DOI:10.13718/j.cnki.xsxb.2023.04.002

基于非凸惩罚函数的高维协方差矩阵的建模[①]

杨小卜

兰州财经大学 统计学院,兰州 730030

摘要: 近年来,随着金融数据爆炸式的增长与数据存储能力的提高,高维与高频金融数据的建模以及其在投资组合中的应用引起了人们广泛的关注.本文聚焦于高维协方差矩阵的建模问题.首先,基于 VAR-LASSO 模型引入 SCAD 惩罚函数与 MCP 惩罚函数替换 LASSO 惩罚函数,分别提出了 VAR-SCAD 模型与 VAR-MCP 模型.其次,在理论层面证明了 VAR-SCAD 模型与 VAR-MCP 模型参数的 Oracle 性质,弥补了 VAR-LASSO 模型参数不满足 Oracle 性质这一缺点,提高了模型的估计精确性.最后,通过实际频率为5分钟的高频股票数据,构建已实现协方差矩阵与投资组合进行实证分析.通过实证分析可以发现,VAR-SCAD 模型与 VAR-MCP 模型在测试精确性方面的表现要优于 VAR-LASSO 模型,VAR-SCAD 模型与 VAR-MCP 模型构建的投资组合的收益率高于 VAR-LASSO 模型构建的投资组合,其中 VAR-MCP 模型构建的投资组合的收益率最高.

关 键 词: 高频数据;高维已实现协方差矩阵;VAR-SCAD 模型;VAR-MCP 模型

中图分类号: O212 **文献标志码:** A **文章编号:** 1000-5471(2023)04-0013-10

Modeling of High-dimensional Covariance Matrix Based on Non-convex Penalty Function

YANG Xiaobo

School of Statistics, Lanzhou University of Finance and Economics, Lanzhou 730030, China

Abstract: In recent years, with the explosive growth of financial data and the improvement of data storage capacity, the modeling of high-dimensional and high-frequency financial data and its application in investment portfolios have attracted a lot of attention. This paper focuses on the modeling problem of high-dimensional covariance matrix. Firstly, this paper introduces the SCAD penalty function and MCP penalty function to replace the LASSO penalty function based on the VAR-LASSO model, and proposes the VAR-SCAD model and the VAR-MCP model, respectively. Secondly, the Oracle property of the parameters of the VAR-SCAD model and VAR-MCP model is proved at the theoretical level to compensate for the shortcoming that the parameters of the VAR-LASSO model do not satisfy the Oracle property and to improve the estimation accuracy of the models. Finally, the realized covariance matrix is constructed from the actual frequency of 5-minute high-frequency stock data and the portfolio is analyzed empirically. The empirical analysis reveals that the VAR-SCAD model and the VAR-MCP model perform better than the VAR-LASSO mod-

① 收稿日期:2022-09-07
基金项目:国家自然科学基金项目(71961013);甘肃省自然科学基金项目(20JR5RA204).
作者简介:杨小卜,硕士研究生,主要从事复杂数据的研究.

el in terms of test accuracy, and the returns of the portfolios constructed by the VAR-SCAD model and the VAR-MCP model are higher than those of the portfolios constructed by the VAR-LASSO model, with the returns of the portfolios constructed by the VAR-MCP model highest.

Key words: high-frequency data; high-dimensional realized covariance matrix; VAR-SCAD model; VAR-MCP model

随着信息技术的发展与数据可获取性的提高，金融资产的维度也随即呈现出爆炸增长的趋势. 金融资产维度的增加会给资产协方差矩阵的估计带来困难，甚至会导致病态协方差矩阵的产生. 而资产协方差矩阵又是投资组合理论的基础，其估计的精确与否最终会直接干扰与影响投资组合模型.

为了对高维协方差矩阵进行精准的估计，学者们进行了许多相关的研究，主要的思想是对矩阵进行稀疏与降维处理. 如文献[1-2]通过引入不同的门限函数把总体协方差矩阵的一些非对角线元素替换为 0，在保留对角线元素的基础上来避免维数诅咒. 文献[3-4]使用因子模型来降低数据维度，提升其估计效率. 上述 2 类方法虽然可以提升协方差矩阵的估计精确性，但没有从时间变化的角度去分析与研究协方差矩阵，没有将不同时间段的协方差矩阵看成是一组时间序列，也没有考虑因时间变化而产生的信息对于协方差矩阵估计的影响.

文献[5]发现协方差矩阵随着时间的推移有着较强的自回归结构，并且发现预期收益与这一变化密切相关. 随着这一发现，学者们开始将时间序列的估计方法运用于协方差矩阵的估计中，如文献[6]提出的条件自回归威沙特(CAW)模型、文献[7]提出的结合 DCC 框架的双不对称 GARCH-MIDAS 模型、文献[8]提出的 DCC-GARCH 模型. 但上述模型所估计的协方差矩阵维度较低，没有考虑高维的情况. 此外，高频数据也逐渐成为近些年来的研究热点. 文献[9]的研究表明低频协方差矩阵转化为高频协方差矩阵的经济价值是巨大的，但是文中也只是讨论了低维资产配置的问题. 文献[10]提出了基于高频数据的高维协方差矩阵估计模型——VAR-LASSO 模型，该模型在 VAR 模型的基础上，引入 LASSO 惩罚函数，对待估向量的参数进行稀疏处理，来提高模型估计的精确性. 通过实证分析发现，该模型的估计精确性明显优于传统的 DCC 模型与 EWMA 模型. 虽然 VAR-LASSO 模型可以估计高维协方差矩阵，但由于 LASSO 惩罚函数不满足 Oracle 性质，当面对较大的真实未知参数时，会产生较大的估计误差，导致参数估计为有偏估计.

因此，本文拟借鉴文献[10]提出的 VAR-LASSO 方法，尝试对高维已实现协方差矩阵进行建模. 为了克服 LASSO 惩罚函数不满足 Oracle 性质的缺陷，将满足 Oracle 性质的 SCAD 惩罚函数与 MCP 惩罚函数引入 VAR 模型中，提出 VAR-SCAD 模型与 VAR-MCP 模型，并证明两种模型估计参数的 Oracle 性质. 最后用实际高频股票数据对高维已实现协方差矩阵进行建模，使用文献[11]提出的特征值替换方法来确保预测协方差矩阵的正定性，并对模型构建最小方差投资组合模型，探究其在实际投资组合中的应用.

