多元响应线性回归模型的马氏 Mallows 模型平均方法改进

赖鑫渝,张立欣,黄振生
南京理工大学 理学院,南京 210094

摘　要:针对多元响应线性回归模型,提出了修改的马氏 Mallows 模型平均(MMMAc)方法。为了更充分地利用多元响应变量之间的相关性信息从而更好地提高预测精度,组合权重选择准则的构造同样考虑了马氏距离预测风险,并通过构造 Wishart 分布,推导出预测损失的无偏估计作为权重的选择准则,最终得到的 MMMAc 准则相比马氏 Mallows 模型平均(MMMA)准则增加了一个偏差矫正项,减小了对预测损失估计的偏差,因此通过最小化该准则得到的权重估计能更接近不可得的理论最优组合权重;最后,模拟对比实验验证了 MMMAc 方法的优势:MMMAc 估计具有与 MMMA 估计同样的渐进最优性,因此两者的表现在大样本情形下没有太大差异,然而,由于修改后的权重选择准则为预测损失的无偏估计,因此在样本量不足的情形下,MMMAc 方法的预测表现更佳。
关键词:多元响应线性回归;模型平均方法;Mallows 准则;Wishart 分布
中图分类号:O212.1　　**文献标识码**:A　　doi:10.16055/j.issn.1672-058X.2023.0002.014

Improvement of Mahalanobis Mallows Model Averaging Method for Multivariate Response Linear Regression Models

LAI Xinyu, ZHANG Lixin, HUANG Zhensheng
School of Science, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract: A corrected Mahalanobis Mallows model averaging (MMMAc) method was proposed for the multivariate response linear regression model. In order to make full use of the correlation information among multivariate response variables and improve the prediction accuracy, Mahalanobis distance prediction risk was also considered in the construction of combinational weight selection criteria. And by constructing the Wishart distribution, the unbiased estimate of predicted loss was derived as the selection criterion of weight. Compared with the original MMMA criterion, the resulting MMMAc added a bias correction term, which reduced the bias in the prediction loss estimation. Therefore, the weight estimation obtained by minimizing this criterion was closer to the theoretical optimal combination weight that is not available. Finally, the simulation and comparison experiments verified the advantages of the MMMAc method: the MMMAc estimation has the same asymptotic optimality as the MMMA estimation, so there is no significant difference in their performance in the large sample case; however, since the modified weight selection criterion is an unbiased estimate of the prediction loss, the MMMAc method performs better in the case of insufficient sample size.
Keywords: multivariate response linear regression; model averaging methods; Mallows criterion; Wishart distribution

1　引　言

在回归分析中,模型选择是一个重要的环节,如为了提高预测精度,Mallows[1] 基于预测均方误差损失提出了 Cp 模型选择准则,准则值最小的候选模型确定为最优模型。然而,模型选择方法忽略了选择过程的不确定性,可能得到与真实数据产生过程相去甚远的模

收稿日期:2021-12-29　**修回日期**:2022-02-14　**文章编号**:1672-058X(2023)02-0094-05
基金项目:全国统计科学研究重大项目(2018LD01).
作者简介:赖鑫渝(1996—),女,福建泉州人,硕士研究生,从事模型平均方法的推断及应用问题研究.
引用格式:赖鑫渝,张立欣,黄振生. 多元响应线性回归模型的马氏 Mallows 模型平均方法改进.[J]. 重庆工商大学学报(自然科学版),2023,40(2):94—98.
LAI Xinyu, ZHANG Lixin, HUANG Zhensheng. Improvement of Mahalanobis Mallows model averaging method for multivariate response linear regression models[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2023, 40(2):94—98.

型。与传统的模型选择不同,模型平均方法不指定唯一的最优模型作为后续分析的依据,而是通过对每一个候选模型的估计或预测进行加权组合,从而得到更稳健的结果,同时避免了选择到一个糟糕的模型。Giraud[2]通过一些理论证明,探讨了在一般设置下,模型平均方法享有模型选择方法所有好的性质。

