

## 基于高维非稀疏条件偏相关系数的估计研究

杨静颖<sup>1</sup>, 晏梅<sup>2</sup>

1. 电子科技大学 数学科学学院, 成都 511731

2. 云南师范大学 数学学院, 昆明 650500

**摘要:**目的 针对不同偏相关系数的估计方法, 提出在高维非稀疏条件下不同偏相关系数估计方法的算法性能、估计准确性和效率的探讨方法。方法 现有 Pcor 估计方法的研究主要关注高维数据和稀疏性假设下偏相关关系的存在性, 但是, 在非稀疏条件下, Pcor 估计方法的算法效率和估计准确性研究较为缺乏。本研究综合考虑了适用于非稀疏条件的偏相关系数估计方法, 并采用正则化方法处理相应的高维回归模型, 进一步探索估计方法对偏相关系数的估计性能和效率, 为验证不同算法的估计表现, 进行了大量的数值模拟实验, 并分析了股票市场中的实际数据。结果 在高维非稀疏条件下, 无偏自适应 LASSO 和渐进无偏 MCP 在偏相关系数的估计中表现都很出色。结论 在高维非稀疏条件下, 偏相关系数的估计方法与高维稀疏条件下呈现出相似的特点: 当 Pcor 为负值时, 估计较为准确; 当 Pcor 为正值时, 估计存在一定的偏差。在正则化方法的选择上, 无偏自适应 LASSO 和渐进无偏的 MCP 方法综合表现都优于相应的有偏 LASSO 方法, 特别地, 在小样本量下, 自适应 LASSO · RES 算法表现较优, 而在大样本量下, MCP · REG2 较好, 其中, REG2 方法在 Pcor 取正值时效果最好。值得注意的是, 相较于稀疏条件下控制变量得到有效控制, 在非稀疏条件下控制变量的干扰和影响增多, 因此当非稀疏条件越趋近于稀疏条件时, 算法误差越低, 效率越高; 在适当的非稀疏性条件下, 无偏自适应 LASSO · RES 和渐进无偏 MCP · REG2 算法都表现良好, 也有较好的鲁棒性和稳定性; 在较强非稀疏性条件下自适应 LASSO · RF 算法表现最好。

**关键词:** 偏相关系数; 高维数据; 非稀疏条件; 正则化方法; LASSO; 自适应 LASSO; MCP

**中图分类号:** O212.1 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2025.0003.015

### Estimation Study of Partial Correlation Coefficients Based on High-dimensional Non-sparse Conditions

YANG Jingying<sup>1</sup>, YAN Mei<sup>2</sup>

1. School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 511731, China

2. School of Mathematics, Yunnan Normal University, Kunming 650500, China

**Abstract: Objective** This study explores algorithm performance, estimation accuracy, and efficiency of different partial correlation coefficient estimation methods under high-dimensional non-sparse conditions. **Methods** Existing research on Pcor estimation methods primarily focuses on the existence of partial correlation relationships in high-dimensional data under sparse assumptions. However, research on algorithm efficiency and estimation accuracy of Pcor estimation methods under non-sparse conditions is relatively lacking. This study first comprehensively considered partial correlation coefficient estimation methods applicable to non-sparse conditions and employed regularization methods to handle corresponding high-dimensional regression models. Further exploration was conducted to investigate the estimation methods' performance and

**收稿日期:** 2023-09-08 **修回日期:** 2023-11-03 **文章编号:** 1672-058X(2025)03-0118-09

**基金项目:** 云南省科技计划基础研究项目(202201AU070051); 云南师范大学博士科研启动项目(2020ZB014); 云南省现代分析数学及其应用重点实验室(202302AN360007)。

**作者简介:** 杨静颖(1989—), 女, 山西晋中人, 博士研究生, 从事时间序列分析统计研究。

**通信作者:** 晏梅(1988—), 女, 云南曲靖人, 讲师, 博士, 从事非参数统计、高维数据分析研究。myanxy920@hotmail.com。

**引用格式:** 杨静颖, 晏梅. 基于高维非稀疏条件偏相关系数的估计研究[J]. 重庆工商大学学报(自然科学版), 2025, 42(3): 118-126.

YANG Jingying, YAN Mei. Estimation study of partial correlation coefficients based on high-dimensional non-sparse conditions [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(3): 118-126.

efficiency regarding partial correlation coefficients. To verify the estimation performance of different algorithms, extensive numerical simulation experiments were conducted, and real data from the stock market were analyzed. **Results** Under high-dimensional non-sparse conditions, both unbiased adaptive LASSO and asymptotically unbiased MCP performed excellently in estimating partial correlation coefficients. **Conclusion** Under high-dimensional non-sparse conditions, partial correlation coefficient estimation methods exhibit similar characteristics to those under high-dimensional sparse conditions: accurate estimation when Pcor is negative and some bias when Pcor is positive. In terms of regularization method selection, the comprehensive performance of unbiased adaptive LASSO and asymptotically unbiased MCP methods is superior to the corresponding biased LASSO methods. Specifically, under small sample sizes, the performance of the adaptive LASSO · RES algorithm is superior, while under large sample sizes, MCP · REG2 performs better, with REG2 being most effective when Pcor is positive. It is worth noting that controlling variables is more challenging and impactful under non-sparse conditions while controlling variables are effectively controlled under sparse conditions. Therefore, as non-sparse conditions approach sparse conditions, algorithmic errors decrease, and efficiency increases. Under appropriate non-sparse conditions, unbiased adaptive LASSO · RES and asymptotically unbiased MCP · REG2 algorithm perform well, exhibiting good robustness and stability. Under stronger non-sparse conditions, the adaptive LASSO · RF algorithm performs the best.

**Keywords:** partial correlation coefficient; high-dimensional data; non-sparse conditions; regularization methods; LASSO; adaptive LASSO; MCP

## 1 引言

偏相关系数 (Pcor) 是统计分析中的一个重要指标,指两个随机变量  $X$  和  $Y$  在消除控制变量集合  $Z$  中随机变量影响后的相关关系,记为  $\rho_{XY|Z}$ <sup>[1-2]</sup>。偏相关系数本质上是在移除  $X$  和  $Z$  之间的联系后,量化  $X$  和  $Y$  之间剩余的内在相关性。因此,与识别直接相关关系的皮尔逊相关系数相比,偏相关系数能够区分出两个随机变量之间的相关性是由直接因果关系产生,还是由中间变量诱导产生的。

偏相关系数广泛应用于金融<sup>[3-4]</sup>、基因学<sup>[5]</sup>以及其他领域的应用研究中<sup>[6-7]</sup>。Kenett 等<sup>[3-4]</sup>将 Pcor 应用于金融市场中股票之间关联关系的研究;Xia 等<sup>[8]</sup>将 Pcor 用于检测高斯图形模型中条件因变量的配对情况;王珊等<sup>[9]</sup>将 Pcor 应用于稻谷重金属污染的研究。更多的应用详见参考文献[10-13]。

