

参赛队号：

2021 年（第七届）
全国大学生统计建模大赛

参赛学校：天津财经大学

论文题目：基于网络的高维数据动态相依性测度及其信
息含量研究

参赛队员：贺于蘋 周鑫 吴林娜

指导老师：袁铭

目 录

一、引言.....	1
二、文献综述.....	2
(一) 国内外研究综述.....	2
(二) 文献评述.....	2
(三) 本文主要创新.....	4
三、研究思路.....	5
四、研究方法.....	6
(一) R-vine Copula 及其参数估计	6
(二) 基于 Graphical-Lasso 下 R-vine 结构选择的改进	8
(三) 基于网络特征的相依性描述.....	11
(四) 基于稀疏 Group-Lasso 的预测模型.....	13
五、实证研究.....	14
(一) 数据描述.....	14
(二) 基于 Graphical-Lasso 的 R-vine Copula 模型估计.....	15
(三) 稳健性检验.....	19
(四) 基于加权复杂网络的信息提取.....	20
(五) 基于回归的信息含量研究.....	24
六、研究结论.....	28
参考文献.....	29
附录一 股票基本信息.....	34
致 谢.....	36

表目录

表 1	pair-copula 函数对照表	8
表 2	第一窗口期各股收益率分布拟合统计量	16
表 3	第二窗口期各股收益率分布拟合统计量	16
表 4	第三窗口期各股收益率分布拟合统计量	16
表 5	第一窗口期藤 Copula 拟合结果	18
表 6	第二窗口期藤 Copula 拟合结果	18
表 7	第三窗口期藤 Copula 拟合结果	18
表 8	稳健性检验	19
表 9	市场收益率研究特征子集表	24
表 10	各指标对沪深 300 指数收益率的回归结果	25
表 11	市场波动率研究特征子集表	26
表 12	各指标对沪深 300 指数波动率的回归结果	27

图目录

图 1	沪深 300 指数收盘价	15
图 2	第一窗口期、第二窗口期、第三窗口期相关系数热力图	15
图 3	不同惩罚因子对应的无向图结构	17
图 4	最佳惩罚因子及其无向图结构	17
图 5	第一窗口期藤结构前两层可视化	20
图 6	第二窗口期藤结构前两层可视化	21
图 7	第三窗口期藤结构前两层可视化	22
图 8	各网络指标与市场收益率及波动率走势图	23

摘要

资产价格之间的动态相依性通常蕴含着股市波动的信息，在金融产品定价、投资组合、市场风险管控等方面得到了广泛应用，因此对其进行测度一直是学界研究的热点问题。由于资产价格之间的相依性具有复杂性、高维性以及及时变性等特点，目前学界尚没有完美的测度方案，因此本文聚焦于金融资产相依性的测度问题，结合滑窗技术与藤 Copula 模型构建了一个测度金融资产动态相依性的新方法，并对其进行了稳健性检验和信息含量研究，为该研究领域提供了新思路。

在方法构建上，本文利用核密度估计沪深 300 成分股的边缘分布，通过高维 R 藤 Copula 模型构建其动态相依结构。为了解决高维下的计算复杂度问题，在传统的 Dißmann 算法上引入 Graphical-lasso 算法改进藤 Copula 的模型选择与估计过程。在信息含量研究上，本文将藤结构视作加权复杂网络，取藤结构中更为显著的前两棵树并采用加权中心性等相关指标描述其网络特征。随后引入传统预测因子建立稀疏 Group-lasso 回归模型，分别对市场收益率和波动率进行预测，同时将该预测模型与以总体简单平均相关系数作为相依性测度指标的预测模型进行对比，进而验证本文构建的动态相依结构所蕴含的市场信息是否有所提高。

基于上述过程得出以下结论：第一，本文基于藤 Copula 模型对金融资产相依性进行测度，改进了藤 Copula 模型结构的选择，提高了估计效率与模型优度。第二，股市在不同时期，股票间整体与局部显示出不同特点的相依关系。第三，在各网络特征指标中，尾部相关系数对于市场震荡反映更加敏感。第四，利用本文方法构建的网络特征指标对市场收益率与波动率解释力的提升比总体简单平均相关系数更加显著，即该模型挖掘了更多的市场信息。

关键词：R 藤 Copula；动态相依；Graphical-Lasso；加权网络；稀疏 Group-Lasso

Abstract

The dynamic dependence between asset prices usually contains the information of stock market fluctuations, which has been widely used in financial product pricing, portfolio, market risk management and control, so its measurement has been a hot issue in academic research. Due to the complexity, high-dimensional and time-varying characteristics of the dependence between asset prices, there is no perfect measurement scheme in the academic circle. Therefore, this paper focuses on the measurement of the dependence of financial assets, which provides a new idea for this research field, constructs a new method to measure the dynamic dependence of financial assets by combining the sliding window technology and the rattan Copula model, and studies its robustness test and information content.

In terms of method construction, this paper uses kernel density to estimate the marginal distribution of CSI 300 component stocks, and constructs its dynamic dependent structure through high-dimensional R-vine Copula model. In order to solve the problem of computational complexity in high dimension, the traditional Dißmann algorithm is introduced to improve Copula's model selection and estimation process. In the research of information content, the rattan structure is regarded as a weighted complex network, and the first two trees in the rattan structure are selected, and the weighted centrality and other related indicators are used to describe the network characteristics. Then, the traditional predictors are introduced to establish a sparse Group-Lasso regression model to predict the market return and volatility respectively. At the same time, the prediction model is compared with the prediction model with the overall simple average correlation coefficient as the dependency measure index to verify whether the market information contained in the dynamic dependency structure constructed in this paper is improved.

Based on the above process, the following conclusions are drawn: firstly, this paper measures the dependence of financial assets based on vine-copula model, improves the choice of vine-copula model structure, improves the estimation efficiency and model superiority. Second, in different periods of the stock market, the whole and part of the stock market show different characteristics of dependence. Third, among the network characteristics, tail correlation coefficient is more sensitive to market shocks. Fourth, the explanatory power of the network characteristic index constructed by this method to the market return and volatility is more significant than the overall simple average correlation coefficient, that is, the model excavates more market information.

Key Word: R-Vine Copula; Dynamic Dependence; Graphical-Lasso ; Weighted Network; Sparse Group-Lasso

基于网络的高维数据动态相依性测度及其信息含量研究

一、引言

在经济数字化的背景下，各行业金融资产之间的协同性日益增强，资产价格之间的动态相依性通常蕴含着市场收益与市场风险信息。其相依性的测度在金融产品定价、资产配置、市场风险管控等方面有着广泛的应用。但现有研究中，学界主要将目光放在较少数量的金融资产相依性的探究中，而很少关注到以数字化经济为背景加速发展及金融运转经营积累下来的大量数据资源，及其积聚而成的高维资产价格之间的相依关系。

现有文献表明，由于金融资产价格间的相依性具有复杂性、高维性以及时变性等特点，作为金融资产相依性测度经典代表的线性相关系数失效，如何确定高维变量的联合分布形式，以及怎样找到较线性相关系数更有效的相依性度量，成为了资产相依性研究中的两个重要问题。目前，学界对于金融资产的相依性关系和结构仍没有较好的测度方法。传统模型如 GARCH 模型族、Copula 模型，其对资产相依性的度量存在约束性较强、无法捕捉尾部相依性、维数灾难等问题。针对 GARCH 模型族和 Copula 模型的局限性，Bedford 和 Cooke(2001)提出了藤 Copula 模型，藤 Copula 利用树结构（藤结构）解决了维数灾难问题且在描述高维变量的联合分布时具有更强的灵活性，成功地用于金融市场复杂相依结构的度量问题（ref），但该方法仍存在计算复杂度高、非全局最优导致的误差传播等问题。

在实证研究领域，目前学界主要在测度资产收益率间相依性的基础上，利用模型得到的动态相关系数或者估计得到的联合概率密度函数考察不同资产或者不同市场之间的风险溢出效应，并对投资组合的 VaR、ES 等指标进行预测。然而这些研究绝大多数情况下只考虑了几种资产，并且在相依性上只考虑了关联水平，而没有考虑到关联结构蕴含的额外信息。

由此，本文尝试利用滑窗技术和藤 Copula 模型描述沪深 300 成分股的动态相依结构，并考察动态相依结构是否含有能够对市场波动率预测的信息。由于确定藤 Copula 模型的最优藤结构的计算负担会随资产个数而急剧增加，本文拟采用 Muller 和 Czado（2019）提出的基于 Graphical-lasso 的优化方法对确定最优藤结构的过程进行改进。与此同时，为了能使模型中的预测变量既能够反映相依水平也能反映相依结构，本文将估计得到的藤结构视为网络，并利用网络特征参数描述收益率的相依性。与一般基于收益率相关系数建立的网络相比，由于藤结构中包含了众多树，可以更精细的捕捉资产收益率之间的相依结构。

综上，本文跳出对于金融资产相依性测度的传统研究框架，探究了高维下资产间动态相依性的精准测度，赋予了其对市场收益和风险预测新的驱动力，有助于提高我国金融市场的风险防范能力和综合监管水平，加强市场稳定性，对于我国金融体制的运营管理具有重要意义。

二、文献综述

（一）国内外研究综述

近年来，针对描述资产价格之间的动态相依性和相依结构，相继涌现出 GARCH 模型族、Copula、藤 Copula 等模型及方法，但是现有方法仍存在不同层面的问题待解决。聂巧平等（2020）总结了 GARCH 模型族各类方法的优缺点，如 DCC-GARCH（2001）能够对大型相关系数矩阵进行估计，但是该模型由于约束较多无法单独描述尾部相依性和捕获市场间信息的非对称性。Sklar（1959）首次提出 Copula 理论，证明了 Copula 函数能够有效描述收益率之间的非对称、非正态的尾部信息，且能够捕捉到收益率序列的厚尾特征，但 Copula 模型在实际应用中存在维数灾难问题，因此该模型仅适用于较低维数的情形。针对现有研究相关模

型的局限性, Bedford 和 Cooke (2001) 提出了藤 Copula 模型, 该模型在描述高维变量的联合分布时具有更强的灵活性, 且通过构建藤结构成功应用于金融市场复杂相依结构的度量, 因此该模型更适用于高维变量的情况。

