3　平均损失意义下线性模型中几类有偏估计的比较

上一章我们在均方误差和均方误差矩阵准则下探讨了异方差与自回归线性模型中回归系数向量的几类有偏估计的优良性，本章将在经典的 Gauss-Markov 模型中考虑基于广义马氏损失函数或马氏损失函数的平均损失准则下回归系数向量的几类有偏估计的优良性，并结合均方误差准则和平均损失准则探讨 $r-k$ 估计和 $r-d$ 估计中的岭参数 k 和 Liu 参数 d 的选择问题及不同情况下最小二乘估计的最优替代估计的选择问题.

3.1　引　言

考虑经典的 Gauss-Markov 模型

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon} \tag{3.1}$$

其中 $\boldsymbol{y}$ 为 $n\times 1$ 观测向量；$\boldsymbol{X}$ 为 $n\times p$ 列满秩样本资料矩阵；$\boldsymbol{\beta}$ 为 $p\times 1$ 未知系数向量；$\boldsymbol{\varepsilon}$ 为 n 维随机误差向量，且满足其均值向量为 $E(\boldsymbol{\varepsilon})=\boldsymbol{0}$，协方差阵为 $\mathrm{Cov}(\boldsymbol{\varepsilon})=\sigma^2\boldsymbol{I}$，$\boldsymbol{I}$ 为 n 阶单位矩阵，$\sigma^2>0$ 为未知参数.

有关未知系数向量 $\boldsymbol{\beta}$ 的估计中，最常用的是最小二乘估计

$$\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{3.2}$$

然而，当解释变量之间存在复共线性时，最小二乘估计往往不稳定，有时其均方误差特别大，有时其符号与问题的实际意义不符，有时其在高维空间的某些方向上严重偏离实际值. 为克服复共线性带来的种种影响，一种方法是考虑主成分估计（Massy，1965）或将主成分估计与其他估计相结合而产生的新估计，如 Baye 和 Parker（1984）结合岭估计和主成分估计提出的 $r-k$ 估计，Kaçıranlar 和 Sakallioğlu（2001）结合 Liu 估计和主成分估计提出的 $r-d$

估计，Özkale（2009b）提出的约束主成分估计，Xu 和 Yang（2011）提出的约束 $r-k$ 估计和约束 $r-d$ 估计，Chang 和 Yang（2012）结合主成分估计和两参数估计提出的两参数主成分估计等.

而关于未知系数向量 β 各类不同估计的优良性比较分析，最重要的是选择合适的损失函数. 通常情况下，我们之所以选择二次损失函数作为评价准则，是因为二次损失函数既能反映估计偏差的大小，又能反映估计方差的大小，同时二次损失函数是凸函数，具有良好的分析性质. 然而，当估计量具有不同的协方差阵时，二次损失函数不能有效地消除各估计量在协方差阵上的差异带来的影响. 于是，一些学者考虑加权二次损失函数或 Fisherian 损失函数，然而若加权矩阵或 Fisherian 损失函数中正定矩阵选择不合适，将导致评价结果不可信（Peddada et al.，1989）. Peddada et al.（1989）运用正态总体的分类思想，导出了 Fisherian 损失函数和马氏损失函数，并得到了岭估计在基于马氏损失函数的平均损失准则和 Pitman 准则下相对于最小二乘估计不可容许的充要条件. 基于马氏损失函数的平均损失准则能消除各估计量因协方差阵不同带来的影响，适合于有偏估计与有偏估计或有偏估计与无偏估计之间的比较，但不适合无偏估计与无偏估计之间的比较.

Massy（1965），Baye 和 Parker（1984），Sarkar（1996），Kaçıranlar 和 Sakallioğlu（2001），Özkale 和 Kaciranlar（2007a），Özkale（2009b），Xu 和 Yang（2011）和 Chang 和 Yang（2012）已在均方误差准则或均方误差矩阵准则下对上述各类估计的优良性做了比较分析，本章将在基于广义马氏损失函数的平均损失准则下探讨上述各类估计的优良性.

3.2 估计的优良性分析

受 Peddada et al.（1989）的启发，Üstündağ 和 Sakallioğlu（2012）在基于马氏损失函数的平均损失准则下，对 $r-k$ 估计与最小二乘估计以及主成分估计与岭估计的优良性做了比较分析. 然而，从下文讨论中可以看出，当 $r=p$ 时，$r-k$ 估计即简化为岭估计，因而 Üstündağ 和 Sakallioğlu（2012）中的定理 2.3 可以看成 Peddada et al.（1989）一文中第 3581 页中的特例，而当 $0<r<p$ 时，由于 $r-k$ 估计的协方差阵是奇异的，因而 Üstündağ 和

Sakallioğlu（2012）中第 2821 页给出马氏损失函数根本不适合 $r-k$ 估计.

为此，我们先将马氏损失函数进行推广.

Üstündağ 和 Sakallioğlu（2012）给出的马氏损失函数的一般形式为：

$$L(\hat{\beta})=(\hat{\beta}-\beta)'(\mathrm{Cov}(\hat{\beta}))^{-1}(\hat{\beta}-\beta) \tag{3.3}$$

易知，当估计 $\hat{\beta}$ 的协方差阵不可逆时，马氏损失函数（3.3）是不适合的.

推广后的广义马氏损失函数的一般形式为：

$$L_M(\hat{\beta})=(\hat{\beta}-\beta)'(\mathrm{Cov}(\hat{\beta}))^{+}(\hat{\beta}-\beta) \tag{3.4}$$

其中 $(\mathrm{Cov}(\hat{\beta}))^{+}$ 表示估计 $\hat{\beta}$ 的协方差阵的 Moore-Penrose 广义逆. 易知广义马氏损失函数（3.4）既适合估计 $\hat{\beta}$ 的协方差阵非奇异的情况，又适合估计 $\hat{\beta}$ 的协方差阵奇异的情况.