如何为每一个候选模型选择合适的组合权重是模型平均方法中的一个重要问题,Hjort 和 Claeskens[3]系统地介绍了模型平均方法,并提出了频率模型平均框架。不同于贝叶斯框架下的模型平均方法,频率模型平均方法(FMA)由数据驱动得到组合权重的估计,例如 Buckland 等[4]将模型选择领域的 AIC 和 BIC 信息准则推广应用到模型平均方法中,提出了 smoothed-AIC 和 smoothed-BIC 权重选择准则,即通过信息值确定候选模型的组合权重。然而,为了更好地提高预测精度,Hansen[5]将模型选择领域中的著名的 Mallows Cp 准则推广到了模型平均领域,针对经典的线性回归模型提出了 Mallows 模型平均准则,该准则由两项构成:第一项为模型平均估计的残差平方和,即对拟合优度的度量;第二项为每个候选模型中解释变量个数的加权组合,可看作对模型复杂度的一个加权惩罚项。由此可见,通过最小化 Mallows 模型平均准则得到的权重估计在拟合的偏差与模型复杂度导致的方差之间寻求平衡,方法一经提出便得到了广泛的关注;Wan 等[6]扩展了 Hansen[5]的工作,放松了 Mallows 模型平均方法要求离散权重和嵌套候选模型(Nested Models)这一约束;此外,Mallows 模型平均方法还被推广用于解决其他不同背景下模型选择带来的不确定性问题,例如在非参数回归背景下,Zhu 等[7]针对变系数部分线性模型提出了相应的 Mallows 模型平均方法,与参数情形不同的是,该方法在候选模型的构造上考虑了变量的选择和变量的归属两层不确定性;又如在时间序列背景下,Liao 等[8]针对无穷阶自回归模型提出了广义 Mallows 模型平均方法,并证明了在一定条件下,相应的模型平均估计是渐进最优的。

然而,虽然近年来基于 Mallows 准则的模型平均方法取得了大量杰出的研究成果,但是其中大部分方法都针对一元响应变量情形。随着科技和互联网技术的发展,越来越多的场景需要解决由同一组因子预测多个目标变量的任务,这一类回归问题称为多元多重回归(Multivariate Multiple Regression)。

Zhu 等[9]首次成功尝试将模型平均方法推广到多元响应线性回归上,并基于马氏距离预测风险提出了马氏 Mallows 模型平均准则,该准则在形式上与 Hansen[5]提出的 Mallows 模型平均准则类似,均是由对拟合优度的度量和对模型复杂度的惩罚这两项构成。由于推导的 Mallows 准则包含未知的协方差矩阵,因此与 Hansen[5]的做法相同,Zhu 等[9]用最大候选模型的样本协方差矩阵作为插入估计来直接得到最终的权重选择准则。然而,为了得到预测风险更无偏的估计作为权重的选择准则,使得通过最小化该准则得到的权重估计更接近最小化真实预测风险的理论最优组合权重,从而提高模型平均估计的预测精度,与 Hansen[5]和 Zhu[9]的做法不同,本文通过构造 Wishart 分布(即χ^2分布在多元上的推广),得到了改进的马氏 Mallows 模型平均准则(MMMAc),该准则为马氏距离预测风险的无偏估计。模拟实验验证了在样本量不足的情形下,基于该准则的模型平均方法有更好的表现。

2　模型设置和参数估计

给定 p 维响应变量和 q 维解释变量的 n 组观测样本,考虑如下经典多元响应线性模型:

$$\boldsymbol{Y}=\boldsymbol{X\Theta}+\boldsymbol{E}=\boldsymbol{\mu}+\boldsymbol{E}$$

其中,$\boldsymbol{Y}$ 为和 $\boldsymbol{X}$ 分别为 $n\times p$ 维因变量矩阵和 $n\times q$ 维解释变量矩阵,$\boldsymbol{\Theta}$ 为由待估参数组成的 $q\times p$ 维系数矩阵,$\boldsymbol{E}$ 为 $n\times p$ 维误差矩阵,$\boldsymbol{E}$ 的每一行为独立同多元正态分布的 p 维变量,均值为 0,协方差矩阵为 $\sum$ 。

设 $\vartheta_j, j=1,2,\cdots;J$ 为整数集 $\vartheta=\{1,2,\cdots,q\}$ 的 J 个子集,对于上述模型,用如下 J 个子集候选模型来拟合:$\boldsymbol{Y}=\boldsymbol{X}_j\boldsymbol{\Theta}_j+\boldsymbol{E}=\boldsymbol{\mu}_j+\boldsymbol{E}, j=1,2,\cdots J$;其中 $\boldsymbol{X}_j$ 为 $\boldsymbol{X}$ 由 ϑ_j 中 k_j 个整数索引的列所组成的 $n\times k_j$ 维矩阵,而 $\boldsymbol{\Theta}_j$ 则为 $\boldsymbol{\Theta}$ 由 ϑ_j 索引的行组成的 $k_j\times p$ 维矩阵。