藤 Copula 模型在近年得到了快速发展, Bedford 和 Cooke (2002) 根据藤结构多元联合分布的多种分解方式提出了正则藤 (R 藤)。由于 R 藤结构众多, Dißmann 等人 (2013) 采用传统的启发式搜索来提升查找效率并减少查找次数, 然而这种算法对 R 藤结构的选择仍存在一些缺点, 如样本 Kendall' τ 相关系数得到的最大生成树不能保证最佳对数似然的结构, 并且误差会由低阶树向高阶树传播并累积。Müller (2018) 提出用 Lasso 计算 R 藤结构和 R 藤矩阵的正则化路径, 允许基于单一结构选取灵活阈值来降低模型复杂度, 虽解决了模型依赖于最大生成树的问题, 但是一个 n 维的 R 藤需要进行拟合和存储, 伴随着同样的计算复杂性。Hobæk Haff 等人 (2016) 试图将其他无向图形结构与 R 藤相联系以便进行结构选择, 同时, Kovács 和 Szántai (2016) 提出了一种计算 k -截断 R 藤的算法。然而利用在高维空间中寻找稀疏无向图形结构的方法, 找到所谓的和弦覆盖团大小最多 $k+1$ 已知是 NP 完全的。

在实证研究领域, 国内学者更加注重对少数资产间的风险溢出建立评估模型, 并对投资组合的相关指标进行预测, 没有考虑到资产的相依结构。如吕亮雯 (2006) 利用 DCC-MVGARCH 模型来描述资产价格之间的动态相依性, 并由此对 VaR 进行估算, 研究发现动态相依性可以较好的捕捉市场间的波动关系。彭选华 (2019) 融合了 SV-M-t 模型和 Copula 理论, 以上证综合指数等数据为样本, 计算系统重要性股市风险溢出, 研究发现该模型理论结合能较好地度量出该风险溢出的非对称性。

在对市场收益率及波动率的研究中, 姜美宏 (2021) 基于 Fama-French 三因子回归模型, 建立以沪深 300 指数作为市场投资组合, 发现以沪深指数收益率作

为预测市场组合收益率的模型效果更好。刘珊珊（2019）利用回归分析研究股票收益率问题，证明了我国 A 股上市公司股票收益率的影响因素。而郝建阳（2021）等人以沪深 300 指数的波动率为研究对象来代表股市的波动率进行研究。张琳等（2020）研究了宏观信息对股票市场收益率及其波动的影响，发现 PMI 显著影响股市收益率。雷立坤等人（2018）利用自回归条件异方差混频数据抽样模型，证明了经济政策不确定性能够很好地解释我国股市波动率。

（二）文献评述

在描述资产价格之间的动态相依性和相依结构问题上，GARCH 模型族、Copula 等方法会有存在约束性较强、无法捕捉尾部相依性、维数灾难等问题。而藤 Copula 模型通过构建高维资产的联合分布解决了以上问题，但是藤 Copula 模型由于其结构复杂，已有的搜索选择算法均存在计算复杂度高、非全局最优导致的误差传播等问题。

在资产收益率间相依性的实证研究领域，目前国内学者主要考察不同资产或者不同市场之间的风险溢出效应，少有学者对股市收益及波动进行预测。此外目前所做的研究绝大多数情况下只考虑了几种资产，即低维数据，而少有学者利用高维数据进行相关研究，且在相依性上大部分文章只考虑了关联水平而没有考虑关联结构。

（三）本文主要创新

总结以往研究，本文具有以下创新点：

1. 从高维视角下探究资产动态相依水平，充分利用大量数据资源，驱动信息测度。
2. 在藤 Copula 模型的基础上，本文提出一种有效算法，改进藤 Copula 模型

结构的选择，提高了估计效率与模型优度。

3. 实证领域受限于维度问题，从而缺乏对资产间相依结构的考虑。而本文利用加权复杂网络去描述资产相依结构，充分挖掘资产信息含量。

三、研究思路

本文尝试利用滑窗技术和藤 Copula 模型描述沪深 300 成分股的动态相依结构，并考察动态相依结构是否含有能够对市场收益率和波动率预测的信息。由于确定藤 Copula 模型的最优藤结构的计算负担会随资产个数而急剧增加，本文采用 Muller 和 Czado (2019) 提出的 Graphical-lasso 优化方法对确定最优藤结构的过程进行改进，并且对其进行稳健性检验以验证改进效果。同时，为了能使模型中的预测变量既能够反映相依水平也能反映相依结构，本文将估计得到的藤结构视为网络，并利用网络特征参数描述收益率的相依性。与一般基于收益率相关系数建立的网络相比，由于藤结构中包含了众多树，可以更精细的捕捉资产收益率之间的相依结构。其主要行文思路如下：

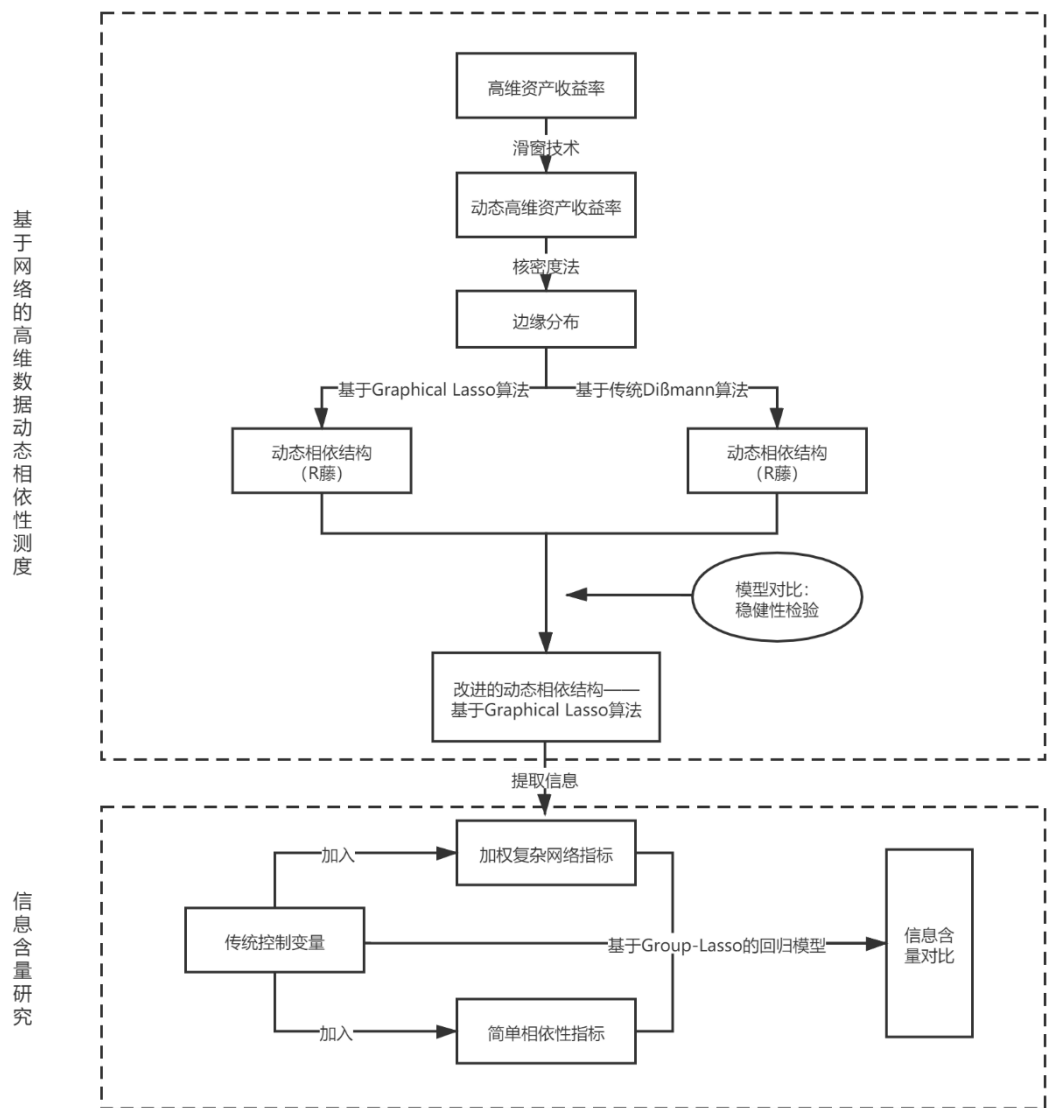
首先，利用核密度估计沪深 300 成分股的边缘分布，通过高维 R 藤 Copula 模型构建其动态相依结构；

其次，在传统的 Dißmann 算法上引入 Graphical-lasso 算法改进藤 Copula 的模型选择与估计过程，并通过稳健性检验，进一步验证模型结构的优化，以及计算成本和时间的降低；

接着，将藤结构视作加权复杂网络，其权重分别由 τ 、上尾、下尾相关系数构成，并对藤结构中的前两棵树采用加权中心性、接近中心性等相关指标描述其网络特征；

最后，引入传统预测因子分别对市场收益率与波动率建立回归模型，将通过 Group-lasso 筛选后的网络指标与总体简单平均相关系数分别加入模型进行对比，

来验证模型所蕴含的市场信息。



四、研究方法

(一) R-vine Copula 及其参数估计

为了描述高维资产收益率之间复杂的相依性，学者们提出了条件 Copula 函数 (Patton, 2006)，它根据 Sklar 定理连接多个收益率的边缘分布构造多元联合分布，进而可以得到不同个体之间的相依结构，为解决多元 Copula 带来的维度

“爆炸”，Bedford 提出可以利用条件相关随机变量的概率密度分解对多元 Copula 构造基于条件 Copula 的藤曼模型，即图形建模工具藤（vine）。在藤结构的基础上，由于多元联合分布分解成为一系列成对 Copula 函数时， n 个变量有 $\frac{n!}{2} \times 2^{\frac{n-2}{2}}$ 种分解方式，进而有不同的藤结构。因此 Bedford 和 Cooke（2002）提出了正则藤（R-vine，R 藤），R 藤的定义如下：