定义 3.1　设 $\hat{\beta}_1, \hat{\beta}_2$ 为 β 的两个不同估计，则 $\hat{\beta}_1$ 在基于广义马氏损失函数（3.4）的平均损失准则下优于 $\hat{\beta}_2$ 当且仅当

$$E(L_M(\hat{\beta}_1))\leqslant E(L_M(\hat{\beta}_2)) \tag{3.5}$$

为下文讨论的方便，我们再引入一些简单的记号和基本概念.

记 $\boldsymbol{T}=(\boldsymbol{T}_r \vdots \boldsymbol{T}_{p-r})$，满足

$$\boldsymbol{T}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}=\boldsymbol{\Lambda}$$

其中 $0<r\leqslant p$，$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p)$，$\lambda_1\geqslant\cdots\geqslant\lambda_p$ 表示 $\boldsymbol{X}'\boldsymbol{X}$ 的顺序特征值，则 Chang 和 Yang（2012）在模型（3.1）中提出的两参数主成分估计为：

$$\hat{\beta}_r(k,d)=\boldsymbol{T}_r(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r+\boldsymbol{I}_r)^{-1}(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r+d\boldsymbol{I}_r)(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r+k\boldsymbol{I}_r)^{-1}\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{y} \tag{3.6}$$

其中 $k>0$，$0<d<1$.

从式（3.6）可以看出，主成分估计、$r-k$ 估计、$r-d$ 估计和两参数估计可以看成两参数主成分估计的特例，即

主成分估计：

$$\hat{\beta}_r=\hat{\beta}_r(0,1)=\boldsymbol{T}_r(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{y}$$

$r-k$ 估计：

$$\hat{\boldsymbol{\beta}}_r(k)=\hat{\boldsymbol{\beta}}_r(k,1)=\boldsymbol{T}_r(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r+k\boldsymbol{I}_r)^{-1}\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{y}$$

$r-d$ 估计：

$$\hat{\boldsymbol{\beta}}_r(d)=\hat{\boldsymbol{\beta}}_r(0,d)=\boldsymbol{T}_r(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r+\boldsymbol{I}_r)^{-1}(\boldsymbol{I}_r+d(\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r)^{-1})\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{y}$$

两参数估计：

$$\hat{\boldsymbol{\beta}}(k,d)=\hat{\boldsymbol{\beta}}_p(k,d)=(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X}+d\boldsymbol{I})(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

采用与 Xu 和 Yang（2011）相同的方法，两参数主成分估计、$r-k$ 估计和 $r-d$ 估计又可写成：

$$\hat{\boldsymbol{\beta}}_r(k,d)=\boldsymbol{T}_r\boldsymbol{T}_r'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X}+d\boldsymbol{I})(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}=\boldsymbol{T}_r\boldsymbol{T}_r'\hat{\boldsymbol{\beta}}(k,d)$$
$$\hat{\boldsymbol{\beta}}_r(k)=\boldsymbol{T}_r\boldsymbol{T}_r'(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}=\boldsymbol{T}_r\boldsymbol{T}_r'\hat{\boldsymbol{\beta}}(k)$$
$$\hat{\boldsymbol{\beta}}_r(d)=\boldsymbol{T}_r\boldsymbol{T}_r'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(\boldsymbol{I}+d(\boldsymbol{X}'\boldsymbol{X})^{-1})\boldsymbol{X}'\boldsymbol{y}=\boldsymbol{T}_r\boldsymbol{T}_r'\hat{\boldsymbol{\beta}}(d)$$

其中 $\hat{\boldsymbol{\beta}}_k=(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}$ 为岭估计，$\hat{\boldsymbol{\beta}}_d=(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(\boldsymbol{I}+d(\boldsymbol{X}'\boldsymbol{X})^{-1})\boldsymbol{X}'\boldsymbol{y}$ 为 Liu 估计.

记

$$\boldsymbol{S}=\boldsymbol{X}'\boldsymbol{X}\ ,\ \boldsymbol{\Lambda}_r=\operatorname{diag}(\lambda_1,\cdots,\lambda_r)\ ,\ \boldsymbol{\Lambda}_{p-r}=\operatorname{diag}(\lambda_{r+1},\cdots,\lambda_p)$$
$$\boldsymbol{S}_r(1)=\boldsymbol{\Lambda}_r+\boldsymbol{I}_r\ ,\ \boldsymbol{S}_r(k)=\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r\quad 及\quad \boldsymbol{S}_r(d)=\boldsymbol{\Lambda}_r+d\boldsymbol{I}_r$$

接下来，我们计算上述各类估计基于广义马氏损失函数的平均损失. 考虑到其他估计可以看成两参数主成分估计的特例，我们先计算两参数主成分估计基于广义马氏损失函数的平均损失. 易知，两参数主成分估计 $\hat{\boldsymbol{\beta}}_r(k,d)$ 的偏差向量和协方差阵为：

$$\operatorname{Bias}(\hat{\boldsymbol{\beta}}_r(k,d))=(\boldsymbol{T}_r\boldsymbol{S}_r^{-1}(1)\boldsymbol{S}_r(d)\boldsymbol{S}_r^{-1}(k)\boldsymbol{\Lambda}_r\boldsymbol{T}_r'-\boldsymbol{I})\boldsymbol{\beta}\tag{3.7}$$
$$\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))=\sigma^2\boldsymbol{T}_r\boldsymbol{S}_r^{-1}(1)\boldsymbol{S}_r(d)\boldsymbol{S}_r^{-1}(k)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(k)\boldsymbol{S}_r(d)\boldsymbol{S}_r^{-1}(1)\boldsymbol{T}_r'\tag{3.8}$$

而协方差阵 $\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))$ 的 Moore-Penrose 广义逆为：

$$\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))^{+}=\sigma^{-2}\boldsymbol{T}_r\boldsymbol{S}_r(1)\boldsymbol{S}_r^{-1}(d)\boldsymbol{S}_r(k)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r(k)\boldsymbol{S}_r^{-1}(d)\boldsymbol{S}_r(1)\boldsymbol{T}_r'\tag{3.9}$$

