山西大学学报(自然科学版)48(5):900－910,2025
Journal of Shanxi University(Nat. Sci. Ed. )

**DOI**:10.13451/j.sxu.ns.2024025

# 随机限制回归模型的自适应最小绝对收缩和选择算子方法

靳琴琴,彭伟,廖天自,孙宝*
(太原科技大学 应用科学学院,山西 太原 030024)

**摘 要**:随机限制最小绝对收缩和选择算子(M-Lasso)方法可以在变量选择的同时使用随机的先验信息,但是该方法基于的最小绝对收缩和选择算子(Lasso)对每一个系数的惩罚是等权重的,这可能会导致某些重要的信息被过度压缩。为此,本文提出随机限制自适应Lasso(Ma-Lasso)方法。该方法赋予系数不同的权重,并且在变量选择的同时使用了随机先验信息,可提高估计的精度。通过数值实验结果分析发现,该方法在稀疏模型上表现出比其他方法更小的均方误差,并且在发现率、真实发现率以及真实模型选择次数比例方面也具有一定的优势。最后,通过将该方法应用于贵州茅台各季度财报数据和股票价格数据中,发现Ma-Lasso所构建的模型的贝叶斯信息准则(Bayesian Information Criterion,BIC)值相较于M-Lasso方法下降了约5%,进一步验证了它的优越性。

**关键词**:稀疏模型;变量选择;均方误差;随机先验信息
**中图分类号**:O212　　**文献标志码**:A　　**文章编号**:0253-2395(2025)05-0900-11

# Adaptive Least Absolute Shrinkage and Selection Operator Method for Stochastic Limit Regression Model

JIN Qinqin, PENG Wei, LIAO Tianzi, SUN Bao*
(*School of Applied Science, Taiyuan University of Science and Technology, Taiyuan* 030024, *China*)

**Abstract:** The Mixed least absolute shrinkage and selection operator (M-Lasso) method can use random prior information at the same time as variable selection, but the minimum absolute shrinkage and selection operator (Lasso) based on this method is equally weighted for each coefficient, which may cause some important information to be overcompressed. To address this issue, this paper proposes the stochastic constrained adaptive Lasso (Ma-Lasso) method. The method assigns different weights to the coefficients and has oracle properties. It uses stochastic prior information along with variable selection, which can improve the precision of the estimation. The analysis of numerical experimental results reveals that the method exhibits a smaller mean square error on the sparse model than the other methods, and also has some advantages in terms of the discovery rate, the true discovery rate, and the ratio of the number of times the true model is selected. Finally, by applying Ma-Lasso to the quarterly financial data and stock price data of Kweichow Moutai, it is found that the BIC value of the model constructed by this method decreases by about 5% compared with M-Lasso method, which further verifies its superiority.

**Key words:** sparse model; variable selection; mean squared error; random prior information

**收稿日期**:2023-09-15;**接受日期**:2024-03-07
**基金项目**:山西省高校科技创新项目(2021L324);山西省基础研究计划青年科学研究项目(202103021223282);太原科技大学博士科研启动基金(20212014)
**作者简介**:靳琴琴(1986—),女,山西洪洞人,博士,讲师,主要研究方向为基因大数据研究。E-mail: qinqinjin@tyust.edu.cn
* **通信作者**:孙宝(SUN Bao),E-mail:2009031@tyust.edu.cn
**引文格式**:靳琴琴,彭伟,廖天自,等.随机限制回归模型的自适应最小绝对收缩和选择算子方法[J].山西大学学报(自然科学版),2025,48(5):900-910. DOI:10.13451/j.sxu.ns.2024025.

## 0 引言

最小二乘法(Ordinary Least Square,OLS)是最常用的线性回归模型系数的估计方法。它通过最小化结果变量的预测值与观测值之间的误差来估计回归模型中的参数。该方法可以对当前数据集进行无偏估计,但容易导致过拟合现象。当误差项为异方差或者相关时,研究人员可以用广义最小二乘法(Generalized Least Squares,GLS)来估计参数;而如果模型参数存在线性限制,则可以使用受限最小二乘法(Restricted Least Squares,RLS)[1]。

但是当数据间存在多重共线性时,最小二乘法不再是一个有效估计[2-5]。为此,Hoerl 和 Kennard 提出岭回归方法[6],这是能够进行共线性数据分析的一种有偏估计回归方法。它可通过缩小回归系数来减少预测误差,以缓解过度拟合,但它不能精确地将系数压缩为0,不进行协变量选择,除非调节参数$\lambda$趋向于$\infty$。Tibshirani[7]提出的最小绝对收缩和选择算子(Least Absolute Shrinkage and Selection Operator,Lasso)方法正好克服了这一点,当调节参数$\lambda$足够大时就会迫使一些系数估计为0,从而实现变量选择和系数估计[8-9]。Lasso 在生物信息学和经济学方面应用广泛[10-12],但是对于特征数大于样本个数的数据集,Lasso 方法最多可以选择$n$个变量。Zou 等[13]将岭回归与 Lasso 的惩罚函数结合提出带有 L1 范数与 L2 范数惩罚项的弹性网(Elastic net)方法,该方法可以用来解决变量个数大于样本个数的情况。Zou[14]还在 Lasso 的基础上提出了自适应 Lasso 方法(Adaptive Lasso,A-Lasso),该方法在 L1 范数惩罚下引入数据自适应的权重,比 Lasso 算法具有更准确的变量选择和模型预测能力。

针对回归模型中随机先验限制问题,Theil[15]提出了混合估计方法(Mixed Estimator,ME)。该方法将 OLS 和 GLS 相结合,同时使用样本的信息和额外的先验信息(如方差比等)进行估计。通过这种方式,ME 方法可以提高估计的准确性和稳定性,特别是在样本容量较小的情况下,更能体现出优势。但当数据特征大于样本量的时候,传统的解决随机限制的 ME 模型并不能得到很好的效果。为了解决这一问题,Guler 等[16]将 ME 与 Lasso 进行结合,提出了混合 Lasso(Mixed-Lasso,M-Lasso)估计。该方法很好地解决了随机限制在数据特征大于样本量的数据集下 ME 不能准确估计的问题。

但由于 M-Lasso 基于的 Lasso 方法对每一个系数给与的惩罚权重是等同的,这就不可避免的会导致对于重要变量的惩罚过重,对于不重要变量惩罚过轻,因而这可能就会导致模型的性能下降。为此,本文将 M-Lasso 基于的 Lasso 方法改进为 A-Lasso 方法。通过引入 A-Lasso 方法,可以更准确地为不同系数分配不同的惩罚权重,从而平衡对重要变量和不重要变量的惩罚程度,并提升模型的性能和解释能力。另外,本文还通过仿真实验验证了在稀疏模型上,与其他方法相比,该方法具有更小的均方误差。最后本文还通过实际应用,进一步验证了该方法在稀疏模型上的有效性。