- I. 一个包含 n 个变量的 R-vine 结构 v 由 $n-1$ 棵树 $T = (T_1, T_2, \dots, T_{n-1})$ 组成， T_i 的节点集合和边集合分别为 N_i 、 E_i ， $(i=1, 2, \dots, n-1)$ ；
 - II. T_1 的节点集合为 $N_1 = \{N_1, N_2, \dots, N_n\}$ ， T_1 的边的集合为 $E_1 = \{E_1, E_2, \dots, E_{n-1}\}$ ；
 - III. 对于 $(i=2, \dots, n-1)$ ，树 T_i 的节点集为第 $i-1$ 棵树 T_{i-1} 的边集，即 $N_i = E_{i-1}$ ；
 - IV. 当 T_i 中的两条边共用一个节点时，则两条边在树 T_{i+1} 中相连。
- 具有 n 个随机变量的 R 藤 Copula 模型的联合分布概率密度函数为：

$$c(F_1(x_1), F_2(x_2) \dots, F_n(x_n)) = \prod_{i=1}^n \prod_{e \in E_i} c_{j(e), k(e) | D(e)} \left(F(x_{j(e)} | x_{D(e)}), F(x_{k(e)} | x_{D(e)}) \right) \quad (1)$$

其中 e 为 E_i 中连接随机变量 R 藤 Copula 的每条边，且 $e = j(e), k(e) | D(e)$ ， $c_{j(e), k(e) | D(e)}$ 为 pair-copula 的概率密度函数。

在参数估计之前需要确定 R 藤模型的结构，由于 R 藤结构众多，Dißmann 等（2013）提出采用启发式方法的搜索最优模型结构，其具体步骤为：首先，对于每一对节点计算 Kendall' τ 相关系数。然后根据最大生成树算法，选择 Kendall' τ 相关系数的绝对值之和最大的 $d-1$ 条边。最后，在选定的边上估计 pair-copula。对于藤结构中后续的树，将上一颗树的边作为下一颗树的节点，计算节点之间的条件 Kendall' τ 相关系数，继续根据最大生成树算法确定树结构，在 $d-1$ 次迭代之后，就能确定 R 藤结构。

可供选择的 pair-copula 函数主要分为对称与分对称函数两类，如下表：

表 1 pair-copula 函数对照表

Copula 函数类型	分类	缩写
Gaussian	对称	G
Student	对称	t
Survival Clayton	对称	SC
Gumbel	非对称	G
Survival Gumbel	非对称	SG
Frank	非对称	F
Multivariate normal	对称	N

其选择标准一般是基于GIC准则的，即

$$GIC(\Theta, X) = -2\mathcal{L}(\Theta, X) + \log(\log(n)) \log(p) p \quad (2)$$

在确定 pair-copula 函数后采用两阶段极大似然法对 R 藤 Copula 模型参数进行估计。即首先估计出边缘分布函数的参数 $\theta_n, n = 1, 2, \dots, N$,然后将它们的估计值 $\hat{\theta}_n, n = 1, 2, \dots, N$ 作为已知数带入 Copula 函数中，进而估计得到 Copula 函数中的参数 θ_c 的值。

(二) 基于 Graphical-Lasso 下 R-vine 结构选择的改进

针对传统 Dißmann 算法对 R 藤结构的选择仍存在一些缺点。一方面无法有效解决高维下的计算复杂度问题，另一方面相依结构的估计存在误差传播问题。因此本文采用 Graphical-lasso 方法（Friedman, 2008）搜索最优藤结构。

具体地，记 $X = (x_{i,j})_{i=1,\dots,T,j=1,\dots,N}$ 为收益率序列组成的矩阵，并且服从 N 维 0 期望的多元正态分布，其方差协方差矩阵为 Σ 。定义 S 为 X 的样本协方差矩阵，则稀疏图的估计问题可以表示为找到一个矩阵 Ω 使公式（3）最大化。但是这通常不会产生零项，因此不会使图稀疏。而图 lasso 方法通过引入惩罚项解决了这一问题。

$$\log(\det(\Omega)) + \text{tr}(S\Omega) \quad (3)$$

$$\log(\det(\Omega)) + \text{tr}(S\Omega) + \lambda \|\Omega\|_1 \quad (4)$$

其中 λ 为惩罚因子，当 $\lambda \rightarrow \infty$ 时， Ω 中的元素均被收缩至 0，此时图中将不存

在任何边； $\lambda \rightarrow 0$ 时， Ω 中的元素均是非零的，图结构仍然是完全连通的。对于 λ 的取值问题，可行的搜索策略是从一个较小的 λ 开始，按照一定规则生成一系列 λ_j ，($j = 1, 2, \dots, J$)；对于每一个 λ_j 都求解公式(4)的最大化问题，在此基础上，我们采用一种并行的估计策略，具体地，对于一个固定的 $\lambda > 0$ ，存在一个对应的图结构 $G_\lambda = (V, E^\lambda)$ ，从样本逆协方差矩阵 $\hat{\Omega}^\lambda$ 中计算出与 λ 有关的连通分量，使 $\hat{\Omega}^\lambda$ 和等效的 E^λ 得到分割顶点 $V = \bigcup_{i=1}^P V_i^\lambda$ ，每个 V_i^λ ($i = 1, \dots, P$)是基于边集合 E^λ 的连通分量。因此，可以将图 lasso 问题分解为可以完全彼此独立执行的 p 个并行任务，大大降低了计算复杂度。

对于超高维 R 藤，我们需要将数据集的维度通过连通分量分割成若干组，随后对每个组中的图结构进行估计。在这些组中，由于利用了全局 Markov 性质，因此我们可以以更高的精度在更小的维度上估计 R 藤结构。具体的，给定数据 $X \in \mathbb{R}^{n \times d}$ ，我们将 $V = \{1, \dots, d\}$ 的 j 个不相交的组转换为 P_j 个连通分量：

$$V^j = \bigcup_{i=1}^{P_j} V_i^j, j = 1, \dots, J \quad (5)$$

这由一系列图形结构表示：

$$G_1 = \left(V^1 = \bigcup_{i=1}^{P_1} V_i^1, \bigcup_{i=1}^{P_1} E_i^1 \right), \dots, G_J = \left(V^J = \bigcup_{i=1}^{P_J} V_i^J, \bigcup_{i=1}^{P_J} E_i^J \right) \quad (6)$$

在大多数实际应用中， $J=15$ 或 $J=30$ ，序列 $G_j, j = 1, \dots, J$ 可被认为对应于 $\lambda > 0$ 的 J 个不同罚参数的 Graphical Lasso 解的路径。定义：

$$\delta_j = \max_{i=1, \dots, P_j} |V_i^j| \quad (7)$$

为连通分量上维数最大的子集，且 $\delta_j < d$ ，因此我们可以通过 $T = \arg \max_{j=1, \dots, J} \delta_j$ ($\delta_j < d_T$) 选择相应的图形结构和关联的分区 V_T ，从而计算序列 $\{\lambda_1, \dots, \lambda_J\}$ 的 Graphical Lasso 解的路径。

对于特定的 $\lambda = T$ ，我们定义关于 G_T 上的一个连通分量为一个连通子图 $H = (W, F)$ ，其节点集 W 的节点个数为 v ，随后在边集 F 上估计一个 R-vine，为了搜

寻第一棵树的结构，Dißmann 算法将每一对节点 (j,l) 的 Kendall 系数作为权重 $\mu_{j,l}$ 构造最大生成树，但是这样不能保证最大生成树可以得到符合最佳对数似然的结构，因此我们对每一对节点 (j,l) 估计一个 pair-copula 密度函数 $c_{j,l}$ 作为其权重，由于 $v < d$ 且 H 为稀疏无向图，因此其计算复杂度会远小于 $\frac{v(v-1)}{2}$ ，且构造的最大生成树最大程度上减少了误差。

对于 λ 的选择，过小的 λ 使得图结构密集且连通分量少，计算复杂度增加，而过大的 λ 使得图结构稀疏，游离的节点增加，削弱藤结构的总体关联性。以上我们限制了 λ 的下限，即保证了连通分量上的最大子集维度，接着我们需要限制 λ 的上限，定义关联水平 $k_F > 0$ 为 G_T 中需要估计的 R 藤的数量，它决定了数据集在给定 λ 下具有强内部关联性的连通分量数量。例如 $k_F=1$ ，则 G_T 存在两种情况，一种为只存在一个连通分量，另一种为只存在一个 R 藤结构，其余均为游离点。

综上，在对 λ_j ， $(j = 1, 2, \dots, J)$ 的搜索过程中，通过设定阈值 d_T 与 k_F 分割数据集，在每个连通分量上估计 R-vine 结构并基于以下规则构造维度为 d 的联合 R 藤下三角矩阵：

- ①节点按编号由大到小排列在主对角线上。
- ②主对角线上的每个节点与矩阵倒数第一行的相应节点组成第一棵树的边，且矩阵倒数第一行的节点构成条件集。
- ③主对角线上的元素和矩阵倒数第二行的相应节点对应，且倒数第一行和倒数第二行的元素构成条件集，并以此类推。

此过程基于连通分量之间的弱关联性假设，并且有研究证明，随着维度 d 的增加，其关联性减弱，因此总体关联性可由各连通分量的内部关联性近似。

(三) 基于网络特征的相依性描述

为了充分描述 n 种资产之间的相依结构, 需要全面考虑它们之间的相关系数, 但是当 N 很大时, 它们之间的相关系数共有 $\frac{n(n-1)}{2}$ 个。直接将这些相关系数加入预测模型, 显然会使模型陷入维数灾难。针对这一问题, 本文基于 n 种资产之间的相关系数矩阵建立加权复杂网络, 并基于复杂网络的网络特征描述相依结构的总体特征。本文采用的网络特征指标有以下四种:

