线性模型中回归系数和误差方差同时的经验Bayes估计及其优良性 *

陈　玲[1,2]　　韦来生[1]

([1]中国科学技术大学统计与金融系, 合肥, 230026; [2]安徽大学数学科学学院, 合肥, 230039)

摘　要

在线性模型中, 当先验分布中超参数部分未知时, 构造了回归系数和误差方差的同时参数型经验Bayes估计(PEBE). 在均方误差矩阵(MSEM)准则下, 讨论了回归系数的PEBE相对于最小二乘估计(LSE)的优良性; 在均方误差(MSE)准则下讨论了误差方差的PEBE相对于其LSE的优良性. 当先验分布中超参数全部未知时, 重新构造了回归系数和误差方差的同时PEBE, 并给出了它们在MSE准则下相对LSE优良性的模拟结果.

关键词: 线性模型, 参数型经验Bayes估计, 最小二乘估计, 均方误差(矩阵)准则, 模拟结果.

学科分类号: O212.1.

§1. 引　言

设有如下正态线性回归模型

$$Y_{m\times 1} = X_{m\times p}\beta_{p\times 1} + e_{m\times 1}, \qquad e \sim N_m(0, \sigma^2 I), \tag{1.1}$$

其中$R(X) = p < m$, β和$\sigma^2 > 0$皆为未知参数.

回归系数和误差方差的最小二乘估计(LSE)分别为

$$\widehat{\beta} = (X'X)^{-1}X'Y, \tag{1.2}$$

$$\widehat{\sigma}^2 = \|Y - X\widehat{\beta}\|^2/(m-p). \tag{1.3}$$

众所周知, 上式给出的β和σ^2的LSE具有一些良好的性质: 如$\widehat{\beta} \sim N_p(\beta, \sigma^2(X'X)^{-1})$, $(m-p)\widehat{\sigma}^2/\sigma^2 \sim \chi^2_{m-p}$, $\widehat{\beta}$和$\widehat{\sigma}^2$相互独立, 它们还分别为β和σ^2的一致最小方差无偏估计(UMVUE). 这些性质奠定了LSE的重要地位. 但在处理包含有大量自变量的回归问题时, 设计阵X常常接近病态(即自变量之间存在线性关系), 其性质较差. 近几十年来, 众多统计学者致力于改进LSE, 提出了一些新的估计, 其中最重要的一类是有偏估计, 如岭估计、广义岭估计、主成份估计、特征根估计和Stein压缩估计等.

*国家自然科学基金(11071232, 11171001)、安徽大学青年科研基金(2010KJQN1002)和安徽大学博士科研启动经费(023033190168)资助.

本文2011年3月11日收到, 2012年7月24日收到修改稿.

改进LS估计的另一种方法是Bayes方法. 线性模型中参数的Bayes估计问题在文献中已有一些研究了. Box和Tiao (1973)假定模型(1.1)中β和σ^2皆服从无信息先验分布时, 导出了β和σ^2的Bayes估计; 王松桂(1987)在假定β和σ^2服从正态–逆Gamma先验时导出了β的Bayes估计; Broemling (1985)假定$\tau=1/\sigma^2$, 在β和τ服从正态–Gamma先验时, 导出了β和τ的Bayes估计. Wei和Zhang (2007)讨论了回归系数β的线性Bayes估计及其优良性问题. 本文将在β和σ^2服从正态–逆Gamma先验时, 同时导出β和σ^2的Bayes估计, 在此基础上利用历史样本构造β和σ^2的经验Bayes估计, 并研究其小样本性质.

假定参数β和σ^2的先验分布为正态–逆Gamma分布, 即$\beta|\sigma^2\sim N_p(\mu,k^{-1}\sigma^2I_p)$, $\sigma^2\sim\Gamma^{-1}(r/2,\lambda/2)$, 其中$k^{-1}$为$\beta$的先验方差与模型(1.1)中样本方差之比, 根据经验或专业知识, 可以假定k^{-1}已知. 由此可知, (β,σ^2)的联合先验密度为

$$\begin{aligned}\pi(\beta,\sigma^2)&=p_1(\beta|\sigma^2)p_2(\sigma^2)\\&=M_1\cdot(\sigma^2)^{-[(p+r)/2+1]}\exp\Big\{-\frac{1}{2\sigma^2}[k(\beta-\mu)'(\beta-\mu)+\lambda]\Big\},\end{aligned}\tag{1.4}$$

其中$r>0$为已知的超参数, $\mu_{p\times1}$和$\lambda>0$为未知的超参数,

$$M_1=\Big(\frac{k}{2\pi}\Big)^{p/2}\cdot\frac{(\lambda/2)^{r/2}}{\Gamma(r/2)}.$$

模型(1.1)中给定$\theta=(\beta,\sigma^2)$时Y的条件密度为

$$\begin{aligned}f(y|\theta)&=(2\pi\sigma^2)^{-m/2}\exp\Big\{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)]\Big\}\\&=(2\pi\sigma^2)^{-m/2}\exp\Big\{-\frac{1}{2\sigma^2}[(m-p)\widehat{\sigma}^2+(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)]\Big\}.\end{aligned}\tag{1.5}$$

由(1.4)和(1.5)可知, $\theta=(\beta,\sigma^2)$的联合后验密度为

$$\begin{aligned}\pi(\theta|y)&=\pi(\beta,\sigma^2|y)\propto f(y|\beta,\sigma^2)\pi(\beta,\sigma^2)\\&\propto(\sigma^2)^{-[(m+p+r)/2+1]}\exp\Big\{-\frac{1}{2\sigma^2}[(m-p)\widehat{\sigma}^2+\lambda+H]\Big\}\\&\propto(\sigma^2)^{-[(m+p+r)/2+1]}\exp\Big\{-\frac{1}{2\sigma^2}[c_0+(\beta-\widetilde{\beta})'\widetilde{\Sigma}^{-1}(\beta-\widetilde{\beta})]\Big\}.\end{aligned}\tag{1.6}$$

此处符号“$\propto$”表示“正比于”, 而

$$\begin{aligned}H&=(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)+k(\beta-\mu)'(\beta-\mu)\\&=(\beta-\widetilde{\beta})'\widetilde{\Sigma}^{-1}(\beta-\widetilde{\beta})+\widehat{\beta}'X'X\widehat{\beta}+k\mu'\mu-\widetilde{\beta}'\widetilde{\Sigma}^{-1}\widetilde{\beta}\\&=(\beta-\widetilde{\beta})'\widetilde{\Sigma}^{-1}(\beta-\widetilde{\beta})+(\widehat{\beta}-\mu)'A^{-1}(\widehat{\beta}-\mu),\end{aligned}\tag{1.7}$$

$$c_0=(m-p)\widehat{\sigma}^2+(\widehat{\beta}-\mu)'A^{-1}(\widehat{\beta}-\mu)+\lambda,\tag{1.8}$$

$$\widetilde{\beta}=(X'X+kI_p)^{-1}(X'X\widehat{\beta}+k\mu)=\widetilde{\Sigma}(X'X\widehat{\beta}+k\mu),\tag{1.9}$$

$$\widetilde{\Sigma}=(X'X+kI_p)^{-1},\qquad A^{-1}=k\widetilde{\Sigma}X'X=[(X'X)^{-1}+k^{-1}I_p]^{-1}.\tag{1.10}$$

将(1.6)添加正则化常数, 得到β和σ^2的联合后验密度

$$\pi(\beta,\sigma^2|y)=M_2\cdot(\sigma^2)^{-[(m+p+r)/2+1]}\exp\Big\{-\frac{1}{2\sigma^2}[c_0+(\beta-\widetilde{\beta})'\widetilde{\Sigma}^{-1}(\beta-\widetilde{\beta})]\Big\},\tag{1.11}$$

其中

$$M_2=(2\pi)^{-p/2}|\widetilde{\Sigma}|^{-1/2}\cdot\Big(\frac{c_0}{2}\Big)^{(m+r)/2}\Big/\Gamma\Big(\frac{m+r}{2}\Big).$$

将联合后验密度(1.11)对σ^2积分, 得到β的边缘后验密度

$$\pi_1(\beta|y)=M_3\cdot\Big|\frac{c_0\widetilde{\Sigma}}{m+r}\Big|^{-1/2}\Big[1+\frac{1}{m+r}(\beta-\widetilde{\beta})'\Big(\frac{c_0\widetilde{\Sigma}}{m+r}\Big)^{-1}(\beta-\widetilde{\beta})\Big]^{-[p+(m+r)]/2},\tag{1.12}$$

其中

$$M_3=\Gamma\Big(\frac{p+m+r}{2}\Big)\Big/\Big\{[\pi(m+r)]^{p/2}\Gamma\Big(\frac{m+r}{2}\Big)\Big\}.$$

易见β的边缘后验分布是均值为$\widetilde{\beta}$、相关阵为$c_0\widetilde{\Sigma}/(m+r)$而自由度为$m+r$的$p$元$t$分布, 记为$Mt_p(\widetilde{\beta},c_0\widetilde{\Sigma}/(m+r),m+r)$.

将联合后验密度(1.11)对β积分, 得到σ^2的边缘后验密度

$$\pi_2(\sigma^2|y)=\frac{(c_0/2)^{(m+r)/2}}{\Gamma\Big(\dfrac{m+r}{2}\Big)}(\sigma^2)^{-[(m+r)/2+1]}\exp\Big\{-\frac{c_0}{2\sigma^2}\Big\},\tag{1.13}$$

故σ^2的边缘后验分布是逆Gamma分布$\Gamma^{-1}((m+r)/2,c_0/2)$.