近年来,高维偏相关系数的计算方法也取得了显著进展。在高维度低样本量条件下,Peng 等<sup>[14]</sup>假设偏相关系数矩阵稀疏,提出应用稀疏回归技术来计算偏相关系数的高效计算方法;基于基因共表达网络构建的想法,Ha 等<sup>[15]</sup>提出一种三步法估算高维偏相关矩阵的统计框架;Khare 等<sup>[16]</sup>提出一种基于伪似然法的稀疏高维图形模型选择新方法;白贵书<sup>[17]</sup>总结归纳了稀疏条件下高维偏相关系数的现有算法,并研究了在稀疏条件下的算法效率。但是,以上研究的高维偏相关系数估计都是集中在稀疏条件下实施的,重点关注在稀疏性假设下偏相关关系的存在性,即注重于偏相关系数是否等于零的判别。在非稀疏条件下 Pcor 估计方法的研究相对较少,并且很少涉及算法效率

和估计准确性。

尽管许多高维问题在本质上是稀疏的<sup>[14-18]</sup>,但是在金融研究中经常出现非稀疏的协方差矩阵<sup>[19]</sup>。例如,同一行业的所有股票都与其他类别的股票有共同的相关性,因此在相同行业因素影响下,同一行业板块的股票往往存在共同涨跌的情况,即对应非稀疏的协方差矩阵和偏相关系数矩阵<sup>[20]</sup>。更准确的非稀疏偏相关系数矩阵估计可以给出更多股票之间在剔除市场影响后的内在关联。然而,非稀疏条件的复杂性使得在非稀疏条件下 Pcor 估计方法的算法效率研究非常少。

鉴于此,本文旨在研究特定非稀疏条件下,现有高维偏相关系数估计方法的有效性、算法性能和估计准确性,并与在稀疏条件下的算法表现进行比较。通过数值模拟实验和实例分析,文章进一步探讨了在非稀疏条件下高维偏相关系数估计的可行性和合理性,且不同算法的性能分析可以为在非稀疏条件下的新算法研究提供参考方向。

文章结构安排如下:第二章先介绍偏相关系数的基础知识和计算方法分析,然后引入正则化方法,提出高维偏相关系数的几种可应用于非稀疏条件的估计算法。在仿真模拟实验中,第三章讨论不同偏相关系数计算方法的估计误差和算法效率。第四章实现了基于股票数据的偏相关系数矩阵实际案例分析。第五章总结研究成果并提出未来的研究方向。

## 2 计算方法分析与算法设计

### 2.1 偏相关系数的计算方法

本章介绍了偏相关系数的定义、传统低维计算方法、逆矩阵计算方法和回归系数计算方法。

### 2.1.1 偏相关系数的定义

设  $X$  和  $Y$  是两个随机变量,且  $Z=(Z_1, Z_2, \dots, Z_p)$  是  $p$  个随机变量的组合,称为控制变量。则在控制变量  $Z$  的条件下,将  $X$  和  $Y$  分别对  $Z$  进行如下线性回归:

$$\begin{cases} X = a_0 + \sum_{i=1}^p a_i Z_i + \varepsilon \\ Y = b_0 + \sum_{i=1}^p b_i Z_i + \eta \end{cases} \quad (1)$$

利用线性回归得到的残差  $\varepsilon$  和  $\eta$ , 可以给出  $X$  和  $Y$  之间的偏相关系数,定义如下:

$$\rho_{XY|Z} = \rho_{\text{corr}}(\varepsilon, \eta) = \frac{\text{cov}(\varepsilon, \eta)}{\sqrt{\text{var}(\varepsilon)} \sqrt{\text{var}(\eta)}}$$

其中,  $\rho_{\text{corr}}$  是两个随机变量相关系数,  $\text{cov}(\cdot, \cdot)$  是两个随机变量的协方差,  $\text{var}(\cdot)$  是随机变量的方差,偏相关系数的阶数是指控制变量的个数  $p$ 。

在低维情况 ( $p < n$ ) 下,传统方法就是利用最小二乘法(OLS)计算式(1)中回归模型的残差  $\varepsilon$  和  $\eta$ , 进而由残差的相关系数来计算偏相关系数  $\rho_{\text{corr}}$ 。

### 2.1.2 Pcorr 的逆矩阵算法

设协方差矩阵为  $\Sigma = \text{cov}(U)$ , 且协方差矩阵的逆矩阵  $\Omega = (\omega_{ij})_{(p+2) \times (p+2)} = \Sigma^{-1}$  为浓度矩阵, 其中  $U = [U_1, U_2]$ ,  $U_1 = [X, Y]$ ,  $U_2 = [Z_1, \dots, Z_p]$ , 则, Cox 等<sup>[21]</sup> 证明了偏相关系数的逆矩阵计算公式如下,

$$\rho_{XY|Z} = -\frac{\omega^{12}}{\sqrt{\omega^{11} \omega^{22}}} \quad (2)$$

### 2.1.3 Pcorr 的回归系数算法

考虑如下线下回归模型:

$$\begin{cases} X = \lambda_0 Y + \sum_{i=1}^p \lambda_i Z_i + \zeta \\ Y = \gamma_0 X + \sum_{i=1}^p \gamma_i Z_i + \tau \end{cases}$$

其中,  $\zeta$  和  $\tau$  是回归误差。依据 Peng 等<sup>[14]</sup> 所证明的结论, 得到  $\lambda_0$  和  $\gamma_0$  的计算公式如下:

$$\lambda_0 = \rho_{XY|Z} \sqrt{\frac{\omega^{22}}{\omega^{11}}}, \gamma_0 = \rho_{XY|Z} \sqrt{\frac{\omega^{11}}{\omega^{22}}}$$

其中,  $\omega^{11}$ 、 $\omega^{12}$ 、 $\omega^{21}$ 、 $\omega^{22}$  的定义与式(2)中相同。继而可推导出  $\lambda_0 \gamma_0 = \rho_{XY|Z}^2$ , 则可以给出偏相关系数的回归系数计算公式如下:

$$\rho_{XY|Z} = \text{sign}(\lambda_0) \sqrt{\lambda_0 \gamma_0}$$

## 2.2 高维数据的正则化方法

设数据  $\{x_i, y_i, z_i\}_{i=1}^n$  是  $X, Y, Z$  的  $n$  个独立同分布样本, 不妨假设数据是中心化的, 考虑如下矩阵形式的线性回归模型组:

$$\begin{cases} X = Za + \varepsilon \\ Y = Zb + \eta \end{cases} \quad (3)$$

其中,  $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T$  和  $\eta = [\eta_1, \dots, \eta_n]^T$  为回归残差,  $X = [x_1, \dots, x_n]^T$ ,  $Y = [y_1, \dots, y_n]^T$ ,  $Z = (z_{ij})_{n \times p} = [Z_1, \dots, Z_n]^T$ 。

在高维 ( $p > n$ ) 情况下, 传统的偏相关系数估计方法不再适用, 因此需要引入处理高维数据的线性回归方法。本文在传统线性回归模型中引入惩罚函数, 使用正则化方法在控制变量集合中选取一个较小的子集来解释模型。