## 1 模型及Ma-Lasso方法

考虑如下广义线性模型:

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

其中$\boldsymbol{y}^*=\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{y}$,$\boldsymbol{X}^*=\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}$,$\boldsymbol{\varepsilon}^*=\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{\varepsilon}$。$\boldsymbol{y}$是$n\times 1$维的可观测的向量,$\boldsymbol{X}$是$n\times p$维列满秩矩阵,$\boldsymbol{\beta}$是$p\times 1$维未知参数向量,$\boldsymbol{\varepsilon}$是$n\times 1$维误差向量,其中$\boldsymbol{\varepsilon}$服从均值为0,协方差为$\sigma^2\boldsymbol{V}$的正态分布,$\boldsymbol{V}$是给定的正定矩阵。将 OLS 应用于方程(1)中,得到广义最小二乘估计[17],具体公式如下所示:

$$\hat{\boldsymbol{\beta}}^{\mathrm{GLS}}=\arg\min_{\beta}\mathrm{RSS}=\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{y}=\boldsymbol{S}^{-1}\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{y}^*, \tag{2}$$

其中残差平方和$\mathrm{RSS}=\dfrac{1}{2}\sum_{i=1}^{n}\left(y_i^*-\sum_{j=1}^{p}x_{ij}^*\beta_j\right)^2$,并且$\boldsymbol{S}=\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)$。

若存在对$\boldsymbol{\beta}$的先验线性限制如下所示：

$$\boldsymbol{r}=\boldsymbol{R\beta}, \tag{3}$$

其中$\boldsymbol{R}$是一个$m\times p$维先验矩阵，$\boldsymbol{r}$是已知的$m\times 1$维常数向量，并且$\boldsymbol{R}$的秩$\mathrm{rank}(\boldsymbol{R})=m\leqslant p$。在公式(3)中给出的限制下，得到RLS模型估计如下：

$$\hat{\boldsymbol{\beta}}^{\mathrm{RLS}}=\hat{\boldsymbol{\beta}}^{\mathrm{GLS}}-\boldsymbol{S}^{-1}\boldsymbol{R}^{\mathrm{T}}\left\{\boldsymbol{RS}^{-1}\boldsymbol{R}^{\mathrm{T}}\right\}^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\mathrm{GLS}}-\boldsymbol{r}\right)。 \tag{4}$$

然而得到的先验信息可能具有一些随机性，这些随机性来自先前的统计分析或不确定理论所获得的先验估计[18]。先验信息中的不确定性模型如下：

$$\boldsymbol{r}=\boldsymbol{R\beta}+\boldsymbol{\phi}, \tag{5}$$

其中$\boldsymbol{\phi}$是$m\times 1$维误差向量，其均值为0，协方差为$\sigma^2\boldsymbol{W}$，$\boldsymbol{W}$是已知的正定矩阵。

令$\boldsymbol{y}_A=\left(\boldsymbol{y}^{\mathrm{T}}\quad \boldsymbol{r}^{\mathrm{T}}\right)^{\mathrm{T}}$，$\boldsymbol{X}_A=\left(\boldsymbol{X}^{\mathrm{T}}\quad \boldsymbol{R}^{\mathrm{T}}\right)^{\mathrm{T}}$，$\boldsymbol{u}=\left(\boldsymbol{\epsilon}^{\mathrm{T}}\quad \boldsymbol{\phi}^{\mathrm{T}}\right)^{\mathrm{T}}$，那么线性模型就可写成如下公式：

$$\boldsymbol{y}_A=\boldsymbol{X}_A\boldsymbol{\beta}+\boldsymbol{u}, \tag{6}$$

令$\boldsymbol{T}=\begin{pmatrix}\boldsymbol{V} & \boldsymbol{0}\\ \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{W}\end{pmatrix}$是已知的正定矩阵，将GLS应用到上式中，得到ME估计，估计式如下[18]：

$$\hat{\boldsymbol{\beta}}^{\mathrm{ME}}=\left(\boldsymbol{X}_A^{\mathrm{T}}\boldsymbol{T}^{-1}\boldsymbol{X}_A\right)^{-1}\boldsymbol{X}_A^{\mathrm{T}}\boldsymbol{T}^{-1}\boldsymbol{y}_A=\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}+\boldsymbol{R}^{\mathrm{T}}\boldsymbol{W}^{-1}\boldsymbol{R}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{y}+\boldsymbol{R}^{\mathrm{T}}\boldsymbol{W}^{-1}\boldsymbol{r}\right)。 \tag{7}$$

然而上述方法并不能用于稀疏模型，Lasso估计可用于估计稀疏模型，具体估计式如下：

$$\hat{\boldsymbol{\beta}}^{\mathrm{Lasso}}(\lambda)=\arg\min_{\beta}\left\{\mathrm{RSS}+\lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\}, \tag{8}$$

其中$\lambda$是一个非负的调优参数，用于控制回归系数压缩的程度，数值越大则惩罚力度越强。参数$\lambda$使用贝叶斯信息准则(Bayesian Information Criterion，BIC)方法进行选择，具体公式如下：

$$\mathrm{BIC}(\lambda)=\log\left(\hat{\sigma}^2\right)+\widehat{df}\frac{\log(n)}{n}, \tag{9}$$

其中$\hat{\sigma}^2=\dfrac{\mathrm{RSS}}{n}$，$\widehat{df}$是非零系数的个数，$\lambda=2^{k-15}$，其中$k=1,2,3,\cdots,20$。通过给出$\lambda$计算BIC值，取值最小的BIC值所对应的调优参数的值即为模型的最优$\lambda$。

Zou[14]通过在$\sum_{j=1}^{p}\left|\beta_j\right|$惩罚项前增加权重，对Lasso进行改进，提出自适应Lasso方法(Adaptive-Lasso)[14]。在该方法中，惩罚项变成了$\sum_{j=1}^{p}\hat{\boldsymbol{w}}\left|\beta_j\right|$，其中$\hat{\boldsymbol{w}}=\dfrac{1}{\left|\beta_j\right|^{\gamma}}$，$\gamma$是一个任意的非负数。公式如下：

$$\hat{\boldsymbol{\beta}}^{\mathrm{A-Lasso}}(\lambda)=\arg\min_{\beta}\left\{\mathrm{RSS}+\lambda\sum_{j=1}^{p}\hat{\boldsymbol{w}}\left|\beta_j\right|\right\}。 \tag{10}$$

采用A-Lasso进行变量选择，能够使得对响应变量重要的变量更易进入模型，而对其不重要的变量更易被剔除，并且在更好地实现变量选择的同时也能够有效减小系数估计的偏差。因此相比于Lasso方法，自适应Lasso方法也更适用于观测指标数$p$和样本量$N$的比值非常大的情况。

