钺, 黄劲东, 范文雷. 全最小二乘法及其在参数估计中的应用[J]. 自动化学报, 1995, 21(121): 40-47.

[3] 顾启泰. 正交最小二乘算法及其应用[J]. 清华大学学报: 自然科学版, 1996, 36(3): 106-112.

[4] 王福昌, 曹慧荣, 朱红霞. 经典最小二乘与全最小二乘法及其参数估计[J]. 统计与决策, 2009(1): 16-17.

[5] MARKOVSKY I, VAN HUFFEL S. Overview of total least-squares methods[J]. Signal Processing, 2007, 87: 2283-2302.

[6] 孔祥玉, 韩崇昭, 魏瑞轩, 等. 一种全最小二乘算法及其在非线性滤波器中的应用[J]. 信号处理, 2005, 7(2): 140-143.

[7] 王宏伟, 刘向利, 马广富. 基于正交最小二乘估计的非线性系统模糊辨识[J]. 计算机学报, 2004, 27(8): 1143-1146.

[8] 李德强, 黄莎白. 基于正交最小二乘法的小波网络在系统辨识中的应用[J]. 控制与决策, 2003, 18(3): 378-381.

[9] 黄宜军, 刘宏兵, 熊炎. QR-ROLS 算法在小波网络构造中的仿真研究[J]. 系统仿真学报, 2007, 19(9): 2101-2104.

Total Least Square Estimator of Principal Components under Multicollinearity

LIU Wen-li, LÜ Shu-long, LIANG Fei-bao

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

Abstract: The total least square estimator of principal components is proposed to avoid inversing the singular matrix under explanatory variables of multi-dimensions with multicollinearity. Under certain conditions, we prove by numerous test that it has smaller mean square errors (MSE) than least square estimator (LSE).

Keywords: least square estimator; total least square estimator; principal components; multicollinearity

(责任编辑: 陈志贤 英文审校: 张金顺, 黄心中)