

# 考虑股市网络因子的多因子选股策略研究

戴云飞, 朱宗元

浙江财经大学数据科学学院, 浙江 杭州

收稿日期: 2024年7月2日; 录用日期: 2024年7月23日; 发布日期: 2024年8月7日

## 摘要

本文基于中国A股市场数据, 选取四个节点网络指标构建了网络因子, 将其与传统的Fama-French三因子模型相结合, 提出了新的四因子选股模型。实证研究表明, 该模型比三因子模型更能解释收益率的变动。基于四因子模型进行的选股策略的回测结果显示, 网络因子构建的投资组合在收益率、风险和最大回撤率等方面均优于上证A股指数。

## 关键词

Fama-French三因子模型, 复杂网络, 因子分析, 量化选股

# Research on Multi-Factor Stock Selection Strategy Considering Network Factors in the Stock Market

Yunfei Dai, Zongyuan Zhu

School of Data Sciences, Zhejiang University of Finance and Economics, Hangzhou Zhejiang

Received: Jul. 2<sup>nd</sup>, 2024; accepted: Jul. 23<sup>rd</sup>, 2024; published: Aug. 7<sup>th</sup>, 2024

## Abstract

This paper, based on data from China's A-share market, constructs a network factor using four network indicators and combines it with the traditional Fama-French three-factor model to propose a new four-factor stock selection model. Empirical research shows that this model can better explain variations in returns compared to the three-factor model. Back-testing results of the stock selection strategy based on the four-factor model indicate that the investment portfolio constructed with the network factor outperforms the Shanghai A-share index in terms of returns, risk, and maximum drawdown rate.

## Keywords

Fama-French Three-Factor Model, Complex Networks, Factor Analysis, Quantitative Stock Selection

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着科技的不断进步和金融业的持续创新, 中国的金融市场也在经历着深刻的变革。早在上世纪五六十年代, 量化投资便初见雏形。随着计算机技术的发展, 分析师和学者开始利用计算机处理大量市场数据, 以寻找投资机会。然而, 由于当时计算机技术尚不成熟, 大规模数据计算依然困难, 量化投资未能迅速发展。随着计算机技术的进一步普及和金融市场的国际化, 量化投资逐步进入更为系统和正式的阶段。

与此同时, 复杂网络的相关理论也随着计算机科学技术的广泛应用而从图论发展而来。Watts 和 Strongatz (1998) [1]提出了小世界网络模型, Barabási 和 Albert (1999) [2]提出了复杂网络的无标度特性, 也被称为 BA (Barabási-Albert)无标度网络模型。小世界网络和无标度网络的发现成为复杂性科学的标志性事件, 这两篇文章发表后在短时间内获得了大量引用, 占据了网络科学引文的核心位置, 推动了网络科学理论的深入发展。

股票市场天然具有复杂网络的特点, 越来越多的学者开始利用复杂网络理论对中国股票市场进行研究。本文根据 2013 年 7 月至 2023 年 6 月中国股市收益率数据构建了月度复杂网络, 并将节点网络指标与传统的 Fama-French 三因子模型相结合, 构建了新的四因子模型, 以探究各网络指标因子对模型的贡献。这为推动金融研究的创新和金融市场的发展提供了新思路, 对适应数字化时代的经济环境和应对市场的变革具有重要意义。

## 2. 文献综述

### 2.1. 复杂网络在股市中的应用

学者们的分析研究发现, 股票市场与复杂系统有着高度的相似性, 因此, 越来越多的人开始认为股票市场是一个复杂系统。随着这一认识的深入, 人们开始重视利用复杂网络理论来解释金融市场中的各种问题。如吴婕等(2022) [3]在刚结束的新冠疫情环境中, 利用复杂网络理论研究了股票市场波动性, 并提出将关节点定向攻击的方法运用于股票网络, 最终挖掘出在风险状态下需要重点监控的股票, 以防风险的大规模传播。徐玉华等(2022) [4]通过复杂网络理论建立了我国股票一级行业指数收益率模型, 研究了我国股市各行业板块之间的风险传染机制, 并运用最小树形图法研究了 2008 年金融危机、2015 年股灾及 2020 疫情常态化 3 个阶段的我国股票市场风险传导路径。胡宁(2021) [5]也类似地运用最小生成树的方法构建了上证 50 股票网络, 利用特征向量中心性思想研究了节点的系统性风险, 对于股票市场的监管和投资而言, 都具有一定的现实意义。