将式（3.7），（3.8）与（3.9）代入下式可得：

$$\begin{aligned}E(L_M(\hat{\boldsymbol{\beta}}_r(k,d)))&=E((\hat{\boldsymbol{\beta}}_r(k,d)-\boldsymbol{\beta})'(\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d)))^{+}(\hat{\boldsymbol{\beta}}_r(k,d)-\boldsymbol{\beta}))\\&=\operatorname{Bias}(\hat{\boldsymbol{\beta}}_r(k,d))'\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))^{+}\operatorname{Bias}(\hat{\boldsymbol{\beta}}_r(k,d))+\end{aligned}$$

$$
\begin{aligned}
&\operatorname{tr}(\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))^{+})\\
&=\boldsymbol{\beta}'(\boldsymbol{T}_r\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(k)\boldsymbol{S}_r(d)\boldsymbol{S}_r^{-1}(1)\boldsymbol{T}_r'-\boldsymbol{I})\operatorname{Cov}(\hat{\boldsymbol{\beta}}_r(k,d))^{+}\\
&\quad(\boldsymbol{T}_r\boldsymbol{S}_r^{-1}(1)\boldsymbol{S}_r(d)\boldsymbol{S}_r^{-1}(k)\boldsymbol{\Lambda}_r\boldsymbol{T}_r'-\boldsymbol{I})\boldsymbol{\beta}+\boldsymbol{r}\\
&=\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r(\boldsymbol{\Lambda}_r-\boldsymbol{S}_r(1)\boldsymbol{S}_r^{-1}(d)\boldsymbol{S}_r(k))(\boldsymbol{I}_r-\\
&\quad\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r(k)\boldsymbol{S}_r^{-1}(d)\boldsymbol{S}_r(1))\boldsymbol{T}_r'\boldsymbol{\beta}+\boldsymbol{r}\\
&=\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\\
&\quad\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d)\boldsymbol{T}_r'\boldsymbol{\beta}+\boldsymbol{r}
\end{aligned}\tag{3.10}
$$

将 $d=1$ 代入式（3.10）可得

$$E(L_M(\hat{\boldsymbol{\beta}}_r(k)))=E(L_M(\hat{\boldsymbol{\beta}}_r(k,1)))=\sigma^{-2}k^2\boldsymbol{\beta}'\boldsymbol{T}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{T}_r'\boldsymbol{\beta}+r \tag{3.11}$$

同样我们可以得到如下结论：

$$E(L_M(\hat{\boldsymbol{\beta}}_r(d)))=\sigma^{-2}(1-d)^2\boldsymbol{\beta}'\boldsymbol{T}_r\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(d)\boldsymbol{T}_r'\boldsymbol{\beta}+r \tag{3.12}$$

$$E(L_M(\hat{\boldsymbol{\beta}}_r))=E(L_M(\hat{\boldsymbol{\beta}}_r(0,1)))=r \tag{3.13}$$

$$E(L_M(\hat{\boldsymbol{\beta}}(k,d)))=\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}((k+1-d)\boldsymbol{\Lambda}+k\boldsymbol{I}_p)^2\boldsymbol{S}^{-1}(d)\boldsymbol{\Lambda}^{-1}\boldsymbol{S}^{-1}(d)\boldsymbol{T}'\boldsymbol{\beta}+p \tag{3.14}$$

$$E(L_M(\hat{\boldsymbol{\beta}}_k))=E(L_M(\hat{\boldsymbol{\beta}}_p(k,1)))=\sigma^{-2}k^2\boldsymbol{\beta}'\boldsymbol{T}\boldsymbol{\Lambda}^{-1}\boldsymbol{T}'\boldsymbol{\beta}+p \tag{3.15}$$

$$E(L_M(\hat{\boldsymbol{\beta}}_d))=\sigma^{-2}(1-d)^2\boldsymbol{\beta}'\boldsymbol{T}\boldsymbol{S}^{-1}(d)\boldsymbol{\Lambda}\boldsymbol{S}^{-1}(d)\boldsymbol{T}'\boldsymbol{\beta}+p \tag{3.16}$$

$$E(L_M(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}))=E(L_M(\hat{\boldsymbol{\beta}}_p(0,1)))=p \tag{3.17}$$

定理 3.1 两参数主成分估计 $\hat{\boldsymbol{\beta}}_r(k,d)$ 在基于广义马氏损失函数的平均损失准则下优于最小二乘估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 的充要条件是：

$$p-r\geqslant\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d)\boldsymbol{T}_r'\boldsymbol{\beta} \tag{3.18}$$

证明　必要性. 由式（3.19）易得.

充分性.

$$
\begin{aligned}
&E(L_M(\hat{\boldsymbol{\beta}}_r(k,d)))-E(L_M(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}))\\
&=r-p+\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d)\boldsymbol{T}_r'\boldsymbol{\beta}
\end{aligned}\tag{3.19}
$$

定理证毕.

将 $d=1$ 和 $k=0$ 依次代入式（3.18），我们可得到如下推论.

推论 3.1　$r-k$ 估计 $\hat{\beta}_r(k)$ 在基于广义马氏损失函数的平均损失准则下优于最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的充要条件是：

$$p-r \geqslant \sigma^{-2}k^2\boldsymbol{\beta}'\boldsymbol{T}_r\boldsymbol{\Lambda}_r^{-1}\boldsymbol{T}_r'\boldsymbol{\beta} \tag{3.20}$$

推论 3.2　$r-d$ 估计 $\hat{\beta}_r(d)$ 在基于广义马氏损失函数的平均损失准则下优于最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的充要条件是：

$$p-r \geqslant \sigma^{-2}(1-d)^2\boldsymbol{\beta}'\boldsymbol{T}_r\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(d)\boldsymbol{T}_r'\boldsymbol{\beta} \tag{3.21}$$

注 3.1　事实上，在基于广义马氏损失函数的平均损失意义下，将 $k=0$ 和 $d=1$ 同时代入式（3.18），我们有主成分估计 $\hat{\beta}_r$ 总是优于最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$；将 $r=p$ 和 $d=1$ 同时代入式（3.18），我们有最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 总是优于岭估计 $\hat{\beta}_k$；将 $r=p$ 和 $k=0$ 同时代入式（3.18），我们有最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 总是优于 Liu 估计 $\hat{\beta}_d$.