对于每个候选模型 $j=1,2,\cdots,J$,与 Zhu[7]相同,考虑回归系数矩阵的最小二乘估计,即 $\hat{\boldsymbol{\Theta}}_j=(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{Y}$,令 $\boldsymbol{P}_j=\boldsymbol{X}_j(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^{\mathrm{T}}$,则有 $\hat{\boldsymbol{\mu}}_j=\boldsymbol{X}_j\hat{\boldsymbol{\Theta}}_j=\boldsymbol{P}_j\boldsymbol{Y}$。关于多元响应线性模型的最小二乘回归,更多详细的推导过程和讨论可参考 Izenman[10]以及 Zhu[9]。与模型选择方法不同,本文考虑组合每个候选模型的估计。给定集合 $H=\{\omega\in[0,1]^J:\sum_{j=1}^J\omega_j=1\}$ 中的一个元素 $\omega=(\omega_1,\omega_2,\cdots,\omega_J)^{\mathrm{T}}$ 作为模型平均的权重向量,则$\boldsymbol{\mu}$ 相应的模型平均估计为

$$\hat{\boldsymbol{\mu}}(\boldsymbol{\omega})=\sum_{j=1}^{J}\omega\hat{\boldsymbol{\mu}}_j=\sum_{j=1}^{J}\omega_j\boldsymbol{P}_j\boldsymbol{Y}\equiv\boldsymbol{P}(\omega)\boldsymbol{Y}$$

组合权重的选择是模型平均方法一个重要的环节,前人提供了许多方法,其中一个著名的方法是 Hansen[5]提出的,该方法通过最小化基于预测风险推导的 Mallows 模型平均准则来得到组合权重估计,由此得到的权重估计渐进等价于最小化真实预测风险的理论最优权重。而针对多元响应的情形,为了更好地利用多元响应变量之间的相关性信息,考虑如下基于马氏距离的预测风险:

$R(\boldsymbol{\omega})=\boldsymbol{E}_{\boldsymbol{Y}}\boldsymbol{E}_{\boldsymbol{Y}_F}[\mathrm{tr}\{\Sigma^{-1}(\boldsymbol{Y}_F-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{Y}_F-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))\}]$,其中 $\boldsymbol{Y}_F$ 为与 $\boldsymbol{Y}$ 相互独立且同分布的 $n\times p$ 维随机矩阵,可以看作因变量未来的观测矩阵。理论上,最小化上述预测风险的权重为理想的最优权重,然而由于其中包含未知量 $\boldsymbol{Y}_F$,因此无法直接通过最小化

该风险得到权重估计。一个可行的办法是得到预测风险的无偏估计作为权重选择准则,Mallows Cp 准则最早便是作为预测风险的渐进无偏估计提出来的,因此该风险也被称为“Mallows 型风险”(Mallows-type risk),见 Fujikoshi 和 Satoh[11]相关的讨论。针对一元响应线性回归模型,Hansen[5]提出的 Mallows 模型平均准则由两项构成:第一项为模型平均估计的残差平方和,即对拟合优度的度量;第二项为每个候选模型中解释变量个数的加权组合,可看作对模型复杂度的一个加权惩罚项。而针对多元响应线性回归模型,Zhu[9]基于 Mallows 型预测风险推导的 Mallows 型权重选择准则与 Hansen[5]的 Mallows 权重选择准则形式上相似。

由此可以看出,不论是针对经典的一元响应线性回归模型还是针对多元响应线性回归模型,基于预测误差推导出的 Mallows 模型平均准则均在拟合的偏差和模型复杂度导致的方差之间寻求平衡。然而由上述风险推导的准则包含未知的协方差矩阵 $\sum$,因此 Zhu[9]利用最大候选模型来估计 $\sum$,即用一个插入估计 $\hat{\sum}_{\vartheta}=1/(n-k_{\vartheta})(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}_{\vartheta})'(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}_{\vartheta})$ 来替换准则中的 $\sum$,得到最终的 MMMA 权重选择准则为

$$\hat{\boldsymbol{C}}_n(\boldsymbol{\omega})=\mathrm{tr}\{\hat{\sum}_{\vartheta}^{-1}(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))\}+2p\mathrm{tr}(\boldsymbol{P}(\boldsymbol{\omega}))。$$

3 修改的 Mallows 权重准则

与 Hansen[5]和 Zhu[9]通过一个插入估计得到最终权重选择准则的做法不同,受启发于 Fujikoshi 和 Satoh[11],本节的目标是通过构造 Wishart 分布来推导预测风险的无偏估计,得到修改的 Mallows Cp 模型平均准则。模型平均估计的马氏 Mallows 型预测风险可写为

$$\begin{aligned}R(\boldsymbol{\omega})=&\boldsymbol{E}_Y\boldsymbol{E}_{Y_F}[\mathrm{tr}\{\sum^{-1}(\boldsymbol{Y}_F-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{Y}_F-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))\}]=\\&\boldsymbol{E}_Y\boldsymbol{E}_{Y_F}[\mathrm{tr}\{\sum^{-1}((\boldsymbol{X\Theta}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{X\Theta}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))+\\&\boldsymbol{E}'\boldsymbol{E}-2(\hat{\boldsymbol{\mu}}(\boldsymbol{\omega})-\boldsymbol{X\Theta})'\boldsymbol{E})\}]=\\&\boldsymbol{E}_Y[\mathrm{tr}\{(\boldsymbol{X\Theta}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{X\Theta}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))\sum^{-1}\}]+\\&np\end{aligned}$$