将Lasso模型中加入先验信息，可以提高估计的准确性和稳定性。Guler等提出M-Lasso[16]，可用于具有随机限制的大数据集。他将ME和Lasso结合，以希望在选择正确模型的同时减少均方误差(Mean Squared Error，MSE)。M-Lasso模型如下所示：

$$\hat{\boldsymbol{\beta}}^{\mathrm{M-Lasso}}(\lambda)=\arg\min_{\beta}\left\{\frac{1}{2}\boldsymbol{u}^{*\mathrm{T}}\boldsymbol{u}^{*}+\lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\}, \tag{11}$$

其中$\boldsymbol{u}^{*}=\boldsymbol{T}^{-\frac{1}{2}}\boldsymbol{u}=\left(\boldsymbol{\epsilon}^{*\mathrm{T}}\quad \boldsymbol{\phi}^{*\mathrm{T}}\right)^{\mathrm{T}}$，$\boldsymbol{T}^{-\frac{1}{2}}=\begin{pmatrix}\boldsymbol{V}^{-\frac{1}{2}} & \boldsymbol{0}\\ \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{W}^{-\frac{1}{2}}\end{pmatrix}$，$\boldsymbol{u}=\left(\boldsymbol{\epsilon}^{\mathrm{T}}\quad \boldsymbol{\phi}^{\mathrm{T}}\right)^{\mathrm{T}}$。

A-Lasso 有着比 Lasso 更为精确的模型选择能力,且具有 oracle 性质。受 M-Lasso 的启发,本文将 ME 与 A-Lasso 结合,提出混合自适应 Lasso 估计(Ma-Lasso),希望其在具备随机限制的大数据集上有着比 M-Lasso 更为优秀的表现。为了得到 Ma-Lasso 估计,本文对线性模型进行变换:

$$\boldsymbol{y}_{A^*}=\boldsymbol{X}_{A^*}\boldsymbol{\beta}+\boldsymbol{u}^*, \tag{12}$$

其中 $\boldsymbol{y}_{A^*}=\boldsymbol{T}^{-\frac{1}{2}}\boldsymbol{y}_A=\left(\boldsymbol{y}^{*\mathrm{T}}\quad \boldsymbol{r}^{*\mathrm{T}}\right)^{\mathrm{T}}$, $\boldsymbol{X}_{A^*}=\boldsymbol{T}^{-\frac{1}{2}}\boldsymbol{X}_A=\left(\boldsymbol{X}^{*\mathrm{T}}\quad \boldsymbol{R}^{*\mathrm{T}}\right)^{\mathrm{T}}$。将变换后的 $\boldsymbol{y}_{A^*}$, $\boldsymbol{X}_{A^*}$, $\boldsymbol{u}^*$ 代入 A-Lasso 模型中,得到如下估计:

$$\hat{\boldsymbol{\beta}}^{\mathrm{Ma-Lasso}}(\lambda)=\arg\min_{\beta}\left\{\mathrm{RSS}^*+\lambda\sum_{j=1}^{p}\hat{\boldsymbol{w}}\left|\beta_j\right|\right\}, \tag{13}$$

其中 $\mathrm{RSS}^*=\dfrac{1}{2}\sum_{i=1}^{n}\left(y_{A^*i}-\sum_{j=1}^{p}x_{A^*ij}\beta_j\right)^2$。其求解算法如下所示。

| 算法　求解 Ma-Lasso 估计 |
|---|
| 1. 由 $\boldsymbol{r}=\boldsymbol{R}\beta+\boldsymbol{\phi}$,得到随机限制 $\boldsymbol{r}$,其中 $\boldsymbol{\phi}\sim N\left(\boldsymbol{0},\sigma^2\boldsymbol{W}\right)$; |
| 2. 将 $\boldsymbol{y},\boldsymbol{X}$ 进行变换:$\boldsymbol{X}_{A^*}=\boldsymbol{T}^{-\frac{1}{2}}\boldsymbol{X}_A=\left(\boldsymbol{X}^{*\mathrm{T}}\quad \boldsymbol{R}^{*\mathrm{T}}\right)^{\mathrm{T}}$, $\boldsymbol{y}_{A^*}=\boldsymbol{T}^{-\frac{1}{2}}\boldsymbol{y}_A=\left(\boldsymbol{y}^{*\mathrm{T}}\quad \boldsymbol{r}^{*\mathrm{T}}\right)^{\mathrm{T}}$; |
| 3. 求解最优调优参数 $\lambda$: |
| a) 由 $\lambda=2^{k-15}, k=1,2,3,\cdots,20$ 得到 $\lambda$ 的集合, |
| b) 通过公式:$\mathrm{BIC}(\lambda)=\log\left(\hat{\sigma}^2\right)+\widehat{df}\dfrac{\log(n)}{n}$,计算每一个 $\lambda$ 对应的 BIC 值, |
| c) 选择 BIC 值最小时对应的 $\lambda$ 为最优调优参数 $\lambda$。 |
| 4. 计算 Ma-Lasso 的系数值。将 $\boldsymbol{X}_{A^*}$, $\boldsymbol{y}_{A^*}$,以及最优调优参数 $\lambda$ 代入下面模型, $$\hat{\boldsymbol{\beta}}^{\mathrm{Ma-Lasso}}(\lambda)=\arg\min_{\beta}\left\{\mathrm{RSS}^*+\lambda\sum_{j=1}^{p}\hat{\boldsymbol{w}}\left\|\beta_j\right\|\right\}。$$ |
| 5. 输出 $\hat{\boldsymbol{\beta}}^{\mathrm{Ma-Lasso}}$。 |

## 2　仿真实验

本节通过蒙特卡罗实验来比较上述所有方法的 MSE。使用 Tibshirani[7] 提到的仿真数据进行了四组实验,如下所示:

实验一:在这组实验中,生成样本量 $n=20$ 且 $\boldsymbol{\beta}=(3,1.5,0,0,2,0,0,0)^{\mathrm{T}}$ 的数据。预测因子 $\boldsymbol{X}$ 服从均值为 $\boldsymbol{0}$,协方差为 $\boldsymbol{\Sigma}$ 的多元正态分布。$\boldsymbol{\Sigma}$ 是对角元素为 1,非对角元素为 $0.5^{|i-j|}$ 的矩阵。误差项是由均值为 0,标准差为 3 的正态分布生成的。

实验二:对于每一个 $j$,有 $\beta_j=0.85$,其余参数与实验一相同。

实验三:$\beta_1=5$, $\beta_j=0\ \left(j=2,3,\cdots,8\right)$,误差项是由均值为 0,标准差为 2 的正态分布生成的,其余参数与实验一相同。