由多元t分布的性质(见文献[5])和(1.12)可知, 在二次损失$L(d,\beta)=(d-\beta)'(d-\beta)$下, β的Bayes估计为后验均值, 即

$$\widehat{\beta}_{\rm BE}=\mathsf{E}(\beta|y)=\widetilde{\beta}=\widetilde{\Sigma}(X'X\widehat{\beta}+k\mu)=\widehat{\beta}-k\widetilde{\Sigma}(\widehat{\beta}-\mu).\tag{1.14}$$

由(1.13)可知, 在加权平方损失$L(d,\sigma^2)=(d-\sigma^2)^2/\sigma^4$下, σ^2的Bayes估计为

$$\begin{aligned}\widehat{\sigma}^2_{\rm BE}&=\frac{\mathsf{E}(\sigma^{-2}|Y)}{\mathsf{E}(\sigma^{-4}|Y)}=\frac{(m+r)/c_0}{(m+r)(m+r+2)/c_0^2}=\frac{c_0}{m+r+2}\\&=\frac{1}{m+r+2}[(m-p)\widehat{\sigma}^2+(\widehat{\beta}-\mu)'A^{-1}(\widehat{\beta}-\mu)+\lambda].\end{aligned}\tag{1.15}$$

对刻度参数σ^2选用上述加权平方损失函数的原因是它具有不变性, 它不会随度量单位的变化而改变.

由于超参数μ,λ未知, 故估计量(1.14)和(1.15)是无实用价值的. 这就导致引入参数型经验Bayes (PEB)方法, 利用历史样本对这些未知超参数作出估计, 从而获得参数β,σ^2的Bayes两步估计, 文献上称为参数型经验Bayes估计(PEBE).

本文第二节将在先验分布中部分超参数未知时, 构造参数β和σ^2同时的PEBE. 第三节将研究回归系数β的PEBE在均方误差矩阵(MSEM)准则下相对于LSE的优良性, 第四节将研究回归模型中误差方差的PEBE在均方误差(MSE)准则下相对于LSE的优良性. 第五节将在先验分布中的超参数全部未知时, 重新构造回归系数与误差方差同时的PEBE. 由于很难获得PEBE优良性的理论结果, 故用模拟的方法讨论了PEBE相对于LSE的优良性.

§2. 回归参数β和σ^2的PEBE的构造

经验Bayes方法最初是由Robbins (1955)提出的. 按Morris (1983)的分法, 经验Bayes (EB)方法可分为参数型经验Bayes (PEB)和非参数经验Bayes (NPEB)两类.

当先验分布中含有未知的超参数, 此时的Bayes估计因含有这些未知的超参数而无法使用. 我们将利用历史样本对这些未知的超参数作出估计, 从而得到Bayes两步估计, 即参数型经验Bayes估计(PEBE). 对PEBE, 通常研究它在MSE(或MSEM)准则下的优良性、可容许性和稳健性等小样本性质(见Ghosh等(1989); Wei和Trenkler (1995); 和Zhang等(2005)等).

对NPEB估计, 常常假定先验分布为未知, 但先验分布的某些矩存在. 此时参数的Bayes估计常可表为概率密度函数及其偏导数的函数, 因此, NPEB估计一般是基于历史样本, 通过利用非参数方法对概率密度及其偏导数作出估计而得到. NPEB估计一般具有良好的大样本性质(见Singh (1985); Wei和Zhang (1995); 和Wei (1998)).

本文将采用PEB方法构造有关参数的估计量, 并讨论其优良性.

在EB方法的结构中, 设$\{Y^{(i)},\beta^{(i)},\sigma_i^2\}$, $i=1,\ldots,n+1$为相互独立的随机向量对, 此处$\{Y^{(n+1)},\beta^{(n+1)},\sigma_{n+1}^2\}=\{Y,\beta,\sigma^2\}$. 向量$Y^{(i)},\beta^{(i)},\sigma_i^2$与(1.1)中的$Y,\beta,\sigma^2$相同; 假定$\theta_i=(\beta^{(i)},\sigma_i^2)$, $i=1,\ldots,n$和$\theta=(\beta,\sigma^2)$是相互独立不可观察的随机向量, 且$\beta^{(i)}|\sigma_i^2\sim N_p(\mu,k^{-1}\sigma_i^2I)$, $\sigma_i^2\sim\Gamma^{-1}(r/2,\lambda/2)$; 假定$Y^{(1)},\ldots,Y^{(n)}$和$Y^{(n+1)}=Y$是相互独立的可观察的样本, 它们具有共同的边缘分布. 通常称$Y^{(1)},\ldots,Y^{(n)}$为历史样本, 称Y为当前样本. 当$i=1,\ldots,n+1$时, 令

$$\widehat{\beta}^{(i)}=(X'X)^{-1}X'Y^{(i)},\qquad \widehat{\sigma}_i^2=\|Y^{(i)}-X\widehat{\beta}^{(i)}\|^2/(m-p),\tag{2.1}$$

其中$\widehat{\beta}^{(n+1)}=\widehat{\beta}$, $\widehat{\sigma}_{n+1}^2=\widehat{\sigma}^2$, 如(1.2)和(1.3)所示. 由$Y^{(1)},\ldots,Y^{(n)}$和$Y=Y^{(n+1)}$是相互独立的样本, 可知$\widehat{\beta}^{(i)}$, $i=1,\ldots,n+1$相互独立. 同理, $\widehat{\sigma}_i^2$, $i=1,\ldots,n+1$也相互独立. 由于

$$\mathsf{E}(\widehat{\beta}^{(i)})=\mathsf{E}(\widehat{\beta})=\mathsf{E}[\mathsf{E}(\widehat{\beta}|\theta)]=\mathsf{E}(\beta)=\mu,\qquad i=1,\ldots,n,\tag{2.2}$$

故定义

$$\widehat{\mu}_n=\frac{1}{n}\sum_{i=1}^n\widehat{\beta}^{(i)}\tag{2.3}$$

为μ的估计量.

同样由于$\mathsf{E}(\widehat{\sigma}_i^2)=\mathsf{E}(\widehat{\sigma}^2)=\mathsf{E}[\mathsf{E}(\widehat{\sigma}^2|\theta)]=\mathsf{E}(\sigma^2)=\lambda/(r-2)$, 因此定义

$$\widehat{\lambda}_n=\frac{r-2}{n}\sum_{i=1}^n\widehat{\sigma}_i^2=(r-2)\overline{\sigma}_n^2\tag{2.4}$$

为λ的估计量, 此处$\overline{\sigma}_n^2=n^{-1}\sum\limits_{i=1}^n\widehat{\sigma}_i^2$.

将(1.14)和(1.15)中的未知的超参数μ和λ分别用它们的估计量(2.3)和(2.4)代替, 得到β和σ^2的PEBE 分别为

$$\widehat{\beta}_{\text{EB}} = \widehat{\beta} - (k^{-1}X'X + I)^{-1}(\widehat{\beta} - \widehat{\mu}_n) \triangleq \widehat{\beta} - \Sigma^{-1}(\widehat{\beta} - \widehat{\mu}_n), \tag{2.5}$$

$$\widehat{\sigma}^2_{\text{EB}} = \frac{(m-p)\widehat{\sigma}^2 + (\widehat{\beta} - \widehat{\mu}_n)'A^{-1}(\widehat{\beta} - \widehat{\mu}_n) + \widehat{\lambda}_n}{m+r+2}, \tag{2.6}$$

其中$\Sigma^{-1} = (k^{-1}X'X + I_p)^{-1}$.

设

$$\widetilde{\theta} = (\theta_1, \ldots, \theta_n, \theta), \qquad \widetilde{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_n^2, \sigma^2), \tag{2.7}$$

其中$\theta_i = (\beta^{(i)}, \sigma_i^2)$, $i = 1, \ldots, n$; $\theta = (\beta, \sigma^2)$.

本文以下的计算中E_*表示关于$Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}$和$(Y, \theta)$的联合边缘分布计算的.

§3.　回归系数β的PEBE在MSEM准则下相对于LSE的优良性

为讨论$\widehat{\beta}_{\text{EB}}$在MSEM准则下相对于$\widehat{\beta}$的优良性, 首先给出这一准则的定义.

定义 3.1　设$\widehat{\eta}$为参数向量η的估计量, 称$M(\widehat{\eta}) = \mathsf{E}[(\widehat{\eta} - \eta)(\widehat{\eta} - \eta)']$为$\widehat{\eta}$的均方误差矩阵(MSEM), 称$\text{MSE}(\widehat{\eta}) = \mathsf{E}[(\widehat{\eta} - \eta)'(\widehat{\eta} - \eta)]$为$\widehat{\eta}$的均方误差(MSE).

设$\widehat{\eta}_1$和$\widehat{\eta}_2$为参数向量η的两个不同的估计量, 若$M(\widehat{\eta}_2) - M(\widehat{\eta}_1) \geq 0$ (或$\text{MSE}(\widehat{\eta}_2) - \text{MSE}(\widehat{\eta}_1) > 0$), 则称$\widehat{\eta}_1$在MSEM(或MSE)准则下优于$\widehat{\eta}_2$.

以下我们讨论在MSEM准则下β的PEBE相对于LSE的优良性.

定理 3.1　设β的LSE和PEBE分别由(1.2)和(2.5)给出, 则对任意固定的$n \geq 1$有

$$M(\widehat{\beta}) - M(\widehat{\beta}_{\text{EB}}) \geq 0,$$

即在MSEM准则下, $\widehat{\beta}_{\text{EB}}$优于$\widehat{\beta}$. 特别, 当$n > 1$时有$M(\widehat{\beta}) - M(\widehat{\beta}_{\text{EB}}) > 0$.