1 模型构建

1.1 符号

对于任意向量 $\boldsymbol{x} \in \mathbb{R}^p$，$\| \boldsymbol{x} \| = \sqrt{\sum_{i=1}^{p} x_i^2}$ 和 $\| \boldsymbol{x} \|_{\ell_1} = \sum_{i=1}^{p} | x_i |$ 分别表示向量 $\boldsymbol{x}$ 的 ℓ_2 范数与 ℓ_1 范数. $\mathscr{A} = \{ i : \boldsymbol{\beta}_i^* \neq \boldsymbol{0} \}$ 表示所有非零 $\boldsymbol{\beta}_i^*$ 的 i 的集合，$|\mathscr{A}|$ 称为其基数.

1.2 已实现协方差矩阵

与传统的低频数据相比，高频数据会包含更为细致与丰富的数据信息，其估计值也更接近真实的协方差矩阵. 本文采用文献[12]提出的方法来构造已实现协方差矩阵(RCOV，简记为 $\boldsymbol{R}$). 假设一个投资组合具有 n 维资产，则其已实现协方差矩阵构造为

$$\boldsymbol{R}_t = \begin{pmatrix} R_{11t} & \cdots & R_{1nt} \\ \vdots & & \vdots \\ R_{n1t} & \cdots & R_{nnt} \end{pmatrix}$$

已实现协方差矩阵的结构与一般的协方差矩阵相类似，其对角线元素为方差项，非对角线元素为协方

差项. R_{ijt} 的具体构造方法可以参见文献[12]，后文中记 R_t 为 $\boldsymbol{\Sigma}_t^{\mathrm{RCOV}}$.

1.3　VAR-LASSO 模型的构建

设 $\boldsymbol{\Sigma}_t^{\mathrm{RCOV}}$ 表示 t 时刻 n 维资产的已实现协方差矩阵，定义 $\boldsymbol{y}_t=\mathrm{vech}(\boldsymbol{\Sigma}_t^{\mathrm{RCOV}})$ 为矩阵拉直算子，返回一个长度为 $k=\dfrac{n(n+1)}{2}$ 的向量，其元素为 $\boldsymbol{\Sigma}_t^{\mathrm{RCOV}}$ 的上三角或下三角元素，$\boldsymbol{\Sigma}_t^{\mathrm{RCOV}}$ 的自回归模型为

$$\boldsymbol{y}_t=\sum_{i=1}^{p}\boldsymbol{\Phi}_i^{*}\boldsymbol{y}_{t-i}+\boldsymbol{\varepsilon}_t \qquad t=1,\cdots,T \tag{1}$$

其中 $\boldsymbol{\Phi}_i^{*}$ 为一个 $k\times k$ 维的参数矩阵，$\boldsymbol{\varepsilon}_t\sim N(0,\boldsymbol{\Sigma})$ 且相互独立，p 表示 VAR 模型的滞后阶数.

观察(1)式可以发现，自回归模型中的待估参数的个数 k 以 n^2 的速度增加，会产生大量的待估参数，导致最小二乘估计惩罚函数的精确度大幅度下降. 为了解决这一问题，文献[11]使用文献[13]提出的 LASSO 惩罚函数来提升估计的精确度. 为了方便模型的估计与计算，在使用 LASSO 惩罚函数之前，可以先对模型进行如下的改写：

令 $\boldsymbol{Z}_t=(\boldsymbol{y}_{t-1}^{\mathrm{T}},\cdots,\boldsymbol{y}_{t-p}^{\mathrm{T}})^{\mathrm{T}}$ 为 $kp\times 1$ 维的解释变量所组成的向量，$\boldsymbol{Z}=(\boldsymbol{Z}_T,\cdots,\boldsymbol{Z}_1)^{\mathrm{T}}$ 为 $T\times kp$ 维的协变量矩阵. $\boldsymbol{y}_i=(y_{T,i},\cdots,y_{1,i})^{\mathrm{T}}(i=1,\cdots,k)$ 为第 i 个观测变量的 $T\times 1$ 维向量，$\boldsymbol{\varepsilon}_i=(\varepsilon_{T,i},\cdots,\varepsilon_{1,i})^{\mathrm{T}}$ 是对应的误差向量. 定义 $\boldsymbol{y}=(\boldsymbol{y}_1^{\mathrm{T}},\cdots,\boldsymbol{y}_k^{\mathrm{T}})^{\mathrm{T}}$，$\boldsymbol{\varepsilon}=(\boldsymbol{\varepsilon}_1^{\mathrm{T}},\cdots,\boldsymbol{\varepsilon}_k^{\mathrm{T}})^{\mathrm{T}}$，设 $\boldsymbol{X}=\boldsymbol{I}_N\otimes\boldsymbol{Z}$，则(1)式可以改写为

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^{*}+\boldsymbol{\varepsilon} \tag{2}$$

变换模型之后，引入 LASSO 惩罚函数，该惩罚函数不但可以将不重要的参数压缩为 0，同时还可以同步进行特征选择，实现数据的降维. 文献[10]中模型(1)的参数 $\boldsymbol{\beta}^{*}$ 通过最小化

$$L(\boldsymbol{\beta})=\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2+T\lambda_T\sum_{i=1}^{k^2p}\|\beta_i\|_{\ell_1} \tag{3}$$

进行估计.

1.4　新模型的提出

为了改进 LASSO 惩罚函数不满足 Oracle 性质的缺点，本文使用满足无偏性的 MCP 惩罚函数与 SCAD 惩罚函数来对 VAR-LASSO 模型进行改进.

1.4.1　SCAD 惩罚函数与 MCP 惩罚函数

文献[14]基于 LASSO 惩罚函数不满足 Oracle 性质提出了如下的 SCAD 惩罚函数：

$$p_{\lambda,a}(\beta_i)^{\mathrm{SCAD}}=\lambda_T\beta_i I(|\beta_i|<\lambda_T)+\frac{a\lambda_T\beta_i-\dfrac{\beta_i^2+\lambda_T^2}{2}}{a-1}+$$

$$I(\lambda_T\leqslant|\beta_i|<a\lambda_T)+\frac{(a+1)\lambda_T^2}{2}I(|\beta_i|\geqslant\lambda_T) \tag{4}$$

其中 $a>2$，$\lambda_T\geqslant 0$. 为了更进一步了解其惩罚的背后含义，可以对(4)式求导，得

$$p'_{\lambda,a}(\beta_i)^{\mathrm{SCAD}}=\lambda_T\left[I(|\beta_i|\leqslant\lambda_T)+\frac{(a\lambda_T-|\beta_i|)_+}{(a-1)\lambda_T}I(|\beta_i|\geqslant\lambda_T)\right] \tag{5}$$

可以发现，当 $|\beta_i|\leqslant\lambda_T$ 时，SCAD 惩罚函数与 LASSO 惩罚函数拥有相同的惩罚力度，但是随着 $|\beta_i|$ 的增加，SCAD 惩罚函数的惩罚力度会逐渐降低，当 $|\beta_i|>a\lambda_T$ 时，惩罚力度降为 0，这保证了较大的 $|\boldsymbol{\beta}|$ 不会被过度地惩罚，确保了较大参数估计的无偏性.

