第25卷 第3期 石家庄铁道大学学报(自然科学版) Vol. 25 No. 3
2012年9月 JOURNAL OF SHIJIAZHUANG TIEDAO UNIVERSITY (NATURAL SCIENCE) Sep. 2012

线性回归模型的一个新的随机约束岭估计

司童文

(华盛顿大学 艺术与科学学院,美国 西雅图 98195-3765)

摘要:研究带随机先验信息的线性回归模型,提出一种新的随机约束岭估计,得到新的估计在均方误差矩阵意义下优于混合估计、普通岭估计、随机混合估计(Yalian,2008)和另一随机混合估计(富月,2003)的充要条件。理论结果与数值实验都表明:新的估计在均方误差矩阵意义下的性能是优良的。

关键词:岭估计;混合估计;随机约束岭估计;随机混合估计;均方误差矩阵

中图分类号:O211.5 **文献标识码:** A **文章编号:** 2095-0373(2012)03-0107-04

0 引言

回归系数的参数估计问题是应用统计和计量经济领域的热门问题[1-2]。通常,人们利用样本信息来估计回归系数。然而在许多实际问题中,人们还利用一些先验信息来估计参数。可以从计量经济学理论、以往的试验或基于贝叶斯观点得到一些先验信息。引入正确的先验信息有助于改进估计。Theil et al[3]对附加了随机线性约束的样本模型提出了未知参数的随机约束最小二乘估计(通常称为“混合估计”),并证明了在均方误差意义下无论约束正确与否混合估计都优于无约束的最小二乘估计。

计量数据往往因存在复共线性而不易处理,因此在病态模型中的参数估计变得十分重要。Hoerl et al[4]发现在工程领域中复共线性也是普遍存在的。为解决这个问题,采用 $C(k) = X'X + kI_p$ 的形式而非 C 形式来估计回归系数 β。此方法即是著名的岭估计。

文献[5]、文献[6]通过两种不同的方法得到了混合估计和岭估计的组合估计,分别称为随机混合岭估计(OMRE)和岭型混合估计(ORME)。通过另一种方法提出了一个新的随机约束岭估计,它依然是混合估计和岭估计的组合估计,并得到了在均方误差阵意义下新估计优于混合估计和岭估计的充要条件。

1 模型和估计

考虑如下线性回归模型

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

式中,$\boldsymbol{y}$ 是 $n \times 1$ 的可观测向量;$\boldsymbol{X}$ 是 $n \times p$ 的列满秩矩阵;$\boldsymbol{\beta}$ 是 $p \times 1$ 的未知参数;$\boldsymbol{\varepsilon}$ 是均值为0,协方差矩阵为 $\sigma^2\boldsymbol{I}$ 的 $n \times 1$ 的误差向量。

另外,对于模型(1)给出关于 $\boldsymbol{\beta}$ 的如下形式的先验信息

$$\boldsymbol{r} = R\boldsymbol{\beta} + \boldsymbol{e} \tag{2}$$

式中,$\boldsymbol{r}$ 是 $j \times 1$ 的可观测随机向量;$\boldsymbol{R}$ 是 $j \times p$ 的行满秩矩阵;$\boldsymbol{e}$ 是 $j \times 1$ 均值为0,协方差矩阵为 $\sigma^2\boldsymbol{W}$ 的随机向量,此处 $\boldsymbol{W}$ 是已知的正定矩阵。因此,除了均值,约束(2) 并不完全保持。但下面的讨论均假定约束(2) 成立。同时假定 $\boldsymbol{\varepsilon}$ 与 $\boldsymbol{e}$ 是相互独立的。

混合估计定义为如下形式

$$\hat{\boldsymbol{\beta}}_{OME} = (\boldsymbol{S} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R})^{-1}(\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{r}) \tag{3}$$

这里 $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}$。利用 $(\boldsymbol{S} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R})^{-1} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{R}'(\boldsymbol{W} + \boldsymbol{R}\boldsymbol{S}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{S}^{-1}$ 和 $(\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{R}'(\boldsymbol{W} +$