证明:　由MSEM的定义和(2.5)可知

$$\begin{aligned} M(\widehat{\beta}_{\text{EB}}) &= \mathsf{E}_*[(\widehat{\beta} - \beta) - \Sigma^{-1}(\widehat{\beta} - \widehat{\mu}_n)][(\widehat{\beta} - \beta) - \Sigma^{-1}(\widehat{\beta} - \widehat{\mu}_n)]' \\ &= M(\widehat{\beta}) + \Sigma^{-1}\mathsf{E}_*[(\widehat{\beta} - \widehat{\mu}_n)(\widehat{\beta} - \widehat{\mu}_n)']\Sigma^{-1} - \Sigma^{-1}\mathsf{E}_*[(\widehat{\beta} - \widehat{\mu}_n)(\widehat{\beta} - \beta)'] \\ &\quad - \mathsf{E}_*[(\widehat{\beta} - \beta)(\widehat{\beta} - \widehat{\mu}_n)']\Sigma^{-1} \\ &= M(\widehat{\beta}) + \Sigma^{-1}J_1\Sigma^{-1} - \Sigma^{-1}J_2 - J_2'\Sigma^{-1}, \end{aligned} \tag{3.1}$$

由于对$i = 1, \ldots, n$, 有$\mathsf{Cov}\,[\mathsf{E}(\beta^{(i)}|\sigma_i^2)] = \mathsf{Cov}\,(\mu) = 0$, 因此

$$\begin{aligned} \mathsf{Cov}\,(\widehat{\beta}^{(i)}) &= \mathsf{E}[\mathsf{Cov}\,(\widehat{\beta}^{(i)}|\theta_i)] + \mathsf{Cov}\,[\mathsf{E}(\widehat{\beta}^{(i)}|\theta_i)] = \mathsf{E}[\sigma_i^2(X'X)^{-1}] + \mathsf{Cov}\,(\beta^{(i)}) \\ &= \mathsf{E}(\sigma_i^2)\cdot(X'X)^{-1} + \mathsf{E}[\mathsf{Cov}\,(\beta^{(i)}|\sigma_i^2)] + \mathsf{Cov}\,[\mathsf{E}(\beta^{(i)}|\sigma_i^2)] \\ &= \mathsf{E}(\sigma_i^2)\cdot(X'X)^{-1} + \mathsf{E}(\sigma_i^2)\cdot k^{-1}I = \frac{\lambda}{r-2}\cdot[(X'X)^{-1} + k^{-1}I]. \end{aligned} \tag{3.2}$$

根据(2.2)可知, $\mathsf{E}\widehat{\beta}=\mathsf{E}\widehat{\mu}_n=\mu$, 且$\widehat{\beta}$与$\widehat{\mu}_n$边缘独立, 故有

$$\begin{aligned} J_1 &= \mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)(\widehat{\beta}-\widehat{\mu}_n)']=\mathsf{Cov}\,(\widehat{\beta}-\widehat{\mu}_n) \\ &= \mathsf{Cov}\,(\widehat{\beta})+\mathsf{Cov}\,(\widehat{\mu}_n)=\frac{n+1}{n}\cdot\frac{\lambda}{r-2}[(X'X)^{-1}+k^{-1}I]. \end{aligned} \tag{3.3}$$

下面计算J_2.

$$\begin{aligned} J_2 &= \mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)(\widehat{\beta}-\beta)']=\mathsf{E}_*\{(\widehat{\beta}-\widehat{\mu}_n)[(\widehat{\beta}-\widehat{\mu}_n)+(\widehat{\mu}_n-\beta)]'\} \\ &= \mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)(\widehat{\beta}-\widehat{\mu}_n)']+\mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)(\widehat{\mu}_n-\beta)']=J_{21}+J_{22}, \end{aligned} \tag{3.4}$$

其中

$$J_{21} = J_1=\frac{n+1}{n}\cdot\frac{\lambda}{r-2}[(X'X)^{-1}+k^{-1}I], \tag{3.5}$$

$$\begin{aligned} J_{22} &= \mathsf{E}_*[(\widehat{\mu}_n-\beta)(\widehat{\beta}-\widehat{\mu}_n)']=\mathsf{E}_{\widetilde{\theta}}\big\{\mathsf{E}[(\widehat{\mu}_n-\beta)(\widehat{\beta}-\widehat{\mu}_n)'|\widetilde{\theta}]\big\} \\ &= \mathsf{E}_{\widetilde{\theta}}\big\{\mathsf{E}_{Y^{(1)},\ldots,Y^{(n)}}[\mathsf{E}((\widehat{\mu}_n-\beta)(\widehat{\beta}-\widehat{\mu}_n)'|\widetilde{\theta},Y^{(1)},\ldots,Y^{(n)})]\big\} \\ &= \mathsf{E}_{\widetilde{\theta}}\big\{\mathsf{E}_{Y^{(1)},\ldots,Y^{(n)}}[((\widehat{\mu}_n-\beta)(\beta-\widehat{\mu}_n)'|\widetilde{\theta})]\big\} \\ &= -\mathsf{E}_{\widetilde{\theta}}\big\{\mathsf{E}_{Y^{(1)},\ldots,Y^{(n)}}[((\widehat{\mu}_n-\beta)(\widehat{\mu}_n-\beta)'|\widetilde{\theta})]\big\}=-[\mathsf{Cov}\,(\widehat{\mu}_n)+\mathsf{Cov}\,(\beta)] \\ &= -\frac{1}{n}\cdot\frac{\lambda}{r-2}[(X'X)^{-1}+k^{-1}I]-k^{-1}\frac{\lambda}{r-2}I. \end{aligned} \tag{3.6}$$

将(3.5)和(3.6)代入(3.4)得

$$\begin{aligned} J_2 &= \frac{(n+1)\lambda}{n(r-2)}[(X'X)^{-1}+k^{-1}I]-\frac{\lambda}{n(r-2)}[(X'X)^{-1}+k^{-1}I]-\frac{k^{-1}\lambda}{r-2}I \\ &= \frac{\lambda}{r-2}(X'X)^{-1}. \end{aligned} \tag{3.7}$$

将(3.3)和(3.7)代入(3.1)得

$$\begin{aligned} M(\widehat{\beta})-M(\widehat{\beta}_{\mathrm{EB}}) &= \Sigma^{-1}J_2+J_2'\Sigma^{-1}-\Sigma^{-1}J_1\Sigma^{-1}=\Sigma^{-1}(J_2\Sigma+\Sigma J_2'-J_1)\Sigma^{-1} \\ &= \Sigma^{-1}\Big\{\frac{\lambda}{r-2}\Big(1-\frac{1}{n}\Big)[(X'X)^{-1}+k^{-1}I]\Big\}\Sigma^{-1} \\ &= \frac{(n-1)\lambda}{n(r-2)}[k^{-1}X'X+I]^{-1}\cdot[(X'X)^{-1}+k^{-1}I]\cdot[k^{-1}X'X+I]^{-1} \\ &\geq 0. \end{aligned}$$

特别当$n>1$时有$M(\widehat{\beta})-M(\widehat{\beta}_{\mathrm{EB}})>0$. □

由$\mathrm{MSE}(\widehat{\beta})-\mathrm{MSE}(\widehat{\beta}_{\mathrm{EB}})=\mathrm{tr}\{M(\widehat{\beta})-M(\widehat{\beta}_{\mathrm{EB}})\}$和定理3.1立得下列推论:

推论 3.1 在定理3.1的条件下, 有$\mathrm{MSE}(\widehat{\beta})-\mathrm{MSE}(\widehat{\beta}_{\mathrm{EB}})>0$. 即在MSE准则下$\widehat{\beta}_{\mathrm{EB}}$优于$\widehat{\beta}$.

§4.　误差方差的PEBE在MSE准则下相对于LSE的优良性

为了讨论$\widehat{\sigma}^2_{\rm EB}$的优良性, 我们需要以下引理.

引理 4.1　(1) 设X是$p\times 1$的随机向量, $\mathsf{E}(X)=\mu$, $\mathsf{Cov}\,(X)=\Sigma$, B是$p\times p$的常数阵, 则$\mathsf{E}(X'BX)=\mu'B\mu+{\rm tr}(B\Sigma)$.

(2) 设$Z\sim N_p(\nu,\Sigma)$, B是$p\times p$的常数阵, 则$\mathsf{Var}\,(Z'BZ)=4\nu'B\Sigma B\nu+2{\rm tr}(B\Sigma B\Sigma)$.

证明:　(1) 参见王松桂(1987)定理3.3.1.