文献[15]提出的 MCP 惩罚函数也同样满足 Oracle 性质，且在处理特征之间有很高相关性数据时，表现要比 SCAD 惩罚函数更好. MCP 惩罚函数的惩罚项为

$$p_{\lambda,a}(\beta_i)^{\mathrm{MCP}}=\lambda_T\int_0^{|\beta_i|}\left(1-\frac{x}{a\lambda_T}\right)_+\mathrm{d}x \tag{6}$$

其中 $a>1$，$\lambda_T\geqslant 0$.

$$p'_{\lambda,a}(\beta_i)^{\mathrm{MCP}}=\lambda_T\left(1-\frac{|\beta_i|}{a\lambda_T}\right)_+\mathrm{sign}(\beta_i) \tag{7}$$

MCP 惩罚函数与 SCAD 惩罚函数的惩罚逻辑类似，随着 $|\beta_i|$ 的增加，惩罚力度会逐渐地降低到 0.

1.4.2 VAR-SCAD模型与VAR-MCP模型

在VAR-LASSO模型的基础上，将(3)式中的$\lambda_T\|\beta_i\|_{l_1}$项代换为(4)式，则VAR-SCAD模型的参数可通过如下函数估计：

$$L(\boldsymbol{\beta})=\|\boldsymbol{y}-\boldsymbol{X\beta}\|^2+T\sum_{i=1}^{k^2p}p_{\lambda,a}(|\beta_i|)^{\mathrm{SCAD}} \tag{8}$$

将(6)式代换为(3)式中的$\lambda_T\|\beta_i\|_{l_1}$项，可以得出VAR-MCP模型的参数估计函数为

$$L(\boldsymbol{\beta})=\|\boldsymbol{y}-\boldsymbol{X\beta}\|^2+T\sum_{i=1}^{k^2p}p_{\lambda,a}(|\beta_i|)^{\mathrm{MCP}} \tag{9}$$

在(8)式与(9)式中，还有未知的参数a需要进行估计. 不同参数a的取值会直接影响(8)式与(9)式的估计性能. 对于(8)式中的参数a，文献[14]通过蒙特卡洛模拟得出a的最优值约等于3.7，(9)式的a在实际的使用中通常默认为3. 通常使用CV法、L曲线法[16]、AIC信息准则等方法对λ_T进行估计，本文使用CV法进行估计.

1.4.3 VAR-SCAD模型与VAR-MCP模型估计参数的Oracle性质

为了后续估计参数Oracle性质的证明，先给出如下4个正则条件：

1) $\varepsilon_{i,1}$具有有限四阶矩，$i=1,\cdots,k$；

2) $\boldsymbol{C}=E\left(\frac{1}{T}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\right)$是正定矩阵；

3) 当$T\to\infty$时，$\sqrt{T}\lambda_T\to\infty$且$\lambda_T\to 0$；

4) 当$T\to\infty$时，$\frac{2\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\varepsilon}}{\sqrt{T}}$与$\frac{2\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{X}}{T}$均有界.

定理1 在条件1)－4)成立的情况下，当$T\to\infty$时，VAR-SCAD模型的估计参数满足如下性质：

(i) $P(\hat{\boldsymbol{\beta}}_{\mathscr{A}^c}=\boldsymbol{0})\to 1$；

(ii) $\sqrt{T}(\hat{\boldsymbol{\beta}}_{\mathscr{A}}-\boldsymbol{\beta}^*_{\mathscr{A}})\xrightarrow[d]{}N(\boldsymbol{0},\ [(\boldsymbol{I}_k\otimes\boldsymbol{C})_{\mathscr{A}}]^{-1}\ [\boldsymbol{\Sigma}\otimes\boldsymbol{C}]_{\mathscr{A}}\ [(\boldsymbol{I}_k\otimes\boldsymbol{C})_{\mathscr{A}}]^{-1})$.

证 令$\boldsymbol{\beta}=\boldsymbol{\beta}^*+\frac{\boldsymbol{\mu}}{\sqrt{T}}$且$\boldsymbol{\mu}=(\mu_1,\ \cdots,\ \mu_{k^2p})^{\mathrm{T}}$，则(9)式可以改写为

$$L_T(\boldsymbol{\mu})=\left\|\boldsymbol{y}-\boldsymbol{X}\left(\boldsymbol{\beta}^*+\frac{\boldsymbol{\mu}}{\sqrt{T}}\right)\right\|^2+T\sum_{i=1}^{k^2p}p_{\lambda,a}\left(\left|\beta_i^*+\frac{\mu_i}{\sqrt{T}}\right|\right)^{\mathrm{SCAD}}$$

设$\hat{\boldsymbol{\mu}}=\arg\min L(\boldsymbol{\mu})$，则可得$\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}^*+\frac{\hat{\boldsymbol{\mu}}}{\sqrt{T}}$.

定义

$$\begin{aligned}&V_T(\boldsymbol{\mu})=L_T(\boldsymbol{\mu})-L_T(\boldsymbol{0})=\\&\left\|\boldsymbol{y}-\boldsymbol{X}\left(\boldsymbol{\beta}^*+\frac{\boldsymbol{\mu}}{\sqrt{T}}\right)\right\|^2-\|\boldsymbol{y}-\boldsymbol{X\beta}^*\|^2+T\sum_{i=1}^{k^2p}\left(p_{\lambda,a}\left(\left|\beta_i^*+\frac{\mu_i}{\sqrt{T}}\right|\right)^{\mathrm{SCAD}}-p_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}\right)=\\&\boldsymbol{\mu}^{\mathrm{T}}\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}{T}\boldsymbol{\mu}-2\frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\varepsilon}}{\sqrt{T}}+T\sum_{i=1}^{k^2p}\left(p_{\lambda,a}\left(\left|\beta_i^*+\frac{\mu_i}{\sqrt{T}}\right|\right)^{\mathrm{SCAD}}-p_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}\right)\end{aligned}$$

由文献[17]中的定理11.2.1与文献[18]中第十章的定理1可得

$$\boldsymbol{\mu}^{\mathrm{T}}\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}{T}\boldsymbol{\mu}\xrightarrow[p]{}\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}_k\otimes\boldsymbol{C})\boldsymbol{\mu}\qquad\frac{\boldsymbol{X}'\boldsymbol{\varepsilon}}{\sqrt{T}}\xrightarrow[d]{}\boldsymbol{w}\sim N(\boldsymbol{0},\ \boldsymbol{\Sigma}\otimes\boldsymbol{C})$$