假设用 $\mathrm{tr}\{\hat{\sum}^{-1}\boldsymbol{W}\}$ 估计该风险时偏差为 $\boldsymbol{B}$,则上式可写为 $R(\boldsymbol{\omega})=\boldsymbol{E}_Y[\mathrm{tr}\{\hat{\sum}^{-1}\boldsymbol{W}\}]+\boldsymbol{B}$,其中残差平方和 $\boldsymbol{W}=(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))$,$\sum$ 的无偏估计 $\hat{\sum}_{\vartheta}=1/(n-k_{\vartheta})(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}_{\vartheta})'(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}_{\vartheta})$,由此,偏差 $\boldsymbol{B}$ 可以写为

$$\begin{aligned}\boldsymbol{B}=&\boldsymbol{E}_Y[\mathrm{tr}\{\sum^{-1}(\boldsymbol{X\Theta}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))'(\boldsymbol{X\Theta}-\hat{\boldsymbol{\mu}}(\boldsymbol{\omega}))\}]+\\&np-\boldsymbol{E}_Y[\mathrm{tr}\{\hat{\sum}_{\vartheta}^{-1}\boldsymbol{W}\}]\end{aligned}$$

在 Zhu[9]提出的 MMMA 准则中,$\boldsymbol{B}$ 的估计为 $\hat{\boldsymbol{B}}=2p\mathrm{tr}(\boldsymbol{P}(\boldsymbol{\omega}))$,考虑改进这个估计。构造矩阵:

$$\boldsymbol{W}_0=\sum^{-\frac{1}{2}}\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I}_n-\boldsymbol{P}_{\vartheta})\boldsymbol{Y}\sum^{-\frac{1}{2}}$$

$$\boldsymbol{W}_1=\sum^{-\frac{1}{2}}\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_{\vartheta}-\boldsymbol{P}(\omega))\boldsymbol{Y}\sum^{-\frac{1}{2}}=\boldsymbol{D}-\sum_{j=1}^{J}\omega_j\boldsymbol{A}_j$$

$$\boldsymbol{W}_2=\sum_{j=1}^{J}\omega_j\boldsymbol{C}_j=\sum^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{X\Theta})'\boldsymbol{P}(\omega)(\boldsymbol{Y}-\boldsymbol{X\Theta})\sum^{-\frac{1}{2}}$$

$$\boldsymbol{\Omega}_1=\sum^{-\frac{1}{2}}\boldsymbol{\Theta}'\boldsymbol{X}'(\boldsymbol{P}_{\vartheta}-\boldsymbol{P}(\boldsymbol{\omega}))\boldsymbol{X\Theta}\sum^{-\frac{1}{2}}$$

$$\boldsymbol{\Omega}_2=\sum^{-\frac{1}{2}}\boldsymbol{\Theta}'\boldsymbol{X}'\boldsymbol{P}(\omega)(\boldsymbol{P}(\omega)-\boldsymbol{I})\boldsymbol{X\Theta}\sum^{-\frac{1}{2}}$$

其中,$\boldsymbol{A}_j=\sum^{-\frac{1}{2}}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_j\boldsymbol{Y}\sum^{-\frac{1}{2}},j=1,2,\cdots,J$; $\boldsymbol{C}_j=\sum^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{X\Theta})'\boldsymbol{P}_j(\boldsymbol{Y}-\boldsymbol{X\Theta})\sum^{-\frac{1}{2}},\boldsymbol{j}=1,2,\cdots,\boldsymbol{J}$; $\boldsymbol{D}=\sum^{-\frac{1}{2}}\boldsymbol{Y}'\boldsymbol{P}_{\vartheta}\boldsymbol{Y}\sum^{-\frac{1}{2}}$,$\boldsymbol{P}_{\vartheta}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$,$\boldsymbol{P}_j=\boldsymbol{X}_j(\boldsymbol{X}_j'\boldsymbol{X}_j)\boldsymbol{X}_j'$。

由 Muirhead[12]中 Wishart 分布的定义容易得到定理 1:

定理 1 令 $\boldsymbol{W}_0,\boldsymbol{D},\boldsymbol{A}_j,\boldsymbol{C}_j,j=1,2,\cdots,J$ 为由上述式子定义的随机矩阵,则它们分别服从非中心 Wishart 分布:$W_p(\boldsymbol{I}_p,n-k_{\vartheta},\boldsymbol{\Theta}'\boldsymbol{X}'(\boldsymbol{I}_n-\boldsymbol{P}_{\vartheta})\boldsymbol{X\Theta})$,$W_p(I_p,k_{\vartheta},\boldsymbol{\Theta}'\boldsymbol{X}'\boldsymbol{P}_{\vartheta}\boldsymbol{X\Theta})$,$W_p(I_p,k_j,\boldsymbol{\Theta}'\boldsymbol{X}'\boldsymbol{P}_j\boldsymbol{X\Theta})$ 以及 Wishart 分布:$W_p(\boldsymbol{I}_p,k_j),j=1,2,\cdots,J$。