(2) 由王松桂(1987)引理4.1.1作适当的变换, 可知引理结论成立.　□

引理 4.2　设$\sigma_1^2,\ldots,\sigma_n^2$, $\sigma^2=\sigma_{n+1}^2\overset{\rm i.i.d.}{\sim}\Gamma^{-1}(r/2,\lambda/2)$, 则对$i=1,\ldots,n+1$有

$$(1)\ \ \mathsf{E}(\sigma_i^2)=\mathsf{E}(\sigma^2)=\frac{\lambda}{r-2},\qquad \mathsf{E}(\sigma_i^4)=\mathsf{E}(\sigma^4)=\frac{\lambda^2}{(r-2)(r-4)},$$

$$\mathsf{Var}\,(\sigma_i^2)=\mathsf{Var}\,(\sigma^2)=\mathsf{E}(\sigma^4)-\mathsf{E}(\sigma^2)=\frac{2\lambda^2}{(r-2)^2(r-4)};\tag{4.1}$$

$$(2)\ \ \mathsf{E}\Big(\sum_{i=1}^n\sigma_i^2\Big)^2=\frac{n\lambda^2[2+n(r-4)]}{(r-2)^2(r-4)};\tag{4.2}$$

$$(3)\ \ \mathsf{E}\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^n\sigma_i^2\Big)^2=\frac{\lambda^2[n^3(r-2)+(2n^2+n)(r-4)+2]}{n^3(r-2)^2(r-4)}.\tag{4.3}$$

证明:　(1) 由$\sigma_1^2,\ldots,\sigma_n^2$, $\sigma^2=\sigma_{n+1}^2\overset{\rm i.i.d.}{\sim}\Gamma^{-1}(r/2,\lambda/2)$, 易知(4.1)成立.

(2) 由(4.1)可知

$$\begin{aligned}\mathsf{E}\Big(\sum_{i=1}^n\sigma_i^2\Big)^2&=\mathsf{E}\Big(\sum_{i=1}^n\sigma_i^4+2\sum_{i<j}\sigma_i^2\sigma_j^2\Big)=n\mathsf{E}(\sigma^4)+n(n-1)\mathsf{E}(\sigma_1^2\sigma_2^2)\\&=\frac{n\lambda^2}{(r-2)(r-4)}+n(n-1)\Big(\frac{\lambda}{r-2}\Big)^2=\frac{n\lambda^2[2+n(r-4)]}{(r-2)^2(r-4)}.\end{aligned}$$

(3) 由(4.1)和(4.2)可知

$$\begin{aligned}\mathsf{E}\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^n\sigma_i^2\Big)^2&=\mathsf{E}(\sigma^4)+\frac{2}{n^2}\mathsf{E}\Big(\sigma^2\sum_{i=1}^n\sigma_i^2\Big)+\frac{1}{n^4}\mathsf{E}\Big(\sum_{i=1}^n\sigma_i^2\Big)^2\\&=\frac{\lambda^2}{(r-2)(r-4)}+\frac{2}{n}\Big(\frac{\lambda}{r-2}\Big)^2+\frac{\lambda^2[2+n(r-4)]}{n^3(r-2)^2(r-4)}\\&=\frac{\lambda^2[n^3(r-2)+(2n^2+n)(r-4)+2]}{n^3(r-2)^2(r-4)}.\end{aligned}$$　□

引理 4.3　设$\widehat{\sigma}_i^2$, $i=1,\ldots,n$与$\widehat{\sigma}_{n+1}^2=\widehat{\sigma}^2$由(2.1)给出, $\widehat{\lambda}_n$由(2.4)给出, 则有

$$(1)\ \ \mathsf{E}_*(\widehat{\sigma}^2)=\mathsf{E}_*(\widehat{\sigma}_i^2)=\frac{\lambda}{r-2},\qquad \mathsf{E}_*(\widehat{\sigma}^4)=\mathsf{E}_*(\widehat{\sigma}_i^4)=\frac{\lambda^2(m-p+2)(r-2)}{(m-p)(r-2)^2(r-4)};$$

$$(2)\ \ \mathsf{E}_*(\widehat{\lambda}_n)=\lambda,\qquad \mathsf{Var}\,(\widehat{\lambda}_n)=\frac{2\lambda^2(m-p+r-2)}{n(m-p)(r-4)}.$$

证明： (1) 由于给定$\theta=(\beta,\sigma^2)$时, $(m-p)\widehat{\sigma}^2/\sigma^2\sim\chi^2_{m-p}$, 因此由(4.1)可知

$$
\begin{aligned}
\mathsf{E}_*(\widehat{\sigma}_i^2) &= \mathsf{E}_*(\widehat{\sigma}^2)=\mathsf{E}[\mathsf{E}(\widehat{\sigma}^2|\theta)]=\mathsf{E}(\sigma^2)=\frac{\lambda}{r-2},\\
\mathsf{Var}\,(\widehat{\sigma}_i^2) &= \mathsf{Var}\,(\widehat{\sigma}^2)=\mathsf{E}[\mathsf{Var}\,(\widehat{\sigma}^2|\theta)]+\mathsf{Var}\,[\mathsf{E}(\widehat{\sigma}^2|\theta)]\\
&= \mathsf{E}\Big(\frac{2\sigma^4}{m-p}\Big)+\mathsf{Var}\,(\sigma^2)=\frac{2\lambda^2(m-p+r-2)}{(m-p)(r-2)^2(r-4)},
\end{aligned}
\tag{4.4}
$$

故有

$$
\mathsf{E}_*(\widehat{\sigma}_i^4)=\mathsf{E}_*(\widehat{\sigma}^4)=\mathsf{Var}\,(\widehat{\sigma}^2)+[\mathsf{E}_*(\widehat{\sigma}^2)]^2=\frac{\lambda^2(m-p+2)(r-2)}{(m-p)(r-2)^2(r-4)}.
$$

(2) 由(1)和(2.4)可得$\mathsf{E}_*(\widehat{\lambda}_n)=[(r-2)/n]\sum\limits_{i=1}^{n}\mathsf{E}_*(\widehat{\sigma}_i^2)=\lambda$; 再由(4.4)可知

$$
\mathsf{Var}\,(\widehat{\lambda}_n)=(r-2)^2\mathsf{Var}\,(\overline{\sigma}_n^2)=\frac{(r-2)^2}{n}\mathsf{Var}\,(\widehat{\sigma}_1^2)=\frac{2\lambda^2(m-p+r-2)}{n(m-p)(r-4)}. \qquad \square
$$

引理 4.4 设$\widehat{\beta}^{(i)},\widehat{\sigma}_i^2\ (i=1,\ldots,n+1)$由(2.1)定义, $\widehat{\mu}_n,\widehat{\lambda}_n$由(2.3)和(2.4)定义, $\widetilde{\sigma}^2$由(2.7)式给出, 则

(1) 给定$\widetilde{\sigma}^2$时, $\widehat{\beta}^{(i)}|\widetilde{\sigma}^2$的边缘条件分布为$N_p(\mu,\sigma_i^2A)$, 它们相互独立, 且$\widehat{\beta}^{(i)}|\widetilde{\sigma}^2$与$\widehat{\sigma}_i^2|\widetilde{\sigma}^2$条件独立;

(2) $\mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)]=p\Big(1+\dfrac{1}{n}\Big)\dfrac{\lambda}{r-2}$;

(3) $\mathsf{Var}\,[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)]=\dfrac{\lambda^2}{n^3(r-2)^2(r-4)}\{n^3a_1+n^2b_1+nc_1+d_1\}$,

其中A^{-1}由(1.10)给出, $a_1=2p^2+2p(r-2)$, $b_1=4p(r-4)$, $c_1=2p(r-4)$, $d_1=2p(p+2)$.

证明： (1) 注意到$\theta_i=(\beta^{(i)},\sigma_i^2)$, 由$\widehat{\beta}^{(i)}|\theta_i\sim N_p(\beta^{(i)},\sigma_i^2(X'X)^{-1})$和先验分布$\beta^{(i)}|\sigma_i^2\sim N_p(\mu,(\sigma_i^2/k)I_p)$可知, $\widehat{\beta}^{(i)}|\sigma_i^2$的边缘条件密度为

$$
\begin{aligned}
p(\widehat{\beta}^{(i)}|\sigma_i^2) &= \int p(\widehat{\beta}^{(i)}|\beta^{(i)},\sigma_i^2)p_1(\beta^{(i)}|\sigma_i^2)\mathrm{d}\beta^{(i)}\propto\int(\sigma_i^2)^{-p}\exp\Big\{-\frac{H_i}{2\sigma_i^2}\Big\}\mathrm{d}\beta^{(i)}\\
&\propto (\sigma_i^2)^{-p/2}\exp\Big\{-\frac{(\widehat{\beta}^{(i)}-\mu)'A^{-1}(\widehat{\beta}^{(i)}-\mu)}{2\sigma_i^2}\Big\},
\end{aligned}
$$

易见$\widehat{\beta}^{(i)}|\sigma_i^2\sim N_p(\mu,\sigma_i^2A)$, 此处$A^{-1}$由(1.10)给出, 而$H_i$的表达式与(1.7)中的$H$类似, 只要将其中的$\widehat{\beta}$、$\beta$和$\widetilde{\beta}$分别用$\widehat{\beta}^{(i)}$、$\beta^{(i)}$和$\widetilde{\beta}^{(i)}$代替.

注意到$\widehat{\sigma}_i^2|\theta_i$的分布与$\beta^{(i)}$无关, 且与$\widehat{\beta}^{(i)}|\theta_i$独立, 于是

$$
\begin{aligned}
f(\widehat{\beta}^{(i)},\widehat{\sigma}_i^2|\sigma_i^2) &= \int f(\widehat{\beta}^{(i)},\widehat{\sigma}_i^2|\theta_i)\pi(\beta^{(i)}|\sigma_i^2)\mathrm{d}\beta^{(i)}=\int f(\widehat{\beta}^{(i)}|\theta_i)f(\widehat{\sigma}_i^2|\sigma_i^2)\pi(\beta^{(i)}|\sigma_i^2)\mathrm{d}\beta^{(i)}\\
&= \int f(\widehat{\beta}^{(i)}|\beta^{(i)},\sigma_i^2)\pi(\beta^{(i)}|\sigma_i^2)\mathrm{d}\beta^{(i)}\cdot f(\widehat{\sigma}_i^2|\sigma_i)=f(\widehat{\beta}^{(i)}|\sigma_i^2)\cdot f(\widehat{\sigma}_i^2|\sigma_i^2).
\end{aligned}
$$

因此给定σ_i^2时, 即给定$\widetilde{\sigma}^2$时, $\widehat{\beta}^{(i)}$和$\widehat{\sigma}_i^2$是条件独立的.