则

$$\boldsymbol{\mu}^{\mathrm{T}}\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}{T}\boldsymbol{\mu}-2\frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\varepsilon}}{\sqrt{T}}\xrightarrow[d]{}\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}_k\otimes\boldsymbol{C})\boldsymbol{\mu}-2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w}$$

记$\mathscr{T}=T\sum_{i=1}^{k^2p}\left(p_{\lambda,a}\left(\left|\beta_i^*+\frac{\mu_i}{\sqrt{T}}\right|\right)^{\mathrm{SCAD}}-p_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}\right)$. 当$T\to\infty$时，对$T$使用中值定理，可得

$$\mathscr{T}=T\sum_{i=1}^{k^2p}\left(p_{\lambda,a}\left(\left|\beta_i^*+\frac{\mu_i}{\sqrt{T}}\right|\right)^{\mathrm{SCAD}}-p_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}\right)\to\sqrt{T}\sum_{i=1}^{k^2p}p'_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}\mathrm{sgn}(\beta_i^*)\mu_i$$

令 $\sum_{i=1}^{k^2p}\mathscr{T}_i=\sum_{i=1}^{k^2p}p'_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}\mathrm{sgn}(\beta_i^*)\mu_i$，展开 $p'_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}$，则可得

$$\mathscr{T}_i=\sqrt{T}\mu_i\lambda_T\left[I(|\beta_i^*\leqslant\lambda_T|)+\frac{(a\lambda_T-|\beta_i^*|)_+}{(a-1)}I(|\beta_i^*|>\lambda_T)\right]\mathrm{sgn}(\beta_i^*)$$

当 $|\beta_i^*|\geqslant a\lambda_T$ 时，$p'_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}=0$，则 $\mathscr{T}_i=0$.

当 $\lambda_T\leqslant|\beta_i^*|<a\lambda_T$ 时，由于 $\lambda_T\to0$，有 $|\beta_i^*|\to0$. 得出 $p'_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}=0$，则 $\mathscr{T}_i=0$.

当 $0<|\beta_i^*|\leqslant\lambda_T$ 且 $\mu_i\neq0$ 时，$p'_{\lambda,a}(|\beta_i^*|)^{\mathrm{SCAD}}=\mathrm{sgn}(\beta_i)\lambda_T$. 结合假设 2) 可得 $\mathscr{T}_i=\infty$.

当 $0<|\beta_i^*|\leqslant\lambda_T$ 且 $\mu_i=0$ 时，可得 $\mathscr{T}_i=0$.

综上所述，由 Slutsky 定理可得

$$V_T(\boldsymbol{\mu})\xrightarrow[d]{}V(\boldsymbol{\mu})=\begin{cases}\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}_k\otimes\boldsymbol{C})\boldsymbol{\mu}-2\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w} & \mu_i=0\text{ 且 }i\in\mathscr{A}^c,\\ \infty & \text{否则}\end{cases}$$

由于 $V_T(\boldsymbol{\mu})$ 是一个凸函数，且 $V(\boldsymbol{\mu})$ 有唯一最小值点 $(\boldsymbol{I}_k\otimes\boldsymbol{C})^{-1}\boldsymbol{w}$，可得

$$\hat{\boldsymbol{\mu}}_{\mathscr{A}}\xrightarrow[d]{}N(\boldsymbol{0},\ [(\boldsymbol{I}_k\otimes\boldsymbol{C})_{\mathscr{A}}]^{-1}\ [\boldsymbol{\Sigma}\otimes\boldsymbol{C}]_{\mathscr{A}}\ [(\boldsymbol{I}_k\otimes\boldsymbol{C})_{\mathscr{A}}]^{-1})$$

又因为 $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}^*+\frac{\hat{\boldsymbol{\mu}}}{\sqrt{T}}$，可得 $\hat{\boldsymbol{\mu}}=\sqrt{T}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\in O_p(1)$，则

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{\mathscr{A}}-\boldsymbol{\beta}^*_{\mathscr{A}})\xrightarrow[d]{}N(\boldsymbol{0},\ [(\boldsymbol{I}_k\otimes\boldsymbol{C})_{\mathscr{A}}]^{-1}\ [\boldsymbol{\Sigma}\otimes\boldsymbol{C}]_{\mathscr{A}}\ [(\boldsymbol{I}_k\otimes\boldsymbol{C})_{\mathscr{A}}]^{-1})$$

定理 1(ii) 得证，下面证明定理 1(i).

假定对于 $i\in\mathscr{A}^c$，$\hat{\beta}_i\neq0$，由(9) 式可得

$$\frac{2x_i^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\sqrt{T}}+\sqrt{T}\sum_{i=1}^{k^2p}p'_{\lambda,a}(|\hat{\beta}_i|)^{\mathrm{SCAD}}\mathrm{sgn}(\hat{\beta}_i)=0$$

当 $|\hat{\beta}_i|\geqslant a\lambda_T$ 时，$p'_{\lambda,a}(|\hat{\beta}_i|)^{\mathrm{SCAD}}=0$，则

$$\frac{2x_i^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\sqrt{T}}=\frac{2x_i^T[\boldsymbol{\varepsilon}-\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)]}{\sqrt{T}}=\frac{2x_i^T\boldsymbol{\varepsilon}}{\sqrt{T}}-\frac{2x_i^T\boldsymbol{X}}{T}\sqrt{T}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)$$

由假设 4) 可得

$$\frac{2x_i^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\sqrt{T}}\to O_p(1)$$

当 $0<|\hat{\beta}_i|\leqslant\lambda_T$ 时，$p'_{\lambda,a}(|\hat{\beta}_i|)^{\mathrm{SCAD}}=\mathrm{sgn}(\hat{\beta}_i)\lambda_T$，则

$$\left|\sqrt{T}\sum_{i=1}^{k^2p}p'_{\lambda,a}(|\hat{\beta}_i|)^{\mathrm{SCAD}}\mathrm{sgn}(\hat{\beta}_i)\right|=\left|\sqrt{T}\ \mathrm{sgn}(\hat{\beta}_i)\lambda_T\right|=\sqrt{T}\lambda_T\to\infty$$

当 $\lambda_T\leqslant|\hat{\beta}_i|<a\lambda_T$ 时，有

$$\left|\sqrt{T}\sum_{i=1}^{k^2p}p'_{\lambda,a}(|\hat{\beta}_i|)^{\mathrm{SCAD}}\mathrm{sgn}(\hat{\beta}_i)\right|=\left|\sqrt{T}\frac{2a\lambda_T-|\hat{\beta}_i|}{(a-1)}\mathrm{sgn}(\hat{\beta}_i)\right|=\sqrt{T}\frac{2a\lambda_T-|\hat{\beta}_i|}{(a-1)}\leqslant$$

$$\sqrt{T}\frac{(2a-1)\lambda_T}{(a-1)}\to\infty$$

综上所述，可以得

$$p(\hat{\beta}_i\neq0)\leqslant p\left[\frac{2x_i^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\sqrt{T}}+\sqrt{T}\sum_{i=1}^{k^2p}p'_{\lambda,a}(|\hat{\beta}_i|)^{\mathrm{SCAD}}\mathrm{sgn}(\hat{\beta}_i)=0\right]\to0$$