在式(3)的线性回归模型中, 传统最小二乘法的回归系数估计量为  $\hat{a} = (Z^T Z)^{-1} Z^T X$  和  $\hat{b} = (Z^T Z)^{-1} Z^T Y$ 。但在高维  $p > n$  情况下,  $Z^T Z$  不可逆, 则最小二乘方法失效。因此, 引入惩罚函数  $p_\lambda(\beta)$ , 得到回归系数的正则化估计量如下:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left( \frac{1}{n} \|Y - Z\beta\|^2 + p_\lambda(\beta) \right)$$

这里惩罚函数的选择是多样的, 如 LASSO 惩罚函数<sup>[22-23]</sup>、自适应 LASSO 惩罚函数<sup>[24]</sup>、SCDA 惩罚函数<sup>[25]</sup> 和 MCP 惩罚函数<sup>[26]</sup>。本文中用的是 LASSO 和 MCP 这两种惩罚函数, 如表 1 所示。

表 1 正则化方法的惩罚函数

Table 1 Penalty functions of regularisation methods

正则化方法 (R 包)	惩罚函数
LASSO (glmnet)	$p_\lambda(\beta) = \lambda \ \beta\ _1$
自适应 LASSO (msgps)	$p_\lambda(\beta) = \lambda \sum_{j=1}^p \omega_j  \beta_j $ ( $\omega_j = 1/ \hat{\beta}_j^{\text{LASSO}} ^\gamma, \gamma > 0$ )
MCP (ncvreg)	$p_\lambda(\beta) = (a\lambda - \beta)_+ / a, (a > 1)$

注 1:  $\lambda$  和  $a$  均为惩罚函数的调节参数,  $\omega_j$  是自适应权重,  $\hat{\beta}_j^{\text{LASSO}}$  是由 LASSO 方法估计出的回归系数,  $\gamma$  是自适应权重阶数,  $I(\cdot)$  是指标函数。

注 2: 正则化 LASSO 和 MCP 方法的惩罚函数中惩罚参数  $\lambda$  是由“glmnet”和“ncvreg”R 包中的 CV 交叉验证函数“cv.glmnet”和“cv.ncvreg”选取的最优惩罚参数<sup>[27-29]</sup>。而自适应 LASSO 方法惩罚函数中的参数  $\lambda$  和  $\gamma$  是由“msgps”R 包中 BIC 准则选取的最优参数<sup>[30-31]</sup>。

## 2.3 高维偏相关系数估计方法

本章介绍了 3 种高维偏相关系数的估计算法。使用的正则化技术是在第二章介绍的 LASSO、自适应 LASSO 和 MCP。下列方法都是先试用正则化技术估计式(3)中线性回归模型组, 得到估计的回归系数  $\hat{a}$  和  $\hat{b}$ , 再计算样本估计量  $\hat{X} = Z\hat{a}$  和  $\hat{Y} = Z\hat{b}$  及相应的残差估计量  $\hat{\varepsilon} = X - \hat{X}$  和  $\hat{\eta} = Y - \hat{Y}$ 。

### 2.3.1 RES 算法

RES 算法是使用正则化技术后, 直接以偏相关系

数的定义来计算偏相关系数,如式(4):

$$\rho_{XYZ} = \frac{\sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\varepsilon})(\hat{\eta}_i - \bar{\eta})}{\sqrt{\sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\varepsilon})^2 \sum_{i=1}^n (\hat{\eta}_i - \bar{\eta})^2}} \quad (4)$$

其中,  $\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i; \bar{\eta} = \frac{1}{n} \sum_{i=1}^n \hat{\eta}_i$ 。

### 2.3.2 RF 算法

为了消除 LASSO 的估计偏差, RF 算法利用最小二乘法来重新计算回归系数和回归残差。设  $Z_X, Z_Y$  分别是选取  $\hat{a}, \hat{b}$  中非零系数对应变量的新矩阵。再利用最小二乘法重新估计回归系数,新估计的回归系数为  $\hat{a}_{RF} = (Z_X^T Z_X)^{-1} Z_X^T X$  和  $\hat{b}_{RF} = (Z_Y^T Z_Y)^{-1} Z_Y^T Y$ 。接着,更新的残差估计量为  $\hat{\varepsilon}_{RF} = X - \hat{X}_{RF} = X - Z_X \hat{a}_{RF}$  和  $\hat{\eta}_{RF} = Y - \hat{Y}_{RF} = Y - Z_Y \hat{b}_{RF}$ 。最后,将新残差估计量带入式(4),可得 RF 算法所得的偏相关系数。

### 2.3.3 REG2 算法

为了更彻底地消除控制变量  $Z$  的影响, REG2 算法在式(3)的线性回归模型基础上,再构造回归模型,进一步检测  $X$  和  $Y$  中  $Z$  的共同影响因子。所构造的新回归模型如下:

$$\begin{cases} X = c_1 \hat{X} + c_2 \hat{Y} + \varepsilon' \\ Y = d_1 \hat{X} + d_2 \hat{Y} + \eta' \end{cases} \quad (5)$$

其中,  $\varepsilon'$  和  $\eta'$  为回归误差。然后,对式(5)中线性回归模型组用最小二乘估计得到新估计量  $\hat{X}_{Reg2}$  和  $\hat{Y}_{Reg2}$ 。接着,计算新的残差估计量  $\hat{\varepsilon}_{Reg2} = X - \hat{X}_{Reg2}$  和  $\hat{\eta}_{Reg2} = Y - \hat{Y}_{Reg2}$ 。最后,将新残差估计量带入式(4),可得 REG2 算法所得的偏相关系数。

## 3 数值模拟

本章的数值模拟选取两组高维非稀疏条件下基于标准正态分布生成的样本,来研究上述算法在高维非稀疏条件下的估计效率。首先,取  $z$  和  $e_i$  都独立服从相同的标准正态分布,对每个  $i=1,2,\dots,p$ ,都利用  $Z_i = 0.4z + e_i$  生成  $n$  个独立同分布的样本,即控制变量  $Z = [Z_1, \dots, Z_p]$  的  $np$  个样本。接着,生成变量  $X$  和  $Y$  的样本如下:

$$\begin{cases} X = Za + \varepsilon \\ Y = Zb + \zeta \end{cases} \quad (6)$$

其中,  $\zeta = (\omega\varepsilon + \eta) / \sqrt{1 + \omega^2}$ ;  $\varepsilon$  和  $\eta$  都独立服从相同的标准正态分布。则所求偏相关系数为  $\rho_{XYZ} = \omega / \sqrt{1 + \omega^2}$ ,即不同的 Pcor 都可通过给定  $\omega$  的取值来实现。在稀疏条件下,一般会选择式(6)中线性回归模型组的回归系数  $a, b$  中向量的分量少数非零,如  $a = -b = [0.1, 0.2, 0.3, 0, \dots, 0]^T$ ,即控制变量中只有  $Z_1, Z_2, Z_3$