定理 3.2　两参数主成分估计 $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}(\boldsymbol{k},\boldsymbol{d})$ 在基于广义马氏损失函数的平均损失准则下总是优于两参数估计 $\hat{\boldsymbol{\beta}}(\boldsymbol{k},\boldsymbol{d})$.

证明　必要性由式（3.22）易得.

充分性.

$$\begin{aligned}&E(L_M(\hat{\boldsymbol{\beta}}(k,d)))-E(L_M(\hat{\boldsymbol{\beta}}_r(k,d)))\\&=p-r+\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_{p-r}((k+1-d)\boldsymbol{\Lambda}_{p-r}+k\boldsymbol{I}_{p-r})^2\boldsymbol{S}_{p-r}^{-1}(d)\boldsymbol{\Lambda}_{p-r}^{-1}\boldsymbol{S}_{p-r}^{-1}(d)\boldsymbol{T}_{p-r}'\boldsymbol{\beta}\geqslant 0\end{aligned} \tag{3.22}$$

定理证毕.

注 3.2　事实上，在基于广义马氏损失函数的平均损失意义下，将 $d=1$ 代入式（3.22），我们有 $r-k$ 估计 $\hat{\beta}_r(k)$ 总是优于岭估计 $\hat{\beta}_k$；将 $k=0$ 代入式（3.22），我们有 $r-d$ 估计 $\hat{\beta}_r(d)$ 总是优于 Liu 估计 $\hat{\beta}_d$.

可参考 Hu 和 Jiewu（2013）.

定理 3.3　$r-k$ 估计 $\hat{\beta}_r(k)$ 和 $r-d$ 估计 $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}(d)$ 在基于广义马氏损失函数的平均损失准则下总是优于两参数主成分估计 $\hat{\beta}_r(k,d)$.

证明　$r-k$ 估计 $\hat{\beta}_r(k)$ 和两参数主成分估计 $\hat{\beta}_r(k,d)$ 基于广义马氏损失函数的平均损失的差为：

$$\boldsymbol{E}(L_M(\hat{\boldsymbol{\beta}}_r(k)))-\boldsymbol{E}(L_M(\hat{\boldsymbol{\beta}}_r(k,d)))$$

$$=\boldsymbol{\sigma}^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r[k^2\boldsymbol{\Lambda}_r^{-1}-((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d)]\boldsymbol{T}_r'\boldsymbol{\beta} \quad (3.23)$$

现在我们证明，对所有的 $k>0$ 和 $0<d<1$，有

$$\Delta_1=k^2\boldsymbol{\Lambda}_r^{-1}-((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d)\leqslant 0 \quad (3.24)$$

$$\Delta_1\leqslant 0\Leftrightarrow k^2\boldsymbol{\Lambda}_r^{-1}\leqslant((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d) \quad (3.25)$$

$$\Leftrightarrow k^2\boldsymbol{I}_r\leqslant((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{S}_r^{-1}(d) \quad (3.26)$$

$$\Leftrightarrow (d-1)k\leqslant(1-d)\lambda_j,\ j=1,2,\cdots,r \quad (3.27)$$

同样的方法，我们可以得到

$$E(L_M(\hat{\boldsymbol{\beta}}_r(d)))-E(L_M(\hat{\boldsymbol{\beta}}_r(k,d)))=\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r\Delta_2\boldsymbol{T}_r'\boldsymbol{\beta}$$

其中

$$\Delta_2=(1-d)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(d)-((k+1-d)\boldsymbol{\Lambda}_r+k\boldsymbol{I}_r)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r^{-1}\boldsymbol{S}_r^{-1}(d)\leqslant\mathbf{0}$$

对所有的 $k>0$ 和 $0<d<1$ 成立.

定理证毕.

定理 3.4　主成分估计 $\hat{\boldsymbol{\beta}}_r$ 在基于广义马氏损失函数的平均损失准则下总是优于两参数主成分估计 $\hat{\boldsymbol{\beta}}_r(k,d)$.

推论 3.3　主成分估计 $\hat{\boldsymbol{\beta}}_r$ 在基于广义马氏损失函数的平均损失准则下总是优于 $r-k$ 估计 $\hat{\boldsymbol{\beta}}_r(k)$ 和 $r-d$ 估计 $\hat{\boldsymbol{\beta}}_r(d)$.

注 3.3　尽管满足某些条件时，两参数主成分估计在均方误差准则下优于两参数估计、$r-k$ 估计、$r-d$ 估计和主成分估计，但从定理 3.2～3.4 我们可以看出，两参数主成分估计在基于广义马氏损失函数的平均损失准则下相对于两参数估计、$r-k$ 估计、$r-d$ 估计和主成分估计是不可容许的.

定理 3.5　在基于广义马氏损失函数的平均损失准则下，当 $k\geqslant 1-d$ 时，$r-d$ 估计 $\hat{\boldsymbol{\beta}}_r(d)$ 优于 $r-k$ 估计 $\hat{\boldsymbol{\beta}}_r(k)$；当 $k<1-d$ 且 $\lambda_1\leqslant\dfrac{kd}{1-d-k}$ 时，$r-d$ 估计 $\hat{\boldsymbol{\beta}}_r(d)$ 优于 $r-k$ 估计 $\hat{\boldsymbol{\beta}}_r(k)$；当 $k<1-d$ 且 $\lambda_r\geqslant\dfrac{kd}{1-d-k}$ 时，$r-k$ 估计 $\hat{\boldsymbol{\beta}}_r(k)$ 优于 $r-d$ 估计 $\hat{\boldsymbol{\beta}}_r(d)$.