注意到由于误差矩阵 E 各行之间相互独立,因此对任意常数矩阵 $\boldsymbol{A}\in\boldsymbol{R}^{n\times n}$,易证得 $\boldsymbol{E}(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{E})=\mathrm{tr}(\boldsymbol{A})\sum$。由定理 1 和 Wishart 分布的性质以及 Fujikoshi 和 Satoh[11]中的定理 2 可以得到:

$$E[\mathrm{tr}(W_2)]=p\mathrm{tr}(P(\omega))=\boldsymbol{P}\sum_{j=1}^{J}\omega_jk_j$$

$$\begin{aligned}E[\mathrm{tr}(\sum^{-1}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\boldsymbol{Y})]=\\p\mathrm{tr}\{\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\}+\mathrm{tr}(\boldsymbol{\Omega}_2)\end{aligned}$$

$$E[\mathrm{tr}(\boldsymbol{W}_0^{-1})]=\frac{1}{n-k_{\vartheta-p-1}}\boldsymbol{I}_p$$

和 $\boldsymbol{E}[\mathrm{tr}(\boldsymbol{W}_1)]=\mathrm{tr}(\boldsymbol{\Omega}_1)+\boldsymbol{p}\mathrm{tr}\{\boldsymbol{P}_{\vartheta}-\boldsymbol{P}(\boldsymbol{\omega})\}$。由此,$\boldsymbol{B}$ 可以写为

$$\begin{aligned}\boldsymbol{B}=&np+\boldsymbol{E}_Y[\mathrm{tr}\{\boldsymbol{W}_2+\boldsymbol{\Omega}_1+\sum^{-1}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\boldsymbol{Y}\}]-\\&(\boldsymbol{n}-\boldsymbol{k}_{\vartheta})\boldsymbol{E}_Y[\mathrm{tr}\{\boldsymbol{I}_p+\boldsymbol{W}_0^{-1}\boldsymbol{W}_1+\\&\sum^{-1}\boldsymbol{W}_0^{-1}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\boldsymbol{Y}\}]=\\&\frac{1}{\boldsymbol{n}-\boldsymbol{k}_{\vartheta}-\boldsymbol{p}-1}[(2\boldsymbol{n}-2\boldsymbol{k}_{\vartheta}-\boldsymbol{p}-1)\cdot\sum_{j=1}^{J}\boldsymbol{\omega}_j\cdot\boldsymbol{p}\cdot\boldsymbol{k}_j-\\&(\boldsymbol{p}+1)\boldsymbol{p}\boldsymbol{k}_{\vartheta}-(\boldsymbol{p}+1)\mathrm{tr}(\boldsymbol{\Omega}_1)-(p+1)\mathrm{tr}(\boldsymbol{\Omega}_2)-\\&(p+1)p\mathrm{tr}(\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I}))]\end{aligned}$$

而 $\mathrm{tr}(\boldsymbol{\Omega}_1)$ 和 $\mathrm{tr}(\boldsymbol{\Omega}_2)$ 的无偏估计则可以通过如下结果得到:

$$\begin{aligned}\boldsymbol{E}[\mathrm{tr}\{((\boldsymbol{Y}-\hat{\boldsymbol{Y}}_{\vartheta})^{\mathrm{T}}(\boldsymbol{Y}-\hat{\boldsymbol{Y}}_{\vartheta}))^{-1}(\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_{\vartheta}-\boldsymbol{P}(\boldsymbol{\omega}))\boldsymbol{Y})\}]=\\\boldsymbol{E}[\mathrm{tr}(\boldsymbol{W}_0^{-1}\boldsymbol{W}_1)]=\frac{1}{n-k_{\vartheta}-p-1}\{p(k_{\vartheta}-\sum_{j=1}^{J}\omega_jk_j)+\mathrm{tr}(\boldsymbol{\Omega}_1)\}\end{aligned}$$

同理有,

$$\boldsymbol{E}[\mathrm{tr}\{((\boldsymbol{Y}-\hat{\boldsymbol{Y}}_{\boldsymbol{\vartheta}})'(\boldsymbol{Y}-\hat{\boldsymbol{Y}}_{\boldsymbol{\vartheta}}))^{-1}(\boldsymbol{Y}'\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\boldsymbol{Y})\}]=\frac{1}{n-k_{\vartheta}-p-1}\{p\mathrm{tr}(\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I}))+\mathrm{tr}(\boldsymbol{\Omega}_2)\}$$