与  $X, Y$  有相关关系。在非稀疏性条件下,选择式(6)中回归系数  $a$  和  $b$  的分量均非零,即控制变量  $Z_1, \dots, Z_p$  与  $X, Y$  均有相关关系。同时,还需要使得这些复杂的非稀疏性对算法造成的干扰是可控的,以保证算法的稳定性。所以,假设当  $p \rightarrow \infty$  时,回归系数  $a$  和  $b$  渐进收敛到 0。本文在表 2 中的情形 1 和情形 2 中取  $a = -b$ ,设置渐进收敛速率为  $O\left(\frac{\sigma^i}{\sqrt{i}}\right)$  和  $O\left(\frac{1}{i}\right)$ ,并增加调节参数  $\sigma, \theta$  来调节控制非稀疏系数的收敛速率。这里的调节参数  $\sigma$  和  $\theta$  就是调节控制变量  $Z_1, \dots, Z_p$  对  $X, Y$  的影响,即当  $\sigma$  和  $\theta$  增大时,每个控制变量  $Z_1, \dots, Z_p$  对  $X, Y$  的影响增大,所产生数据的非稀疏性越强,对 Pcor 估计方法的干扰就越强。后文将选择不同的调节参数  $\sigma$  和  $\theta$ ,即不同非稀疏条件,计算不同算法下的偏相关系数 Pcor,以观察不同算法的估计效率和在不同非稀疏程度下的算法敏感性。

表 2 非稀疏条件下的回归系数  $a$

Table 2 Regression coefficients  $a$  under non-sparse conditions

情形	回归系数 $a$	调节参数
情形 1	$\left[ \sigma, \frac{\sigma^2}{\sqrt{2}}, \dots, \frac{\sigma^{p/2}}{\sqrt{p/2}}, -\sigma, -\frac{\sigma^2}{\sqrt{2}}, \dots, -\frac{\sigma^{p/2}}{\sqrt{p/2}} \right]$	$\sigma \in (0, 1]$
情形 2	$\theta \left[ 1, \frac{1}{2}, \dots, \frac{1}{p/2}, -1, -\frac{1}{2}, \dots, -\frac{1}{p/2} \right]$	$\theta \in (0, 1]$

在数值模拟中,偏相关系数的取值为  $\rho_{XYZ} = -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$ ,3 组样本量和控制变量的阶数为  $n=50, p=200, 400, 800; n=150, p=600, 1200, 2000; n=300, p=1200, 2000, 3000$ 。对每一组  $(n, p)$ ,选取不同算法计算不同真实 Pcor 取值下的 Pcor 估计值,以此来对比不同算法的效率和表现。取最具代表性的数据来分析,选择调节参数  $\sigma=0.3, 0.6$  和  $\theta=0.3, 0.6$ 。本文通过重复 100 次偏相关系数估计实验,选取 LASSO、MCP 和自适应 LASSO (ALASSO) 正则化方法,计算出 Pcor 估计值的平均偏差 (MEAN BIAS) 和均方误差 (MSE) 来分析不同 Pcor 估计算法的表现。

表 3 和表 4 是在不同  $(n, p)$  取值和不同估计算法下, Pcor 估计值的均方误差均值,并用下划线标注了在相同  $(n, p)$  取值下的最小均方误差位置。数值模拟发现:所有算法 Pcor 估计值的均方误差,会随样本量  $n$  的增大而减小,随着控制变量阶数  $p$  的增大而增大。无偏自适应 LASSO 和渐进无偏 MCP 正则化方法的综合表现均优于有偏 LASSO。当样本量较小 ( $n=50$ ) 时,无偏 ALASSO 方法表现较优。对较小的调节参数 ( $\sigma=0.3$  或  $\theta=0.3$ ), ALASSO · RES 方法的均方误差最小,此时 MCP 和 LASSO · RES 方法的均方误差也相对较小。对较大的调

节参数( $\sigma=0.6$  或  $\theta=0.6$ ), ALASSO · REG2 方法的均方误差最小。这可能是由于样本量和控制变量的影响都较小,使无偏自适应 LASSO 方法的算法表现最好,但渐进无偏 MCP 方法效果也较好。当样本量较大( $n=150, 300$ )时,渐进无偏 MCP 方法相较于有偏 LASSO 体现出明显优势,特别是 REG2 方法综合表现最好,此时无

偏 ALASSO(RES 和 REG2)算法表现也较好,但略弱于 MCP 方法。这是由于 REG2 方法相较于 RES 方法,在两次回归后可以更彻底地找到控制变量  $Z$  在  $X$  和  $Y$  中的影响因子。但在小部分情况下会出现 RES 优于 REG2 的情况,往往这时两种方法的均方误差差距很小,可能是正则化方法在减少估计误差方面起了主导作用。

表 3 情形 1 中 Pcor 估计值的均方误差均值  
Table 3 Mean square errors of the Pcor estimates in Situation 1

$\times 10^2$

均方误差		$\sigma=0.3$									$\sigma=0.6$							
$n$	$p$	LASSO			ALASSO			MCP			LASSO			ALASSO			MCP	
		RES	REG2	RF	RES	REG2	RF	RES	REG2	RES	REG2	RF	RES	REG2	RF	RES	REG2	
50	200	3.57	5.83	5.78	<u>3.10</u>	5.45	11.1	3.68	3.72	12.2	8.75	13.2	7.47	<u>5.67</u>	9.12	8.66	6.26	
50	400	4.01	8.71	6.64	<u>3.65</u>	7.63	14.8	4.23	4.82	13.7	10.5	14.7	9.26	8.29	13.5	10.3	<u>7.45</u>	
50	800	4.12	9.88	5.95	<u>4.01</u>	10.2	19.1	4.39	5.46	16.9	12.4	17.4	10.9	10.2	17.0	13.6	<u>9.16</u>	
150	600	1.47	1.41	2.49	0.79	1.70	5.81	1.04	<u>0.74</u>	3.78	3.00	2.55	2.04	2.19	3.98	<u>1.79</u>	1.86	
150	1 200	1.82	2.56	3.15	<u>0.99</u>	2.83	9.41	1.45	1.07	4.84	4.16	3.01	2.33	3.28	6.79	<u>1.95</u>	2.05	
150	2 000	2.03	3.01	3.29	1.31	4.17	12.4	1.60	<u>1.23</u>	6.06	5.47	3.81	2.71	4.39	9.18	<u>2.48</u>	2.85	
300	1 200	0.62	0.55	0.66	0.33	0.99	4.30	0.30	<u>0.29</u>	1.82	1.40	0.81	0.78	0.90	2.04	0.74	<u>0.66</u>	
300	2 000	0.64	0.77	0.72	0.38	1.52	6.51	0.29	<u>0.28</u>	2.07	2.05	1.29	0.89	1.31	3.50	<u>0.82</u>	0.83	
300	3 000	0.73	0.78	0.85	0.48	2.19	8.73	<u>0.29</u>	0.31	2.62	2.50	1.45	1.04	1.84	5.05	1.02	<u>1.00</u>	

表 4 情形 2 中 Pcor 估计值的均方误差均值  
Table 4 Mean square errors of the Pcor estimates in Situation 2