证明　考虑 $E(L_M(\hat{\boldsymbol{\beta}}_r(d)))$ 和 $E(L_M(\hat{\boldsymbol{\beta}}_r(k)))$ 的差值

$$E(L_M(\hat{\boldsymbol{\beta}}_r(d)))-E(L_M(\hat{\boldsymbol{\beta}}_r(k)))$$

$$=\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{T}_r[(1-d)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(d)-\boldsymbol{k}^2\boldsymbol{\Lambda}_r^{-1}]\boldsymbol{T}_r'\boldsymbol{\beta} \qquad (3.28)$$

所以 $E(L_M(\hat{\boldsymbol{\beta}}_r(d)))-E(L_M(\hat{\boldsymbol{\beta}}_r(k)))\leqslant 0$ 的一个充分条件是

$$(1-d)^2\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(d)\leqslant k^2\boldsymbol{\Lambda}_r^{-1} \qquad (3.29)$$

而式（3.29）又等价于

$$(1-d-k)\lambda_j\leqslant kd,\ j=1,2,\cdots,r \qquad (3.30)$$

定理证毕.

推论 3.4　在基于广义马氏损失函数的平均损失准则下，当 $k\geqslant 1-d$ 时，Liu 估计 $\hat{\boldsymbol{\beta}}_d$ 优于岭估计 $\hat{\boldsymbol{\beta}}_k$；当 $k<1-d$ 且 $\lambda_1\leqslant\dfrac{kd}{1-d-k}$ 时，Liu 估计 $\hat{\boldsymbol{\beta}}_d$ 优于岭估计 $\hat{\boldsymbol{\beta}}_k$；当 $k<1-d$ 且 $\lambda_r\geqslant\dfrac{kd}{1-d-k}$ 时，岭估计 $\hat{\boldsymbol{\beta}}_k$ 优于 Liu 估计 $\hat{\boldsymbol{\beta}}_d$.

Özkale 和 Kaçıranlar（2007b）为克服压缩估计的缺点，提出了一类两参数估计：

$$\hat{\boldsymbol{\beta}}_{kd}=(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X}+kd\boldsymbol{I})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \qquad (3.31)$$

其中 $k>0$，$0<d<1$.

从式（3.31）可以看出，最小二乘估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$、岭估计 $\hat{\boldsymbol{\beta}}_k$ 和 Liu 估计 $\hat{\boldsymbol{\beta}}_d$ 都可以看成两参数估计 $\hat{\boldsymbol{\beta}}_{kd}$ 的特例，即

最小二乘估计：

$$\lim_{d\to 1}\hat{\boldsymbol{\beta}}_{kd}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}=\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

最小二乘估计：

$$\lim_{k\to 0}\hat{\boldsymbol{\beta}}_{kd}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}=\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

岭估计：

$$\lim_{d\to 0}\hat{\boldsymbol{\beta}}_{kd}=(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}=\hat{\boldsymbol{\beta}}_k$$

Liu 估计：

$$\hat{\boldsymbol{\beta}}_{kd}=(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X}+d\boldsymbol{I})\boldsymbol{X}'\boldsymbol{y}=\hat{\boldsymbol{\beta}}_d\ （k=1\text{时}）$$

在本节的最后，我们将讨论两参数估计 $\hat{\beta}_{kd}$ 在基于马氏损失函数（3.3）的平均损失函数准则下，相对于最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$、岭估计 $\hat{\beta}_k$ 以及 Liu 估计 $\hat{\beta}_d$ 的优良性.

定义 3.2　设 $\hat{\beta}_1, \hat{\beta}_2$ 为 β 的两个不同估计，则 $\hat{\beta}_1$ 在基于马氏损失函数（3.3）的平均损失准则下优于 $\hat{\beta}_2$ 当且仅当

$$E(L(\hat{\beta}_1)) \leqslant E(L(\hat{\beta}_2)) \tag{3.32}$$

定理 3.6　两参数估计 $\hat{\beta}_{kd}$ 在基于马氏损失函数（3.3）的平均损失函数准则下总是优于岭估计 $\hat{\beta}_k$.

证明　易知

$$L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta}) = \sigma^{-2}\boldsymbol{y}'\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{y} - 2\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_k\boldsymbol{S}_{kd}^{-1}\boldsymbol{X}'\boldsymbol{y} + \sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_k\boldsymbol{S}_{kd}^{-1}\boldsymbol{S}\boldsymbol{S}_{kd}^{-1}\boldsymbol{S}_k\boldsymbol{\beta}$$

$$L(\hat{\boldsymbol{\beta}}_k, \boldsymbol{\beta}) = \sigma^{-2}\boldsymbol{y}'\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{y} - 2\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_k\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{y} + \sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_k\boldsymbol{S}^{-1}\boldsymbol{S}_k\boldsymbol{\beta}$$

考虑 $E(L(\hat{\beta}_{kd}))$ 和 $E(L(\hat{\beta}_k)$ 的差值，因为

$$E(L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta})) = \sigma^{-2}k^2(d-1)^2\boldsymbol{\beta}'\boldsymbol{S}_{kd}^{-1}\boldsymbol{S}\boldsymbol{S}_{kd}^{-1}\boldsymbol{\beta} + p$$
$$E(L_M(\hat{\boldsymbol{\beta}}_k, \boldsymbol{\beta})) = \sigma^{-2}k^2\boldsymbol{\beta}'\boldsymbol{S}^{-1}\boldsymbol{\beta} + p$$

所以

$$\begin{aligned} E(L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta})) - E(L(\hat{\boldsymbol{\beta}}_k, \boldsymbol{\beta})) &= \sigma^{-2}k^2\boldsymbol{\beta}'[(d-1)^2\boldsymbol{S}_{kd}^{-1}\boldsymbol{S}\boldsymbol{S}_{kd}^{-1} - \boldsymbol{S}^{-1}]\boldsymbol{\beta} \\ &= \sigma^{-2}k^2\boldsymbol{\beta}'\boldsymbol{S}_{kd}^{-1}((d-1)^2\boldsymbol{S} - \boldsymbol{S}_{kd}\boldsymbol{S}^{-1}\boldsymbol{S}_{kd})\boldsymbol{S}_{kd}^{-1}\boldsymbol{\beta} \\ &= -\sigma^{-2}k^2\boldsymbol{\beta}'\boldsymbol{S}_{kd}^{-1}(d(2-d)\boldsymbol{S} + 2kd\boldsymbol{I} + k^2d^2\boldsymbol{S}^{-1})\boldsymbol{S}_{kd}^{-1}\boldsymbol{\beta} < 0 \end{aligned}$$

定理证毕.