复杂网络理论在股票市场研究中的应用不仅丰富了学术研究的内容, 还为投资者提供了更多有效的工具和方法, 促进了股票市场的健康发展和投资环境的改善。

## 2.2. 多因子模型的发展

多因子模型的发展历程可以追溯到上个世纪 50 年代, 先是 Markowitz (1952) [6] 提出了均值方差理论, 为投资组合管理提供了重要的框架。在此基础上, Sharp (1964) [7] 提出了资本资产定价模型(Capital Asset Pricing Model, CAPM), 它假设资产收益率只与市场因素相关, 这一定价模型已经成为现代资产定价问题的基础和核心。随着对市场行为更深入的理解, 人们开始意识到单因子模型难以解释所有的资产收益率波动, 学者们开始提出多因子模型, Ross (1976) [8] 提出套利定价理论, 而后 Fama 和 French (1992) [9] 提出了著名的 Fama-French 三因子模型, 将股票收益率的波动分解为市场因子、公司规模因子和账面市值比因子, 这一模型在解释和预测股票收益率方面取得了显著的成果, 并成为学术界和实践中的重要工具。而在之后, Fama-French 在三因子模型的基础上开始考虑其他因素对资产收益率的影响, 例如市场动量因子与投资因子, 并于 2015 年再次提出了 Fama-French 五因子模型[10]。从三因子模型到五因子模型的发展, 也标志着对资产收益率影响因素更为全面的认识和理解, 为投资者和学者提供了更有效的工具来解释和预测资产收益率的波动。

自五因子模型提出以来, 被多个市场证实其拥有更强的解释力度。然而, 当学者们验证其在中国股市的表现时, 得到的结果却不相同。有学者认为五因子模型的效果显著优于三因子模型, 如李志冰等(2017) [11] 以 1994 年 7 月至 2015 年 8 月 A 股上市公司为样本, 考察了五因子模型在中国股市不同时期的应用, 发现五因子模型有非常强的解释能力, 比 CAPM、三因子模型及 Carhart 四因子模型表现更好。也有学者持相反意见, 如赵胜民等(2016) [12] 通过实证分析发现我国股市市值效应和价值效应明显, 而 RMW 和 CMA 两个因子却无法诠释股票组合的回报率。除了对模型的检验之外, 还有很多学者尝试对模型进行改进。如赵合(2023) [13] 将投资者情绪作为新因子与三因子模型结合, 发现可以对收益进行一定的解释; 潘映锦(2024) [14] 则在三因子模型中依次加入 ESG 因子和它的三个二级因子, 构建出四个四因子模型, 论证了我国 A 股市场上 ESG 因子的有效性。

## 2.3. 文献评述

综合来看, 复杂网络理论和多因子模型在股票市场研究中的应用可以互为补充。复杂网络理论通过构建股票网络和分析网络指标, 为识别市场风险点和研究风险传染机制提供了新视角。而多因子模型则通过不断引入新的因子, 提升了对股票收益率波动的解释力和预测力。故本文暂不对五因子模型的适用性进行验证, 旨在通过建立股票网络、计算网络指标并构建网络指标因子的方法, 在三因子模型的基础上加入网络因子形成四因子模型, 检验股市网络指标对 A 股收益的解释程度, 并以此构建投资组合, 评估其效果。

## 3. 数据的来源与处理

本文的数据来源于 Wind 数据库和 CSMAR 国泰安数据库。研究对象为沪深股市的所有 A 股股票, 交易周期为 2013 年 7 月 1 日至 2023 年 6 月 30 日, 共计 10 年, 包含 2435 个交易日。构建网络所需的收益率数据来自 CSMAR 数据库的沪深 A 股考虑现金红利再投资的日个股回报率数据; 而实证研究所需的股票月度收益、总市值与市净率数据均来源于 Wind 数据库。本章的数据处理采用 Python 语言进行。

本文将提取的交易日数据按每月一个周期划分为 120 个窗口, 以构建 120 个时序排列的股市网络, 从而更好地研究股市网络的动态演化性质。在构建网络之前, 由于某些数据存在缺失或异常, 还需对数据进行相应的处理。具体步骤如下:

- 1) 以一个月为一个周期, 清除每周期内数据缺失占比超过 20% 的股票;
- 2) 缺失数据占比未超过 20% 的股票, 通过前项后项填充法补充缺失数据;