实验四:样本量为 $n=100$ 且 $\boldsymbol{\beta}=\left(\underbrace{0,\cdots,0}_{10},\underbrace{2,\cdots,2}_{10},\underbrace{0,\cdots,0}_{10},\underbrace{2,\cdots,2}_{10}\right)^{\mathrm{T}}$ 预测因子 $\boldsymbol{X}$ 由 $x_{ij}=z_{ij}+z_i$ 生成,其中 $z_{ij}$ 和 $z_i$ 是独立的标准正态分布变量。误差项是由均值为 0,标准差为 15 的正态分布生成。

误差项由两个不同过程产生:(1)White Noise(WN)误差:$\boldsymbol{\varepsilon}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{V})$, $\boldsymbol{V}=\boldsymbol{I}_n$。(2)一阶自回归模型(AR(1))的误差:$\varepsilon_i=0.5\varepsilon_{i-1}+e_i$,其中 $e_i$ 是由均值为 0,方差为 $\sigma_e^2=\sigma^2\left(1-0.5^2\right)$ 的正态分布生成的。即 $\boldsymbol{\varepsilon}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{V})$,其中 $\boldsymbol{V}=\left(0.5^{|i-j|}\right)$。

随机限制

$$\boldsymbol{r}=\boldsymbol{R}\boldsymbol{\beta}+\boldsymbol{\phi}; \tag{14}$$

$$\boldsymbol{\phi} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{W}\right), \tag{15}$$

其中$\boldsymbol{\phi}$的方差有两种：

(i)$\sigma^2\boldsymbol{W}=0.1\boldsymbol{I}_m$(Loose 先验信息)，

(ii)$\sigma^2\boldsymbol{W}=0.05\boldsymbol{I}_m$(Tight 先验信息)。

限制矩阵$\boldsymbol{R}$有两种：

(1) 当$p=8$时，$\boldsymbol{R}=\begin{bmatrix}(1 \quad 0 \quad 0 \quad 0) & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & (1 \quad 0 \quad 0 \quad 0)\end{bmatrix}_{2\times 8}$；

当$p=40$时，$\boldsymbol{R}=\begin{bmatrix}(1 \quad 0 \quad 0 \quad 0) & \cdots & \boldsymbol{0}^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}^{\mathrm{T}} & \cdots & (1 \quad 0 \quad 0 \quad 0)\end{bmatrix}_{10\times 40}$。

(2) 当$p=8$时，$\boldsymbol{R}=\begin{bmatrix}(1 \quad 0) & \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & (1 \quad 0) & \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} & (1 \quad 0) & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} & (1 \quad 0)\end{bmatrix}_{4\times 8}$；

当$p=40$时，$\boldsymbol{R}=\begin{bmatrix}(1 \quad 0) & \boldsymbol{0}^{\mathrm{T}} & \cdots & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & (1 \quad 0) & \cdots & \boldsymbol{0}^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} & \cdots & (1 \quad 0)\end{bmatrix}_{20\times 40}$。

其中第一种方案中$m=\frac{p}{4}$，第二种方案中$m=\frac{p}{2}$。

模拟过程中，参数$\boldsymbol{\beta}$，$\boldsymbol{X}$，$\boldsymbol{R}$在仿真前设定。在每组实验中，$\boldsymbol{\epsilon}$和$\boldsymbol{\phi}$由上述方案生成，$\boldsymbol{y}$和$\boldsymbol{r}$由公式(1)和(14)生成。通过GLS，RLS，ME，Lasso，A-Lasso，M-Lasso，Ma-Lasso方法进行参数估计。

本文通过均方误差(Mean Squared Error，MSE)，发现百分比(Discovery Percentage，DP)，真实发现率(True Discovery Percentage，TDP)，选择真实模型的次数比例(Proportion of Times that the True model is Selected，PTTS)作为评价指标[1]。公式如下：

$$\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}\right)=\frac{1}{10\,000}\sum_{H=1}^{10\,000}\left(\hat{\boldsymbol{\beta}}_H-\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\hat{\boldsymbol{\beta}}_H-\boldsymbol{\beta}\right), \tag{16}$$

$$\mathrm{DP}\left(\hat{\boldsymbol{\beta}}\right)=\frac{1}{10\,000}\sum_{H=1}^{10\,000}\frac{\#\left(\hat{\beta}_{H,j}=0/\beta_j=0\right)}{\#\left(\beta_j=0\right)}100, \tag{17}$$

$$\mathrm{TDP}\left(\hat{\boldsymbol{\beta}}\right)=\frac{1}{10\,000}\sum_{H=1}^{10\,000}\frac{\#\left(\hat{\beta}_{H,j}\neq 0/\beta_j\neq 0\right)}{\#\left(\beta_j\neq 0\right)}100, \tag{18}$$

其中$\hat{\boldsymbol{\beta}}_H$是第$H$次重复试验得到的估计值，$\#(\cdot)$是满足条件的次数。DP和TDP用于测量稀疏(或非稀疏)系数被正确估计的实际百分比。PTTS是当实际参数为零时，估计的模型参数为零，当实际参数不为零时，其不等于零的次数百分比，即满足所有零估计为零，所有非零估计为非零的次数百分比。

本文使用R语言(版本4.2.1)进行编程模拟，实验重复次数为10 000，研究结果如表1—表8所示。

表1—表8显示了不同方法在不同先验限制下的各种性能的结果。

表1和表2的结果对比表明，在一阶自回归模型(AR(1))的误差下的MSE明显小于White Noise(WN)误差下的MSE。Ma-Lasso在MSE方面展现出了显著的优势。其MSE不仅比M-Lasso小0.2，而且比其他方法减少了30%甚至更多。这表明在处理某些类型的数据时，Ma-Lasso具有更

高的预测精度和稳定性。此外,从DP的结果来看,虽然Ma-Lasso不是最优的,但它在正确估计零系数的能力上与M-Lasso相差不大,并且明显优于其他方法。这进一步证明了Ma-Lasso在变量选择上的有效性。另外,从PTTS的结果上看,Ma-Lasso在带白噪声误差的随机限制下,其性能略低于M-Lasso,而在带有AR(1)误差的随机限制下,其性能与M-Lasso的性能相同。

**表1** 实验一[$n$=20, $p$=8, White Noise (WN)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 1** The results of various performance indicators obtained in experiment one [$n$=20, $p$=8, White Noise (WN) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 12.453 | — | — | — | 9.825 | — | — | — |
| RLS | 2 | 7.583 | — | — | — | 4.570 | — | — | — |
| RLS | 4 | 3.310 | — | — | — | 2.256 | — | — | — |
| ME | 2 | 7.531 | — | — | — | 4.560 | — | — | — |
| ME | 4 | 3.261 | — | — | — | 2.236 | — | — | — |
| Lasso | — | 7.415 | 56.11 | 90.34 | 8.00 | 4.708 | 64.79 | 95.41 | 25 |
| M-Lasso | 2 | 2.973 | 69.28 | 93.83 | 31.00 | 1.926 | 74.59 | **97.98** | 32 |
| M-Lasso | 4 | 1.945 | **82.70** | **94.12** | **44** | 1.419 | **84.16** | 95.87 | **33** |
| A-Lasso | — | 7.544 | 72.31 | 86.20 | 14 | 6.305 | 70.47 | 86.37 | 13 |
| Ma-Lasso | 2 | 2.947 | 78.67 | 87.69 | 22 | 1.672 | 72.20 | 94.35 | 18 |
| Ma-Lasso | 4 | **1.768** | 79.97 | 89.55 | 29 | **1.216** | 76.60 | 94.08 | 23 |