1. 加权中心性

加权中心性测量的是网络整体的相关性, 加权中心性越大代表网络相关关系越强, 反之则相关关系越弱。对于无向图, 取得每个点的边, 将边的权重求和, 即为该点的加权重。加权重总和除以节点数即为加权中心性:

$$W_i = \frac{\sum_{i=1}^n e_i}{n} \quad (8)$$

其中 e_i 为每个点的加权重, $e_i = \sum_{j=1}^m d_j$, m 为节点的度, d_j 为边权, n 为总节点数。

2. 接近中心性

接近中心性反应某一节点与其他节点之间的接近程度。如果一个节点到网络中各点的距离越短, 那么这个点传播信息的时候就越不会受制于其他点。因此利用节点到网络中其他所有节点的最短路径之和, 并进行归一化, 来表示平均最短距离, 平均距离越小, 接近中心性应越大, 于是再将其倒数定义为该节点的接近中心性:

$$C_i = \frac{1}{d_i} = \frac{n-1}{\sum_{j \neq i} d_{ij}} \quad (9)$$

其中 $d_i = \frac{1}{n-1} \sum_{j \neq i} d_{ij}$, d_{ij} 为节点 i 与节点 j 的边权。

3. 中介中心性

一个节点的中介中心性强调该点在其他点之间的调节能力、控制能力指数和

中介调节效应,它是指网络中经过某点并连接这两点的最短路径占这两点之间的最短路径线总数之比。该值越大,节点作为中介在网络中其他任意两个节点之间建立间接连接的边数越多,越能通过网络向更多个体提取、传递信息,其计算公式为:

$$B_i = \sum_{j < k} \frac{n_{ijk}}{n_{jk}} \quad (10)$$

其中, n_{ijk} 指在节点 j 和 k 之间,通过节点 i 建立间接连接的路径线数目, n_{jk} 表示节点 j 和 k 总的连接路径线数目。

4.特征向量中心性

一个节点的重要性不仅取决于与其直接相连节点的数量,还取决于这些节点的重要性,即节点的“质量”。特征向量中心性是测量节点对网络影响的一种方式。针对连接数相同的节点,相邻节点分数更高的节点会比相邻节点分数更低的节点分数高,依据此原则给所有节点分配对应的分数。特征向量得分较高意味着该节点与许多自身得分较高的节点相连接。

记 x_i 是节点 i 的重要性度量值, 则:

$$EC(i) = x_i = c \sum_{j=1}^n a_{ij} x_j \quad (11)$$

其中, c 为一个比例常数, $a_{ij}=1$ 当且仅当 i 与 j 相连, 否则为 0。

需要说明的是,由于通过 copula 函数可以得到变量间的 Kendall' τ 相关系数、上尾相关系数和下尾相关系数, 因此本文分别基于三种相关系数建立复杂网络, 并通过四种网络特征指标描述资产之间的相依结构。与此同时, 由于 vine-Copula 的藤结构具有多个层次, 每一层次的相依结构可能具有独特的信息含量, 因此在后文的实证研究种, 本文也将分别提取不同层次相依结构的网络特征。

（四）基于稀疏 Group-Lasso 的预测模型

考察动态相依结构的信息含量就是要基于相依结构建立预测模型，并考察相依结构是否能够改进对收益率和波动率的预测。对于收益率和波动率，本文根据已有研究文献分别构建预测模型公式（12）和公式（13）：

$$Ret_t = \beta_0 + \beta_1 Corr_{t-1} + \gamma Factor_{t-1} + \mu_t \quad (12)$$

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 Corr_{t-1} + \gamma Factor_{t-1} + \mu_t \quad (13)$$

其中 Ret_t 为沪深 300 指数的收益率， RV_t 为沪深 300 指数的已实现波动率， $Corr_t$ 为表示资产相依结构的代理变量，也即上文给出的表示相依结构的网络特征指标，作为比较本文也将所有资产之间的相关系数均值作为相依结构的代理变量。模型中的 $Factor_t$ 为其他可能影响收益率或者波动率的因子。

需要注意的是，由于本文基于大量的网络特征指标描述资产之间的相依结构，因此预测模型可能存在比较严重的多重共线性问题，又为了控制两个回归模型在其他预测因子上保持一致，需要单独对网络指标变量组进行筛选，针对这种情况，本文拟采用稀疏 Group-Lasso 方法进行变量选择。线性回归下，其损失函数为：

$$\hat{\beta}^{SGL} = \arg_{\beta} \min \left\{ \frac{1}{2} \|y - \sum_{i=1}^m X^{(i)} \beta^{(i)}\|_2^2 + (1 - \alpha) \lambda \sum_{i=1}^m \|\beta^{(i)}\|_2 + \alpha \lambda \|\beta\|_1 \right\} \quad (14)$$

矩阵 $X^{(i)}$ 表示与第 i 组匹配的特征矩阵， $\beta^{(i)}$ 表示于其对应的系数向量， λ 为惩罚项参数， $\alpha \in [0, 1]$ ，是调整组内与组间 Lasso 惩罚项的调和参数，由于传统预测因子的选择已经剔除了多重共线性，故最终我们可以得到稀疏的网络特征指标与所有的预测因子。

五、实证研究

（一）数据描述

本文选取沪深 300 指标股中流通市值最大的 80 支股票作为研究对象并计算其日对数收益率，样本期间为 2013 年 1 月 4 日到 2020 年 12 月 31 日，数据来源为 Wind 数据。与此同时，为了能捕捉股票间动态相依性，本文采用滑窗技术来研究股票在样本期间内相依水平和相依结构的动态变化。将滑动窗口宽度设为 120 天，步长为一个月。上述股票的基本信息详见附录一。

为了减小实证结果呈现的繁琐性，特选取 2013 上半年数据为第一窗口期、2016 上半年数据为第二窗口期、2020 下半年数据为第三窗口期作为示例展示实证结论。

图 1 展示了沪深 300 收盘价在整个样本区间内的走势图，图 2 为三个不同时间段的股票间相关系数热力图。其中图 2（1）对应的时期股市整体处于一个较为平稳的状态，在这个期间各个股票间的相关系数普遍较小，大多在 0.2 到 0.4 之间；图 2（2）对应的时期股市经历了剧烈动荡，这一期间内股票间的相依性显著强于股市平静期，相关系数大多处于 0.4 到 0.6 之间；图 2（3）对应的时期为股市的上涨期，这一期间内的股票相关系数与平稳期无明显差异。故股市在不同时期显示出的相依性具有一定的信息含量。

值得注意的是，一些股票的相依性表现出独特性。例如股市在股灾时期相较于平稳时期，整体相依性显著增强的情况下，万科 A，格力电器，闻泰科技等股票与其他股票间的相关系数都显著低于平均水平。TCL 科技，中信特钢，上海机场与交通银行等股票之间，无论是在股市平稳还是动荡时期都表现出相对较大的相依关系。因此对于某些股票来说，股市整体状况对其相依性的影响并不显著，如果仅考虑股票之间的总体相依水平可能会损失大量信息。

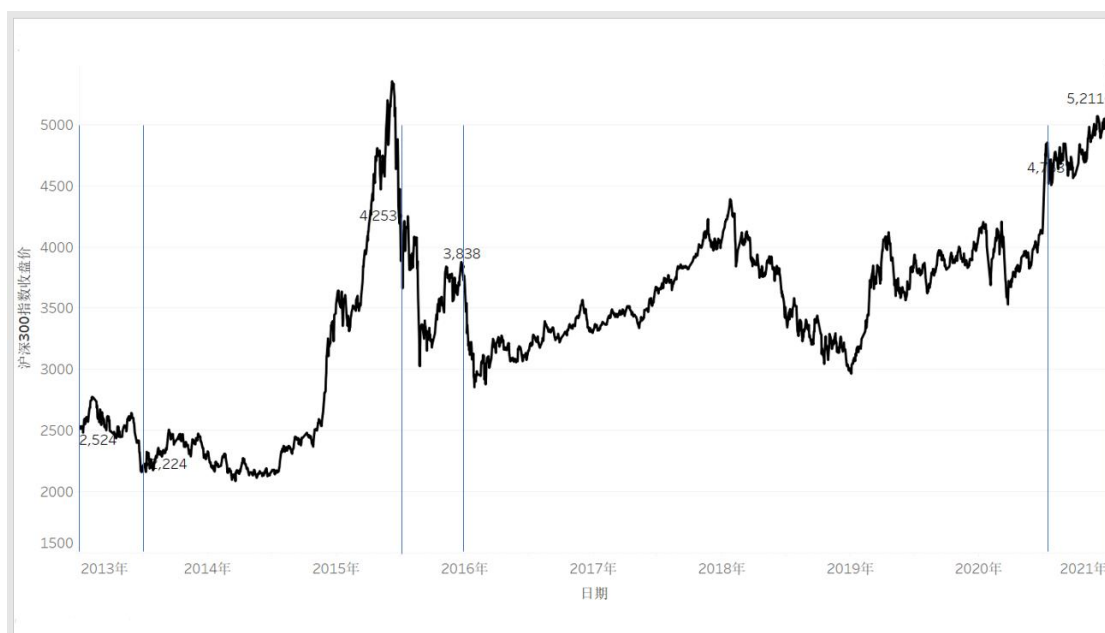


图1 沪深300指数收盘价

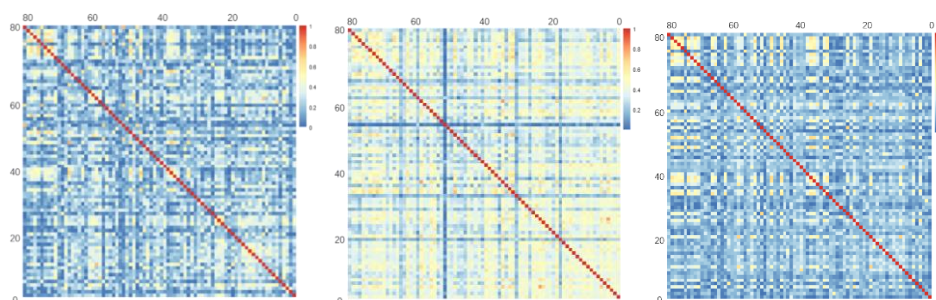


图2 第一窗口期、第二窗口期、第三窗口期相关系数热力图

(二) 基于 Graphical-Lasso 的 R-vine Copula 模型估计

在建立藤 Copula 模型之前需要对收益率序列的边缘分布进行估计，本文采用核函数方法进行拟合，进一步对经过概率积分变换后的边缘分布进行 KS 检验，如下表，检验结果显示大部分 p 值小于 0.05，序列服从均匀分布，满足建立藤 Copula 模型的要求。

表 2 第一窗口期各股收益率分布拟合统计量

2013	贵州茅台	工商银行	...	中航沈飞	复星医药
$D = \max_{n1} S_{n1}(X) - S_{n2}(X) $	0.0556	0.0518	...	0.0476	0.0304
p 值	0.4211	0.5142	...	0.6225	0.9750

表 3 第二窗口期各股收益率分布拟合统计量

2016	贵州茅台	工商银行	...	中航沈飞	复星医药
$D = \max_{n1} S_{n1}(X) - S_{n2}(X) $	0.0423	0.0528	...	0.0399	0.0450
p 值	0.6394	0.5783	...	0.6582	0.8642