- 3) 清除市净率为负数的股票;
- 4) 为了避免股票市场的壳效应, 删除各周期市值最小的 30% 股票(Liu 等, 2019) [15];
- 5) 因 Wind 数据库无直接的账面市值比数据, 因此本文采用市净率的倒数作为股票的账面市值比。

## 4. 方法设计

### 4.1. 研究思路

本文首先根据中国 A 股个股收益率数据, 采用阈值法构建月度复杂网络, 并计算和整理网络各节点的指标值。接着, 计算 Fama-French 三因子模型的传统三因子, 并基于网络指标计算新的因子, 之后进行因子的适用性分析, 包括因子的平稳性检验和变量相关性检验, 以确保没有冗余因子。然后, 将网络因子引入传统的 Fama-French 三因子模型中, 构建新的四因子模型, 并进行 GRS 检验与回归分析, 比较三因子模型与四因子模型的效果。最后, 根据四因子模型结果构建投资组合, 并评估投资组合的表现。

### 4.2. 股市网络的构建

#### 1) 计算相关系数矩阵

股市网络构建的方法有很多, 本文则采用吴翎燕等(2013) [16]的方法, 基于股票收益率的相关系数构建网络, 通过捕捉股票收益率之间的相关性来揭示资产之间的协同效应, 有助于理解股票市场的整体动态。具体方法是先计算每只股票收益率之间的相关系数, 将股票作为节点, 两只股票之间的相关系数则构成了网络中的连边, 从而构建一个无向网络。本文选择使用皮尔逊相关系数来构建股票收益率之间的网络。

#### 2) 通过邻接矩阵构建网络

虽然每两个股票收益率都可以计算相关系数, 但并不是每两个节点之间都有连边。本文采取阈值法对相关系数矩阵进行处理, 构建网络邻接矩阵, 从而生成股市网络。记两股票节点  $i$  与  $j$  之间的相关系数  $\rho_{ij}$ , 给定阈值  $\theta$ 。因为网络只考虑两只股票相关性的强弱, 不考虑方向, 因此选择绝对值  $|\rho_{ij}|$  进行比较。若  $|\rho_{ij}|$  大于阈值  $\theta$ , 则两股票之间记有一条连边, 同时邻接矩阵对应值记为 1, 如果小于则没有, 邻接矩阵对应值记为 0。同时, 网络中不允许自环, 因此网络邻接矩阵对角线元素  $a_{ii} = 0$ 。从而得到邻接矩阵  $A$ , 最后通过邻接矩阵构建出无权无向的股市网络模型。具体判定公式如下:

$$a_{ij} = \begin{cases} 1 & |\rho_{ij}| > \theta \text{ and } i \neq j \\ 0 & \text{else} \end{cases} \quad (1)$$

阈值法构建网络方法较为简单直观, 但对阈值的选取要求比较严格。阈值设置过高, 网络会冗杂, 阈值设置过低, 可能会丢失重要的节点。因此, 本文采用每期相关系数均值作为网络阈值, 尽可能保证信息的完整性和准确性。最终构建共 120 个周期的网络, 每周期都有其各自的阈值。

根据每期阈值构建 120 个周期的股市网络, 从而可以计算出每个周期、各个节点的网络指标数值。本文选取度值、介数、聚集系数及接近中心性四个网络指标作为主要研究对象。这四个指标在网络分析中是比较常用且具有代表性的指标, 且相对容易计算。本文涉及的所有网络指标计算均采用 Python 语言中的 igraph 模块进行。

### 4.3. 因子模型的构建

#### 1) 模型的构建

前小节已构建了网络并计算相应的网络指标, 接着, 选取网络的度值、介数、聚集系数和接近中心性四个常见的网络指标, 分别构建不同的网络因子, 并将其作为 Fama-French 三因子模型的第四个因子,

以期提高模型的有效性, 并最终比较基于各模型选股的有效程度。

经典的 Fama-French 三因子模型可以由如下公式表述:

$$R_{it} - R_{ft} = a_i + b_i(R_{mt} - R_{ft}) + s_iSMB_t + h_iHML_t + \varepsilon_{it} \quad (2)$$