(2) 由(1)可知, $\widehat{\beta}|\widetilde{\sigma}^2 \sim N_p(\mu, \sigma^2 A)$, 类似有$\widehat{\mu}_n|\widetilde{\sigma}^2 \sim N_p\Big(\mu, \Big((1/n^2)\sum\limits_{i=1}^{n}\sigma_i^2\Big)A\Big)$, 且它们相互独立, 因此$(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2 \sim N_p\Big(0, \Big(\sigma^2+(1/n^2)\sum\limits_{i=1}^{n}\sigma_i^2\Big)A\Big)$, 故由引理4.1,

$$
\begin{aligned}
\mathsf{E}[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2] &= \mathrm{tr}\Big\{A^{-1}\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big)A\Big\} \\
&= p\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big), \qquad (4.5) \\
\mathsf{Var}\,[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2] &= 2\mathrm{tr}\Big\{\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big)^2[A^{-1}A]^2\Big\} \\
&= 2p\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big)^2. \qquad (4.6)
\end{aligned}
$$

因此, 由(4.5)和(4.1)可知

$$
\begin{aligned}
\mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)] &= \mathsf{E}\{\mathsf{E}[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2]\} \\
&= p\mathsf{E}\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big) = p\Big(1+\frac{1}{n}\Big)\frac{\lambda}{r-2}.
\end{aligned}
$$

(3) 由于

$$
\begin{aligned}
\mathsf{Var}\,[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)] &= \mathsf{E}\{\mathsf{Var}\,[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2]\} \\
&\quad +\mathsf{Var}\,\{\mathsf{E}[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2]\} \\
&= J_1+J_2, \qquad (4.7)
\end{aligned}
$$

由(4.6)和(4.3)可知,

$$
\begin{aligned}
J_1 &= \mathsf{E}\Big\{2p\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big)^2\Big\} \\
&= \frac{2p\lambda^2}{n^3(r-2)^2(r-4)}[n^3(r-2)+(2n^2+n)(r-4)+2], \qquad (4.8)
\end{aligned}
$$

由(4.5)和(4.1)可知

$$
J_2 = \mathsf{Var}\,\Big\{p\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\Big)\Big\} = \frac{2p^2\lambda^2}{(r-2)^2(r-4)}\Big(1+\frac{1}{n^3}\Big), \qquad (4.9)
$$

将(4.8)和(4.9)代入(4.7), 引理4.4 (3)得证, 证毕. □

引理 4.5 设$\widehat{\sigma}^2$和$\widehat{\sigma}_{\mathrm{EB}}^2$分别由(1.3)和(2.6)给出, 则有

$$
\mathsf{E}_*(\widehat{\sigma}^2-\widehat{\sigma}_{\mathrm{EB}}^2)^2 = \frac{\lambda^2}{n^3(m+r+2)^2(r-2)^2(r-4)}\{n^3a+n^2b+nc+d\},
$$

其中

$$
\begin{aligned}
a &= 2(p+r+14)(r-2)+\frac{2(p+r+2)^2(r-2)}{m-p},\\
b &= -4p(r-4)+\frac{2(m-p+r-2)(r-2)^2}{m-p},\\
c &= (p^2+2p)(r-4)+4p(r-2), \qquad d=2p^2+4p.
\end{aligned}
\tag{4.10}
$$

证明: 由(2.6)可知

$$
\begin{aligned}
\mathsf{E}_*(\widehat{\sigma}^2-\widehat{\sigma}^2_{\rm EB})^2 &= \frac{1}{(m+r+2)^2}\mathsf{E}_*[(p+r+2)\widehat{\sigma}^2-(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)-\widehat{\lambda}_n]^2\\
&= \frac{1}{(m+r+2)^2}(Q_1+Q_2+Q_3-Q_4-Q_5+Q_6).
\end{aligned}
\tag{4.11}
$$

利用引理4.3 (1)可得

$$
Q_1=(p+r+2)^2\mathsf{E}_*(\widehat{\sigma}^4)=\frac{\lambda^2[n^3(p+r+2)^2(m-p+2)(r-2)]}{n^3(m-p)(r-2)^2(r-4)}.
\tag{4.12}
$$

由引理4.4可知

$$
\begin{aligned}
Q_2 &= \mathsf{Var}\,[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)]+[\mathsf{E}_*(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)]^2\\
&= \frac{\lambda^2(p^2+2p)}{n^3(r-2)^2(r-4)}\{n^3(r-2)+(2n^2+n)(r-4)+2\}.
\end{aligned}
\tag{4.13}
$$

由引理4.3 (2)可知

$$
\begin{aligned}
Q_3 &= \mathsf{E}_*(\widehat{\lambda}^2_n)=\mathsf{Var}\,(\widehat{\lambda}_n)+(\mathsf{E}_*\widehat{\lambda}_n)^2=\frac{2\lambda^2(m-p+r-2)}{n(m-p)(r-4)}+\lambda^2\\
&= \frac{\lambda^2[2n^2(m-p+r-2)(r-2)^2+n^3(m-p)(r-2)^2(r-4)]}{n^3(m-p)(r-2)^2(r-4)}.
\end{aligned}
\tag{4.14}
$$

注意$\widehat{\sigma}^2$与$\widehat{\lambda}_n$是边缘独立的, 由引理4.3可得

$$
\begin{aligned}
Q_4 &= 2(p+r+2)\mathsf{E}_*(\widehat{\sigma}^2\cdot\widehat{\lambda}_n)=2(p+r+2)\mathsf{E}_*(\widehat{\sigma}^2)\cdot\mathsf{E}_*(\widehat{\lambda}_n)\\
&= \frac{2(p+r+2)\lambda^2}{r-2}=\frac{\lambda^2[2n^3(p+r+2)(r-2)(r-4)]}{n^3(r-2)^2(r-4)}.
\end{aligned}
\tag{4.15}
$$

由引理4.4可知, 当$\widehat{\sigma}^2$给定时, $\widehat{\sigma}^2,\widehat{\beta}$和$\widehat{\mu}_n$是条件独立, 并结合(4.5)可得

$$
\begin{aligned}
Q_5 &= 2(p+r+2)\mathsf{E}\big\{\mathsf{E}(\widehat{\sigma}^2|\widehat{\sigma}^2)\cdot\mathsf{E}[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widehat{\sigma}^2]\big\}\\
&= 2(p+r+2)\mathsf{E}\Big\{\sigma^2\cdot p\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^n\sigma_i^2\Big)\Big\}\\
&= \frac{2(p+r+2)p\lambda^2}{(r-2)(r-4)}+\frac{2(p+r+2)p\lambda^2}{n(r-2)^2}\\
&= \frac{\lambda^2[n^3\cdot 2p(p+r+2)(r-2)+n^2\cdot 2p(p+r+2)(r-4)]}{n^3(r-2)^2(r-4)}.
\end{aligned}
\tag{4.16}
$$

由引理4.4可知, 当$\widetilde{\sigma}^2$给定时, $\widehat{\lambda}_n, \widehat{\beta}$和$\widehat{\mu}_n$也是条件独立, 由(4.5)和(4.2)可得

$$\begin{aligned} Q_6 &= 2\mathsf{E}\big\{\mathsf{E}(\widehat{\lambda}_n|\widetilde{\sigma}^2)\cdot\mathsf{E}[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2]\big\} \\ &= 2\mathsf{E}\Big\{\frac{r-2}{n}\sum_{i=1}^n\sigma_i^2\cdot p\Big(\sigma^2+\frac{1}{n^2}\sum_{i=1}^n\sigma_i^2\Big)\Big\} \\ &= \frac{2p(r-2)}{n}\Big\{\frac{n\lambda^2}{(r-2)^2}+\frac{\lambda^2[2+n(r-4)]}{n(r-2)^2(r-4)}\Big\} \\ &= \frac{\lambda^2[(n^3+n^2)\cdot 2p(r-2)(r-4)+n\cdot 4p(r-2)]}{n^3(r-2)^2(r-4)}. \end{aligned} \tag{4.17}$$

将(4.12) – (4.17)代入(4.11), 合并同类项, 获得引理4.5的结论. □

引理 4.6　设$\widehat{\sigma}^2$和$\widehat{\sigma}^2_{\rm EB}$由(1.3)和(2.6)定义, 则

$$\mathsf{E}_*[(\widehat{\sigma}^2-\widehat{\sigma}^2_{\rm EB})(\widehat{\sigma}^2-\sigma^2)] = \frac{2\lambda^2(p+r+2)(r-2)}{(m+r+2)(m-p)(r-2)^2(r-4)}.$$

证明:　由(2.6)可知,

$$\begin{aligned} &\mathsf{E}_*[(\widehat{\sigma}^2-\widehat{\sigma}^2_{\rm EB})(\widehat{\sigma}^2-\sigma^2)] \\ =& \frac{1}{m+r+2}\mathsf{E}_*\big\{[(p+r+2)\widehat{\sigma}^2-(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)-\widehat{\lambda}_n](\widehat{\sigma}^2-\sigma^2)\big\} \\ =& \frac{1}{m+r+2}\{(p+r+2)\mathsf{E}_*[\widehat{\sigma}^2(\widehat{\sigma}^2-\sigma^2)]-\mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)\cdot(\widehat{\sigma}^2-\sigma^2)] \\ &-\mathsf{E}_*[\widehat{\lambda}_n(\widehat{\sigma}^2-\sigma^2)]\} \\ =& \frac{1}{m+r+2}\{(p+r+2)J_{11}-J_{12}-J_{13}\}. \end{aligned} \tag{4.18}$$