表 4 第三窗口期各股收益率分布拟合统计量

2020	贵州茅台	工商银行	...	中航沈飞	复星医药
$D = \max_{n1} S_{n1}(X) - S_{n2}(X) $	0.0120	0.0153	...	0.0208	0.0148
p 值	0.8758	0.6204	...	0.2456	0.6635

在得到边缘密度后建立 R-vine Copula 模型，图 3 展示了基于图 lasso 的最优 vine 结构的搜索过程。不难看出，随着惩罚因子 λ 的增加，连通分量越多，协方差矩阵越稀疏，图 4 中显示当 $\lambda = 0.375$ 时， $d_T=75$ ， $k_F=2$ ，其结构达到最优。随后对两个连通分量进行 Pair-Copula 拟合并分别构造最大生成树，最后基于下三角矩阵生成联合 R 藤结构。

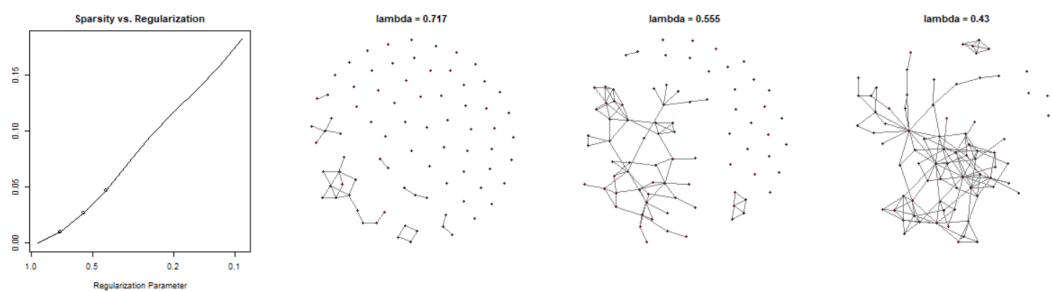


图3 不同惩罚因子对应的无向图结构

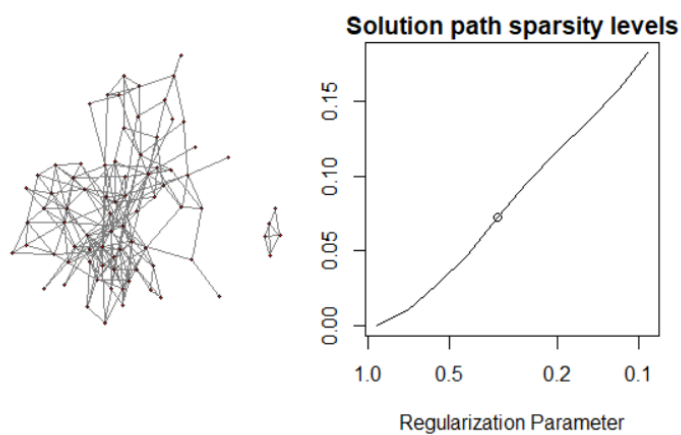


图4 最佳惩罚因子及其无向图结构

如表 5 至表 7 所示，此处估计结果应有 79 层，由于篇幅所限，这里仅报告三个窗口期 T1 和 T2 的前两层 pair-copula 估计结构。表中 Edge 代表藤结构的边，数字代表股票编号。par 为相应二元 Copula 函数的参数，par2 为相应二元 Copula 函数的自由度。 τ 代表秩相关系数， λ^u 和 λ^d 分别表示上下尾相关系数。

表 5 第一窗口期藤 Copula 拟合结果

Tree	Edge	Copula	par	par2	τ	λ^u	λ^d
1	22,71	t	0.91	2.45	0.73	0.72	0.72
	71,62	SG	3.22	0.00	0.69	0.00	0.76
	24,6	t	0.88	6.16	0.68	0.51	0.51
...
2	35,44;49	F	3.23	0.00	0.32	0.00	0.00
	77,10;48	F	2.78	0.00	0.28	0.00	0.00
	7,32;5	t	0.41	3.57	0.27	0.23	0.23
...

表 6 第二窗口期藤 Copula 拟合结果

Tree	Edge	Copula	par	par2	τ	λ^u	λ^d
1	77,57	t	0.94	4.84	0.78	0.69	0.69
	46,22	t	0.94	2.97	0.78	0.74	0.74
	49,32	t	0.91	2.00	0.73	0.73	0.73
...
2	64,60;26	N	0.47	0.00	0.31	0.00	0.00
	51,42;26	F	2.98	0.00	0.30	0.00	0.00
	3,27;14	G	1.39	0.00	0.28	0.35	0.00
...

表 7 第三窗口期藤 Copula 拟合结果

Tree	Edge	Copula	par	par2	τ	λ^u	λ^d
1	32,35	t	0.89	4.06	0.70	0.62	0.62
	62,22	G	3.32	0.00	0.70	0.77	0.00
	7,2	G	3.25	0.00	0.69	0.76	0.00
...
2	49,18;14	F	2.83	0.00	0.29	0.00	0.00
	44,35;32	F	2.81	0.00	0.29	0.00	0.00
	9,2;7	F	2.51	0.00	0.26	0.00	0.00
...

(三) 稳健性检验

本节拟对基于传统 Dißmann 算法与基于 Graphhical-lasso 算法下估计得到 R 藤 Copula 模型进行对比,以探究改进式算法的稳健性及其效率。具体的以 GIC 准则、对数似然函数、估计时间三个维度进行评价,由于越小的 GIC 与越大的对数似然函数值表示的拟合效果越好。因此选取各模型的平均 GIC 与 GIC 四分之一位点的平均值、各模型的平均对数似然函数值与对数似然函数四分之三位点的平均值,以及平均估计时间作为对比模型结果的指标。

表 8 稳健性检验

	GIC_Average	logLik_Average	GIC_1/4Q_Average	logLik_3/4Q_Average	time_Average
Dißmann 算法	-1.061	1.049	-0.009	5.735	282s
Graphhical-lasso 算法	-2.854	3.109	-0.022	38.938	55s

由上表所示,传统算法与 Graphhical-lasso 算法在平均 GIC 和平均对数似然函数值上差异不大,这是由于一方面模型数量较多,另一方面每个模型中需要估计的 Copula 函数众多,而具有强依赖的资产是局部性的,因此整体拉低了平均水平。并且 GIC 四分之一位点的均值也相差不大,说明算法的改进对 pair-Copula 函数形式的选取影响较小。此外值得注意的是,Graphhical-lasso 算法得到的对数似然函数四分之三位点的平均值有明显提高。这说明估计得到的联合分布在相依性较强的局部具有更好的拟合效果,而对比整体均值,进一步显示出改进的算法对于相依结构的识别更灵敏。最后在同等配置的计算资源下,由于改进算法的分块并行估计策略,使得运行时间大大降低,为更高维的计算提高了全新思路。

（四）基于加权复杂网络的信息提取

藤结构中包含了大量的树结构，每一棵树中又包含了节点的位置、边权重以及权重类型等信息，本文将树结构视作一种加权复杂网络，通过引入网络指标充分提取包括位置、权重等网络信息。由于低阶树反映了局部节点的相依性，而高阶树的节点是低阶树的条件边缘分布，因此越高阶的树，其边权重越小，反映范围更广的局部相依性，综合实证发现，在第三层及其更高层的树结构中，边权重的值几乎都不再显著，其值大多在 0 处波动。综上对构建的藤 Copula 模型取前两层树结构进行信息提取。

下图展示的是三个窗口期藤结构的第一层和第二层，图中的数字代表股票编号，节点颜色显示的是节点的度，边的颜色显示的是边权重的大小。因为 τ 相关系数不受 Copula 类型的影响，具有统一可比性，因此以 τ 相关系数作为权重。

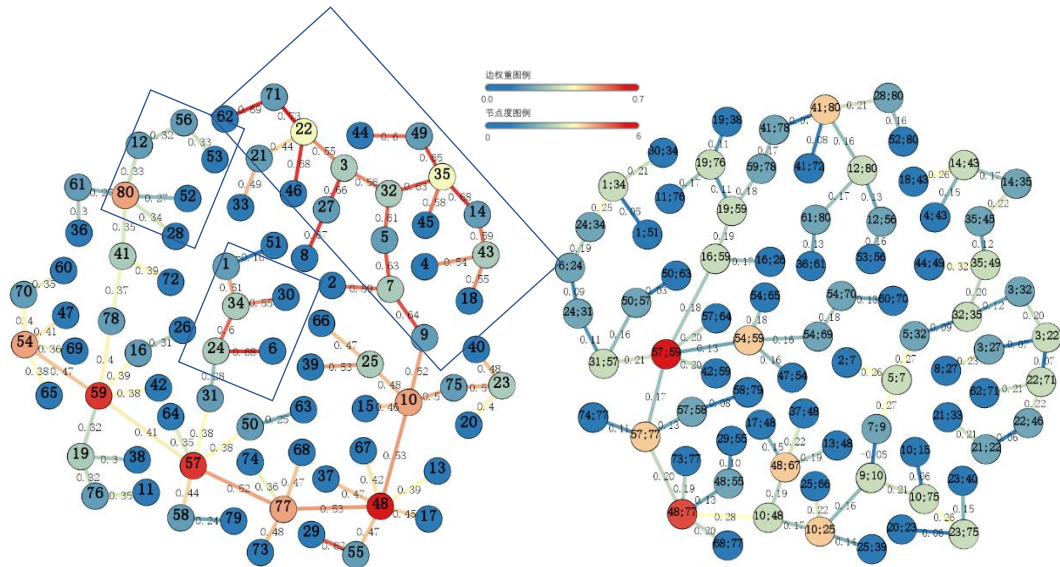


图 5 第一窗口期藤结构前两层可视化

第一窗口期属于股市较为稳定的时期。对于第一颗树，一些中央企业，中国石油（10），中国建筑（48），中国中车（57）等节点出现在了结构的中心位置，

这些节点的度较大，说明在股市平稳期，这些中央企业处于风险传播的中心。如图中方框所示，结构中有几个明显的板块聚集效应。以交通银行（32）为中心的方块内部股票都为证券，银行和保险股等金融机构股票，且相关性较大，相关系数都在 0.6 左右。以洋河股份（34）为中心的方块内部股票都为白酒股，内部相关系数同样在 0.6 左右。以复星医药（80）为中心的方块内部股票都为医药板块的股票，但内部相关程度不高，相关系数在 0.3 左右。第二棵树表示股票间间接的相依关系，可以清楚的看到，第二棵树的相依程度与第一棵树相比明显较弱，但还是有股票间的间接相关系数在 0.3 以上，但都出现在金融行业内部，例如以光大银行（49）为条件，中信银行（44），浦发银行（35）之间的间接风险传染效应为 0.32。