定理 2　在条件 1) — 4) 成立的情况下，当 $T\to\infty$ 时，VAR-MCP 模型的估计参数满足如下性质：

(i) $P(\hat{\boldsymbol{\beta}}_{\mathscr{A}^c}=\mathbf{0})\to 1$；

(ii) $\sqrt{T}(\hat{\boldsymbol{\beta}}_{\mathscr{A}}-\boldsymbol{\beta}^*_{\mathscr{A}})\underset{d}{\to} N(\mathbf{0},\ [(\boldsymbol{I}_k\otimes \boldsymbol{C})_{\mathscr{A}}]^{-1}\ [\boldsymbol{\Sigma}\otimes \boldsymbol{C}]_{\mathscr{A}}\ [(\boldsymbol{I}_k\otimes \boldsymbol{C})_{\mathscr{A}}]^{-1})$.

定理 2 的证明过程与定理 1 类似，区别在于两种模型的惩罚项，结合定理 1 的证明过程可得证定理 2.

2 实证分析

2.1 数据来源与清洗

选取标普 500 指数中的 18 只成分股进行实证分析研究. 为了使选取的股票具有一定的波动性，同时没有明显的牛熊趋势，选取区间为 2011 年 6 月 1 日到 2013 年 6 月 3 日，共有 504 个股票交易日. 使用上文提到的 RCOV 方法，在股票交易时间内，每 5 分钟获得一次对数交易价格，并构建已实现协方差矩阵，一共可以得到 504 个已实现协方差矩阵. 并对 504 个已实现协方差矩阵进行数据清洗. 具体方法如下：对任意的一个已实现协方差矩阵，如果其中某一个元素的值大于其本身所在矩阵所有元素均值的正负 3 倍，则对该元素与该元素所在矩阵进行标记；如果某一已实现协方差矩阵有 30% 以上的元素被标记，则删除这个被标记的矩阵，用其相邻 5 个未被标记的矩阵的均值来代替这个被标记矩阵. 条件 1) 中 $\boldsymbol{\varepsilon}=\boldsymbol{y}-\hat{\boldsymbol{y}}$，因为使用股票收盘价数据，所以模型误差项不会出现无穷的情况，结合后文的表 1 也可以得到这样的结果，从而具有有限四阶矩. 对于条件 2)，因为选用实际的股票数据不会有任意两种股票的收益率在 504 个交易日内完全一样，这导致股票之间的协变量是非线性相关的，从而使得矩阵 $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ 中没有共线的行与列，所以 $\boldsymbol{C}$ 正定.

2.2 测试误差

记$\hat{\boldsymbol{e}}_{T+h}=\mathrm{vech}(\hat{\boldsymbol{\Sigma}}_{T+h}-\boldsymbol{\Sigma}_{T+h})$，$\hat{\boldsymbol{\Sigma}}^v_{T+h}$ 为$\hat{\boldsymbol{e}}_{T+h}$ 按行组成的矩阵，其中 $h\geqslant 1$ 为向后预测的区间长度，在实证分析中将 h 分别设置为 1,5,10,20，以研究 h 对模型测试精度的影响. T_1 与 T_2 分别为测试开始与结束的时间. 在对比 3 种模型的测试精度时，采用 ℓ_2 测试误差(ℓ_2)、平均最大绝对误差(AMaxE)、F 范数(F) 与平均测试误差中位数(AMedE) 这 4 种指标.

$$\ell_2=\|\hat{\boldsymbol{e}}_{T+h}\| \qquad \mathrm{AMaxE}=\frac{1}{T_2-T_1+1}\sum_{T=T_1}^{T_2}\max(|\hat{\boldsymbol{e}}_{T+h}|)$$

$$F=\|\hat{\boldsymbol{\Sigma}}^v_{T+h}\|_F \qquad \mathrm{AMedE}=\frac{1}{T_2-T_1+1}\sum_{T=T_1}^{T_2}\mathrm{median}(\hat{\boldsymbol{e}}_{T+h})$$

上述 4 种指标越小，代表模型的估计精确度越高，ℓ_2 测试误差对比如图 1、图 2 所示.

将清洗后的数据划分为训练集与测试集两个部分，其中训练集 404 天，测试集 100 天. 在 VAR 模型的拟合中，默认 VAR 模型的滞后阶数为 1.