定理 3.7　当 $0<k<1$ 时，两参数估计 $\hat{\beta}_{kd}$ 在基于马氏损失函数（3.3）的平均损失函数准则下总是优于 Liu 估计 $\hat{\beta}_d$，当 $k>1$ 时，Liu 估计 $\hat{\beta}_d$ 在基于马氏损失函数（3.3）的平均损失函数准则下总是优于两参数估计 $\hat{\beta}_{kd}$.

证明　易知

$$L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta}) = \sigma^{-2}\boldsymbol{y}'\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{y} - 2\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_k\boldsymbol{S}_{kd}^{-1}\boldsymbol{X}'\boldsymbol{y} + \sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_k\boldsymbol{S}_{kd}^{-1}\boldsymbol{S}\boldsymbol{S}_{kd}^{-1}\boldsymbol{S}_k\boldsymbol{\beta}$$

$$\boldsymbol{L}(\hat{\boldsymbol{\beta}}_d, \boldsymbol{\beta}) = \sigma^{-2}\boldsymbol{y}'\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{y} - 2\sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_1\boldsymbol{S}_d^{-1}\boldsymbol{X}'\boldsymbol{y} + \sigma^{-2}\boldsymbol{\beta}'\boldsymbol{S}_1\boldsymbol{S}_d^{-1}\boldsymbol{S}\boldsymbol{S}_d^{-1}\boldsymbol{S}_1\boldsymbol{\beta}$$

考虑 $E(L(\hat{\beta}_{kd}))$ 和 $E(L(\hat{\beta}_d)$ 的差值，因为

$$k^2 \boldsymbol{S}_{kd}^{-1}\boldsymbol{S}\boldsymbol{S}_{kd}^{-1} - \boldsymbol{S}_d^{-1}\boldsymbol{S}\boldsymbol{S}_d^{-1} = (k-1)\boldsymbol{S}^2(k\boldsymbol{S}_{kd}^{-2}\boldsymbol{S}_d^{-1} + \boldsymbol{S}_{kd}^{-1}\boldsymbol{S}_d^{-2})$$

所以

$$\begin{aligned} E(L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta})) - E(L(\hat{\boldsymbol{\beta}}_d, \boldsymbol{\beta})) &= \sigma^{-2}(d-1)^2 \boldsymbol{\beta}'(k^2 \boldsymbol{S}_{kd}^{-1}\boldsymbol{S}\boldsymbol{S}_{kd}^{-1} - \boldsymbol{S}_d^{-1}\boldsymbol{S}\boldsymbol{S}_d^{-1})\boldsymbol{\beta} \\ &= \sigma^{-2}(d-1)^2(k-1)\boldsymbol{\beta}'\boldsymbol{S}^2(k\boldsymbol{S}_{kd}^{-2}\boldsymbol{S}_d^{-1} + \boldsymbol{S}_{kd}^{-1}\boldsymbol{S}_d^{-2})\boldsymbol{\beta} \end{aligned}$$

显然，当 $0<k<1$ 时，有

$$E(L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta})) - E(L(\hat{\boldsymbol{\beta}}_d, \boldsymbol{\beta})) < 0$$

当 $k>1$ 时，有

$$E(L(\hat{\boldsymbol{\beta}}_{kd}, \boldsymbol{\beta})) - E(L(\hat{\boldsymbol{\beta}}_d, \boldsymbol{\beta})) > 0$$

定理证毕.

定理 3.8　最小二乘估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 在基于马氏损失函数（3.3）的平均损失函数准则下总是优于两参数估计 $\hat{\boldsymbol{\beta}}_{kd}$.

推论 3.5　最小二乘估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 在基于马氏损失函数（3.3）的平均损失函数准则下总是优于岭估计 $\hat{\boldsymbol{\beta}}_k$ 和 Liu 估计 $\hat{\boldsymbol{\beta}}_d$.

注 3.4　尽管满足某些条件时，由 Özkale 和 Kaçıranlar（2007b），Farebrother（1976）和 Sakallıoğlu et al. （1996）可知，两参数估计在均方误差准则下优于岭估计 $\hat{\boldsymbol{\beta}}_k$、Liu 估计 $\hat{\boldsymbol{\beta}}_d$ 和最小二乘估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$，但从定理 3.6～3.8 我们可以看出，两参数估计、岭估计 $\hat{\boldsymbol{\beta}}_k$ 和 Liu 估计 $\hat{\boldsymbol{\beta}}_d$ 在基于马氏损失函数的平均损失准则下相对于最小二乘估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 是不可容许的.

3.3　参数选择

在实际应用中，有偏估计中未知参数的选取是一个很重要的问题，本节我们将结合均方误差准则和基于广义马氏损失函数的平均损失准则来讨论 $r-k$ 估计和 $r-d$ 估计中岭参数 k 和 Liu 参数 d 的选择问题.

设 $\hat{\boldsymbol{\beta}}_1$ 和 $\hat{\boldsymbol{\beta}}_2$ 为 $\boldsymbol{\beta}$ 的两个不同估计，则 $\hat{\boldsymbol{\beta}}_1$ 在均方误差准则下优于 $\hat{\boldsymbol{\beta}}_2$ 当且仅当

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_1) \leqslant \mathrm{MSE}(\hat{\boldsymbol{\beta}}_2)$$

其中

$$\mathrm{MSE}(\hat{\boldsymbol\beta}_i)=\mathrm{tr}(\mathrm{cov}(\hat{\boldsymbol\beta}_i))+(E(\hat{\boldsymbol\beta}_i)-\boldsymbol\beta)'(E(\hat{\boldsymbol\beta}_i)-\boldsymbol\beta),i=1,2$$