其中,  $R_{it}$  为  $t$  时期标的资产的收益率;  $R_{ft}$  为  $t$  时期的市场无风险收益率;  $R_{mt}$  为  $t$  时期的市场收益率;  $SMB_t$  为  $t$  时期低市值组合与高市值组合收益率之差;  $HML_t$  为  $t$  时期高账面市值比组合与低账面市值比组合收益率之差;  $a_i$  为截距项;  $b_i$ 、 $s_i$  和  $h_i$  为其对应因子的因子暴露;  $\varepsilon_{it}$  为 0 均值残差项。

加入网络因子后的四因子模型表达式如下:

$$R_{it} - R_{ft} = a_i + b_i(R_{mt} - R_{ft}) + s_iSMB_t + h_iHML_t + n_iNET_t + \varepsilon_{it} \quad (3)$$

其中,  $NET_t$  为  $t$  时期高网络因子组合与低网络因子组合收益率之差;  $n_i$  则为网络因子的因子暴露。

## 2) 因子的构建

在因子的构建上, Fama-French 三因子模型中, 采用 2\*3 方式构建了规模因子与价值因子。本文也采用类似的方法。将股票按市值分为高市值、低市值两组, 再将这两组分别按照账面市值比与各网络指标值按照 30%、40%、30% 的比例共分为 6 组, 并进行因子计算(表 1~3)。

**Table 1.** Cross-classification table of market value to book-to-market ratio

**表 1.** 市值 - 账面市值比交叉分组表

	低账面市值比	中账面市值比	高账面市值比
小市值	SL	SN <sub>B/M</sub>	SH
大市值	BL	BN <sub>B/M</sub>	BH

**Table 2.** Cross-classification table of market value-network factor

**表 2.** 市值 - 网络因子交叉分组表

	小网络因子值	中网络因子值	大网络因子值
小市值	SS	SN <sub>NET</sub>	SB
大市值	BS	BN <sub>NET</sub>	BB

**Table 3.** 2 × 3 factor construction method

**表 3.** 2 × 3 因子构建方法

因子模型	因子算法
三因子模型	$SMB = (SL + SN_{B/M} + SH)/3 - (BL + BN_{B/M} + BH)/3$
	$HML = (SH + BH)/2 - (SL + BL)/2$
四因子模型	$SMB_{B/M} = (SL + SN_{B/M} + SH)/3 - (BL + BN_{B/M} + BH)/3$
	$SMB_{NET} = (SS + SN_{NET} + SB)/3 - (BS + BN_{NET} + BB)/3$
	$SMB = (SMB_{B/M} + SMB_{NET})/2$
	$HML = (SH + BH)/2 - (SL + BL)/2$
	$NET = (SB + BB)/2 - (SS + BS)/2$

NET 网络因子的计算包括度值网络因子 NETd、介数网络因子 NETb、聚集系数网络因子 NETc 和接近中心性网络因子 NETcc。

最终构建好的因子值描述性统计如下表 4 所示。

**Table 4.** Descriptive statistics of factors

**表 4.** 因子描述性统计

因子名称	mean	var	min	max
SMB	0.51	22.02	-19.49	18.14
SMBd	0.34	22.16	-22.01	17.69
SMBb	0.41	24.16	-23.38	18.71
SMBc	0.42	22.51	-20.36	18.79
SMBcc	0.41	10.45	-16.68	10.64
HML	0.55	35.11	-21.67	23.42
NETd	1.19	11.13	-9.56	14.30
NETb	0.87	5.34	-4.58	11.57
NETc	0.35	10.69	-16.68	10.64
NETcc	0.41	10.45	-16.68	10.64

## 5. 实证分析

### 5.1. 因子的适用性检验

在进行进一步分析之前, 有必要对上述指标的适用性进行检验。

#### 1) 平稳性检验

本文使用的被解释变量以及各因子数据, 均为时间序列数据, 故有必要对各变量的平稳性进行检验。本章采用 ADF 方法检验时间序列的平稳性。传统三因子与各网络因子平稳性检验结果如下表 5 所示。

**Table 5.** Results of stationarity tests

**表 5.** 平稳性检验结果

变量	ADF 值	P 值	结论
MKT	-7.91	0.00	平稳
SMB	-9.98	0.00	平稳
SMBd	-8.64	0.00	平稳
SMBb	-9.45	0.00	平稳
SMBc	-7.28	0.00	平稳
SMBcc	-6.34	0.00	平稳
HML	-11.98	0.00	平稳
NETd	-3.08	0.03	平稳
NETb	-8.76	0.00	平稳
NETc	-8.47	0.00	平稳
NETcc	-3.11	0.03	平稳