$\times 10^2$

均方误差		$\theta=0.3$									$\theta=0.6$							
$n$	$p$	LASSO			ALASSO			MCP			LASSO			ALASSO			MCP	
		RES	REG2	RF	RES	REG2	RF	RES	REG2	RES	REG2	RF	RES	REG2	RF	RES	REG2	
50	200	6.25	8.37	8.20	<u>5.60</u>	8.29	12.6	6.85	6.82	20.5	13.6	22.0	16.4	<u>12.1</u>	13.6	19.3	14.1	
50	400	7.71	10.8	9.75	<u>6.96</u>	10.6	17.1	8.02	8.16	26.2	16.6	26.5	20.2	<u>14.3</u>	17.2	23.5	17.3	
50	800	7.91	12.1	9.44	<u>7.50</u>	13.4	21.0	8.30	9.14	29.0	19.9	28.9	22.0	<u>17.2</u>	20.8	26.9	19.7	
150	600	3.99	3.30	5.77	2.61	3.56	7.76	3.59	<u>2.59</u>	11.2	8.63	8.53	8.81	8.23	8.14	8.31	<u>7.81</u>	
150	1 200	4.78	4.54	6.25	3.16	5.33	11.5	4.25	<u>3.07</u>	14.1	10.9	10.4	9.95	10.2	10.8	9.82	<u>9.14</u>	
150	2 000	4.95	4.82	6.45	3.41	6.53	14.1	4.50	<u>3.23</u>	15.2	11.3	10.6	10.0	10.9	12.8	10.1	<u>9.17</u>	
300	1 200	2.43	1.84	2.56	1.41	2.27	5.53	1.69	<u>1.38</u>	7.08	6.06	<u>4.71</u>	5.50	5.34	5.13	5.03	4.79	
300	2 000	2.70	2.16	2.83	1.61	3.28	8.31	1.87	<u>1.47</u>	8.32	7.21	5.47	5.86	6.33	6.79	5.64	<u>5.39</u>	
300	3 000	2.98	2.69	3.31	1.86	4.16	10.5	2.08	<u>1.63</u>	9.28	8.67	6.14	6.28	7.40	8.58	6.23	<u>5.99</u>	

表 5 给出了不同( $n, p$ )取值、不同 Pcor 取值和不同估计算法下,100 次重复试验的情形 2 中 Pcor 估计值和 Pcor 真实取值之间的偏差均值情况,并用下划线标明了在相同的( $n, p$ )、 $\theta$  和真实 Pcor 取值条件下最小平均偏差绝对值对应项。数值模拟发现:当 Pcor 取值小于零时,估计 Pcor 和真实 Pcor 的偏差较小;当 Pcor 取值大于零时,估计 Pcor 和真实 Pcor 的偏差均为负值,且绝对值较大,即在高维非稀疏矩阵上本文所列举的 8 种算法对 Pcor 小于零的估计值较为准

确,但对 Pcor 大于零的估计值都低于真实 Pcor 且误差较大。当  $\theta$  增大,即控制变量影响增强时,估计 Pcor 和真实 Pcor 的偏差也会随之增大;但当样本量  $n$  增大时,即使  $p/n$  比值相同的情况下,偏差也会减小。综合观察发现:无偏算法的综合表现都较好,明显优于相应的有偏 LASSO 算法,例如,ALASSO · RES( $\theta=0.3$ )、ALASSO、REG2 ( $\theta=0.6$ ) 和 MCP · REG2 效果较好,其中 REG2 在 Pcor 取正值时效果最好。

表 5 情形 2 中 Pcor 估计值和真实 Pcor 取值之间的平均偏差

Table 5 Average deviations between the estimated values of Pcor and the true values of Pcor in Situation 2

平均偏差		$\theta=0.3$									$\theta=0.6$							
		LASSO			ALASSO			MCP			LASSO			ALASSO			MCP	
		RES	REG2	RF	RES	REG2	RF	RES	REG2	RES	REG2	RF	RES	REG2	RF	RES	REG2	
$(n,p) = (50,200)$	-0.9	<u>1.5</u>	10.6	3.6	1.7	6.0	25.6	2.8	2.6	0.8	11.6	14.5	-1.5	0.8	17.5	4.0	1.4	
	-0.6	-5.1	8.3	-2.7	-2.1	11.1	26.8	-3.9	<u>0.3</u>	-11.4	6.4	<u>1.1</u>	-10.9	-3.5	13.2	-6.4	-3.2	
	-0.3	-10.5	<u>3.2</u>	-9.5	-5.7	9.9	18.5	-11.5	-5.3	-20.9	<u>-1.0</u>	-8.2	-18.7	-6.3	6.8	-18.3	-10.7	
	0	-17.3	-9.8	-18.7	-14.2	-4.9	<u>-2.7</u>	-17.4	-12.4	-34.8	-15.9	-28.1	-31.2	-18.9	<u>-10.0</u>	-33.5	-25.3	
	0.3	-24.3	-23.2	-26.4	<u>-21.6</u>	-22.9	-23.8	-25.1	-24.1	-47.0	-31.3	-42.7	-42.1	-32.0	<u>-28.8</u>	-45.2	-35.4	
	0.6	-30.5	-31.1	-33.9	<u>-30.2</u>	-37.8	-43.9	-32.2	-31.9	-60.2	<u>-44.8</u>	-58.9	-54.6	-45.4	-47.5	-58.0	-48.5	
$(n,p) = (300,1200)$	0.9	-39.6	-47.1	-44.3	-38.0	-48.9	-59.1	-41.2	-42.6	-76.4	<u>-64.6</u>	-79.3	<u>-67.7</u>	<u>-62.6</u>	<u>-67.7</u>	-72.9	-64.4	
	-0.9	-0.4	2.8	2.0	0.9	2.5	19.6	<u>-0.2</u>	0.3	-0.9	2.6	6.3	-1.3	-0.9	9.9	-1.0	<u>-0.7</u>	
	-0.6	-5.0	1.8	-2.6	-1.0	4.1	21.1	-4.1	-1.7	-7.8	-0.8	<u>0.1</u>	-7.0	-5.5	8.7	-6.3	-4.4	
	-0.3	-8.7	-1.3	-6.2	-4.4	<u>-0.1</u>	10.7	-6.8	-3.8	-14.6	-7.4	-5.7	-13.0	-11.3	<u>0.7</u>	-12.3	-10.5	
	0	-12.6	<u>-6.8</u>	-11.1	-8.0	-7.5	-2.1	-10.3	-8.1	-21.9	-16.8	-13.8	-19.3	-18.0	<u>-8.9</u>	-18.7	-17.2	
	0.3	-16.6	-13.6	-15.7	-12.2	-16.5	-18.0	-13.5	<u>-11.7</u>	-29.2	-26.5	-21.7	-25.5	-25.3	<u>-20.5</u>	-24.3	-23.5	
0.6	-20.6	-19.1	-20.1	-16.1	-22.7	-30.6	-16.9	<u>-15.7</u>	-36.0	-35.0	<u>-29.5</u>	-31.8	-31.8	-31.4	-30.4	-30.0		
0.9	-25.0	-21.3	-26.0	-19.6	-24.1	-38.4	-20.6	<u>-19.0</u>	-43.9	-41.4	<u>-37.5</u>	-38.5	-38.4	-42.0	-36.6	-36.6		