注:粗体表示在给定评估标准下的最佳估计量;DP表示零系数被正确估计的百分比;TDP表示非零系数被正确估计的百分比;PTTS表示所有非零系数和零系数被正确估计的百分比。在这组实验中,生成样本量$n=20$且$\boldsymbol{\beta}=(3,1.5,0,0,2,0,0,0)^{\mathrm{T}}$的数据。预测因子$\boldsymbol{X}$服从均值为0,协方差为$\boldsymbol{\Sigma}$的多元正态分布。$\boldsymbol{\Sigma}$是对角元素为1,非对角元素为$0.5^{|i-j|}$的矩阵。误差项是由均值为0,标准差为3的正态分布生成的。表中“-”表示此项未测量。表中“-”表示此项未测量。下同。

**表2** 实验一[$n$=20, $p$=8, AR(1)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 2** The results of various performance indicators obtained in experiment one [$n$=20, $p$=8, AR (1) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP(%) | PTTS |
| GLS | — | 12.525 | — | — | — | 7.376 | — | — | — |
| RLS | 2 | 4.368 | — | — | — | 5.136 | — | — | — |
| RLS | 4 | 2.574 | — | — | — | 1.872 | — | — | — |
| ME | 2 | 4.334 | — | — | — | 5.118 | — | — | — |
| ME | 4 | 2.494 | — | — | — | 1.860 | — | — | — |
| Lasso | — | 7.376 | 38.11 | **97.12** | 6 | 5.950 | 48.65 | 92.24 | 7.00 |
| M-Lasso | 2 | 1.732 | 82.64 | 94.95 | 31 | 2.105 | 80.88 | 90.93 | 31 |
| M-Lasso | 4 | 1.520 | **93.12** | 95.39 | **36** | 1.375 | **88.67** | **92.87** | **35** |
| A-Lasso | — | 6.855 | 74.75 | 83.45 | 15 | 4.557 | 69.43 | 74.75 | 9 |
| Ma-Lasso | 2 | 1.556 | 77.17 | 90.92 | 23 | 2.176 | 75.01 | 88.37 | 16 |
| Ma-Lasso | 4 | **1.395** | 83.21 | 90.82 | **36** | **0.998** | 85.58 | 90.26 | 29 |

表3和表4的结果对比表明,在非稀疏模型下,一阶自回归模型的误差下的MSE大于White Noise误差下的MSE。在非稀疏数据或特定条件下,M-Lasso的MSE表现最优,而Ma-Lasso虽然不是最优的,但在除M-Lasso外的其他方法中表现最好。同时,TDP的结果也显示Ma-Lasso在某些条件下表现最优,但在PTTS的结果中,M-Lasso再次优于Ma-Lasso。

表5和表6的结果对比与表1和表2的结果一致,进一步验证了稀疏模型下,一阶自回归模型的误差下的MSE要小于White Noise误差下的MSE。表5和表6的共同结果再次验证了Ma-Lasso

在某些场景下的优势。其MSE比M-Lasso小0.1,并且远小于其他方法。这表明在处理稀疏数据时,Ma-Lasso仍然具有显著的优势。然而,从PTTS的结果来看,无论是带有白噪声误差还是AR(1)误差的随机限制条件下,Ma-Lasso都不再是最优的方法。

**表3**　实验二[$n$=20, $p$=8, White Noise(WN)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 3**　The results of various performance indicators obtained in experiment two [$n$=20, $p$=8, White Noise (WN) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 8.083 | — | — | — | 9.628 | — | — | — |
| RLS | 2 | 4.943 | — | — | — | 6.696 | — | — | — |
| RLS | 4 | 3.796 | — | — | — | 3.278 | — | — | — |
| ME | 2 | 4.913 | — | — | — | 6.699 | — | — | — |
| ME | 4 | 3.755 | — | — | — | 3.231 | — | — | — |
| Lasso | — | 5.536 | — | 72.68 | 11 | 5.836 | — | 70.28 | 9 |
| M-Lasso | 2 | 3.839 | — | 74.26 | 16 | 4.384 | — | 68.76 | 16 |
| M-Lasso | 4 | **3.200** | — | 73.54 | **24** | **2.925** | — | 72.91 | **24** |
| A-Lasso | — | 6.692 | — | 51.36 | 1 | 7.511 | — | 48.98 | 0 |
| Ma-Lasso | 2 | 5.122 | — | 68.58 | 1 | 5.453 | — | 70.53 | 3 |
| Ma-Lasso | 4 | 3.775 | — | **76.56** | 8 | 3.551 | — | **83.51** | 18 |

注:在这组实验中,对于每一个$j$,有$\beta_j=0.85$,其余参数与实验一相同。

**表4**　实验二[$n$=20, $p$=8, AR (1)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 4**　The results of various performance indicators obtained in experiment two [$n$=20, $p$=8, AR (1) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 11.394 | — | — | — | 6.364 | — | — | — |
| RLS | 2 | 8.121 | — | — | — | 3.013 | — | — | — |
| RLS | 4 | 4.400 | — | — | — | 1.975 | — | — | — |
| ME | 2 | 8.025 | — | — | — | 3.011 | — | — | — |
| ME | 4 | 4.314 | — | — | — | 1.951 | — | — | — |
| Lasso | — | 7.219 | — | 76.74 | **30** | 5.523 | — | **85.61** | **42** |
| M-Lasso | 2 | 4.777 | — | 71.98 | 7 | 2.615 | — | 79.61 | 34 |
| M-Lasso | 4 | **3.598** | — | 65.53 | 7 | **1.949** | — | 83.66 | 39 |
| A-Lasso | — | 7.217 | — | 55.8 | 0 | 7.119 | — | 58.28 | 0 |
| Ma-Lasso | 2 | 6.371 | — | 72.71 | 0 | 4.117 | — | 73.11 | 3 |
| Ma-Lasso | 4 | 5.132 | — | **85.09** | 22 | 2.939 | — | 85.45 | 15 |