由上表可以看出, 无论自变量还是因变量, ADF 检验的结果均小于临界值, P 值也均小于 5%, 因此拒绝原假设, 认为解释变量和被解释变量的时间序列数据均为平稳数据。

## 2) 变量相关性检验

为避免加入各网络因子后造成因子冗余, 本节通过因子相关系数矩阵和冗余因子检验两种方法判断因子间是否存在多重共线性(表 6)。

**Table 6.** Correlation matrix of factors

**表 6.** 因子相关系数矩阵

	MKT	SMB	HML
SMB	0.180	-	-
HML	-0.198	-0.543	-
NETD	-0.118	0.076	-0.078
NETB	-0.167	-0.057	-0.079
NETC	-0.286	0.187	0.025
NETCC	-0.132	0.068	-0.071

通过对 Fama-French 三因子模型的市场因子、规模因子和价值因子构建相关系数矩阵, 可以看出, HML 与 SMB 之间的相关系数为-0.543, 处于较高水平, 可能存在一定程度的替代与解释。而其余因子的相关系数低于 0.3, 因此不存在多重共线性问题。

Fama-French 在 2015 年提出了一种冗余因子的检验方法, 即通过各因子的相互回归并观察截距项是否为零, 来检测因子是否冗余。如果截距项为 0, 则该因子是冗余因子, 可以被其他因子取代; 如果截距项不为 0, 则该因子不是冗余因子, 不可被其他因子取代。本节对规模因子 SMB 和价值因子 HML, 在三因子和四因子模型下分别进行了因子冗余检验, 以判断它们的冗余情况(表 7)。

**Table 7.** Redundant factor test results for the three-factor model

**表 7.** 三因子模型冗余因子检验结果

	SMB	HML
MKT	0.0620 (0.064)	-0.1057 (0.080)
SMB		-0.6626*** (0.099)
HML	-0.4173*** (0.062)	
_cons	0.7461** (0.377)	1.0289** (0.473)
N.obs	120	120
adj.R <sup>2</sup>	0.288	0.293

注: 括号内为标准误差; \*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01。

在三因子的因子冗余检验结果中可以看到, 对 SMB 与 HML 两因子的回归结果中, 截距项均显著不为 0, 且调整 R<sup>2</sup> (adj.R<sup>2</sup>)均小于 0.3, 表示因子中 70% 以上的内容无法被其余因子解释。因此, 在三因子模型中不存在冗余因子(表 8、表 9)。

**Table 8.** Redundant factor test results for the four-factor model (SMB)  
**表 8.** 四因子模型冗余因子检验结果(SMB)

	SMB			
MKT	0.0659 (0.065)	0.0481 (0.065)	0.1193* (0.064)	0.0659 (0.065)
HML	-0.4143*** (0.036)	-0.4254*** (0.063)	-0.4109*** (0.060)	-0.4146*** (0.063)
NETD	0.0403 (0.086)			
NETB		-0.1430 (0.126)		
NETC			0.2377*** (0.077)	
NETCC				0.037 (0.087)
_cons	1.0283 (0.713)	-0.1205 (0.849)	1.1407*** (0.385)	1.0047 (0.714)
N.obs	120	120	120	120
adj. $R^2$	0.283	0.290	0.337	0.283

注: 括号内为标准误差; \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ 。

**Table 9.** Redundant factor test results for the four-factor model (HML)  
**表 9.** 四因子模型冗余因子检验结果(HML)

	HML			
MKT	-0.1125 (0.081)	-0.1272 (0.081)	-0.689 (0.085)	-0.1133 (0.081)
SMB	-0.6568*** (0.100)	-0.6672*** (0.089)	-0.6964*** (0.102)	-0.6571*** (0.100)
NETD	-0.0714 (0.109)			
NETB		-0.2608* (0.156)		
NETc			0.1351 (0.103)	
NETcc				-0.0726 (0.109)
_cons	0.5236 (0.904)	-0.5609 (1.062)	1.2711** (0.507)	0.5173 (0.905)
N.obs	120	120	120	120
adj. $R^2$	0.289	0.297	0.297	0.289