在均方误差准则下，关于岭参数 k 和 Liu 参数 d 的选取，一些学者已做了深入的探讨，如 Hoerl and Kennard（1970a）建议取 $K=\dfrac{\hat\sigma^2}{\hat\alpha_{\max}^2}$ 作为岭估计中岭参数 k 的值，Liu（1993）建议取 $D=\dfrac{\sum_{j=1}^{p}(\hat\alpha_j^2-\hat\sigma^2)/(\lambda_j+1)^2}{\sum_{j=1}^{p}(\lambda_j\hat\alpha_j^2+\hat\sigma^2)/(\lambda_j+1)^2\lambda_j}$ 作为 Liu 估计中 Liu 参数 d 的值，其中 $\hat\sigma^2$ 是 σ^2 的无偏估计；$\hat{\boldsymbol\alpha}$ 是 $\boldsymbol\alpha=\boldsymbol T'\boldsymbol\beta$ 的无偏估计；$\hat\alpha_j^2$ 是 $\hat{\boldsymbol\alpha}$ 的 j 个元素；$\hat\alpha_{\max}$ 是 $\hat{\boldsymbol\alpha}$ 的最大元素.

基于希望 $r-k$ 估计在基于广义马氏损失函数的平均损失准则下表现良好且在均方误差准则下表现不差，受 Hoerl 和 Kennard（1970a），Baye 和 Parker（1984）及推论 3.1 的启发，我们建议 $r-k$ 估计中岭参数 k 的取值为：

$$k_{\mathrm{opt}}=\min\{k_0,k_1,k_2\} \tag{3.33}$$

其中 $k_0=\sqrt{\dfrac{\sigma^2(p-r)}{\boldsymbol\beta'\boldsymbol T_r\boldsymbol\Lambda_r^{-1}\boldsymbol T_r'\boldsymbol\beta}}$ ；k_1 表示方程 $\partial\mathrm{MSE}(\hat{\boldsymbol\beta}_0(k))/\partial k=0$ 的正解；$k_2=\dfrac{\sigma^2}{\gamma_{\max}^2}$ ，$\gamma_{\max}$ 是 $\gamma=\boldsymbol T_r'\boldsymbol\beta$ 中的最大元素.

因为 $k_{\mathrm{opt}}\leqslant k_0$ ，由推论 3.1 知

$$E(L_M(\hat{\boldsymbol\beta}_r(k_{\mathrm{opt}})))\leqslant E(L_M(\hat{\boldsymbol\beta}_{\mathrm{OLS}}))$$

另一方面，当 $k=0$ ，我们有

$$\mathrm{MSE}(\hat{\boldsymbol\beta}_r(k))=\mathrm{MSE}(\hat{\boldsymbol\beta}_r)\quad 且\quad \left.\frac{\partial\mathrm{MSE}(\hat{\boldsymbol\beta}_r(k))}{\partial k}\right|_{k=0}=-2\sigma^2\sum_{j=1}^{r}\frac{1}{\lambda_j^{\,2}}<0$$

因此由 $\mathrm{MSE}(\hat{\boldsymbol\beta}_r(k))$ 的连续性可知，在 $k=0$ 附近，$\mathrm{MSE}(\hat{\boldsymbol\beta}_r(k))$ 单调下降. 因而，当 $0<k_{\mathrm{opt}}\leqslant k_1$ 时，我们有

$$\mathrm{MSE}(\hat{\boldsymbol\beta}_r(k_{\mathrm{opt}}))\leqslant\mathrm{MSE}(\hat{\boldsymbol\beta}_r)$$

所以 $r-k$ 估计 $\hat{\boldsymbol\beta}_r(k_{\mathrm{opt}})$ 在基于广义马氏损失函数的平均损失准则下总是优于最小二乘估计 $\hat{\boldsymbol\beta}_{\mathrm{OLS}}$ 且在均方误差准则下表现良好.

同样，基于希望 $r-d$ 估计在基于广义马氏损失函数的平均损失准则下表现良好且在均方误差准则下表现不差，受 Liu（1993）, Kaçıranlar 和 Sakallioğlu（2001）及推论 3.2 的启发，我们建议 $r-d$ 估计中 Liu 参数 d 的取值为：

$$d_{\text{opt}} = \max\{d_0, d_1, d_2\} \tag{3.34}$$

其中 d_0 是方程 $\sqrt{\dfrac{p-r}{\boldsymbol{\beta}'\boldsymbol{T}_r\boldsymbol{S}_r^{-1}(d)\boldsymbol{\Lambda}_r\boldsymbol{S}_r^{-1}(d)\boldsymbol{T}_r'\boldsymbol{\beta}}} = \dfrac{1-d}{\sigma}$ 的解；$d_1 = \max\left(0, \dfrac{\gamma_j^2-\sigma^2}{\sigma^2/\lambda_j+\gamma_j^2}\right)$，$j=1,\cdots,r$；$d_2 = \dfrac{\sum_{j=1}^{r}(\gamma_j^2-\sigma^2)/(\lambda_j+1)^2}{\sum_{j=1}^{r}(\lambda_j\gamma_j^2+\sigma^2)/(\lambda_j+1)^2\lambda_j}$ 是方程 $\partial\text{MSE}(\hat{\boldsymbol{\beta}}_r(d))/\partial d = 0$ 的解；γ_j 是 $\boldsymbol{\gamma} = \boldsymbol{T}_r'\boldsymbol{\beta}, j=1,\cdots,r$ 中的第 j 个元素.