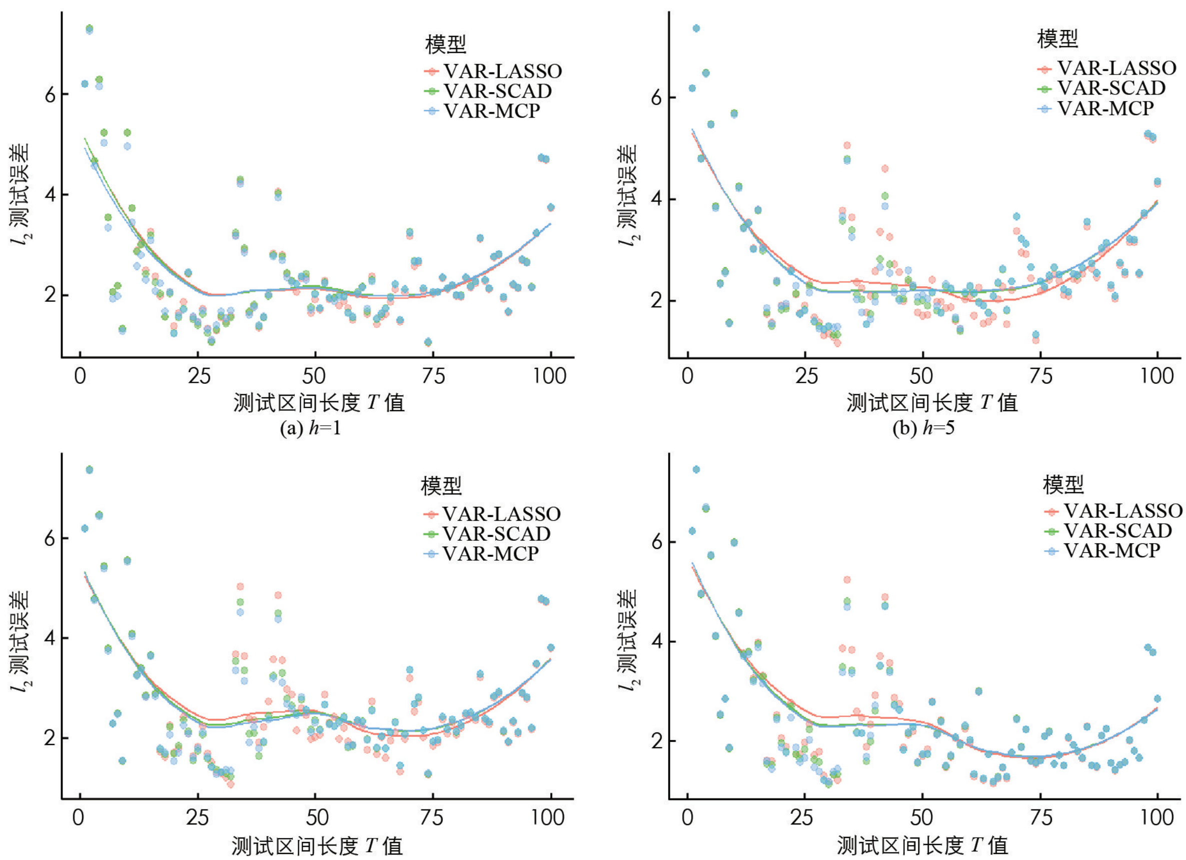
表 1 3 种模型的测试误差比较

模型	h	F	AMaxE	AMedE
VAR-LASSO	1	26.925 0	0.958 6	0.016 9
VAR-SCAD		27.036 8	0.952 3	0.016 8
VAR-MCP		26.663 7	0.944 2	0.016 5
VAR-LASSO	5	29.577 9	1.0249	0.005 5
VAR-SCAD		29.599 3	0.963 3	0.002 3
VAR-MCP		29.572 0	0.956 0	0.001 5
VAR-LASSO	10	29.070 3	1.158 9	0.021 9
VAR-SCAD		28.997 0	1.149 9	0.021 0
VAR-MCP		28.754 7	1.146 3	0.020 5
VAR-LASSO	20	27.992 8	1.056 7	0.016 0
VAR-SCAD		27.503 8	1.041 3	0.013 9
VAR-MCP		27.383 7	1.034 1	0.013 4

观察表 1 可得，VAR-LASSO 模型在 $h=1,5$ 的情况下，F 范数(F)分别为 26.925 0,29.577 9，而

VAR-SCAD 模型仅为 27.036 8，29.599 3，VAR-LASSO 模型的表现要优于 VAR-SCAD 模型. 但在平均最大绝对误差（AMaxE）与平均测试误差中位数（AMedE）这两项指标上，VAR-LASSO 模型均最大，表现逊色于 VAR-MCP 模型与 VAR-SCAD 模型. 当 $h=10,20$ 时，VAR-LASSO 模型的 F 范数（F）分别为 29.070 3 与 27.992 8，拥有 3 种模型中最差的测试精度. 虽然 VAR-SCAD 模型在 $h=1,5$ 时，从指标 F 范数（F）来看，VAR-SCAD 模型的优势相较于 VAR-LASSO 模型不明显，但是 $h=10,20$ 时，从表 1 的 3 种指标来看，VAR-SCAD 模型的测试精确度仅次于 VAR-MCP 模型. 无论 h 取何值，VAR-MCP 模型均拥有最优的估计精确性，3 种评价指标均最小.

更为直观的信息可以从图 1 与图 2 中获得. 观察图 1 可以发现，当 $h=1$ 时，3 种模型的测试误差差异不大，VAR-LASSO 模型与 VAR-SCAD 模型在前期的测试误差几乎重合，VAR-MCP 模型的测试误差与其误差曲线都明显较低，在测试的后期，3 种模型的测试误差十分贴近，3 条线几乎重合. $h=5,10$ 时，3 种模型的测试误差光滑曲线走势相同，在测试前期，VAR-MCP 模型与 VAR-SCAD 模型的测试误差点与其误差曲线较低，但当 $T=50$ 时，VAR-LASSO 模型的表现反而会优于 VAR-MCP 模型与 VAR-SCAD 模型. 当 $h=20$ 时，可以明显看出 VAR-MCP 模型与 VAR-SCAD 模型的测试误差要小于 VAR-LASSO 模型. 不同 h 情况下的 3 种模型的测试误差光滑曲线走势大致相同，都是在测试前期误差较大，测试中期误差会逐渐减小，到了测试后期，误差又会慢慢增大.



(a) h=1　(b) h=5　(c) h=10　(d) h=20

图 1　3 种模型在不同 h 下的误差光滑曲线图

图 2 是小提琴图，图形内部是一个箱线图，箱线图外部颜色的宽度代表着数值分布的密集程度. 由图 2 可知，相较于 VAR-LASSO 模型，VAR-MCP 模型的异常值较小且集中. VAR-SCAD 模型在 $h=1,5$ 时，异常值的大小与分布与 VAR-LASSO 模型类似，而当 $h=10,20$ 时，VAR-SCAD 模型的异常值的大小与分布和 VAR-MCP 模型类似. 同时也可以发现，当 $h=20$ 时，VAR-LASSO 模型的上四分位数偏大，说明 VAR-LASSO 模型较大测试误差值的数量要多于 VAR-SCAD 模型与 VAR-MCP 模型. VAR-SCAD 模型与 VAR-MCP 模型的小提琴图的底部相较于 VAR-LASSO 模型来说更为圆润与宽大，表明这两种模型的

测试误差分布相对集中，模型也更为稳健.