在稀疏条件下,控制变量的干扰和影响可以得到稀疏性的有效控制,而文章中非稀疏条件下控制变量的影响程度由非稀疏参数  $a$  和  $b$  中的调节参数  $\sigma$  和  $\theta$  来控制。调节参数越小,则非稀疏性越弱,就越渐进于稀疏条件,调节参数越大,则非稀疏性越强,控制变量的影响就越复杂。综合 8 种算法在两种情形不同调节参数  $\sigma$  和  $\theta$  下的均方误差表现(表 6),可知不同算法对稀疏性变化的敏感程度,从而分析算法的鲁棒性和稳定性。观察发现:算法的均方误差均会随着调节参数的增大(即非稀疏程度增强)而增加。当非稀疏系数可

控( $\sigma, \theta \leq 0.6$ ),即非稀疏程度可控时,均方误差均控制在一个较小的范围内,且样本量越大,均方误差的波动越小,特别地,无偏 ALASSO · RES 和渐进无偏的 MCP · REG2 算法的误差控制最好,即鲁棒性较好。当非稀疏系数较大( $\sigma, \theta = 0.9$ ),即非稀疏性很强时,均方误差增加了一个数量级,即使样本量增大也无法将均方误差控制在较小范围内,此时的算法表现均较差。但值得注意的是 ALASSO · RF 算法在强非稀疏性条件下表现最好,这对未来在强非稀疏性条件下提高算法效率的研究是个提示。

表 6 情形 1 和情形 2 在不同调节参数  $\sigma$  和  $\theta$  下 Pcor 估计值的均方误差均值

Table 6 Mean square errors of the Pcor estimates under different tuning parameters  $\sigma$  and  $\theta$  in Situations 1 and 2

情形 1		LASSO			ALASSO			MCP		情形 2		LASSO			ALASSO			MCP	
$n$	$\sigma$	RES	REG2	RF	RES	REG2	RF	RES	REG2	$\theta$	RES	REG2	RF	RES	REG2	RF	RES	REG2	
50	0.1	<u>1.34</u>	4.83	2.41	1.73	5.69	12.7	1.47	2.13	0.1	<u>1.43</u>	5.43	2.60	1.89	6.27	13.7	1.61	2.27	
150	0.1	0.44	1.69	0.59	0.71	3.51	9.29	<u>0.43</u>	0.64	0.1	<u>0.54</u>	1.78	0.71	0.85	3.76	9.70	0.54	0.79	
300	0.1	0.24	0.57	0.27	0.42	2.47	7.30	<u>0.24</u>	0.32	0.1	0.31	0.74	0.32	0.51	2.63	7.60	<u>0.31</u>	0.41	
50	0.3	3.90	7.26	5.50	<u>3.59</u>	7.76	14.9	4.18	4.78	0.3	7.29	10.4	9.13	<u>6.68</u>	10.8	16.9	7.73	8.04	
150	0.3	1.68	2.11	2.93	1.03	2.90	9.21	1.29	<u>0.93</u>	0.3	4.57	4.22	6.15	3.06	5.14	11.1	4.11	<u>2.97</u>	
300	0.3	0.68	0.78	0.85	0.40	1.57	6.51	0.31	<u>0.31</u>	0.3	2.71	2.23	2.90	1.63	3.24	8.12	1.88	<u>1.49</u>	
50	0.6	14.5	10.4	14.9	9.22	8.04	13.2	11.1	<u>7.64</u>	0.6	25.2	16.7	25.8	19.6	<u>14.5</u>	17.2	23.2	17.1	
150	0.6	4.85	4.41	3.11	2.36	3.29	6.65	<u>2.05</u>	2.20	0.6	13.5	10.3	9.86	9.59	<u>9.79</u>	10.6	9.40	<u>8.71</u>	
300	0.6	2.14	2.00	1.12	0.90	1.35	3.53	0.84	<u>0.81</u>	0.6	8.22	7.31	5.44	5.88	6.36	6.84	5.63	<u>5.39</u>	
50	0.9	51.3	29.8	40.1	48.6	32.9	<u>23.9</u>	46.7	37.8	0.9	40.8	23.9	30.8	36.4	25.8	<u>20.8</u>	33.3	27.1	
150	0.9	25.9	18.4	14.3	28.2	22.4	<u>13.9</u>	16.1	14.7	0.9	23.2	18.2	15.1	22.0	19.4	<u>13.5</u>	17.4	15.9	
300	0.9	13.4	11.4	6.85	16.2	13.2	<u>7.26</u>	6.00	5.75	0.9	14.6	13.7	8.84	15.0	13.6	9.14	10.0	9.65	

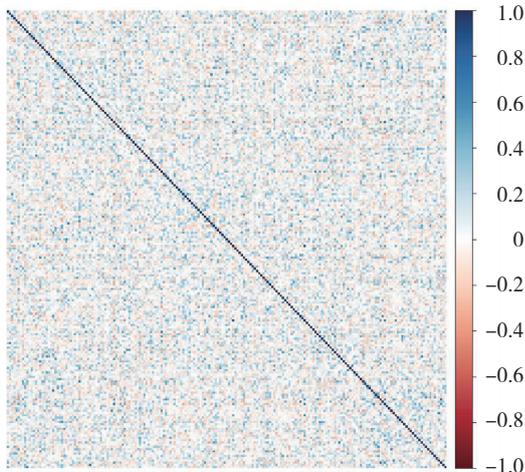
### 4 实例分析与应用

本章选取上海证券交易所中有综合性和代表性的 200 只股票,即控制变量维度  $p = 200$ ,并通过网易财经下载了 2019 年的每日调整收盘价,删除无效数据和缺

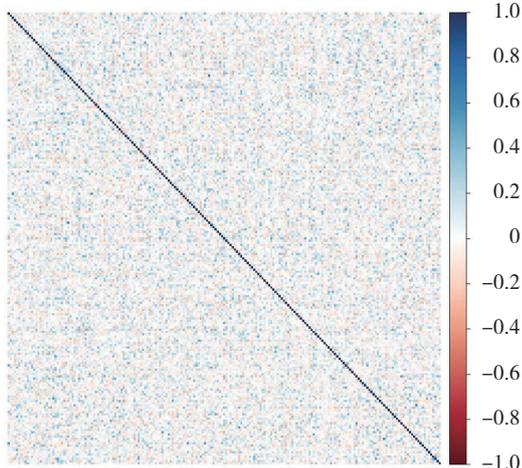
失数据后的样本量为  $n = 160$ 。因此,能选取的样本量和控制变量维度满足高维  $p > n$  的假设。在所选取的 200 只上交所股票中,股票两两之间估计一次 Pcor,其中设置以其余 198 只股票为控制变量,从而得到它们的 Pcor 估