通过简单计算可得:

$$\hat{\boldsymbol{B}}_c=2p\sum_{j=1}^{J}\omega_j k_j-(p+1)\mathrm{tr}\{\boldsymbol{W}_{\boldsymbol{\vartheta}}^{-1}(\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}_{\boldsymbol{\vartheta}}-\boldsymbol{P}(\boldsymbol{\omega}))\boldsymbol{Y}\}-(p+1)\mathrm{tr}\{\boldsymbol{W}_{\boldsymbol{\vartheta}}^{-1}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\boldsymbol{Y})\}$$

其中,$\boldsymbol{W}_{\boldsymbol{\vartheta}}=(\boldsymbol{Y}-\boldsymbol{P}_{\boldsymbol{\vartheta}}\boldsymbol{Y})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{P}_{\boldsymbol{\vartheta}}\boldsymbol{Y})$。综上,即可得到最终修改的马氏 Mallows 权重选择(MMMAc)准则为

$$MC_p(\omega)=n\mathrm{tr}\{\hat{\Sigma}_{\vartheta}^{-1}\boldsymbol{W}\}+2p\sum_{j=1}^{J}\omega_j k_j-(p+1)\sum_{j=1}^{J}\omega_j\mathrm{tr}\{\boldsymbol{W}_{\boldsymbol{\vartheta}}^{-1}(\boldsymbol{W}_{(j)}-\boldsymbol{W}_{\boldsymbol{\vartheta}})\}-(p+1)\mathrm{tr}\{\boldsymbol{W}_{\boldsymbol{\vartheta}}^{-1}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\omega})(\boldsymbol{P}(\boldsymbol{\omega})-\boldsymbol{I})\boldsymbol{Y})\}$$

则相应的权重向量可通过最小化该准则得到,即 $\hat{\boldsymbol{\omega}}=\mathrm{argmin}_{\omega\in H}MC_p(\boldsymbol{\omega})$。

4　模拟实验

上一节提出了修改后的 Mallows 模型平均准则,本节通过接下来的模拟对比实验,将修改后的马氏 Mallows 模型平均(MMMAc)方法与 Zhu[9] 提出的 Mallows 模型平均(MMMA)方法作对比,来探究所提出的 MMMAc 方法在实际数据处理中的表现。

为了更好地探究所提出的方法的表现,模拟实验设置与 Zhu[9] 相同,方便起见,这里将对此做简单的介绍。模拟数据通过等式 $\boldsymbol{Y}=\boldsymbol{X\Theta}+\boldsymbol{E}$ 生成,其中 $\boldsymbol{Y}$ 为 $n\times2$ 维因变量,X 为 $n\times200$ 维解释变量,$\boldsymbol{X}$ 第一列的每一个元素均设为 1 来作为截距项,其他的元素则独立采样于标准正态分布;系数矩阵 $\boldsymbol{\Theta}$ 为 200×2 维矩阵,第 j 行为$(\theta_{j1},\theta_{j2})$,其中 $\theta_{j1}=c\sqrt{2\alpha}j^{-(1/2)\alpha-1/2}$,$\theta_{j2}=c\sqrt{2\alpha}j^{-\alpha-1/10}$,这是 Hansen 模拟实验中的系数设置在多元情形的类似推广;误差矩阵 $\boldsymbol{E}$ 的每一行独立采样于多元正态分布 $MVN(0,\sum)$,其中协方差矩阵设置为$\begin{pmatrix}1 & r_0\\ r_0 & 1\end{pmatrix}$。

上述参数有如下设置:令 r_0 在$\{0,0.3,0.6,0.9\}$之间变化取值,它的取值越大则反映随机误差之间的相关性越强;系数 c 用来控制 R^2,其中 R^2 设置为

$$(VAR(\mu_{j1})+VAR(\mu_{j2}))/(VAR(y_{j1})+VAR(y_{j2}))\approx(VAR(\mu_{j1})+VAR(\mu_{j2}))/(VAR(\mu_{j1})+VAR(\mu_{j2})+VAR(\varepsilon_{j1})+VAR(\varepsilon_{j2})),$$

并且在 0.1~0.9 之间以间隔 0.1 变化,它反映了模型的信噪比,取值越大反映信噪比越大。设置参数 α 在$\{0.5,1.0,1.5\}$中变化取值,每次实验的样本量设置为 $n=25,50,100$,而候选模型个数对应于样本量分别设置为 $M=9,11,14$,第 m 个模型包含前 m 个解释变量。