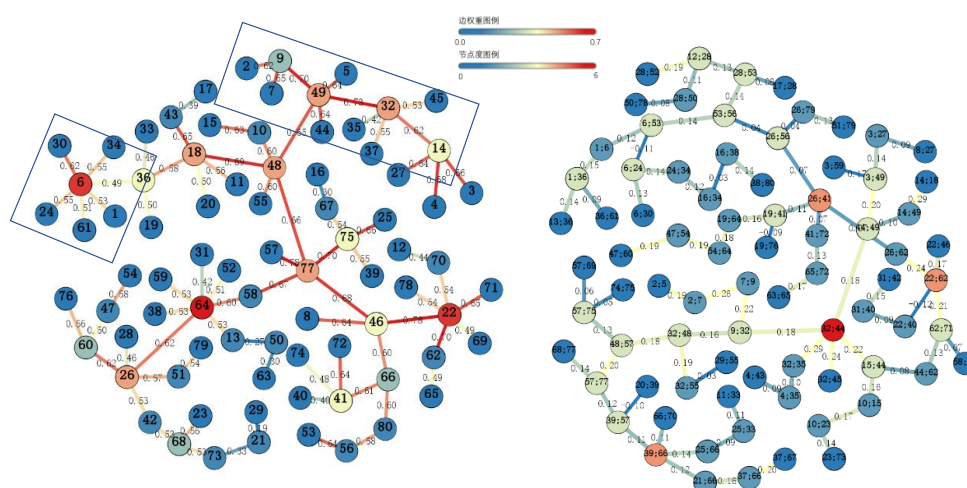


图 6 第二窗口期藤结构前两层可视化

第二窗口期属于股市下行的时期。对于第一颗树，一些金融机构，兴业银行（14），平安银行（18），中信证券（22），东方财富（26）等出现在了结构的中心位置，节点的度较大，说明在股市下行期，金融机构成为了风险传染的中心。如图中方框所示，板块聚集效应较股市平稳期有所减少。银行股票与非银金融股票不再聚集，但银行内部相关性仍然较强，以光大银行（49）为中心的方块内部

股票都为银行股票，相关系数都在 0.6 左右。以五粮液（6）为中心的方块内部股票都属于食品饮料行业，内部相关系数同样在 0.6 左右。股市下行期与股市平稳期的第二棵树相比，间接相依关系虽然有所增强，但股票间的间接相关系数还是大都在 0.3 以下，最大的间接相依关系，出现在以东方财富为条件的汇川技术（60）与用友网络（64）之间。

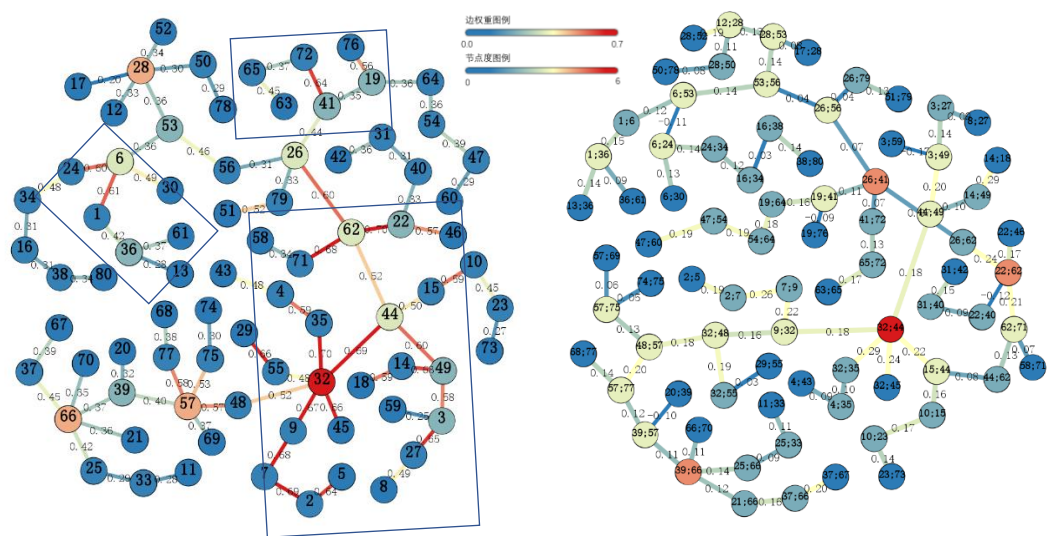


图 7 第三窗口期藤结构前两层可视化

第三窗口期属于股市上行的时期。对于第一颗树，结构中心点与前两个窗口相比明显减少，出现在结构的中心位置的股票来自各个行业，有交通银行（32），爱尔眼科（28），中国中车（57），潍柴动力（66）。如图中方框所示，股市上行的时期仍然出现了板块聚集效应，银行股票与非银金融股票再次以交通银行（32）为中心聚集，相关系数都在 0.6 左右。以五粮液（6）为中心的方块属于食品饮料行业，内部相关系数同样在 0.6 左右。以京东方 A（41）为中心的方块内部股票为电子板块股票，内部相关系数较小在 0.5 左右。股市上行期与股市平稳期的第二棵树相比无明显差异。

接着将每个窗口形成的藤结构视为网络，对其提取前文叙述的网络特征指

标。为了使各指标的走势可以展示在一张图中，对中介中心性进行了处理，具体情况如图 8 所示。

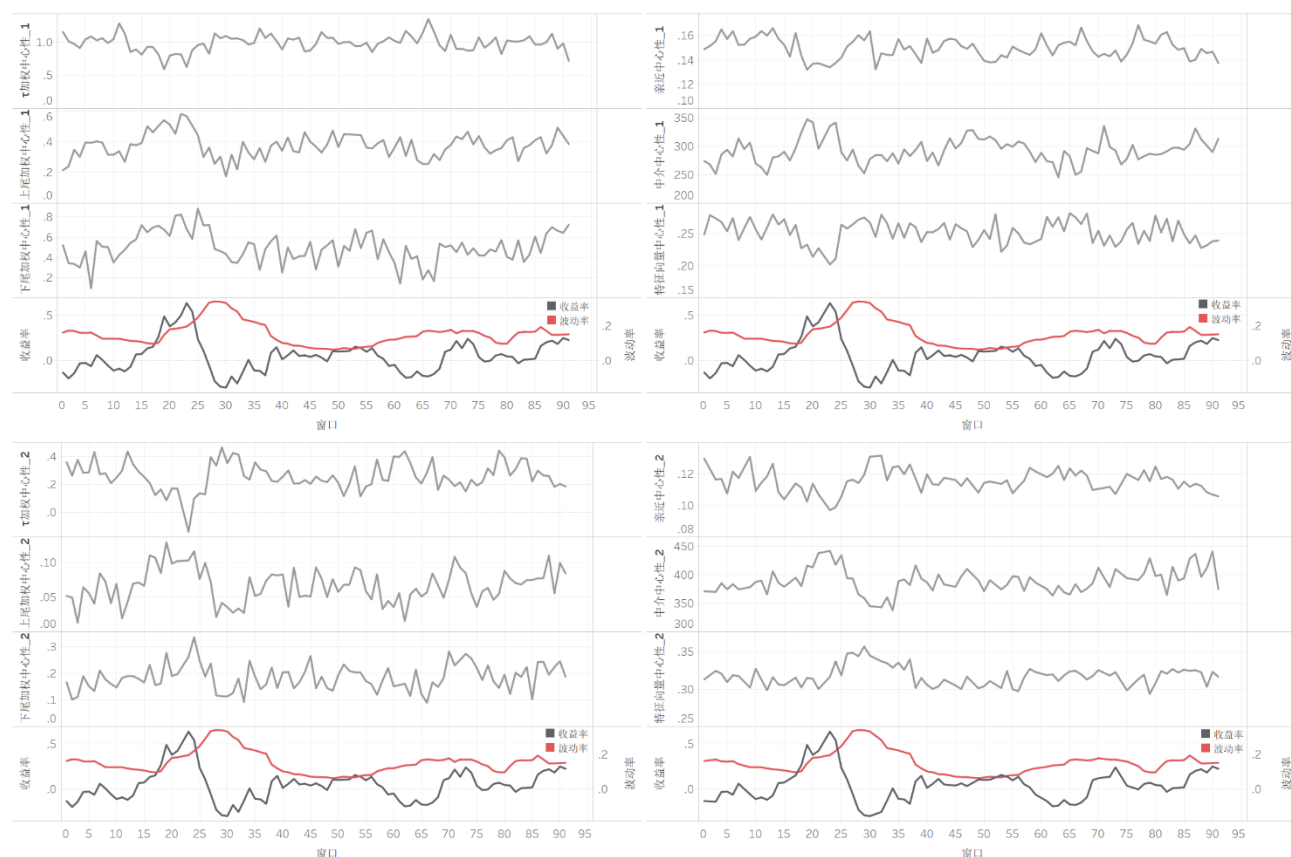


图 8 各网络指标与市场收益率及波动率走势图

在所描述网络特征的指标中，除加权中心性外，其他指标的大小只与网络结构有关而与边权重无关。由图可知，加权中心性的各个指标与沪深 300 收益率之间有着明显相似的变化趋势， τ 加权中心性与收益率序列有着相反的变化特征，而上尾加权中心性和下尾加权中心性则与收益率序列有着相同的变化特征。加权中心性的各个指标与波动率的关系并不显著。其他指标在收益率变化非常剧烈时会出现明显的变化。而对于波动率， T_2 的特征向量中心性与波动率有着显著相似的变化趋势。

（五）基于回归的信息含量研究

1. 市场收益率信息含量研究

本文选取沪深 300 股指的收益率代表股市收益率，作为回归分析的因变量。为对模型质量和测度能力进行评估，本文选取上文提到的 80 支股票的平均相关系数（average_R）作为对比变量，同时依据相关文献选取 Fama-French 三因子作为控制变量，自变量均选取滞后一阶项，建立三个回归模型进行结果对比。