计矩阵。Pcor 估计矩阵显示了任意两只股票之间在剔除其他股票影响后的内在联系,能更好地反映所选两只股票之间的内在相关性,并有效减少市场的影响。

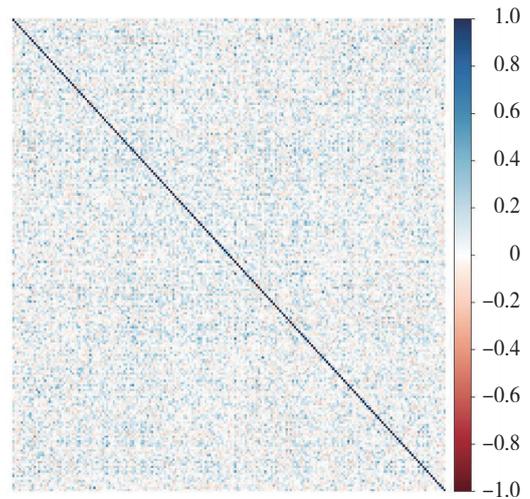
图 1 显示了使用 ALASSO · RES、ALASSO · RF 和 MCP · REG2 方法所估计的 200 只上交所股票的 Pcor 估计矩阵。观察到这几种估计方法均可有效识别股票之间较强的偏相关关系,但对较弱的偏相关性识别可能存在偏差,ALASSO · RES 和 MCP · REG2 对弱偏相关性识别效果更好,其中,ALASSO · RES 估计倾向于 Pcor 小于零,而 MCP · REG2 估计倾向于 Pcor 大于零。



(a) ALASSO · RES



(b) ALASSO · RF



(c) MCP · REG2

图 1 由 ALASSO · RES、ALASSO · RF 和 MCP · REG2 方法对 200 只上交所股票估计的 Pcor 估计矩阵

Fig. 1 Pcor matrix estimated by ALASSO · RES, ALASSO · RF and MCP · REG2 methods on 200 SSE stocks

为了解释剔除市场效应后股票公司之间的相关性,表 7 显示了 100 只上交所股票 ALASSO · RES 方法所估计的 Pcor 估计值中最大前 20 组及对应股票的股票代码、公司名称和所属行业。在研究中发现,由高偏相关系数相互连接的两只股票大多属于同一行业。此外,较高的 Pcor 估计值可能表明公司之间的股权关系或集团公司的从属关系,如中国船舶(600150.SH)是中国重工(601989.SH)旗下的公司。总体而言,在剔除影响市场的其他因素后,高 Pcor 估计值代表了两只股票之间密切相关的内在相关性。

从表 7 的不同算法估计值对比可观察到不同算法的关系:相同正则化方法下,RES 和 REG2 的估计值较接近,且取值普遍大于 RF 算法;对于正则化方法的 Pcor 估计值,ALASSO 方法优于 MCP 方法,MCP 方法优于 LASSO 方法;ALASSO · RES 和 MCP · REG2 表现依然较好,对不同股票之间关系强弱的估计不同。所以,当观察股票之间内在联系时,可以综合考虑 ALASSO · RES 和 MCP · REG2 两种算法的 Pcor 估计值,以获得更多密切关系的解读。

表 7 200 只上交所股票使用 ALASSO · RES 方法的前 20 组 Pcor 估计值及对应股票代码、公司名称和所属行业  
Table 7 Top 20 sets of Pcor estimates using the ALASSO · RES method on 200 SSE stocks with corresponding stock symbols, company names and industries

股票 1			股票 2			ALASSO			MCP			LASSO		
代码	公司	行业	代码	公司	行业	RES	REG2	RF	RES	REG2	RES	REG2	RF	
601288	农业银行	金融	601988	中国银行	金融	0.815	0.800	0.518	0.276	0.273	0.333	0.322	0.295	
600569	安阳钢铁	材料	600808	马钢股份	材料	0.801	0.792	0.392	0.612	0.696	0.081	0.050	0.028	

续表(表 7)

股票 1			股票 2			ALASSO			MCP			LASSO	
代码	公司	行业	代码	公司	行业	RES	REG2	RF	RES	REG2	RES	REG2	RF
600150	中国船舶	工业	601989	中国重工	工业	0.731	0.718	0.359	0.651	0.663	0.707	0.706	0.643
600115	东方航空	工业	601111	中国国航	工业	0.696	0.713	0.411	0.378	0.363	0.421	0.420	0.170
601169	北京银行	金融	601328	交通银行	金融	0.690	0.691	0.268	0.364	0.364	0.299	0.289	0.201
600231	凌钢股份	材料	601328	交通银行	金融	0.683	0.676	0.348	0.097	0.097	0.135	0.132	0.104
601390	中国中铁	工业	601857	中国石油	能源	0.669	0.649	0.406	0.191	0.396	0.178	0.171	0.167
600016	民生银行	金融	601998	中信银行	金融	0.664	0.689	0.355	0.138	0.195	0.238	0.229	0.169
600036	招商银行	金融	601166	兴业银行	金融	0.662	0.618	0.326	0.157	0.258	0.412	0.407	0.323
601008	连云港	工业	601018	宁波港	工业	0.659	0.628	0.411	0.407	0.500	0.459	0.455	0.318
600999	招商证券	金融	601788	光大证券	金融	0.656	0.620	0.279	0.215	0.226	0.175	0.170	0.015
601398	工商银行	金融	601939	建设银行	金融	0.646	0.634	0.330	0.242	0.252	0.430	0.422	0.442
600502	安徽水利	公用事业	600730	中国高科	非必需消费品	0.641	0.614	0.379	0.361	0.362	0.233	0.231	0.142
600101	明星电力	公用事业	600505	西昌电力	公用事业	0.640	0.631	0.641	0.475	0.542	0.619	0.615	0.468
600343	航天动力	工业	600855	航天长峰	工业	0.640	0.592	0.362	0.274	0.271	0.129	0.129	0.008
600808	马钢股份	材料	601328	交通银行	金融	0.636	0.624	0.340	0.156	0.153	0.079	0.070	0.067
601288	农业银行	金融	601328	交通银行	金融	0.630	0.644	0.422	0.157	0.165	0.074	0.064	0.002
600495	晋西车轴	工业	600581	八一钢铁	材料	0.615	0.573	0.223	0.143	0.144	0.188	0.196	0.183
600004	白云机场	工业	600894	广钢股份	材料	0.601	0.627	0.377	0.447	0.472	0.406	0.394	0.318
600088	中视传媒	非必需消费品	600536	中国软件	信息技术	0.600	0.592	0.372	0.398	0.402	0.277	0.270	0.172

### 5 结论与展望

在高维非稀疏条件下,本文所列举的偏相关系数估计算法表现总结如下:当  $P_{cor}$  为负值时,估计值较准确;当  $P_{cor}$  为正值时,估计值均存在一定程度的偏差。无偏自适应 LASSO 和渐进无偏 MCP 正则化的综合表现都明显优于相应的有偏 LASSO 方法。特别地,ALASSO · RES (小样本较优), ALASSO · REG2 和 MCP · REG2 (大样本较优)的算法效果都较好。其中,由于 REG2 方法在两次回归中更彻底地找到控制变量  $Z$  在  $X$  和  $Y$  中的影响因子,REG2 方法在  $P_{cor}$  取正值时效果最好。由于在稀疏条件下控制变量的干扰和影响可以得到有效控制,因此当非稀疏系数收敛速度越快,即非稀疏矩阵越渐进于稀疏矩阵时,各种算法的效率也越高。在适当的非稀疏性条件下,无偏 ALASSO · RES 和渐进无偏 MCP · REG2 算法都表现良好,也有较好的鲁棒性和稳定性。而 ALASSO · RF 算法在强非稀疏性条件下表现最好。