注:在这组实验中,对于每一个$j$,有$\beta_j=0.85$,其余参数与实验一相同。

表7和表8的结果对比表明,在高维数据下,一阶自回归模型的误差下的MSE也大于White Noise误差下的MSE。根据表7和表8的共同结果,M-Lasso在MSE方面再次表现出最优的性能,而Ma-Lasso与其他方法相比存在一定的差距。从TDP的结果来看,Ma-Lasso仍然是最好的方法之一,但在PTTS的结果中,M-Lasso再次优于了Ma-Lasso。

综上所述,在系数稀疏的情况中,Ma-Lasso的效果比M-Lasso的效果好,而其在系数不是稀疏的情况中的效果没有M-Lasso好。这表明在处理具有稀疏解并且带有随机限制的线性模型时,Ma-Lasso比M-Lasso的效果好。并且由于数据的生成具有随机性,MSE的结果虽会在一定范围内波动,但总体来说Ma-Lasso在稀疏模型上的MSE优于其他模型。从DP的结果上来看,Ma-Lasso优于除M-Lasso以外的其他方法。这表明Ma-Lasso在发现真实的0系数能力上与M-Lasso相差无

几,并且优于其他的估计量。从PTTS的结果上看,M-Lasso的结果最优,Ma-Lasso的结果优势不明显但总体还是优于其他方法或与其他方法相差不大。因此,在实际应用中,我们需要根据具体的数据特征和问题背景来选择合适的模型和方法。

**表5** 实验三[$n$=20, $p$=8, White Noise (WN)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 5** The results of various performance indicators obtained in experiment three [$n$=20, $p$=8, White Noise (WN) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 3.649 | — | — | — | 3.947 | — | — | — |
| RLS | 2 | 2.658 | — | — | — | 2.509 | — | — | — |
| RLS | 4 | 1.785 | — | — | — | 1.307 | — | — | — |
| ME | 2 | 2.539 | — | — | — | 2.485 | — | — | — |
| ME | 4 | 1.616 | — | — | — | 1.272 | — | — | — |
| Lasso | — | 1.250 | 78.09 | **100** | 36 | 1.570 | 81.77 | **100** | 30 |
| M-Lasso | 2 | 0.639 | 94.12 | **100** | 73 | 0.538 | 95.42 | **100** | 79 |
| M-Lasso | 4 | 0.423 | **96.18** | **100** | **77** | 0.313 | **98.58** | **100** | **90** |
| A-Lasso | — | 0.744 | 80.26 | **100** | 29 | 1.029 | 82.38 | **100** | 39 |
| Ma-Lasso | 2 | 0.495 | 77.03 | **100** | 24 | 0.560 | 79.88 | **100** | 29 |
| Ma-Lasso | 4 | **0.332** | 78.26 | **100** | 29 | **0.244** | 83.47 | **100** | 35 |

注:在这组实验中,$\beta_1=5,\beta_j=0(j=2,3,\cdots,8)$误差项是由均值为0,标准差为2的正态分布生成的,其余参数与实验一同。

**表6** 实验三[$n$=20, $p$=8, AR(1)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 6** The results of various performance indicators obtained in experiment three [$n$=20, $p$=8, AR (1) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 3.387 | — | — | — | 4.052 | — | — | — |
| RLS | 2 | 2.687 | — | — | — | 2.678 | — | — | — |
| RLS | 4 | 1.870 | — | — | — | 1.013 | — | — | — |
| ME | 2 | 2.575 | — | — | — | 2.667 | — | — | — |
| ME | 4 | 1.654 | — | — | — | 0.979 | — | — | — |
| Lasso | — | 1.398 | 64.35 | **100** | 14 | 1.566 | 60.51 | **100** | 17 |
| M-Lasso | 2 | 0.364 | 95.74 | **100** | 77 | 0.474 | 95.04 | **100** | 80 |
| M-Lasso | 4 | 0.399 | **97.41** | **100** | **81** | 0.338 | **97.09** | **100** | **89** |
| A-Lasso | — | 0.508 | 80.72 | **100** | 34 | 0.660 | 78.77 | **100** | 28 |
| Ma-Lasso | 2 | 0.415 | 78.71 | **100** | 28 | 0.383 | 83.86 | **100** | 38 |
| Ma-Lasso | 4 | **0.228** | 81.60 | **100** | 29 | **0.282** | 83.37 | **100** | 35 |

注:在这组实验中,$\beta_1=5,\beta_j=0(j=2,3,\cdots,8)$误差项是由均值为0,标准差为2的正态分布生成的,其余参数与实验一同。

## 3 实际应用

我们将新方法应用于选择影响股票价格特征的数据集中[19],该数据收集了从2009年6月30日到2021年6月30日的贵州茅台各季度财报数据和股票价格数据。在主要指标中选取了营业收入($X_1$)、营业成本($X_2$)、营业利润($X_3$)、利润总额($X_4$)、所得税费用($X_5$)、净利润($X_6$)、每股收益($X_7$)、货币资金($X_8$)、存货($X_9$)、流动资产合计($X_{10}$)、固定资产净额($X_{11}$)、资产总计($X_{12}$)、流动负债合计($X_{13}$)、负债合计($X_{14}$)、所有者权益合计($X_{15}$)、期初现金及现金等价物余额($X_{16}$)、期末现金及现金等价物余额($X_{17}$)共计17个特征作为影响股票价格的截面数据特征。另外经过对数据进行探索性研究后,发现股票价格呈现一种指数型增长的趋势,而其他特征的增长模式为一般的线性

增长,因此对股票价格$p$取对数。其所构建的模型如下式:

$$\ln(p)=\beta_1X_1+\beta_2X_2+\beta_3X_3+\cdots+\beta_{17}X_{17}+\varepsilon。\tag{19}$$

**表7** 实验四[$n$=100, $p$=40, White Noise(WN)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 7** The results of various performance indicators obtained in experiment four [$n$=100, $p$=40, White Noise (WN) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information (Loose先验信息) | | | | Tight prior information (Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 159.046 | — | — | — | 154.275 | — | — | — |
| RLS | 2 | 99.499 | — | — | — | 90.757 | — | — | — |
| RLS | 4 | 56.789 | — | — | — | 52.266 | — | — | — |
| ME | 2 | 99.512 | — | — | — | 90.808 | — | — | — |
| ME | 4 | 56.707 | — | — | — | 52.274 | — | — | — |
| Lasso | — | 54.865 | 71.17 | 79.66 | 0 | 57.404 | 64.77 | 77.96 | 0 |
| M-Lasso | 2 | 42.365 | 76.13 | 83.47 | 0 | 42.475 | 71.19 | 85.80 | 0 |
| M-Lasso | 4 | **27.887** | 78.14 | **92.53** | 0 | **28.454** | 76.66 | **92.26** | 0 |
| A-Lasso | — | 125.848 | 82.23 | 51.11 | 0 | 126.143 | 82.23 | 49.27 | 0 |
| Ma-Lasso | 2 | 92.442 | 84.52 | 63.98 | 0 | 89.750 | 85.26 | 65.90 | 0 |
| Ma-Lasso | 4 | 58.937 | **84.92** | 79.44 | 0 | 56.922 | **88.08** | 81.47 | 0 |