收稿日期:2012-05-07
作者简介:司童文 男 1989年出生

$RS^{-1}R')^{-1}RS^{-1})R'W^{-1}r = S^{-1}R'(W+RS^{-1}R')^{-1}r$,(3)可以改写为

$$\hat{\boldsymbol{\beta}}_{OME} = \hat{\boldsymbol{\beta}}_{OLSE} + S^{-1}R'(W+RS^{-1}R')^{-1}(r - R\hat{\boldsymbol{\beta}}_{OLSE}) \tag{4}$$

这里$\hat{\boldsymbol{\beta}}_{OLSE} = S^{-1}X'y$为最小二乘估计。

岭估计定义为如下形式

$$\hat{\boldsymbol{\beta}}_{RE}(k) = (S+kI)^{-1}X'y, k>0 \tag{5}$$

令$T(k) = (S+kI)^{-1}S = (I+kS^{-1})^{-1}, (k \geqslant 0)$,并注意到$T(k)$和$S^{-1}$是可交换的,可改写$\hat{\boldsymbol{\beta}}_{RE}(k)$为

$$\hat{\boldsymbol{\beta}}_{RE}(k) = T(k)\hat{\boldsymbol{\beta}}_{OLSE} = T(k)S^{-1}X'y = S^{-1}T(k)X'y \tag{6}$$

Li Yalian et al[5]得到如下形式的组合估计,称为随机混合岭估计(OMRE)

$$\hat{\boldsymbol{\beta}}_{OMRE}(k) = T(k)\hat{\boldsymbol{\beta}}_{OME} = T(k)(S+R'W^{-1}R)^{-1}(X'y+R'W^{-1}r) \tag{7}$$

富月[6]得到另一种形式的组合估计,称为岭型混合估计(ORME)

$$\hat{\boldsymbol{\beta}}_{ORME}(k) = (S+kI+R'W^{-1}R)^{-1}(X'y+R'W^{-1}r) = (C(k)+R'W^{-1}R)^{-1}(X'y+R'W^{-1}r) \tag{8}$$

在此,通过一种新的组合,将$\hat{\boldsymbol{\beta}}_{RE}(k)$代替混合估计中的$\hat{\boldsymbol{\beta}}_{OLSE}$得到一种新的随机约束估计,称为随机约束岭估计(SRRE)

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{SRRE}(k) &= \hat{\boldsymbol{\beta}}_{RE}(k) + S^{-1}R'(W+RS^{-1}R')^{-1}(r - R\hat{\boldsymbol{\beta}}_{RE}(k)) = \\ &S^{-1}T(k)X'y + S^{-1}R'(W+RS^{-1}R')^{-1}(r - RS^{-1}T(k)X'y) = \\ &(S^{-1} - S^{-1}R'(W+RS^{-1}R')^{-1}RS^{-1})(T(k)X'y + R'W^{-1}r) = \\ &(S+R'W^{-1}R)^{-1}(T(k)X'y+R'W^{-1}r)\end{aligned} \tag{9}$$

从定义的形式可以看出,$\hat{\boldsymbol{\beta}}_{SRRE}(k)$是混合估计、岭估计和最小二乘估计的推广。如果$k=0, \hat{\boldsymbol{\beta}}_{SRRE}(0) = \hat{\boldsymbol{\beta}}_{OME}$;如果$R=0, \hat{\boldsymbol{\beta}}_{SRRE}(k) = \hat{\boldsymbol{\beta}}_{RE}(k)$;如果$k=0$和$R=0, \hat{\boldsymbol{\beta}}_{SRRE}(0) = \hat{\boldsymbol{\beta}}_{OLSE}$。

此外,容易计算新估计的期望和方差为

$$E(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = A(T(k)S+R'W^{-1}R)\boldsymbol{\beta} = \boldsymbol{\beta} + A(T(k)-I)S\boldsymbol{\beta} \tag{10}$$

$$Var(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = \sigma^2 A(T(k)ST(k)' + R'W^{-1}R)A \tag{11}$$

这里$A = (S+R'W^{-1}R)^{-1}$。可见$\hat{\boldsymbol{\beta}}_{SRRE}(k)$总是有偏的。

2 在 MSEM 意义下各有偏估计的比较

类似于杨虎等[7]的方法,比较各有偏估计的优劣。此处采用均方误差矩阵作为比较的标准。$\hat{\boldsymbol{\beta}}$的均方误差矩阵定义如下

$$MSEM(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' = Var(\hat{\boldsymbol{\beta}}) + bias(\hat{\boldsymbol{\beta}})bias(\hat{\boldsymbol{\beta}})' \tag{12}$$