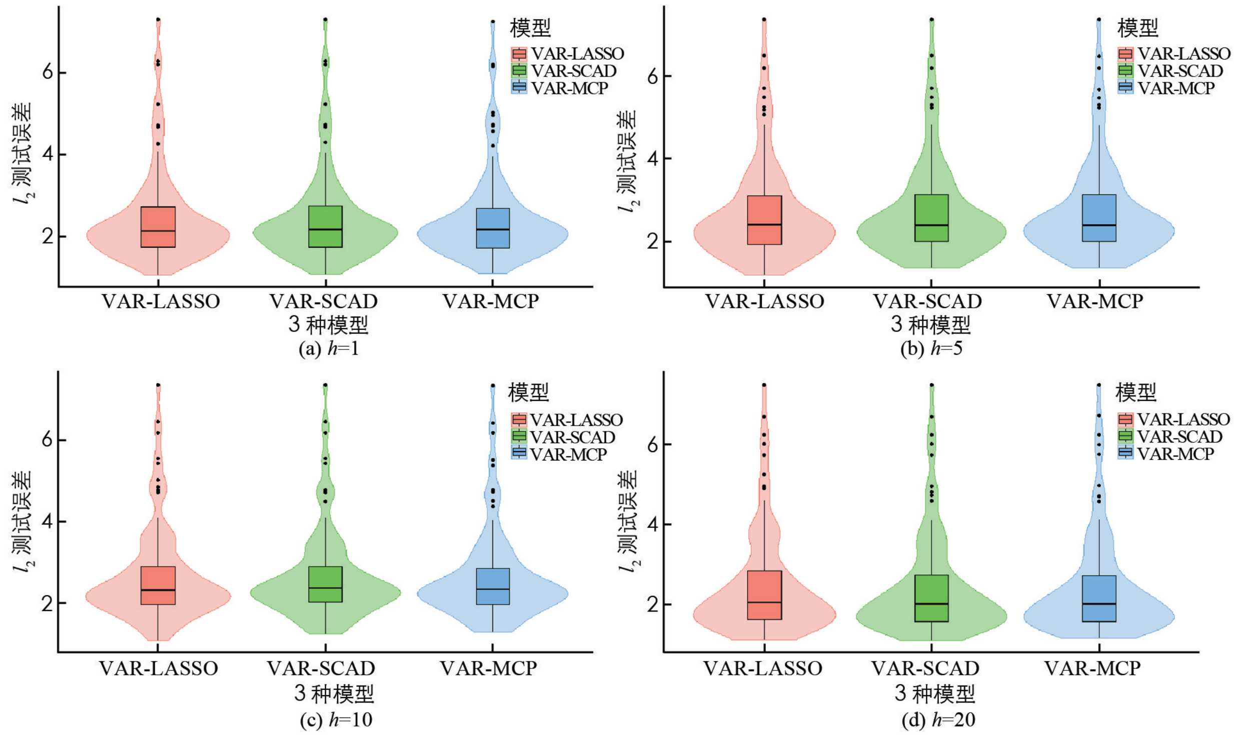


图 2　3 种模型在不同 h 下的小提琴图

综上所述，相较于 VAR-LASSO 模型，VAR-SCAD 模型与 VAR-MCP 模型的测试精确性有着明显的提升，VAR-SCAD 模型与 VAR-MCP 模型的稳健性也要优于 VAR-LASSO 模型. VAR-SCAD 模型在 h 较大的情况下表现较好，VAR-MCP 模型全局表现最优.

2.3　VAR-SCAD 模型表现分析

文献[19]研究指出，在变量具有相关性的情况下，SCAD 惩罚函数的表现会不如 LASSO 惩罚函数，下面从股票相关性角度分析 VAR-SCAD 模型表现不佳的原因.

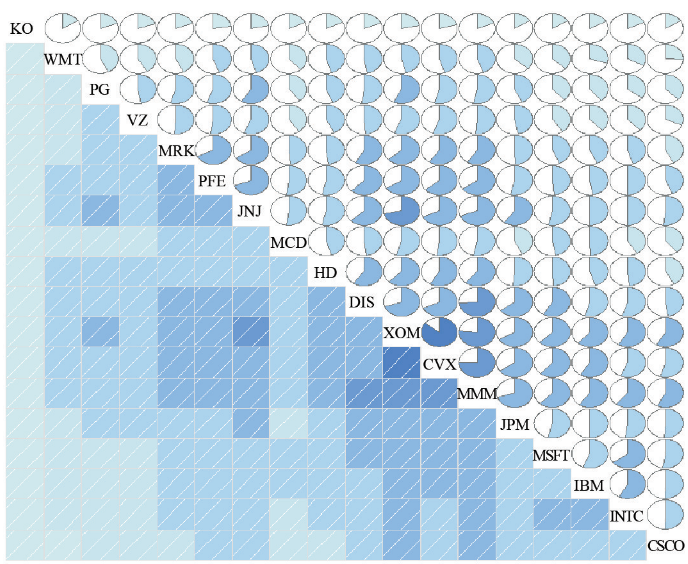


图 3　18 种股票的相关性图

图 3 中间为股票名，上三角单元格饼图的填充程度代表相关系数的大小，下三角单元格为相关系数热力图. 18 支股票的最低相关度为 0.16，51%的股票之间的相关性要高于 0.5，这导致 VAR-SCAD 模型的参

数虽然具有 Oracle 性质，但测试精确性在 h 较小的情况下不如 VAR-LASSO 模型，更不如可以较好处理相关性数据的 VAR-MCP 模型，这也与文献[19]的研究结果保持一致.

3 投资组合中的应用

采用上文 18 种股票的已实现协方差矩阵构建均值—方差投资组合模型. 选取 100 个已实现协方差矩阵进行投资组合模型的模拟研究，在投资组合分析中，将 h 分别设置为 1 与 10.

假定股票交易中没有手续费的产生，无风险收益为 0. 投资组合模型的表达式为

$$\min_w \boldsymbol{w}_{t+1}^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{t+1} \boldsymbol{w}_{t+1} \qquad \text{s. t. } \boldsymbol{w}_{t+1}^{\mathrm{T}} \mathbf{1} = 1,\ \boldsymbol{w}_{t+1}^{\mathrm{T}} \widehat{\boldsymbol{\mu}}_{t+1} = q$$

其中 $\mathbf{1}$ 为元素均为 1 的向量，q 是投资组合所需的预期回报率，$\boldsymbol{w}_{t+1}$ 是投资组合中股票的权重，$\widehat{\boldsymbol{\mu}}_{t+1}$ 为资产的期望收益率向量，其值通过 50 天移动平均取得.

基于 VAR 模型的 3 种模型在估计过程中都没有考虑到$\widehat{\boldsymbol{\Sigma}}_T$ 矩阵的正定性，会导致投资组合模型求解中出现病态协方差矩阵，干扰最优权重 $\boldsymbol{w}_{t+1}$ 的求解. 为了确保$\widehat{\boldsymbol{\Sigma}}_T$ 的正定性，需要对$\widehat{\boldsymbol{\Sigma}}_T$ 采用文献[11] 提出的特征值替换方法对特征值$\lambda \leqslant 0$ 的矩阵进行正定化处理. 首先对$\widehat{\boldsymbol{\Sigma}}_T$ 进行谱分解：$\widehat{\boldsymbol{\Sigma}}_T = \widehat{\boldsymbol{V}}_T^{\mathrm{T}} \widehat{\boldsymbol{\Lambda}}_T \widehat{\boldsymbol{V}}_T$，其中$\widehat{\boldsymbol{\Lambda}}_T$ 为对角线元素为特征值的对角矩阵，$\widehat{\lambda}_{iT}$ 为其对角线元素. 最后设$\lambda_{mT} = \min\{\widehat{\lambda}_{iT} \mid \widehat{\lambda}_{iT} > 0\}$，$\lambda_{mT}$ 是$\widehat{\boldsymbol{\Lambda}}_T$ 中最小的非负特征值. 对于满足$\widehat{\lambda}_{iT} < \lambda_{mT}$ 的特征值，均用λ_{mT} 代替，特征值替换后的对角矩阵用$\check{\boldsymbol{\Lambda}}_T$ 表示，则正定处理后的协方差矩阵$\widehat{\boldsymbol{\Sigma}}_T$ 记为$\check{\boldsymbol{\Sigma}}_T = \widehat{\boldsymbol{V}}_T^{\mathrm{T}} \check{\boldsymbol{\Lambda}}_T \widehat{\boldsymbol{V}}_T$.