基于稀疏 Group-Lasso 方法可以在组水平上选择变量。将 12 个网络特征指标作为一组，经筛选最终得到结果如下表：

表 9 市场收益率研究特征子集表

变量	描述
τ _Weighted_centrality_1	第一棵树的 τ 加权中心性
Uptail_weighted_centrality_2	第二棵树的上尾加权中心性
Lower_weighted_centrality_2	第二棵树下尾加权中心性
Betweenness_centrality	中介中心性
RiskPremium1	Fama-French 三因子
SMB1	Fama-French 三因子
HML1	Fama-French 三因子

通过变量构造三个以沪深 300 指数收益率作为因变量的回归模型，并对 R-藤 Copula 模型指标的拟合程度进行探究，结果见表 10。

表 10 各指标对沪深 300 指数收益率的回归结果

Variable	reg_1	reg_2	reg_3
average_R		-1.198 (-1.156)	
τ _Weighted_centrality_1			-0.591 (-0.326)
Uptail_weighted_centrality_2			-1.510* (-2.082)
Lower_weighted_centrality_2			-0.523 (-1.076)
Betweenness_centrality			-3.618 (-1.189)
RiskPremium1	0.616 (0.329)	0.138 (0.073)	2.326 (1.011)
SMB1	4.205 (1.512)	5.820* (1.893)	-0.020 (-0.005)
HML1	1.572 (0.510)	-2.020 (-0.465)	-11.552 (-1.380)
constant	0.029 (0.595)	0.306 (1.251)	1.354 (1.200)
R^2	0.248	0.337	0.741
Adjusted R^2	0.143	0.172	0.376

注：估计系数对应括号内数据为 t 统计量；“*”、“**”、“***”分别表示在 10%、5%、1% 显著水平下显著

从预测模型结果可以看出，模型调整后的拟合优度 R^2 在 reg_1 模型中为 0.143，当加入了股票平均相关系数这一变量后， R^2 提高至 0.172。值得注意的是，在加入传统用于相依性测度的股票平均相关系数后，回归模型的预测拟合优度变化并不明显，说明传统用于相依性测度的股票简单平均相关系数并不能很好的用于预测金融资产的收益率。对比 reg_2 和 reg_3 的拟合优度 R^2 可以发现，reg_3 的 R^2 提高至 0.376，回归模型可以更好地解释和预测金融资产的收益率。进一步说明藤模型中相依水平及相依结构带来了额外的市场信息。

2. 市场波动率信息含量研究

与上文类似，选取沪深 300 股指的历史波动率作为回归分析的因变量，其中

历史波动率是通过窗口内的沪深 300 指数的每日收益率的标准差乘以交易日天数的平方根计算的。对比变量为 80 支股票的平均相关系数，同时依据相关文献选取传统估算变量货币供应量（M2）、中国宏观经济政策不确定性指数（EPU）、制造业 PMI 指数（PMI_M）作为控制变量，自变量均选取滞后一阶项，建立三个回归模型进行结果对比，其中传统估算变量来源于东方财富网和 Economic Policy Uncertainty 官网。

同样基于稀疏 Group-Lasso 方法筛选变量，最终得到结果如下表：

表 11 市场波动率研究特征子集表

变量	描述
τ _Weighted_centrality_1	第一棵树的 τ 加权中心性
Lower_weighted_centrality_1	第一棵树的下尾加权中心性
Uptail_weighted_centrality_2	第二棵树的上尾加权中心性
Eigenvector_centrality	特征向量中心性
PMI_M	制造业采购经理指数
EPU	经济政策不确定性
M2	货币供应量

通过变量构造三个以沪深 300 指数波动率作为因变量的回归模型，并对 R-藤 Copula 模型指标的拟合程度进行探究，结果见表 12。

表 12 各指标对沪深 300 指数波动率的回归结果

Variable	reg_1	reg_2	reg_3
Volatility_1	0.465** (2.310)	0.300 (1.174)	-0.108 (-0.782)
average_R		0.334 (1.043)	
M2	-0.045** (-3.038)	-0.022 (-0.849)	0.006 (0.368)
EPU	0.000** (-2.593)	0.000 (-1.571)	5.783E-5 (0.351)
PMI_M	1.790E-7* (1.931)	1.333E-7 (1.305)	-3.377E-8 (-0.251)
τ _Weighted_centrality_1			0.530* (2.340)
Lower_weighted_centrality_1			0.000 (-0.002)
Uptail_weighted_centrality_2			0.164 (0.478)
Eigenvector_centrality			0.555 (2.080)
constant	2.196** (2.950)	1.035 (0.774)	-0.466 (-0.598)
R^2	0.377	0.412	0.784
Adjusted R^2	0.348	0.352	0.644

注：估计系数对应括号内数据为 t 统计量；“*”、“**”、“***”分别表示在 10%、5%、1% 显著水平下显著

从预测模型结果可以看出，模型的拟合优度 R^2 在 reg_1 模型中为 0.348，当加入了股票平均相关系数这一变量后， R^2 提高至 0.352。值得注意的是，在加入传统用于相依性测度的股票平均相关系数后，回归模型的预测拟合优度变化并不明显，说明传统用于相依性测度的股票平均相关系数并不能很好的用于预测金融资产的波动率。在 reg_3 中加入本文提出的改进后 R 藤 Copula 模型网络特征指标，对比 reg_2 和 reg_3 的拟合优度 R^2 可以发现，reg_3 的 R^2 提高至 0.644，进一步说明一方面高维资产带来了大量额外的市场信息，另一方面通过加权复杂网络

指标对藤模型信息提取的充分性。

六、研究结论

本文主要研究了高维视角下多资产的相依性结构,并引入一种创新式的算法改进高维结构的选择过程,通过稳健性检验与实证分析验证了模型性能的优异性与结果的可靠性,具体地:

一方面,本文基于 R 藤的高维 Copula 模型去描述多对多的资产依赖关系,当资产数量增加时,藤结构的确定以及节点间连接函数 pair-copula 的估计问题导致计算复杂度呈指数级增长,传统 Dißmann 算法得到的最大生成树不能保证全局最优,以致于误差由低阶树累积并向下传播。为了解决这一问题,引入 Graphical-Lasso 与最大无向图模型对藤结构的选择进行改进,在更小的局部得到维数更低 R 藤结构并进行估计和连接。将最终的藤结构视作加权网络并引入相关指标进行评价描述,以多个维度充分挖掘网络信息含量。

另一方面,以每个窗口期的网络特征指标构建解释变量,并引入传统预测因子对分别市场收益率和波动率建立稀疏 Group-Lasso 回归模型,与总体简单平均相关系数替代指标变量作为对比模型,解决了以网络特征指标为组的变量带来的多重共线性,且进一步控制了对比模型的其他变量一致。通过模型对比验证了资产在高维动态相依结构下,蕴含的市场波动与收益信息。

最终基于实证分析得到以下结论:第一,股市在不同时期的显著性有明显变化,且部分股票的相依性表现独特,即具有局部的信息含量。第二,在各网络特征指标中,加权中心性的指标走势与沪深 300 波动率和收益率相似,市场震荡期间,其他指标出现不规则的变化,尾部相关系数加权中心性比 τ 加权中心性反应更敏感,波动更剧烈。第三,网络特征指标对市场收益率与波动率解释力的提升比总体简单平均相关系数更加显著。

参考文献

- [1] 韦艳华, 张世英. Copula 理论及其在金融分析上的应用 [M]. 北京: 清华大学出版社, 2008
- [2] 张尧庭. 连接函数 (copula) 技术与金融风险分析 [J]. 统计研究, 2002(04): 48-51.
- [3] 杨炘, 陈展辉. 中国股市三因子资产定价模型实证研究 [J]. 数量经济技术经济研究, 2003(12): 137-141.
- [4] 杜子平, 闫鹏, 张勇. 基于“藤”结构的高维动态 Copula 的构建 [J]. 数学的实践与认识, 2009, 39(10): 96-102.
- [5] 郑振龙, 杨伟. 金融资产收益动态相关性: 基于 DCC 多元变量 GARCH 模型的实证研究 [J]. 当代财经, 2012(07): 41-49.
- [6] 张国富, 杜子平. 基于 R 藤 copula 的深圳股票市场系统风险分析 [J]. 金融经济研究, 2014, 29(03): 76-87.
- [7] 叶五一, 李潇颖, 缪柏其. 基于藤 Copula 方法的持续期自相依结构估计及预测 [J]. 中国管理科学, 2015, 23(11): 29-38.
- [8] 龚金国, 邓入侨. 时变 C-Vine Copula 模型的统计推断 [J]. 统计研究, 2015, 32(04): 97-103.
- [9] 韩超, 严太华. 基于高维动态藤 Copula 的汇率组合风险分析 [J]. 中国管理科学, 2017, 25(02): 10-20.
- [10] 杨金海, 范黎波. 全球市场一体化——全球情景下 Fama-French 三因子模型检验 [J]. 技术经济, 2017, 36(06): 109-119.
- [11] 沈忱. 基于 Fama-French 模型的沪深 300 指数效应实证研究 [J]. 时代金融, 2018(21): 155+159.
- [12] 缪晓斌. 基于沪深 300 高频数据的股指波动预测方法研究 [J]. 改革与开

放, 2018(09):5-8.