这里$Var(\hat{\boldsymbol{\beta}})$, $bias(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}$分别表示方差阵和偏差。在均方误差矩阵意义下说$\hat{\boldsymbol{\beta}}_2$优于$\hat{\boldsymbol{\beta}}_1$当且仅当下式成立

$$\Delta(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = MSEM(\hat{\boldsymbol{\beta}}_1) - MSEM(\hat{\boldsymbol{\beta}}_2) \geqslant 0 \tag{13}$$

2.1 随机约束岭估计与混合估计、岭估计的比较

通过(3)、(6)、(9)、(10)和(11),容易得到$\hat{\boldsymbol{\beta}}_{OME}$, $\hat{\boldsymbol{\beta}}_{RE}(k)$和$\hat{\boldsymbol{\beta}}_{SRRE}(k)$的均方误差矩阵

$$MSEM(\hat{\boldsymbol{\beta}}_{OME}) = \sigma^2 A,$$

$$MSEM(\hat{\boldsymbol{\beta}}_{RE}(k)) = \sigma^2 T(k)S^{-1}T(k)' + b_1 b_1',$$

$$MSEM(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = \sigma^2 A(T(k)ST(k)' + R'W^{-1}R)A + b_2 b_2'。$$

这里$b_1 = (T(k)-I)\boldsymbol{\beta}$, $b_2 = bias(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = A(T(k)-I)S\boldsymbol{\beta}$。

为比较$\hat{\boldsymbol{\beta}}_{SRRE}(k)$和$\hat{\boldsymbol{\beta}}_{OME}(k)$, $\hat{\boldsymbol{\beta}}_{RE}(k)$,做如下处理

$$\Delta_1 = MSEM(\hat{\boldsymbol{\beta}}_{OME}) - MSEM(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = \sigma^2 D_1 - b_2 b_2' \tag{14}$$

$$\Delta_2 = MSEM(\hat{\boldsymbol{\beta}}_{RE}(k)) - MSEM(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = \sigma^2 D_2 + b_1 b_1' - b_2 b_2' \tag{15}$$

这里

$$D_1 = A - A(T(k)ST(k)' + R'W^{-1}R)A = A(S - T(k)ST(k)')A \tag{16}$$

$$D_2 = T(k)S^{-1}T(k)' - A(T(k)ST(k)' + R'W^{-1}R)A \tag{17}$$

由此可以得到随机约束岭估计优于混合估计、岭估计的充要条件:

定理 1　在均方误差矩阵意义下,$\hat{\boldsymbol{\beta}}_{SRRE}(k)$ 优于 $\hat{\boldsymbol{\beta}}_{OME}$ 的充要条件是 $\boldsymbol{b}_2'(\sigma^2\boldsymbol{D}_1^{-1})\boldsymbol{b}_2 \leqslant 1$。

定理 2　在均方误差矩阵意义下,$\hat{\boldsymbol{\beta}}_{SRRE}(k)$ 优于 $\hat{\boldsymbol{\beta}}_{RE}(k)$ 的充要条件是 $\boldsymbol{b}_2'(\sigma^2\boldsymbol{D}_2 + \boldsymbol{b}_1\boldsymbol{b}_1')^{-1}\boldsymbol{b}_2 \leqslant 1$。

2.2　随机约束岭估计与随机混合岭估计、岭型混合估计的比较

通过(7)和(8),可以得到 $\hat{\boldsymbol{\beta}}_{OMRE}(k)$ 和 $\hat{\boldsymbol{\beta}}_{ORME}(k)$ 均方误差矩阵如下

$$MSEM(\hat{\boldsymbol{\beta}}_{OMRE}(k)) = \sigma^2 T(k)AT(k)' + \boldsymbol{b}_1\boldsymbol{b}_1' \tag{18}$$

$$MSEM(\hat{\boldsymbol{\beta}}_{ORME}(k)) = \sigma^2(A^{-1} + kI)^{-1}A^{-1}(A^{-1} + kI)^{-1} + \boldsymbol{b}_3\boldsymbol{b}_3' \tag{19}$$