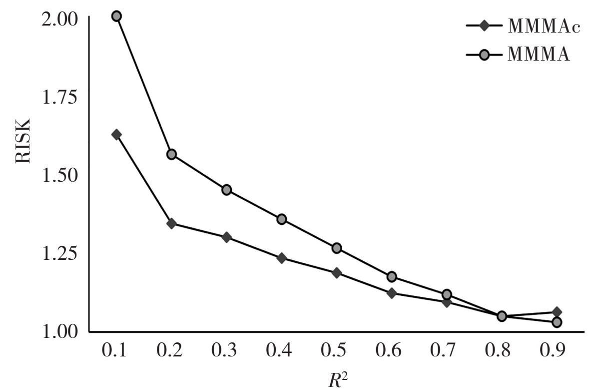
本文提出的修改马氏 Mallows 模型平均(MMMAc)估计通过最小化上一节提出的 MMMAc 准则来得到组合权重的估计,而马氏 Mallows 模型平均(MMMA)估计则通过最小化 MMMA 准则选择权重,为了评估它们的表现,对每个参数组合重复进行 500 次实验。与 Zhu[9] 相同,对每一次实验计算如下损失:

$$R_d=\mathrm{tr}\{\sum{}^{-1}(\hat{\boldsymbol{\mu}}_d(\boldsymbol{\omega})-\boldsymbol{\mu})'(\hat{\boldsymbol{\mu}}_d(\boldsymbol{\omega})-\boldsymbol{\mu})\},d=1,2,\cdots,500。$$

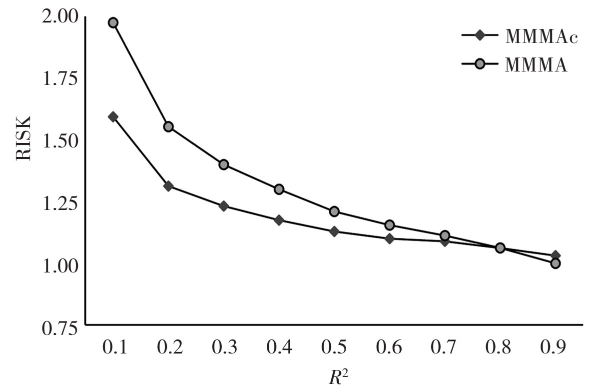
最终的结果取每次实验的平均损失。

模拟实验的计算结果展示在图 1,图 2 中,分别对应于样本量 $n=25,50$ 的情形,样本量为 $n=100$ 的结果与 $n=50$ 类似。为了节省空间,仅展示具有代表性的,即参数 α 设置为 1.5、参数 r_0 分别设置为 0 和 0.6 的结果。

可以观察到,随着 R^2 增加,两个方法的平均损失均呈现下降趋势。这表明随着信噪比的增长,信号强度增大,两者的预测效果均会变好,这符合人们的认知。



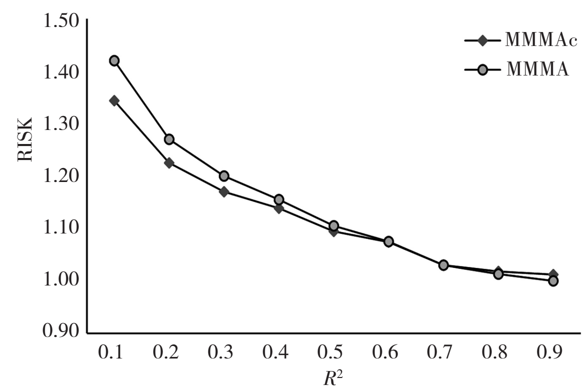
(a) $r=0$



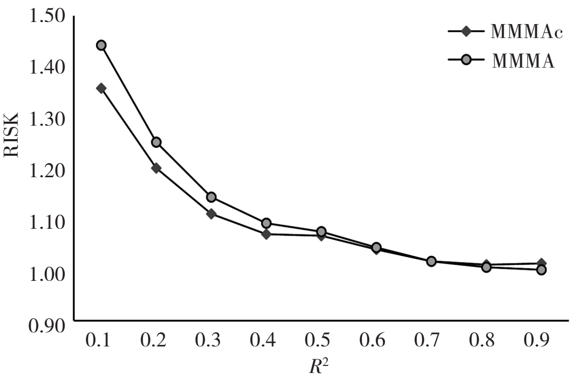
(b) $r=0.6$

图 1　MMMAc 和 MMMA 方法的风险比较($n=25$)

Fig. 1　Risk comparison of MMMAc method and MMMA method($n=25$)



(a) $r=0,n=50$



(b) $r=0.6, n=50$

图 2 MMMAc 和 MMMA 方法的风险比较($n=50$)

Fig. 2 Risk comparison of MMMAc method and MMMA method($n=50$)