注: 括号内为标准误差; \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ 。

在上述对于 SMB 与 HML 的四因子冗余检验结果中可以看到, 多数回归结果的调整  $R^2$  (adj. $R^2$ )均小于 0.3, 表示因子中 70% 以上的内容无法被其余因子解释。只有在一种情况下, 即当 SMB 作为被解释变量, MKT、HML、NETc 作为解释变量时, adj. $R^2$  最高达到 0.337, 但此时的截距项结果显著不为 0。因

此, 可以认为在四因子模型中不存在冗余因子。

## 5.2. 模型比较分析

### 1) GRS 检验

前文已证明各个因子之间相互独立, 排除了多重共线性和冗余因子的可能性。本节将着重验证各个因子与不同资产组合收益率之间的解释关系。Gibbons 等(1989) [17]提出了 GRS 检验方法, GRS 统计量遵循 F 分布, 并假设多个回归截距项均为零。通过检验资产组合的回归截距项是否共同为零, 可以评估模型对投资组合的适用性。GRS 统计量值较小, 表示股票收益率中不能被因子解释的部分较少, 表明模型更为适用; 相反, 值较大, 则意味着模型的适用性较差(表 10)。

Table 10. GRS test results

表 10. GRS 检验结果

	GRS	Mean alpha	adj.R <sup>2</sup>
MKT SMB HML	1.5630662	-0.05214055	0.94363055
MKT SMB HML NETD	1.313939	-0.10767067	0.94514983
MKT SMB HML NETB	1.5046667	-0.09389435	0.94469017
MKT SMB HML NETC	1.5324555	-0.0734451	0.94529814
MKT SMB HML NETCC	1.5303984	-0.08103454	0.94532086

通过 GRS 检验结果可以看出, 与三因子模型相比, 加入网络因子的四因子模型具有更小的 GRS 统计量, 这表明加入网络因子可以提高三因子模型的适用度。分别对比以度值、介数、聚集系数和接近中心性作为网络因子的四因子模型时发现, 当度值作为网络因子时, GRS 统计量最低, 模型的解释能力最优。

### 2) 投资组合回归分析

本文将样本企业按规模和账面市值比从低到高分别分为 5 组, 组成 25 个不同规模和账面市值比的分组, 并进行回归分析。在前文的 GRS 检验中, 结果显示以度值作为网络因子的模型具有最佳的解释能力。因此, 为了避免文章内容的冗杂, 本文仅选取度值作为网络因子的四因子模型进行回归检验。将相同条件下的三因子模型回归结果作为对照组, 对比并分析加入网络因子后的四因子模型分组回归结果(表 11、表 12)。

Table 11. Regression results of 5 × 5 grouping for the three-factor model

表 11. 三因子模型 5 × 5 分组回归结果

规模	账面市值比									
	Low	2	3	4	High	Low	2	3	4	High
	a (截距项)					t(a)				
Small	-0.23	-0.08	0.30**	0.20	0.31	-0.92	-0.49	2.027	1.20	1.59
2	-0.35	0.13	-0.09	0.13	0.05	-1.61	0.811	-0.68	0.305	0.31
3	-0.36*	-0.20	-0.10	-0.22	-0.04	-1.76	-1.41	-0.66	-1.40	-0.24
4	-0.03	-0.06	-0.36**	0.02	-0.01	-0.17	-0.36	-2.19	0.110	-0.05
Big	0.35*	-0.26	-0.14	-0.48***	-0.06	1.79	-1.19	-0.66	-2.63	-0.44