本文只考虑两种渐进收敛非稀疏系数下的高维非稀疏条件,对于不同非稀疏条件偏相关系数估计算法的表现会有所不同,因此后续可以研究更广泛的非稀疏条件下的算法表现,进而考虑给出非稀疏条件下是否能使用估计方法的判别条件。同时,本文只考虑了股票数据中偏相关系数估计问题,实际中偏相关系数

在脑科学、基因科学等领域并不少见。后续可以考虑将这些方法应用到实际应用领域的非稀疏情况,从而扩展到多元复杂数据中的偏相关系数估计。

### 参考文献(References):

- [1] 茆诗松,王静龙,濮晓龙. 高等数理统计[M]. 2 版. 北京: 高等教育出版社, 2006: 83-127.  
MAO Shi-song, WANG Jing-long, PU Xiao-long. Advanced mathematical statistics[M]. 2nd edn. Beijing: Higher Education Press, 2006: 83-127.
- [2] 郑培根,刘梓修. 偏相关系数的定义及计算[J]. 江西财经学院学报, 1986(1): 75-79.  
ZHENG Pei-gen, LIU Zi-xiu. Definition and calculation of partial correlation coefficient [J]. Contemporary Finance & Economics, 1986(1): 75-79.
- [3] KENETT D Y, TUMMINELLO M, MADI A, et al. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market[J]. PLoS One, 2010, 5(12): 15032-15042.
- [4] KENETT D Y, HUANG X, VODENSKA I, et al. Partial correlation analysis: Applications for financial markets [J]. Quantitative Finance, 2015, 15(4): 569-578.
- [5] DE LA FUENTE A, BING N, HOESCHELE I, et al. Discovery of meaningful associations in genomic data using

- partial correlation coefficients[J]. *Bioinformatics*, 2004, 20(18): 3565–3574.
- [6] 李永刚, 包文清. 基于偏相关分析的在校女大学生体质健康发展趋势研究[J]. *淮北师范大学学报(自然科学版)*, 2022, 43(2): 76–80.  
LI Yong-gang, BAO Wen-qing. Research on the development trend of female college students' physical health based on partial correlation analysis [J]. *Journal of Huaibei Normal University (Natural Sciences Edition)*, 2022, 43(2): 76–80.
- [7] EPSKAMP S, FRIED E I. A tutorial on regularized partial correlation networks[J]. *Psychological Methods*, 2018, 23(4): 617–634.
- [8] XIA X, LI J. Copula-based partial correlation screening: A joint and robust approach[J]. *Statistica Sinica*, 2021, 31(1): 421–447.
- [9] 王珊, 苏亮, 刘远立, 等. 皮尔森和偏相关系数模型在稻谷重金属污染程度研究中应用[J]. *中国食品卫生杂志*, 2020, 32(6): 631–635.  
WANG Shan, SU Liang, LIU Yuan-li, et al. Application of Pearson and partial correlation coefficient model in the research of heavy metal pollution in rice[J]. *Chinese Journal of Food Hygiene*, 2020, 32(6): 631–635.
- [10] ALABISO A, SHANG J. High-dimensional linear mixed model selection by partial correlation[J]. *Communications in Statistics-Theory and Methods*, 2023, 52(18): 6355–6380.
- [11] MARRELEC G, KRAINIK A, DUFFAU H, et al. Partial correlation for functional brain interactivity investigation in functional MRI[J]. *NeuroImage*, 2006, 32(1): 228–237.
- [12] FUJITA T, SATO A, YAMAMOTO Y, et al. Relationship between dressing and motor function in stroke patients: A study with partial correlation analysis[J]. *Journal of Physical Therapy Science*, 2015, 27(12): 3771–3774.
- [13] MA S, LI R, TSAI C L. Variable screening via quantile partial correlation [J]. *Journal of the American Statistical Association*, 2017, 112(518): 650–663.
- [14] PENG J, WANG P, ZHOU N, et al. Partial correlation estimation by joint sparse regression models[J]. *Journal of the American Statistical Association*, 2009, 104(486): 735–746.
- [15] HA M J, SUN W. Partial correlation matrix estimation using ridge penalty followed by thresholding and re-estimation[J]. *Biometrics*, 2014, 70(3): 762–770.
- [16] KHARE K, OH S Y, RAJARATNAM B. A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2015, 77(4): 803–825.
- [17] 白贵书. 高维偏相关系数的估计[D]. 成都: 电子科技大学, 2021.  
BAI Gui-shu. Estimation for high-dimensional partial correlation coefficient[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [18] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Sparse inverse covariance estimation with the graphical lasso[J]. *Biostatistics*, 2008, 9(3): 432–441.
- [19] FAN J, FAN Y, LYU J. High dimensional covariance matrix estimation using a factor model[J]. *Journal of Econometrics*, 2008, 147(1): 186–197.
- [20] ONNELA J P, CHAKRABORTI A, KASKI K, et al. Dynamics of market correlations: Taxonomy and portfolio analysis[J]. *Physical Review E*, 2003, 68(52): 056110.
- [21] WERMUTHN. *Multivariate dependencies: Models, analysis and interpretation*[M]. New York: Chapman & Hall CRC, 1996.
- [22] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1996, 58(1): 267–288.
- [23] WANG H. Coordinate descent algorithm for covariance graphical lasso[J]. *Statistics and Computing*, 2014, 24(4): 521–529.
- [24] ZOU H. The adaptive lasso and its oracle properties [J]. *Journal of the American Statistical Association*, 2006, 101(476): 1418–1429.
- [25] FAN J, LI R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. *Journal of the American Statistical Association*, 2001, 96(456): 1348–1360.
- [26] ZHANG C H. Nearly unbiased variable selection under minimax concave penalty[J]. *The Annals of Statistics*, 2010, 38(2): 894–942.
- [27] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization paths for generalized linear models via coordinate descent[J]. *Journal of Statistical Software*, 2010, 33(1): 1–22.
- [28] SIMON N, FRIEDMAN J, HASTIE T, et al. Regularization paths for cox's proportional hazards model via coordinate descent[J]. *Journal of Statistical Software*, 2011, 39(5): 1–13.
- [29] BREHENY P, HUANG J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection[J]. *The Annals of Applied Statistics*, 2011, 5(1): 232–253.
- [30] SCHWARZ G. Estimating the dimension of a model[J]. *The Annals of Statistics*, 1978, 6(2): 461–464.
- [31] FRIEDMAN J H. Fast sparse regression and classification[J]. *International Journal of Forecasting*, 2012, 28(3): 722–738.

责任编辑:李翠薇