此外可以发现,参数 r_0 的变化不会导致两个方法的差距有较大变化,这是由于 MMMAc 准则与 Zhu[9] 提出的 Mallows 模型平均准则类似,均是由基于马氏距离(Mahalanobis distance)的预测风险推导而来的,而马氏距离考虑了多元随机变量之间的相关性,因此在对响应变量相关性信息的利用方面,MMMAc 方法与 MMMA 方法没有太大区别。

在图 1 中,MMMAc 的平均损失总体来说比 MMMA 小,这表明在这个设定下,MMMAc 优于 MMMA。注意到,虽然随着信号强度的增强,两个方法的平均损失均下降,且差距在逐步缩小,但是当信号强度比较弱的时候,MMMAc 相较于 MMMA 方法有较大优势。然而,两个方法的差异仅在图 1 中比较明显,在图 2 中,两条平均损失曲线靠近并接近重合,这表明,在样本量比较大的情形下,两个方法的结果近乎相同。其实这并不是一个令人惊讶的结果,事实上,Zhu[9] 证明了所提出的马氏 Mallows 模型平均方法的渐进最优性,即在概率框架下,随着样本量趋近无穷,通过最小化 Mallows 权重选择准则得到的组合权重渐进等价于不可得的理论最优权重,这为所得到的 Mallows 模型平均估计提供了大样本支持。而修改后的 Mallows 权重选择准则为 Mallows 风险的无偏估计,因此它在样本量不足的情形下也能够提供优秀的估计和预测。

5 结 论

为了更充分地利用响应变量之间的相关性信息,Zhu[9] 基于马氏距离预测风险推导了 Mallows 准则,将准则中未知的协方差矩阵用一个插入估计替换,从而得到了最终的权重选择准则。本文考虑修正 Zhu[9] 提出的 Mallows 准则,通过构造 Wishart 分布推导得到预测风险的无偏估计作为修改后的 Mallows 模型平均准则,该准则相比于 MMMA 准则增加了一个偏差矫正项,减小了对预测风险估计的偏差,因此通过最小化该准则得到的权重估计能更接近不可得的理论最优组合权重。

本文通过一个模拟实验将 MMMAc 方法与 MMMA 方法进行对比,探究了修改后的模型平均方法在实际数据处理中表现。可以发现,当样本量较大的时候,两个方法的表现十分接近,这是因为 MMMA 估计具有渐进最优性,在样本量大的情形下能得到足够优秀的权重估计,因此在该情形下 MMMAc 对 MMMA 估计并不能有更大的改进,从这一点上也验证了 MMMA 方法的大样本性质;而在样本量不足的情形下,MMMAc 相较于 MMMA 方法优势较为明显,尤其是当信噪比较大的时候,这从应用层面证实了 MMMAc 准则相较于 MMMA 准则对预测风险估计的无偏性。虽然大部分关于模型平均方法的理论研究致力于探究估计的大样本性质,但是实际问题中,样本量不足的情形是很普遍的,因此 MMMAc 有理论和实践价值。

参考文献(References):

[1] MALLOWS C L. Some comments on CP[J]. Technometrics, 1973, 15(4): 661—675.

[2] GIRAUD C. Introduction to high-dimensional statistics[M]. New York: Chapman & Hall, CRC Press, 2015.

[3] HJORT N L, CLAEKENS G. Frequentist model average estimators[J]. Journal of the American Statistical Association, 2003, 464(98): 879—899.

[4] BUCKLAND S T, BURNHANM K P, AUGUSTIN N H. Model selection: an integral part of inference[J]. Biometrics, 1997, 53(2): 603—618.

[5] HANSEN B E. Least squares model averaging[J]. Econometrica, 2007, 75(4): 1175—1189.

[6] WAN ATK, ZHANG X, ZOU G. Least squares model averaging by Mallows criterion[J]. Journal of Econometrics, 2010, 156(2): 277—283.

[7] ZHU R, WAN ATK, ZHANG X, et al. A mallows-type model averaging estimator for the varying-coefficient partially linear model[J]. Journal of the American Statistical Association, 2019, 526(114): 882—892.

[8] LIAO J, ZOU G, GAO Y, et al. Model averaging prediction for time series models with a diverging number of parameters[J]. Journal of Econometrics, 2021, 223(1): 190—221.

[9] ZHU R, ZOU G, ZHANG X. Model averaging for multivariate multiple regression models[J]. Statistics, 2018, 52(1): 205—227.

[10] IZENMAN A J. Modern multivariate statistical techniques: regression, classification, and manifold learning[M]. New York, NY: Springer, Springer Texts in Statistics, 2008.

[11] FUJIKOSHI Y, SATOH K. Modified AIC and CP in multivariate linear regression[J]. Biometrika, 1997, 84(3): 707—716.

[12] MUIRHEAD R J. Aspect of multivariate statistical theory[M]. New York: Wiley, 1982.

责任编辑:田 静