因为 $E(L_M(\hat{\boldsymbol{\beta}}_r(d)))$ 在区间$(0,1)$上随着 d 的增大而降低且 $d_{\text{opt}} \geqslant d_0$，我们有

$$E(L_M(\hat{\boldsymbol{\beta}}_r(d_{\text{opt}}))) \leqslant E(L_M(\hat{\boldsymbol{\beta}}_{\text{OLS}}))$$

而当 $d=1$ 且 $\left.\dfrac{\partial\text{MSE}(\hat{\boldsymbol{\beta}}_r(d))}{\partial d}\right|_{d=1} = 2\sigma^2\sum_{j=1}^{r}\dfrac{1}{(\lambda_j+1)\lambda_j} > 0$ 时，有

$$\text{MSE}(\hat{\boldsymbol{\beta}}_r(d)) = \text{MSE}(\hat{\boldsymbol{\beta}}_r)$$

由 $\text{MSE}(\hat{\boldsymbol{\beta}}_r(d))$ 的连续性可知，在 $d=1$ 附近 $\text{MSE}(\hat{\boldsymbol{\beta}}_r(d))$ 单调上升. 因而，当 $d_2 \leqslant d_{\text{opt}} < 1$ 且 d_{opt} 非常接近 1 时，我们有

$$\text{MSE}(\hat{\boldsymbol{\beta}}_r(d_{\text{opt}})) \leqslant \text{MSE}(\hat{\boldsymbol{\beta}}_r)$$

所以 $r-d$ 估计 $\hat{\boldsymbol{\beta}}_r(d_{\text{opt}})$ 在基于广义马氏损失函数的平均损失准则下总是优于最小二乘估计 $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ 且在均方误差准则下表现良好.

然而，k_{opt} 和 d_{opt} 依赖于未知参数 σ^2 和 γ_j，因此，在实际应用中，我们建议用 σ^2 和 γ_j 的无偏估计 $\hat{\sigma}^2$ 和 $\hat{\gamma}_j$ 取代 σ^2 和 γ_j，得到 $\hat{k}_{\text{opt}} = \min\{\hat{k}_0, \hat{k}_1, \hat{k}_2\}$ 和 $\hat{d}_{\text{opt}} = \max\{\hat{d}_0, \hat{d}_1, \hat{d}_2\}$ 作为 k_{opt} 和 d_{opt} 的实际操作值.

注 3.5　由定理 3.3 可知，$r-k$ 估计 $\hat{\boldsymbol{\beta}}_r(k)$ 和 $r-d$ 估计 $\hat{\boldsymbol{\beta}}_r(d)$ 在基于广义马氏损失函数的平均损失准则下总是优于两参数主成分估计 $\hat{\boldsymbol{\beta}}_r(k,d)$，所以，本章我们不考虑两参数主成分估计 $\hat{\boldsymbol{\beta}}_r(k,d)$ 中参数 k 与 d 的选择问题.

最后，基于上面的讨论，我们给出实际应用中最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的最优替代估计的一些建议. 在基于广义马氏损失函数的平均损失准则下，由于主成分估计 $\hat{\beta}_r$ 总是优于 $r-k$ 估计 $\hat{\beta}_r(k)$ 和 $r-d$ 估计 $\hat{\beta}_r(d)$ ，则

$$\mathrm{MSE}(\hat{\beta}_r(\hat{k}_{\mathrm{opt}})) \leqslant \mathrm{MSE}(\hat{\beta}_r) \text{ ，且 } \quad E(L_M(\hat{\beta}_r)) = r$$

综合考虑基于广义马氏损失函数的平均损失准则和均方误差准则，我们建议在下述情形下，选择其他估计作为最小二乘估计的最优替代估计.

情形 1：当 $|E(L_M(\hat{\beta}_r(\hat{k}_{\mathrm{opt}}))) - r| \leqslant \delta$ 且 $\mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}})) \geqslant \mathrm{MSE}(\hat{\beta}_r)$ 时，选择 $\hat{\beta}_r(\hat{k}_{\mathrm{opt}})$ 作为最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的最优替代估计，其中 δ 是由实际工作者确定的实际需求精度.

情形 2：当 $|E(L_M(\hat{\beta}_r(\hat{k}_{\mathrm{opt}}))) - r| \leqslant \delta,\ |E(L_M(\hat{\beta}_r(\hat{d}_{\mathrm{opt}}))) - r| \leqslant \delta$ ，$\mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}})) \leqslant \mathrm{MSE}(\hat{\beta}_r)$ 且 $\mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}})) \leqslant \mathrm{MSE}(\hat{\beta}_r(\hat{k}_{\mathrm{opt}}))$ 时，选择 $\hat{\beta}_r(\hat{d}_{\mathrm{opt}})$ 作为最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的最优替代估计；当 $|E(L_M(\hat{\beta}_r(\hat{k}_{\mathrm{opt}}))) - r| \leqslant \delta$ ，$|E(L_M(\hat{\beta}_r(\hat{d}_{\mathrm{opt}}))) - r| \leqslant \delta$ ，$\mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}})) \leqslant \mathrm{MSE}(\hat{\beta}_r)$ 且 $\mathrm{MSE}(\hat{\beta}_r(\hat{k}_{\mathrm{opt}})) \leqslant \mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}}))$ 时，选择 $\hat{\beta}_r(\hat{k}_{\mathrm{opt}})$ 作为最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的最优替代估计，其中 δ 是由实际工作者确定的实际需求精度.

情形 3：当 $|E(L_M(\hat{\beta}_r(\hat{k}_{\mathrm{opt}}))) - r| > \delta,\ |E(L_M(\hat{\beta}_r(\hat{d}_{\mathrm{opt}}))) - r| \leqslant \delta$ ，且 $\mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}})) \geqslant \mathrm{MSE}(\hat{\beta}_r)$ 时，选择 $\hat{\beta}_r$ 作为最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的最优替代估计；

当 $|E(L_M(\hat{\beta}_r(\hat{k}_{\mathrm{opt}}))) - r| > \delta$ ，$|E(L_M(\hat{\beta}_r(\hat{d}_{\mathrm{opt}}))) - r \leqslant \delta$ 且 $\mathrm{MSE}(\hat{\beta}_r(\hat{d}_{\mathrm{opt}})) \leqslant \mathrm{MSE}(\hat{\beta}_r)$ 时，选择 $\hat{\beta}_r(\hat{d}_{\mathrm{opt}})$ 作为最小二乘估计 $\hat{\beta}_{\mathrm{OLS}}$ 的最优替代估计，其中 δ 是由实际工作者确定的实际需求精度.

而对于其他情形，实际工作者可根据实际情况和选定的评价准则来确定最小二乘估计的最优替代估计.