这里 $\boldsymbol{b}_3 = [(I + kA)^{-1} - I]\boldsymbol{\beta}$ 。

类似上面,可以计算它们的差为

$$\Delta_3 = MSEM(\hat{\boldsymbol{\beta}}_{OMRE}) - MSEM(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = \sigma^2 D_3 + \boldsymbol{b}_1\boldsymbol{b}_1' - \boldsymbol{b}_2\boldsymbol{b}_2' \tag{20}$$

$$\Delta_4 = MSEM(\hat{\boldsymbol{\beta}}_{ORME}) - MSEM(\hat{\boldsymbol{\beta}}_{SRRE}(k)) = \sigma^2 D_4 + \boldsymbol{b}_3\boldsymbol{b}_3' - \boldsymbol{b}_2\boldsymbol{b}_2' \tag{21}$$

$$\Delta_5 = MSEM(\hat{\boldsymbol{\beta}}_{OMRE}) - MSEM(\hat{\boldsymbol{\beta}}_{ORME}(k)) = \sigma^2 D_5 + \boldsymbol{b}_1\boldsymbol{b}_1' - \boldsymbol{b}_3\boldsymbol{b}_3' \tag{22}$$

这里

$$D_3 = T(k)AT(k)' - A(T(k)ST(k)' + R'W^{-1}R)A \tag{23}$$

$$\begin{aligned} D_4 &= (A^{-1} + kI)^{-1}A^{-1}(A^{-1} + kI)^{-1} - A(T(k)ST(k)' + R'W^{-1}R)A = \\ &A[(I + kA)^{-1}A^{-1}(I + kA)^{-1} - (T(k)ST(k) + R'W^{-1}R)]A \end{aligned} \tag{24}$$

$$D_5 = T(k)AT(k)' - (A^{-1} + kI)^{-1}A^{-1}(A^{-1} + kI)^{-1} \tag{25}$$

从而有如下的结论:

定理 3　对于模型(1)和(2),在均方误差矩阵意义下有:

(1) $\hat{\boldsymbol{\beta}}_{SRRE}(k)$ 优于 $\hat{\boldsymbol{\beta}}_{OMRE}(k)$ 的充要条件是 $\boldsymbol{b}_2'(\sigma^2\boldsymbol{D}_3 + \boldsymbol{b}_1\boldsymbol{b}_1')\boldsymbol{b}_2 \leqslant 1$;

(2) $\hat{\boldsymbol{\beta}}_{SRRE}(k)$ 优于 $\hat{\boldsymbol{\beta}}_{ORME}(k)$ 的充要条件是 $\boldsymbol{b}_2'(\sigma^2\boldsymbol{D}_4 + \boldsymbol{b}_3\boldsymbol{b}_3')\boldsymbol{b}_2 \leqslant 1$;

(3) $\hat{\boldsymbol{\beta}}_{ORME}(k)$ 优于 $\hat{\boldsymbol{\beta}}_{OMRE}(k)$ 的充要条件是 $\boldsymbol{b}_3'(\sigma^2\boldsymbol{D}_5 + \boldsymbol{b}_1\boldsymbol{b}_1')\boldsymbol{b}_3 \leqslant 1$。

3　数值实例

为了说明理论结果的正确性,采用文献[8]中的例子,利用 MATLAB 软件进行数值计算,对随机约束岭估计与混合估计、岭估计、随机混合估计以及岭型混合估计的均方误差进行对比比较。取

$$X = \begin{pmatrix} 1.9 & 2.2 & 1.9 & 3.7 \\ 1.8 & 2.2 & 2.0 & 3.8 \\ 1.8 & 2.4 & 2.1 & 3.6 \\ 1.8 & 2.4 & 2.2 & 3.8 \\ 2.0 & 2.5 & 2.3 & 3.8 \\ 2.1 & 2.6 & 2.4 & 3.7 \\ 2.1 & 2.6 & 2.6 & 3.8 \\ 2.2 & 2.6 & 2.6 & 4.0 \\ 2.3 & 2.8 & 2.8 & 3.7 \\ 2.3 & 2.7 & 2.8 & 3.8 \end{pmatrix}, y = \begin{pmatrix} 2.3 \\ 2.2 \\ 2.2 \\ 2.3 \\ 2.4 \\ 2.5 \\ 2.6 \\ 2.6 \\ 2.7 \\ 2.7 \end{pmatrix}。$$