- [13] 雷立坤, 余江, 魏宇, 赖晓东. 经济政策不确定性与我国股市波动率预测研究[J]. 管理科学学报, 2018, 21(06):88-98.
- [14] 刘珊珊. 统计方法在研究股票收益问题中的应用[J]. 山西农经, 2019(21):167-168.
- [15] 彭选华. 基于 DCC-Copula-SV-M-t 模型的股市系统性风险溢出分析[J]. 数理统计与管理, 2019, 38(05):929-939.
- [16] 白正午, 朱淑珍, 张学瑞. 基于动态 t-Copula 模型对我国资本市场的相关性研究[J]. 管理现代化, 2019, 39(06):5-7.
- [17] 王佳, 金秀, 王旭, 李刚. 基于时变 Markov 的 DCC-GARCH 模型最小风险套期保值研究[J]. 中国管理科学, 2020, 28(10):13-23.
- [18] 聂巧平, 胡冰倩. GARCH 模型族及其在金融市场中的应用研究综述[J]. 天津商业大学学报, 2020, 40(02):52-58.
- [19] 孟庆斌, 张永冀, 汪昌云. 中国股市是宏观经济的晴雨表吗?——基于马氏域变模型的研究[J]. 中国管理科学, 2020, 28(02):13-24.
- [20] 张琳, 张军, 王擎. 宏观经济信息发布对股票市场收益率及其波动的影响[J]. 系统工程理论与实践, 2020, 40(06):1439-1451.
- [21] 赵渊, 刘庆尧, 邝俊威, 谢开贵. 电网可靠性评估中相关性变量的非参数 R 藤 Copula 模型[J]. 中国电机工程学报, 2020, 40(03):803-812.
- [22] 何敏园, 李红权. 全球股市间的相依结构与极值风险溢出:基于藤 Copula 的金融复杂性分析[J]. 管理评论, 2020, 32(07):102-110.
- [23] 徐刚刚, 蔡学鹏, 熊尚敏. 藤 Copula 模型在我国股市多资产组合 VaR 预测中的应用[J]. 井冈山大学学报(自然科学版), 2020, 41(02):8-15.
- [24] 朱莞, 周瑛. 基于关联网络的系统性金融风险测度及时变特征研究[J]. 河南

- 科技学院学报, 2021, 41 (01) :8-13.
- [25] 姜美宏. 基于 Fama-French 三因子模型的上证 50 成分股分析[J]. 营销界, 2021 (18) :98-100.
- [26] 邓晓, 杨光艺. 基于分位数预测回归模型的股票市场风险影响因素分析[J]. 统计与决策, 2021, 37 (08) :141-144.
- [27] 郝建阳, 王璐, 张莉, 计玉. 汇率对我国股市波动预测的影响[J]. 武汉理工大学学报(信息与管理工程版), 2021, 43 (02) :137-144.
- [28] 谢铖. 基于 D 藤结构 Copula 函数的风险资产投资决策模型与运用[J]. 统计与决策, 2021, 37 (05) :174-178.
- [29] 韦勇凤, 向一波. 基于 Group-Lasso 方法的非均衡数据信用评级模型[J]. 中国科学院大学学报, 2021, 38 (02) :181-188.
- [30] 吕亮雯. DCC-MVGARCH 模型计算方法研究及在金融市场中的应用[D]. 暨南大学, 2006.
- [31] 左秀霞. 基于树结构 DCC_多元 GARCH 模型的中国股市波动相关性研究[D]. 华中科技大学, 2009.
- [32] 余笑荷. 金融市场相依性建模与风险测度研究[D]. 华东师范大学, 2016.
- [33] 林柯. 藤 Copula 建模理论研究[D]. 天津财经大学, 2017.
- [34] 张阳. 基于尾部风险网络的金融风险传染与防范研究[D]. 吉林大学, 2020.
- [35] 王鑫. 金融机构尾部风险网络关联性及其对系统性风险的影响研究[D]. 上海师范大学, 2020.
- [36] Dependence of Temporty Threshold Shift at 4 ke on Intensity and Time[J]. Journal of Occupational Medicine, 1959, 1(7).
- [37] Tibshirani R . Regression Selection and Shrinkage via the Lasso[J]. R Stat Soc Ser B Stat Methodol. 1994; 58 :267-288.

- [38] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological),1996,58(1).
- [39] Bedford T. and Cooke R.M. Probability density decomposition for conditionally dependent random Variables modeled by vines [J]. Annals of Mathematics and Artificial Intelligence,2001(32):245-268.
- [40] Bedford T. and Cooke R.M. Vines:A new graphical model for dependent random Variables [J]. Annals of Statistics,2002(30):1031-1068.
- [41] Clive W.J. Granger,Timo Teräsvirta,Andrew J. Patton. Common factors in conditional distributions for bivariate time series[J]. Journal of Econometrics,2006,132(1).
- [42] Kjersti Aas,Claudia Czado,Arnoldo Frigessi,Henrik Bakken. Pair-copula constructions of multiple dependence[J]. Insurance Mathematics and Economics,2007,44(2).
- [43]Friedman Jerome,Hastie Trevor,Tibshirani Robert. Sparse inverse covariance estimation with the graphical lasso.[J]. Biostatistics (Oxford, England),2008,9(3).
- [44]Jerome Friedman,Trevor Hastie,Rob Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent[J]. Journal of Statistical Software,2010,33(1).
- [45]Mazumder Rahul,Hastie Trevor. Exact Covariance Thresholding into Connected Components for Large-Scale Graphical Lasso.[J]. Journal of machine learning research : JMLR,2012,13.
- [46] J. Dißmann,E.C. Brechmann,C. Czado,D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns[J]. Computational Statistics and Data Analysis,2013,59.

- [47] Czado,C.,Jeske, S., Hofmann, M., 2013. Selection strategies for regular vine copulae[J].Soc.Fr.Statist.154,174–191.
- [48] Ingrid Hobæk Haff,Kjersti Aas,Arnoldo Frigessi, Virginia Lacal. Structure learning in Bayesian Networks using regular vines[J]. Computational Statistics and Data Analysis,2016,101.
- [49] Kovacs E , T Szántai. On the connection between cherry-tree copulas and truncated R-vine copulas[J]. Kybernetika -Praha-, 2016, 53(3).
- [50] Dominik Müller,Claudia Czado. Representing Sparse Gaussian DAGs as Sparse R-Vines Allowing for Non-Gaussian Dependence[J]. Taylor & Francis,2018,27(2).
- [51] Dominik Müller,Claudia Czado. Selection of sparse vine copulas in high dimensions with the Lasso[J]. Springer US,2019,29(2).
- [52] Dominik Müller,Claudia Czado. Dependence modelling in ultra high dimensions with vine copulas and the Graphical Lasso[J]. Computational Statistics and Data Analysis,2019,137.

附录一 股票基本信息

序号	股票	行业	流通市值	序号	股票	行业	流通市值
1	贵州茅台	食品饮料	25890.2 亿	41	京东方 A	电子	2200.02 亿
2	工商银行	银行	14693.9 亿	42	荣盛石化	化工	1651.23 亿
3	中国平安	非银金融	8498.23 亿	43	宁波银行	银行	2328.62 亿
4	招商银行	银行	10403.2 亿	44	中信银行	银行	1832.03 亿
5	建设银行	银行	694.58 亿	45	民生银行	银行	1748.28 亿
6	五粮液	食品饮料	10281.0 亿	46	招商证券	非银金融	1462.88 亿
7	农业银行	银行	9850.85 亿	47	亿纬锂能	电气设备	1314.28 亿
8	中国人寿	非银金融	6615.64 亿	48	中国建筑	建筑装饰	2113.52 亿
9	中国银行	银行	6976.34 亿	49	光大银行	银行	1654.13 亿
10	中国石油	采掘	6914.07 亿	50	通威股份	电气设备	1414.99 亿
11	海康威视	电子	4656.75 亿	51	航发动力	国防军工	961.31 亿
12	恒瑞医药	医药生物	4811.72 亿	52	长春高新	医药生物	1633.54 亿
13	中国中免	休闲服务	6089.77 亿	53	片仔癀	医药生物	1980.99 亿
14	兴业银行	银行	4570.35 亿	54	赣锋锂业	有色金	814.34 亿
15	中国石化	化工	4166.32 亿	55	保利地产	房地产	1708.05 亿
16	顺丰控股	交通运输	3634.91 亿	56	云南白药	医药生物	700.20 亿
17	长江电力	公用事业	4805.35 亿	57	中国中车	机械设备	1537.52 亿
18	平安银行	银行	4199.41 亿	58	中远海控	交通运输	1398.15 亿
19	立讯精密	电子	2418.14 亿	59	分众传媒	传媒	1391.46 亿
20	万华化学	化工	1522.42 亿	60	汇川技术	电气设备	1312.82 亿
21	格力电器	家用电器	3682.24 亿	61	双汇发展	食品饮料	535.90 亿

22	中信证券	非银金融	2360.43 亿	62	华泰证券	非银金融	1244.91 亿
23	中国神华	采掘	3219.05 亿	63	闻泰科技	电子	869.38 亿
24	泸州老窖	食品饮料	3353.26 亿	64	用友网络	计算机	1169.09 亿
25	三一重工	机械设备	2964.74 亿	65	三安光电	电子	1017.97 亿
26	东方财富	非银金融	1988.46 亿	66	潍柴动力	汽车	805.57 亿
27	中国太保	非银金融	2445.05 亿	67	福耀玻璃	汽车	925.38 亿
28	爱尔眼科	医药生物	2087.13 亿	68	宝钢股份	钢铁	1888.05 亿
29	万科 A	房地产	2886.11 亿	69	中兴通讯	通信	1122.88 亿
30	山西汾酒	食品饮料	2996.27 亿	70	国电南瑞	电气设备	1446.85 亿
31	恒力石化	化工	2050.49 亿	71	海通证券	非银金融	941.58 亿
32	交通银行	银行	1935.07 亿	72	TCL 科技	电子	1283.32 亿
33	海尔智家	家用电器	2031.35 亿	73	中信特钢	钢铁	336.32 亿
34	洋河股份	食品饮料	1985.55 亿	74	上海机场	交通运输	646.46 亿
35	浦发银行	银行	3126.00 亿	75	中国联通	通信	1300.89 亿
36	伊利股份	食品饮料	2344.82 亿	76	歌尔股份	电子	847.02 亿
37	上汽集团	汽车	2349.54 亿	77	中国中铁	建筑装饰	1183.12 亿
38	智飞生物	医药生物	1702.78 亿	78	东方雨虹	建筑材料	894.00 亿
39	海螺水泥	建筑材	2065.05 亿	79	中航沈飞	国防军工	912.65 亿
40	紫金矿业	有色金属	2011.17 亿	80	复星医药	医药生物	808.61 亿

致 谢

在本次研究论文的撰写中，我们得到了指导老师的亲切关怀和精心指导，不论是最开始的定题还是之后的实证研究，老师都给予了我们耐心的指导以及举足轻重的意见，这让我们在模型构建和撰写论文方面都有了较大提高。同时，老师严谨的科学态度、丰富渊博的知识、精益求精的工作作风以及毁人不倦的师者风范也显示了老师高度的敬业精神和职责感。在此，我们对指导老师表示诚挚的感谢以及真心的祝福。

转眼研究生一年的在校生活即将过去，感谢同学们的相遇，让我们在积极向上的氛围中互相学习、互相扶持。在之后的一年中让我们继续保持乐观的态度共同努力，共同积累步入社会和进入工作岗位的经验。同时，感谢学校和学校领导给我们创造的良好学习环境，让我们能够在良好的氛围中完成论文。