由引理4.3和公式(4.1)可知

$$\begin{aligned} J_{11} &= \mathsf{E}_*(\widehat{\sigma}^4)-\mathsf{E}_*(\sigma^2\cdot\widehat{\sigma}^2) = \mathsf{E}_*(\widehat{\sigma}^4)-\mathsf{E}[\sigma^2\mathsf{E}(\widehat{\sigma}^2|\sigma^2)] \\ &= \mathsf{E}_*(\widehat{\sigma}^4)-\mathsf{E}(\sigma^4) = \frac{\lambda^2(m-p+2)(r-2)}{(m-p)(r-2)^2(r-4)}-\frac{\lambda^2}{(r-2)(r-4)} \\ &= \frac{2\lambda^2(r-2)}{(m-p)(r-2)^2(r-4)}. \end{aligned} \tag{4.19}$$

因为给定$\widetilde{\sigma}^2$时, $\widehat{\beta}-\widehat{\mu}_n$与$\widehat{\sigma}^2$条件独立, 并且$\mathsf{E}(\widehat{\sigma}^2-\sigma^2|\widetilde{\sigma}^2)=0$, 因此

$$\begin{aligned} J_{12} &= \mathsf{E}_*[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)\cdot(\widehat{\sigma}^2-\sigma^2)] \\ &= \mathsf{E}\big\{\mathsf{E}[(\widehat{\beta}-\widehat{\mu}_n)'A^{-1}(\widehat{\beta}-\widehat{\mu}_n)|\widetilde{\sigma}^2]\cdot\mathsf{E}[(\widehat{\sigma}^2-\sigma^2)|\widetilde{\sigma}^2]\big\}=0. \end{aligned} \tag{4.20}$$

由于给定$\widetilde{\sigma}^2$时, $\widehat{\lambda}_n$和$\widehat{\sigma}^2$条件独立, 故有

$$J_{13} = \mathsf{E}_*[\widehat{\lambda}_n(\widehat{\sigma}^2-\sigma^2)] = \mathsf{E}\big\{\mathsf{E}[\widehat{\lambda}_n|\widetilde{\sigma}^2]\cdot\mathsf{E}[(\widehat{\sigma}^2-\sigma^2)|\widetilde{\sigma}^2]\big\}=0. \tag{4.21}$$

将(4.19) – (4.21)代入(4.18), 引理4.6得证. □

定理 4.1 设$\widehat{\sigma}^2$与$\widehat{\sigma}^2_{\mathrm{EB}}$分别由(1.3)和(2.6)定义, 则当$r>\max(4,10-p)$和$2n\geq m-p\geq 2$时, $\mathrm{MSE}(\widehat{\sigma}^2)-\mathrm{MSE}(\widehat{\sigma}^2_{\mathrm{EB}})>0$.

证明: 根据定义3.1可得

$$\begin{aligned}\mathrm{MSE}(\widehat{\sigma}^2_{\mathrm{EB}}) &= \mathsf{E}_*(\widehat{\sigma}^2_{\mathrm{EB}}-\sigma^2)^2=\mathsf{E}_*[(\widehat{\sigma}^2-\sigma^2)-(\widehat{\sigma}^2-\widehat{\sigma}^2_{\mathrm{EB}})]^2\\ &= \mathrm{MSE}(\widehat{\sigma}^2)+\mathsf{E}_*(\widehat{\sigma}^2-\widehat{\sigma}^2_{\mathrm{EB}})^2-2\mathsf{E}_*[(\widehat{\sigma}^2-\widehat{\sigma}^2_{\mathrm{EB}})(\widehat{\sigma}^2-\sigma^2)].\end{aligned}$$

由引理4.5和引理4.6, 我们知道

$$\begin{aligned}\mathrm{MSE}(\widehat{\sigma}^2)-\mathrm{MSE}(\widehat{\sigma}^2_{\mathrm{EB}}) &= 2\mathsf{E}_*[(\widehat{\sigma}^2-\widehat{\sigma}^2_{\mathrm{EB}})(\widehat{\sigma}^2-\sigma^2)]-\mathsf{E}_*(\widehat{\sigma}^2-\widehat{\sigma}^2_{\mathrm{EB}})^2\\ &= \frac{\lambda^2}{(m+r+2)^2(r-2)^2(r-4)}\Big[K-a-\frac{b}{n}-\frac{c}{n^2}-\frac{d}{n^3}\Big], \quad (4.22)\end{aligned}$$

其中a,b,c和d由(4.10)定义, 并且

$$K=\frac{4(m+r+2)(p+r+2)(r-2)}{m-p}.$$

因此, 定理4.1成立, 当且仅当

$$K-a-\frac{b}{n}-\frac{c}{n^2}-\frac{d}{n^3}>0.$$

由(4.10)可得

$$\begin{aligned}K-a &= \frac{4[(m-p)+(p+r+2)](p+r+2)(r-2)}{m-p}\\ &\quad -2(p+r+14)(r-2)-\frac{2(p+r+2)^2(r-2)}{m-p}\\ &= 2(p+r-10)(r-2)+\frac{2(p+r+2)^2(r-2)}{m-p}.\end{aligned}$$

显然, 由(4.10)可知$b=-4p(r-4)+2(r-2)^2+2(r-2)^3/(m-p)$. 故由定理条件$2n\geq m-p\geq 2$和$r>\max(4,10-p)$, 可见$p+r-10>0$, $2n/(m-p)\geq 1$, 因此有

$$\begin{aligned}K-a-\frac{b}{n} &= \frac{1}{n}\Big\{\Big[2n(p+r-10)(r-2)+\frac{2n}{m-p}(p+r+2)^2(r-2)\Big]\\ &\quad -\Big[-4p(r-4)+2(r-2)^2+\frac{2(r-2)^3}{m-p}\Big]\Big\}\\ &> \frac{1}{n}\Big\{[(p+4)+(r-2)]^2(r-2)+4p(r-4)-2(r-2)^2-\frac{2(r-2)^3}{m-p}\Big\}\\ &= \frac{1}{n}\Big[(p+4)^2(r-2)+2(p+3)(r-2)^2+4p(r-4)+\frac{(m-p-2)(r-2)^3}{m-p}\Big]\\ &> \frac{1}{n}(p+4)^2(r-2).\end{aligned}$$

由条件$n \geq 1$可知

$$\begin{aligned} & K - a - \frac{b}{n} - \frac{c}{n^2} - \frac{d}{n^3} \\ > & \frac{1}{n^2}[n(p+4)^2(r-2) - (p^2+2p)(r-4) - 4p(r-2)] - \frac{d}{n^3} \\ > & \frac{1}{n^2}[p^2(r-2) + 8p(r-2) - (p^2+2p)(r-4) - 4p(r-2)] - \frac{d}{n^3} \\ = & \frac{1}{n^2}\{p^2[(r-4)+2] + 4p[(r-4)+2] - (p^2+2p)(r-4)\} - \frac{d}{n^3} \\ > & \frac{2p^2+8p}{n^2} - \frac{2p^2+4p}{n^3} > 0. \end{aligned} \tag{4.23}$$

将(4.23)代入(4.22), 定理4.1得证. □

§5. 模拟结果

当先验分布(1.4)中所有的超参数μ, k, λ, r都未知时, 我们先给出k的估计量的构造, 然后用极大似然估计方法结合矩估计方法, 得到r, λ的估计量, 并重新构造β, σ^2的PEBE $\widetilde{\beta}_{\rm EB}$和$\widetilde{\sigma}^2_{\rm EB}$, μ的估计量$\widehat{\mu}_n$仍由(2.3)给出, 然后用模拟的方法来讨论其优良性.

5.1 当μ, k, λ, r都未知时, 经验Bayes估计的构造

为了得到k的估计, 我们设

$$S_n = \frac{1}{n-1}\sum_{i=1}^{n}(\widehat{\beta}^{(i)} - \widehat{\mu}_n)(\widehat{\beta}^{(i)} - \widehat{\mu}_n)', \tag{5.1}$$

令Γ为一n阶正交阵, 其第一列为$(1/\sqrt{n}, \ldots, 1/\sqrt{n})'$, 其余元素为$a_{ij}$. 做变换$(Z_1, \ldots, Z_n) = (\widehat{\beta}^{(1)}, \ldots, \widehat{\beta}^{(n)})\Gamma$, 则

$$Z_1 = \sum_{k=1}^{n}\widehat{\beta}^{(k)}/\sqrt{n} = \sqrt{n}\widehat{\mu}_n, \qquad Z_i = \sum_{k=1}^{n} a_{ki}\widehat{\beta}^{(k)}, \quad i = 2, \ldots, n. \tag{5.2}$$

由于a_{ij}是正交阵Γ的(i,j)元, 故

$$\begin{aligned} \sum_{i=1}^{n} Z_i Z_i' &= \sum_{i=1}^{n}\Big[\Big(\sum_{k=1}^{n} a_{ki}\widehat{\beta}^{(k)}\Big)\Big(\sum_{l=1}^{n} a_{li}\widehat{\beta}^{(l)}\Big)'\Big] = \sum_{k=1}^{n}\sum_{l=1}^{n}\Big(\sum_{i=1}^{n} a_{ki}a_{li}\Big)\widehat{\beta}^{(k)}\widehat{\beta}^{(l)\prime} \\ &= \sum_{k=1}^{n}\sum_{l=1}^{n}\delta_{kl}\widehat{\beta}^{(k)}\widehat{\beta}^{(l)\prime} = \sum_{k=1}^{n}\widehat{\beta}^{(k)}\widehat{\beta}^{(k)\prime}, \end{aligned} \tag{5.3}$$