注:在本组实验中,样本量为$n=100$,预测因子$\boldsymbol{X}$由$x_{ij}=z_{ij}+z_i$生成,其中$z_{ij}$和$z_i$是独立的标准正态分布变量。误差项是由均值为0,标准差为15的正态分布生成。

**表8** 实验四[$n$=100, $p$=40, AR(1)误差]在不同方法和不同先验限制下的各种性能指标的结果

**Table 8** The results of various performance indicators obtained in experiment four [$n$=100, $p$=40, AR (1) Error] using different methods and prior constraints

| 模型估计 | $m$ | Loose prior information(Loose先验信息) | | | | Tight prior information(Tight 先验信息) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | DP/% | TDP/% | PTTS | MSE | DP/% | TDP/% | PTTS |
| GLS | — | 101.493 | — | — | — | 115.051 | — | — | — |
| RLS | 2 | 65.037 | — | — | — | 67.700 | — | — | — |
| RLS | 4 | 40.836 | — | — | — | 37.241 | — | — | — |
| ME | 2 | 64.944 | — | — | — | 67.667 | — | — | — |
| ME | 4 | 40.598 | — | — | — | 37.286 | — | — | — |
| Lasso | — | 57.143 | 61.207 | 78.61 | 0 | 57.883 | 59.75 | 78.86 | 0 |
| M-Lasso | 2 | 39.776 | 66.38 | 90.59 | 1 | 32.317 | 70.25 | 89.25 | 0 |
| M-Lasso | 4 | **28.755** | 70.64 | **95.51** | **4** | **20.208** | 76.07 | **94.74** | 0 |
| A-Lasso | — | 96.101 | 80.91 | 20.01 | 0 | 98.294 | 82.65 | 16.89 | 0 |
| Ma-Lasso | 2 | 91.794 | 87.51 | 15.41 | 0 | 92.000 | 85.04 | 15.83 | 0 |
| Ma-Lasso | 4 | 84.202 | **89.73** | 10.16 | 0 | 85.300 | **86.22** | 14.24 | 0 |

注:在本组实验中,样本量为$n=100$,预测因子$\boldsymbol{X}$由$x_{ij}=z_{ij}+z_i$生成,其中$z_{ij}$和$z_i$是独立的标准正态分布变量。误差项是由均值为0,标准差为15的正态分布生成。

由于数据集没有先验限制,因此本文使用GLS和Lasso的估计值来形成这种先验限制。该限制为$\beta_4-\beta_1-\beta_6+\beta_{14}=1$,$\beta_1+\beta_4-\beta_8+\beta_{13}+\beta_{17}=0.5$。

衡量模型选择功效的BIC准则计算公式如下:

$$\mathrm{BIC}=k\ln(n)-2\ln(L),\tag{20}$$

其中$k$为模型参数的数量,$n$为样本个数,似然函数$L$如下式:

$$L=(\boldsymbol{Y}^*-\boldsymbol{X}^*\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{Y}^*-\boldsymbol{X}^*\boldsymbol{\beta})。\tag{21}$$

根据表9的结果,各个参数估计值的符号是相似的,这表明以上方法在对该数据集进行建模时,

对于参数的影响方向是一致的。另外通过估计值的结果发现，GLS，RLS，ME 的估计值不是稀疏的，表明这些方法并不能很好地提供一个稀疏解。而从 BIC 的值的结果来看，Ma-Lasso 的 BIC 值是最小的，说明 Ma-Lasso 可以很好地应用于所构建的模型，用于选择合适的影响股票价格的特征。从经济学角度来看，通常我们期望利润总额与股票价格之间存在正相关关系，即利润越高，股票价格越高；而所得税费用与股票价格之间通常会存在负相关关系，即所得税费用越高，股票价格越低。然而，根据 GLS 等的结果，这两个关系与通常的预期相反。但是从 Ma-Lasso 的结果来看，营业利润、利润总额以及资产总计对于股票价格有正影响，在一般情况下，我们期望营业利润、利润总额以及资产总计对股票价格具有正相关的影响，即这些指标的增加通常会伴随着股票价格的上涨。这表明 Ma-Lasso 所选出来的影响股票价格的特征与经济理论预期相吻合。

**表9** 各模型的参数估计值及BIC值

**Table 9** Parameter estimates and BIC values of each model

| 参数 | GLS | RLS | ME | Lasso | M-Lasso | A-Lasso | Ma-Lasso |
|---|---|---|---|---|---|---|---|
| 营业收入/万元 | −0.708 | −0.649 | −0.648 | 0 | −0.148 | 0 | 0 |
| 营业成本/万元 | 0.395 | 0.296 | 0.296 | 0 | 0 | 0 | 0 |
| 营业利润/万元 | −2.689 | −4.488 | −4.489 | 0 | 0 | 0 | 0.081 |
| 利润总额/万元 | −27 327 | −16.54 | −16.510 | 0 | 0.632 | 0 | 0.087 |
| 所得税费用/万元 | 6 818.58 | 4.624 | 4.617 | 0 | 0 | 0 | 0 |
| 净利润/万元 | 2 0511 | 16.756 | 16.734 | 0 | 0 | 0 | −0.428 |
| 每股收益/万元 | 0.030 | 0.023 | 0.023 | 0 | 0.052 | 0 | 0 |
| 货币资金/万元 | 0.030 | 0.013 | 0.013 | 0 | 0 | 0 | 0 |
| 存货/万元 | 2.168 | 2.240 | 2.240 | 0.147 | 0 | 0 | 0 |
| 流动资产合计/万元 | 3.475 | 3.491 | 3.491 | 0.160 | 0 | 0 | 0 |
| 固定资产净额/万元 | −0.281 | −0.377 | −0.377 | 0.094 | 0.732 | 0 | 0 |
| 资产总计/万元 | −241.32 | −237.87 | −237.86 | 0 | 0 | 0.667 | 0.888 |
| 流动负债合计/万元 | 4.343 | 16.763 | 16.772 | 0.025 | 0 | 0 | 0 |
| 负债合计/万元 | 46.745 | 33.646 | 33.637 | 0 | 0.002 | 0 | 0.001 |
| 所有者权益合计/万元 | 200.320 | 197.462 | 197.461 | 0 | 0 | 0 | 0 |
| 期初现金及现金等价物余额/万元 | −0.063 | −0.011 | −0.011 | 0 | −0.137 | 0 | 0 |
| 期末现金及现金等价物余额/万元 | 1.036 | 0.940 | 0.940 | 0.210 | 0 | 0 | 0 |
| BIC | 69.064 | 68.925 | 68.924 | 62.152 | 65.329 | 62.159 | 62.045 |