首先计算模型(1)中 $\boldsymbol{\beta}$ 和 σ^2 的最小二乘估计

$$\hat{\boldsymbol{\beta}}_{OLSE} = S^{-1}X'y = (0.6455, 0.0896, 0.1436, 0.1526)', \hat{\sigma}^2_{OLSE} = \frac{\| y - X\hat{\boldsymbol{\beta}}_{OLSE} \|^2}{n - r} = 0.0015。$$

这里 $n = 10, r = rk(X) = 4$ 。

考虑下述随机约束:

$$r = R\beta + e, \boldsymbol{R} = (1, -2, -2, -2), e \sim N(0, \hat{\sigma}^2_{OLSE})。$$

由于β和σ^2是未知的,故在计算随机约束岭估计、混合估计、岭估计、随机混合估计以及岭型混合估计的均方误差时均采用估计的方法,即将β和σ^2用上述对应的最小二乘估计来代替。

由表1可见,无论k何取值,在均方误差矩阵意义下,随机约束岭估计和岭型混合估计均优于混合估计、岭估计和随机混合估计,而随机约束岭估计只比岭型混合估计略微大一些。

表1 各估计的均方误差值

K	*OME*	*RE*	*OMRE*	*ORME*	*SRRE*
0.000 1	0.044 2	0.078 4	0.043 7	0.043 7	0.043 7
0.001 1	0.044 2	0.074 9	0.041 8	0.041 5	0.041 5
0.002 1	0.044 2	0.071 9	0.040 2	0.039 5	0.039 5
0.003 1	0.044 2	0.069 3	0.039 0	0.037 6	0.037 6
0.004 1	0.044 2	0.067 1	0.038 0	0.035 8	0.035 8
0.01	0.044 2	0.060 1	0.036 7	0.028 3	0.028 3
0.05	0.044 2	0.085 8	0.078 2	0.009 4	0.009 6
0.1	0.044 2	0.124 7	0.121 5	0.005 1	0.005 5
0.5	0.044 2	0.208 8	0.208 6	0.006 9	0.007 2

4 结论

对带随机约束的线性回归模型中未知参数 提出了一种新的随机约束岭估计并研究了均方误差矩阵下的优良性,并通过实例验证了理论结果的合理性。理论与数值实验表明本文所提出的新估计在均方误差性能方面是较好的,有着很好的应用价值。

参 考 文 献

[1] Kennedy P. A guide to Econometrics[M]. Cambridge:Massachusetts, 1998.

[2] Rao C R, Toutenburg H. Linear models: least squares and alternative[M]. New York:Springer,1995.

[3] Theil H, Goldberger A S. On pure and mixed statistical estimation in Economics[J]. International Economics Review, 1961, 2: 65-78.

[4] Hoerl A E, Kennard R W. Ridge regression: biased estimation for non-orthogonal problems[J]. Technometrics, 1970, 12: 55-67.

[5] Li Yalian, Hu Yang. A new stochastic mixed ridge estimator in linear regression model[J]. Statistical Papers, 2010, 51(2): 315-323.

[6] 富月. 有随机约束的混合回归模型,随机约束岭型估计[J]. 锦州师范学院学报:自然科学版, 2003, 24(4): 39-41.

[7] Hu Yang,Xu Jianwen. An alternative stochastic restricted Liu estimator in linear regression[J]. Statistical Papers, 2009, 50(3): 639-647.

[8] Gruber MHJ. Improving efficiency by Shrinkage: the James-Stein and ridge regression estimators[M]. New York:Marcel Dekker, Inc. ,1998.

A New Random Restricted Ridge Estimator in Linear Regression Model

Si Tongwen

(College of Arts & Sciences, University of Washington,Seattle, U. S. A 98195-3765)

Abstract: We are concerned with the parameter estimation in linear regression model with random prior information. A new random restricted ridge estimator is introduced and their properties are discussed. The necessary and sufficient conditions that the new estimator is better than the mixed estimator, the ordinary ridge estimator, the stochastic mixed estimator (Yalian Li,2008) and the another random mixed estimator (Yue Fu, 2003) in terms of the mean squared error matrix are obtained. The results of theorems and numerical examples show that the performance of the new estimator is perfect in the terms of the mean squared error matrix.

Key words: ridge estimator; mixed estimator; random restricted ridge estimator; stochastic mixed estimator; mean squared error matrix

(责任编辑 刘宪福)