这里$\delta_{kl} = I_{(k=l)}$. 由(2.2), (5.2)和Γ的构造易知, 当$i, j = 2, \ldots, n$时,

$$\mathsf{E}Z_1 = \sqrt{n}\mu, \qquad \mathsf{E}Z_i = \sum_{k=1}^{n} a_{ki}\mu = \sqrt{n}\mu\Big(\sum_{k=1}^{n} a_{ki}/\sqrt{n}\Big) = 0,$$

由公式(2.1)和(3.2)可知, $\widehat{\beta}^{(i)}$, $i=1,\ldots,n+1$独立, 且

$$\mathsf{Cov}\,(\widehat{\beta}^{(i)})=\frac{\lambda}{r-2}A=\frac{\lambda}{r-2}[(X'X)^{-1}+k^{-1}I],$$

结合(5.2), 我们知道, 当$i,j=2,\ldots,n$时,

$$\begin{aligned}\mathsf{E}(Z_iZ_j') &= \mathsf{E}(Z_i-\mathsf{E}Z_i)(Z_j-\mathsf{E}Z_j)'=\mathsf{E}\Big[\sum_{k=1}^n a_{ki}(\widehat{\beta}^{(k)}-\mu)\cdot\sum_{l=1}^n a_{lj}(\widehat{\beta}^{(l)}-\mu)'\Big]\\ &= \sum_{k=1}^n\sum_{l=1}^n a_{ki}a_{lj}\mathsf{Cov}\,(\widehat{\beta}^{(k)},\widehat{\beta}^{(l)})=\sum_{k=1}^n a_{ki}a_{kj}\mathsf{Cov}\,(\widehat{\beta}^{(k)})\\ &= \delta_{ij}\cdot\frac{\lambda}{r-2}A=\begin{cases}\dfrac{\lambda}{r-2}[(X'X)^{-1}+k^{-1}I], & i=j;\\ 0, & i\neq j,\end{cases}\end{aligned}\tag{5.4}$$

此处$\delta_{ij}=I_{(i=j)}$. 结合(5.2)–(5.3)可知,

$$\begin{aligned}(n-1)S_n &= \sum_{i=1}^n(\widehat{\beta}^{(i)}-\widehat{\mu}_n)(\widehat{\beta}^{(i)}-\widehat{\mu}_n)'=\sum_{i=1}^n\widehat{\beta}^{(i)}\widehat{\beta}^{(i)'}-n\widehat{\mu}_n\widehat{\mu}_n'\\ &= \sum_{i=1}^n Z_iZ_i'-Z_1Z_1'=\sum_{i=2}^n Z_iZ_i'.\end{aligned}$$

由$\widehat{\beta}^{(1)},\ldots,\widehat{\beta}^{(n)}$与$\widehat{\sigma}^2$的独立性可知, $\mathrm{tr}(S_n)$与$\widehat{\sigma}^2$独立, 结合(5.4)可知

$$\begin{aligned}\mathsf{E}\Big[\frac{\mathrm{tr}(S_n)}{\widehat{\sigma}^2}\Big] &= \mathsf{E}[\mathrm{tr}(S_n)]\cdot\mathsf{E}\Big(\frac{1}{\widehat{\sigma}^2}\Big)=\frac{1}{n-1}\mathrm{tr}\Big[\mathsf{E}\Big(\sum_{i=2}^n Z_iZ_i'\Big)\Big]\cdot\mathsf{E}\Big(\frac{r}{\lambda}\cdot\frac{1}{(r/\lambda)\widehat{\sigma}^2}\Big)\\ &= \mathrm{tr}[\mathsf{E}(Z_2Z_2')]\cdot\frac{r}{\lambda}\cdot\mathsf{E}\Big(\frac{1}{F_{m-p,r}}\Big)\\ &= \frac{\lambda}{r-2}\mathrm{tr}[(X'X)^{-1}+k^{-1}I]\cdot\frac{r}{\lambda}\cdot\mathsf{E}(F_{r,m-p})\\ &= \frac{r(m-p)}{(r-2)(m-p-2)}\mathrm{tr}[(X'X)^{-1}+k^{-1}I_p]\\ &= \frac{r(m-p)}{(r-2)(m-p-2)}\Big(\frac{p}{k}+q\Big).\end{aligned}$$

这里, $F_{a,b}$表示一个服从自由度为a,b的F分布的r.v., $q=\mathrm{tr}[(X'X)^{-1}]$. 因此当$r$已知时, k^{-1}的一个无偏估计定义为

$$\frac{1}{p}\Big[\frac{\mathrm{tr}(S_n)}{\widehat{\sigma}^2}\cdot\frac{(r-2)(m-p-2)}{r(m-p)}-q\Big].$$

由于$k^{-1}>0$, 故定义其非负估计为

$$\widehat{k}_n^{*-1}=\frac{1}{p}\Big[\frac{\mathrm{tr}(S_n)}{\widehat{\sigma}^2}\cdot\frac{(r-2)(m-p-2)}{r(m-p)}-q\Big]\vee 0.\tag{5.5}$$

这里, $a\vee b=\max(a,b)$.

当r未知时, 我们根据$U_i=(r/\lambda)\widehat{\sigma}_i^2\sim F(m-p,r)$, 通过R软件中的最优化方法, 最大化$U_i$的似然函数, 可得$r$的极大似然估计$\widetilde{r}_n$, 它没有显式表达. 再将$\widetilde{r}_n$代替(2.4)中的$r$, 得到$\lambda$的矩估计为

$$\widetilde{\lambda}_n=\frac{\widetilde{r}_n-2}{n}\sum_{i=1}^{n}\widehat{\sigma}_i^2.$$

模拟结果显示, 这种构造r,λ的估计量的方法精度更好. 用$\widetilde{r}_n$代替(5.5)中的r, 可得k^{-1}的估计量为

$$\widetilde{k}_n^{-1}=\frac{1}{p}\Big[\frac{\mathrm{tr}(S_n)}{\widehat{\sigma}^2}\cdot\frac{(\widetilde{r}_n-2)(m-p-2)}{\widetilde{r}_n(m-p)}-q\Big]\vee 0. \tag{5.6}$$

因此, 先验分布(1.4)中的超参数全部未知时, β,σ^2的PEBE分别为

$$\widetilde{\beta}_{\mathrm{EB}}=\widehat{\beta}-(\widetilde{k}_n^{-1}X'X+I)^{-1}(\widehat{\beta}-\widehat{\mu}_n), \tag{5.7}$$

$$\widetilde{\sigma}_{\mathrm{EB}}^2=\frac{\widetilde{c}_0}{m+\widetilde{r}_n+2}=\frac{(m-p)\widehat{\sigma}^2+(\widehat{\beta}-\widehat{\mu}_n)'\widetilde{A}_n^{-1}(\widehat{\beta}-\widehat{\mu}_n)+\widetilde{\lambda}_n}{m+\widetilde{r}_n+2}, \tag{5.8}$$

其中

$$\widetilde{A}_n=\widetilde{k}_n^{-1}I_p+(X'X)^{-1},\qquad \widetilde{c}_0=(m-p)\widehat{\sigma}^2+(\widehat{\beta}-\widehat{\mu}_n)'\widetilde{A}_n^{-1}(\widehat{\beta}-\widehat{\mu}_n)+\widetilde{\lambda}_n. \tag{5.9}$$

$\widehat{\mu}_n$仍由(2.3)给出.

5.2 回归系数和误差方差的PEBE的模拟结果

我们通过估计量(5.7), (5.8)在MSE准则下相对于LSE (1.2), (1.3)的模拟结果, 来显示其优良性.

设数据来自下列线性回归模型

$$Y=\beta_0\mathbf{1}+\beta_1X_1+\beta_2X_2+e,\qquad e|\sigma^2\sim N_6(0,\sigma^2I), \tag{5.10}$$

其中参数β,σ^2的先验分布为(1.4), k^{-1}为β的先验方差与模型(1.1)中样本方差之比.

超参数的值分别取为$\mu=(0,0,0)'$, $r=6$, $\lambda=4$, $k=10$, 设计阵为$X_{6\times3}=(\mathbf{1},X_1,X_2)$, 这里$\mathbf{1}$是分量全为1的$6\times1$的向量, $X_1=(1,1,1,0,0,1)'$, $X_2=(1,0,1,0,1,1)'$, 历史样本数为$n=10$. 有了这些值, 我们可以得到(2.1)定义的统计量. 然后, 我们计算$\widetilde{\beta}_{\mathrm{EB}},\widehat{\beta},\widetilde{\sigma}^2_{\mathrm{EB}},\widehat{\sigma}^2$以及它们的MSE值. 重复这个过程1000次, 得到四个估计量的MSE的算术平均值.

为了考察超参数r,λ,k的值和历史样本个数n的大小的变动对MSE值的影响, 我们在每次模拟时, 固定其中的三个参数, 变动一个, 得到回归系数和误差方差的PEBE (5.7)和(5.8)的MSE值, 它们分别由图1−4和图5−8给出.

从图1可见, 不管λ的值如何变化, 回归系数的PEBE的MSE值一直显著的小于LSE的MSE值, 且随着λ的增加, 两者的差别迅速增大, PEBE的优良性越来越显著.