## 4 结论

受 M-Lasso 的启发，本文将 ME 和 A-Lasso 结合，提出 Ma-Lasso 方法，并通过仿真数据比较 GLS，RLS，ME，Lasso，M-Lasso 的 MSE。结果表明，在稀疏模型中，Ma-Lasso 的结果优于上述其他方法。这表明，Ma-Lasso 在处理稀疏模型问题上能给出更精确的参数估计，较其他方法更有优势。最后将 Ma-Lasso 与其他方法一起应用到贵州茅台各季度财报数据和股票价格数据中，从所选特征和 MSE 的结果看，Ma-Lasso 均给出一个较为理想的结果，符合经济理论。这进一步说明 Ma-Lasso 在处理实际问题中有其独有的优越性，可以应用在金融领域等领域。

尽管 Ma-Lasso 在稀疏模型上的表现优于 M-Lasso，并能更好地选择影响股票价格的特征，但是在非稀疏模型上其表现不如 M-Lasso。因此，实际应用中可根据需求进行综合考虑，选择合适的方法。

**参考文献：**

[1] GULER H, GULER E O. Sparsely Restricted Penalized Estimators[J]. *Commun Stat Theory Meth*, 2021, **50**(7): 1656-1670. DOI: 10.1080/03610926.2019.1682164.

[2] DONOHO D L. High-dimensional Data Analysis: The Curses and Blessings of Dimensionality[C]//American Mathematical Society Math Challenges Lecture. Providence, Rhode Island: AMS, 2000, **1**(2000): 1-32.

[3] FAN J Q, HAN F, LIU H. Challenges of Big Data Analysis [J]. *Natl Sci Rev*, 2014, **1**(2): 293-314. DOI: 10.1093/nsr/nwt032.

[4] 李欣. 高维数据的稀疏估计及其应用[D]. 杭州: 浙江大学, 2019.
LI X. Sparse Estimation for High-dimensional Data with Applications[D]. Hangzhou: Zhejiang University, 2019.

[5] 苏锦霞. 基于特征选择的高维数据统计分析[D]. 兰州: 兰州大学, 2018.
SU J X. Statistical Analysis of High-dimensional Data Based on Feature Selection[D]. Lanzhou: Lanzhou University, 2018.

[6] HOERL A E, KENNARD R W. Ridge Regression: Biased Estimation for Nonorthogonal Problems[J]. *Technometrics*, 1970, **12**(1): 55-67. DOI: 10.1080/00401706.1970.10488634.

[7] TIBSHIRANI R. Regression Shrinkage and Selection *via* the Lasso[J]. *J R Stat Soc Ser B Methodol*, 1996, **58**(1): 267-288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

[8] 胡蓉. 基于随机Lasso的Meta分析[D]. 北京: 北京建筑大学, 2019.
HU R. Meta Analysis Based on Random Lasso[D]. Beijing: Beijing University of Civil Engineering and Architecture, 2019.

[9] 王璐, 孙聚波. Lasso回归方法在特征变量选择中的应用[J]. 吉林工程技术师范学院学报, 2021, **37**(12): 109-112. DOI: 10.3969/j.issn.1009-9042.2021.12.032.
WANG L, SUN J B. Application of Lasso Regression Method in the Selection of Feature Variables[J]. *J Jilin Teach Inst Eng Technol*, 2021, **37**(12): 109-112. DOI: 10.3969/j.issn.1009-9042.2021.12.032.

[10] 苏宇腾, 吕世云, 谢文涵, 等. 基于LASSO回归与随机森林算法的2型糖尿病发病风险因素分析[J]. 环境卫生学杂志, 2023, **13**(7): 485-495. DOI: 10.13421/j.cnki.hjwsxzz.2023.07.002.
SU Y T, LYU S Y, XIE W H, *et al*. A Risk Factor Analysis for Type 2 Diabetes Mellitus Based on LASSO Regression and Random Forest Algorithm[J]. *J Environ Hyg*, 2023, **13**(7): 485-495. DOI: 10.13421/j.cnki.hjwsxzz.2023.07.002.

[11] 邢耀. 胖(高)大数据的LASSO Logistic模型构建与应用: 以贷款偿还与糖尿病两数据集为例[D]. 济南: 山东大学, 2022.
XING Y. Construction and Application of LASSO Logistic Model for Fat (High) Big Data-taking Loan Repayment and Diabetes as Examples[D]. Jinan: Shandong University, 2022.

[12] 车前子, 王晶, 白卫国, 等. 基于LASSO回归的骨质疏松肾阳虚状态辨识模型研究[J]. 中华中医药杂志, 2022, **37**(10): 5928-5933.
CHE Q Z, WANG J, BAI W G, *et al*. Study on the State Identification Model of Kidney Yang Deficiency in Osteoporosis Patients Based on LASSO Regression[J]. *China J Tradit Chin Med Pharm*, 2022, **37**(10): 5928-5933.

[13] ZOU H, HASTIE T. Regularization and Variable Selection via the Elastic Net[J]. *J R Stat Soc Ser B Stat Methodol*, 2005, **67**(2): 301-320. DOI: 10.1111/j. 1467-9868.2005.00503.x.

[14] ZOU H. The Adaptive Lasso and Its Oracle Properties[J]. *J Am Stat Assoc*, 2006, **101**(476): 1418-1429. DOI: 10.1198/016214506000000735.

[15] THEIL H. On the Use of Incomplete Prior Information in Regression Analysis[J]. *J Am Stat Assoc*, 1963, **58**(302): 401-414. DOI: 10.1080/01621459.1963.10500854.

[16] GULER H, GULER E O. Mixed Lasso Estimator for Stochastic Restricted Regression Models[J]. *J Appl Stat*, 2021, **48**(13/14/15): 2795-2808. DOI: 10.1080/02664763.2021.1922614.

[17] 刘洪伟, 徐文科. 线性模型的广义最小二乘估计递推算法[J]. 哈尔滨师范大学自然科学学报, 2011, **27**(3): 29-31. DOI: 10.3969/j.issn.1000-5617.2011.03.009.
LIU H W, XU W K. The Recursive Algorithm of Generalized Least Squares Estimation for Linear Model[J]. *Nat Sci J Harbin Norm Univ*, 2011, **27**(3): 29-31. DOI: 10.3969/j.issn.1000-5617.2011.03.009.

[18] CONWAY R N, MITTELHAMMER R C. The Theory of Mixed Estimation in Econometric Modeling[J]. *Stud Econ Finance*, 1986, **10**(1): 79-120. DOI: 10.1108/eb028665.

[19] 赵昊然. 基于LASSO和BP神经网络的股价价格预测研究[D]. 桂林: 广西师范大学, 2022.
ZHAO H R. Research on Stock Price Forecasting Based on LASSO and BP Neural Network[D]. Guilin: Guangxi Normal University, 2022.