为了对比 3 种模型构建的投资组合模型的绩效，选用夏普比率(SP)这一指标对投资组合模型进行评价. 夏普比率表示单位风险所带来的收益，夏普比率值越大表示投资组合的收益越好.

通过表 2 可以发现，在各种情况之下，VAR-MCP 模型构建的投资组合均拥有最大的夏普比率. 在 $h=1$ 时，VAR-LASSO 模型构建的投资组合的夏普比率为 0.101 010 1，仅略大于 VAR-SCAD 模型构建的投资组合的夏普比率. 当 $h=10$ 时，VAR-SCAD 模型构建的投资组合的夏普比率为 0.101 037 2，略小于 VAR-MCP 模型构建的投资组合的夏普比率 0.102 079 6，此时 VAR-LASSO 模型构建的投资组合的表现最差，拥有最小的夏普比率 0.101 003 9.

表 2　3 种模型的投资组合的夏普比率

	VAR-LASSO	VAR-SCAD	VAR-MCP
$h=1$	0.101 010 1	0.101 010 0	0.101 698 3
$h=10$	0.101 003 9	0.101 037 2	0.102 079 6

4 结论

随着信息技术的发展与数据可获取性的提高，金融数据的维度与频率都呈现出快速增长的趋势. 基于高频金融数据，本文在 VAR-LASSO 模型的基础上，将非凸惩罚函数即 MCP 惩罚函数与 SCAD 惩罚函数引入 VAR-LASSO 模型，得到了新的 VAR-SCAD 模型与 VAR-MCP 模型，并证明了 VAR-SCAD 模型与 VAR-MCP 模型估计参数的 Oracle 性质. 使用 VAR-SCAD 模型与 VAR-MCP 模型对高维已实现协方差矩阵进行建模，通过高频股票数据的实证研究，发现将 SCAD 惩罚函数与 MCP 惩罚函数引入 VAR 模型后，较好地克服了 LASSO 惩罚函数面对较大真实未知参数会产生较大估计误差的缺点，VAR-MCP 模型在不同 h 的情况下均拥有最小的测试误差. VAR-SCAD 模型在 h 较大的情况下表现优于 VAR-LASSO 模型，但在 h 较小的情况下，因为股票数据之间的高相关性，VAR-SCAD 模型的表现不如 VAR-LASSO 模型.

最后通过构建均值—方差投资组合模型可以发现，VAR-MCP 模型构建的投资组合可以为投资人带来最高的经济收益. VAR-SCAD 模型构造的投资组合在 h 较大的情况下，表现仅次于 VAR-MCP 模型构造的投资组合.

参考文献:

[1] ROTHMAN A J, LEVINA E, ZHU J. Generalized Thresholding of Large Covariance Matrices [J]. Journal of the American Statistical Association, 2009, 104(485): 177-186.

[2] CAI T, LIU W D. Adaptive Thresholding for Sparse Covariance Matrix Estimation [J]. Journal of the American Statistical Association, 2011, 106(494): 672-684.

[3] FAN J, FAN Y, LV J. High Dimensional Covariance Matrix Estimation Using a Factor Model [J]. Journal of Econometrics, 2008, 147(1): 186-197.

[4] 陈钊, 范剑青, 王丹. 高维因子模型及其在统计机器学习中的应用 [J]. 中国科学(数学), 2020, 50(4): 447-490.

[5] BOLLERSLEV T, ENGLE R F, WOOLDRIDGE J M. A Capital Asset Pricing Model with Time-Varying Covariances [J]. Journal of Political Economy, 1988, 96(1): 116-131.

[6] GOLOSNOY V, GRIBISCH B, LIESENFELD R. The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility [J]. Journal of Econometrics, 2012, 167(1): 211-223.

[7] CANDILA V. Multivariate Analysis of Cryptocurrencies [J]. Econometrics, 2021, 9(3): 28-28.

[8] 周亮. 基于DCC-GARCH模型的协方差矩阵预测 [J]. 统计与决策, 2021, 37(20): 35-38.

[9] FLEMING J, KIRBY C, OSTDIEK B. The Economic Value of Volatility Timing Using "Realized" Volatility [J]. Journal of Financial Economics, 2003, 67(3): 473-509.

[10] CALLOT L A F, KOCK A B, MEDEIROS M C. Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice [J]. Journal of Applied Econometrics, 2017, 32(1): 140-158.

[11] HAUTSCH N, KYJ L M, HAUTSCH N. A Blocking and Regularization Approach to High Dimensional Realized Covariance Estimation [J]. Journal of Applied Econometrics, 2012, 27(4): 625-645.

[12] LUNDE A, SHEPHARD N, SHEPPARD K. Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels and Their Application to Portfolio Choice [J]. Journal of Business & Economic Statistics, 2016, 34(4): 504-518.

[13] TIBSHIRANI R. Regression Shrinkage and Selection Via the Lasso [J]. Journal of the Royal Statistical Society(Series B), 1996, 58(1): 267-288.

[14] FAN J Q, LI R Z. Variable Selection Via Nonconcave Penalized Likelihood and Its Oracle Properties [J]. Journal of the American Statistical Association, 2001, 96(456): 1348-1360.

[15] ZHANG C H. Nearly Unbiased Variable Selection Under Minimax Concave Penalty [J]. The Annals of statistics, 2010, 38(2): 894-942.

[16] 吴炜明, 王延新. 基于L曲线方法的Lasso正则化参数选择 [J]. 西南师范大学学报(自然科学版), 2022, 47(1): 36-42.

[17] BROCKWELL P J, DAVIS R A. Time Series: Theory and Methods [M]. New York: Springer-Verlag, 2015.

[18] CALLOT L A F, KOCK A B. Oracle Efficient Estimation and Forecasting with the Adaptive Lasso and the Adaptive Group Lasso in Vector Autoregressions [J]. Essays in Nonlinear Time Series Econometrics, 2014, 2014: 238-268.

[19] ZOU H, ZHANG H H. On the Adaptive Elastic-net with a Diverging Number of Parameters [J]. Annals of Statistics, 2009, 37(4): 1733-1751.

责任编辑 廖坤