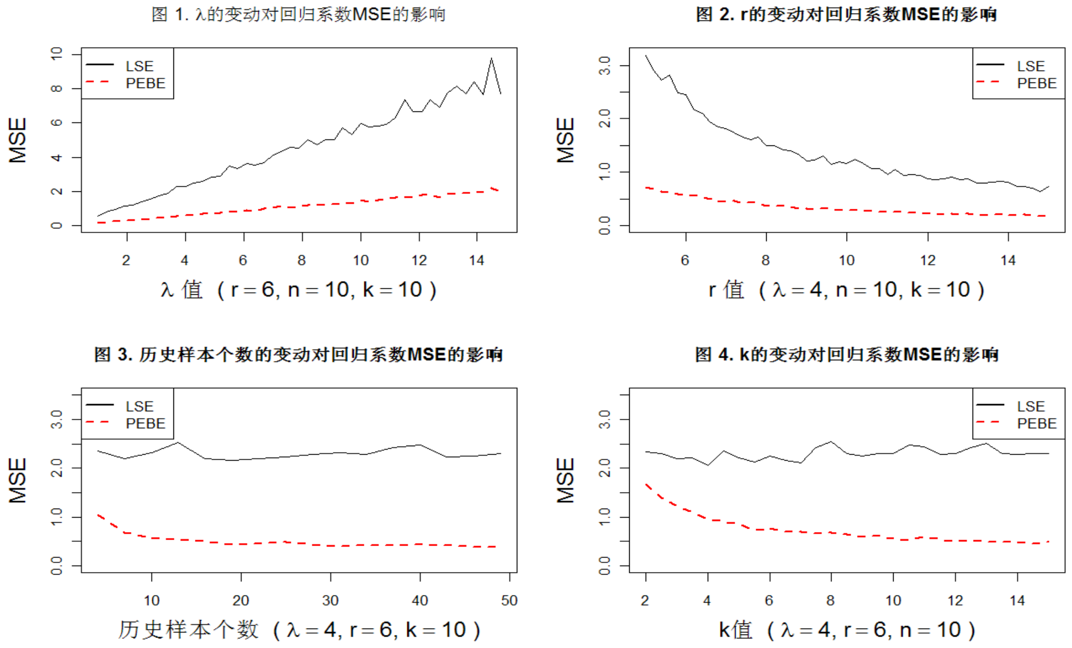

图1−4　回归系数的PEBE和LSE的MSE值

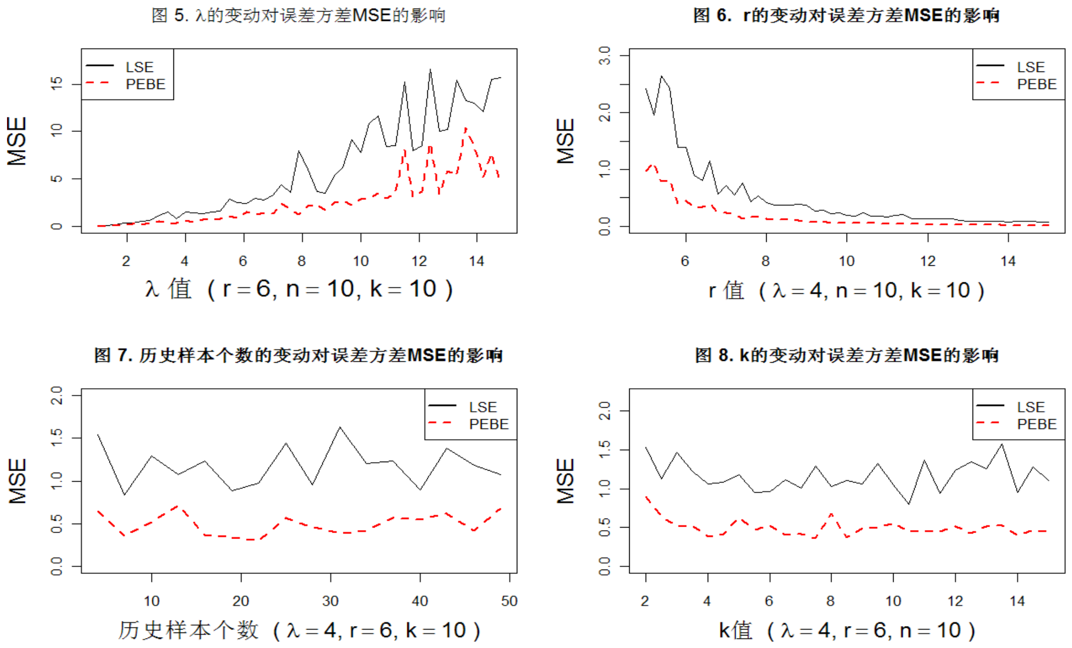

图5−8　误差方差的PEBE和LSE的MSE值

从图2我们发现, 无论如何取定r的值, PEBE的MSE值一直优于LSE的MSE值. 当r较小时, 优良性非常显著. 不过, 随着r的增大, 优良性下降比较明显.

从图3可知, 历史样本个数的变动对PEBE和LSE的MSE的值影响都很小. PEBE的MSE值一直小于LSE的MSE值, 且更稳定. 这说明历史样本个数的变动对PEBE的优良性的影响很小.

图4显示, 无论k的取值如何变化, 回归系数的PEBE的MSE值一直小于LSE的MSE值, LSE的MSE值受k值的影响不大, 但PEBE的MSE值明显随着k取值的增加而减少, 但当$k>10$后, 变化也不明显了. 由于$1/k$表示先验方差与样本方差之比, 因此这个结果与要求β的先验信息相对的精确的事实是一致的.

图5–8给出的是误差方差的PEBE与其LSE的MSE值的模拟结果.

从图5–8可以发现, 所有超参数的值无论如何变化, 误差方差的PEBE的MSE值都小于其LSE的MSE的值, 且有

(1) 当$\lambda<4$时, PEBE与LSE的MSE值差别不大, 但随着λ值的变大, PEBE的MSE值相对于LSE的MSE值的优良性显著增加.

(2) 随着r的变大, PEBE的MSE值相对于LSE的MSE值的优良性显著减小, 且当$r>12$时, 两个估计的MSE的差值接近零.

(3) 历史样本大小n和k的取值的变化对两个估计的MSE值影响不大, PEBE的MSE值一直优于其LSE估计的MSE值, 且稳定性优于LSE.

参 考 文 献

[1] Box, G.P. and Tiao, G.C., *Bayesian Inference in Statistical Analysis*, Massachusetts: Addison-Wesley Press, 1973.

[2] 王松桂, 线性模型的理论及其应用, 合肥: 安徽教育出版社, 1987.

[3] Broemeling, L.D., *Bayesian Analysis of Linear Models*, New York: Marcel Dekker, Inc., 1985.

[4] Wei, L.S. and Zhang, W.P., The superiorities of Bayes linear minimum risk estimation in linear model, *Communications in Statistics: Theory and Methods*, **36**(2007), 917–926.

[5] Kotz, S. and Nadarajah, S., *Multivariate t Distributions and Their Applications*, Cambridge: Cambridge University Press, 2004.

[6] Robbins, H., An empirical Bayes approach to statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**(1955), 157–163.

[7] Morris, C.N., Parametric empirical Bayes inference: theory and applications, *Journal of the American Statistical Association*, **78**(1983), 47–55.

[8] Ghosh, M., Saleh, A.K.Md.E. and Sen, P.K., Empirical Bayes subset estimation in regression models, *Statistics and Decisions*, **7**(1989), 15–35.

[9] Wei, L.S. and Trenkler, G., Mean square error matrix superiority of empirical Bayes estimators under misspecification, *Test*, **4**(1995), 187–205.

[10] Zhang, W.P., Wei, L.S. and Yang, Y.N., The superiority of empirical Bayes estimator of parameters in linear model, *Statistics and Probability Letters*, **72**(2005), 43–50.

[11] Singh, R.S., Empirical Bayes estimation in a multiple linear regression model, *Annals of the Institute of Statistical Mathematics*, **37**(1985), 71–86.

[12] Wei, L.S. and Zhang, S.P., The convergence rates of empirical Bayes estimation in a multiple linear regression model, *Annals of the Institute of Statistical Mathematics*, **47**(1995), 81–97.

[13] Wei, L.S., Convergence rates of empirical Bayes estimation in a class of linear models, *Statistica Sinica*, **8**(1998), 589–605.

The Superiorities of Simultaneous Empirical Bayes Estimation for the Regression Coefficients and Error-Variance in Linear Model

Chen Ling[1,2] Wei Laisheng[1]

([1] *Department of Statistics and Finance, University of Science & Technology of China, Hefei, 230026*)

([2] *School of Mathematical Science, Anhui University, Hefei, 230039*)

When the hyperparameters of prior distribution are partly known in linear model, the simultaneous parametric empirical Bayes estimators (PEBE) of the regression coefficients and error variance are constructed. The superiority of PEBE over the least squares estimator (LSE) of regression coefficients is investigated in terms of the the mean square error matrix (MSEM) criterion, and the superiority of PEBE over LSE of the error variance is discussed under the the mean square error (MSE) criterion. Finally, when all hyperparameters are unknown, the PEBE of regression coefficients and error variance are reconstructed and the superiority of them over LSE under the MSE criterion are studied by simulation methods.

Keywords: Linear model, parametric empirical Bayes estimation, least squares estimation, mean square error (matrix) criterion, simulation results.

AMS Subject Classification: 62C12, 62J05.