续表

		<i>b</i> (MKT 系数)					<i>t</i> ( <i>b</i> )				
Small	0.95***	1.00***	1.02***	0.97***	1.02***	23.10	35.40	40.46	34.54	31.40	
2	1.00***	0.95***	0.98***	1.02***	1.04***	27.74	36.79	39.28	37.70	38.56	
3	0.99***	0.96***	0.98***	1.01***	1.06***	29.33	41.01	37.65	37.69	34.76	
4	0.95***	0.96***	1.04***	1.06***	1.03***	33.67	32.30	37.50	32.47	29.61	
Big	0.97***	1.03***	0.99***	1.19***	0.87***	29.87	28.69	27.82	38.58	35.74	
		<i>s</i> (SMB 系数)					<i>t</i> ( <i>s</i> )				
Small	1.13***	1.18***	1.16***	1.05***	0.95***	19.00	29.14	32.05	25.97	20.40	
2	0.98***	1.06***	0.93***	0.99***	0.81***	18.96	28.57	25.86	25.51	20.78	
3	0.69***	0.74***	0.85***	0.86***	0.70***	14.04	21.86	22.61	22.20	15.94	
4	0.47***	0.48***	0.59***	0.63***	0.57***	11.48	11.18	14.95	13.44	11.31	
Big	-0.42***	-0.17***	-0.06	-0.15***	-0.12***	-8.96	-3.27	-1.07	-3.42	-3.49	
		<i>h</i> (HML 系数)					<i>t</i> ( <i>h</i> )				
Small	-0.39***	-0.29***	-0.06**	0.13***	0.38***	-8.34	-9.042	-2.04	4.03	10.10	
2	-0.58***	-0.41***	-0.11***	0.09***	0.38***	-14.20	-13.78	-3.72	2.791	12.27	
3	-0.64***	-0.41***	-0.14***	0.12***	0.40***	-16.52	-15.32	-4.726	4.02	11.32	
4	-0.71***	-0.44***	-0.15***	0.07***	0.43***	-22.15	-13.07	-4.61	1.73	10.77	
Big	-0.81***	-0.31***	-0.04	0.28***	0.50***	-21.89	-7.54	-1.02	8.04	18.07	
		adj. <i>R</i> <sup>2</sup>					Durbin-Watson				
Small	0.935	0.969	0.970	0.953	0.930	2.614	1.667	1.961	2.136	1.917	
2	0.954	0.972	0.965	0.958	0.947	1.976	2.158	1.667	2.611	1.979	
3	0.952	0.971	0.960	0.952	0.930	2.204	2.111	1.749	1.954	1.942	
4	0.962	0.947	0.950	0.926	0.900	1.790	1.953	2.222	2.037	1.825	
Big	0.932	0.892	0.872	0.927	0.929	1.998	1.505	1.670	1.814	2.287	

注: 括号内的为标准误差; \**p* < 0.1, \*\**p* < 0.05, \*\*\**p* < 0.01。**Table 12.** Regression results of 5 × 5 grouping for the four-factor model with degree as network factor**表 12.** 以度值作为网络因子的四因子模型 5 × 5 分组回归结果

		账面市值比									
规模	Low	2	3	4	High	Low	2	3	4	High	
		<i>a</i> (截距项)					<i>t</i> ( <i>a</i> )				
Small	-0.08	0.10	0.23	0.13	0.17	-0.30	0.64	1.49	0.87	0.95	
2	-0.13	0.05	-0.11	-0.05	0.04	-0.61	0.321	-0.81	-0.28	0.26	
3	-0.34*	-0.44***	-0.18	-0.20	-0.17	-1.68	-3.07	-1.09	-1.29	-0.96	
4	-0.01*	-0.18	-0.42	-0.30	-0.23	-0.06	-0.96	-2.35	-1.63	-1.12	
Big	0.24	-0.21	-0.32	-0.34*	0.04	1.16	-0.96	-1.58	-1.78	0.28	

续表

		<i>b</i> (MKT 系数)					<i>t</i> ( <i>b</i> )				
Small	0.95***	1.00***	0.99***	0.95***	1.02***	22.90	41.23	40.91	38.38	34.35	
2	0.95***	0.96***	0.98***	1.01***	1.03***	27.77	41.65	46.37	38.35	37.66	
3	1.00***	0.95***	1.01***	1.02***	1.06***	30.91	41.23	39.09	40.02	36.43	
4	0.96***	0.95***	1.01***	1.03***	1.03***	34.95	30.96	35.28	34.46	31.33	
Big	0.98***	0.98***	0.98***	1.19***	0.88**	29.08	28.80	30.05	39.21	36.48	
		<i>s</i> (SMB 系数)					<i>t</i> ( <i>s</i> )				
Small	1.18***	1.19***	1.12***	0.95***	0.87***	16.54	28.65	26.79	22.25	17.04	
2	1.03***	1.12***	0.96***	0.92***	0.82***	17.48	28.14	26.68	20.50	17.44	
3	0.75***	0.71***	0.83***	0.87***	0.64***	13.45	18.00	18.72	19.89	12.72	
4	0.48***	0.45***	0.54***	0.52***	0.49***	10.120	8.52	11.02	10.01	8.34	
Big	-0.39***	-0.21***	-0.09	-0.13**	-0.08**	-6.80	-3.56	-1.53	-2.43	-1.99	
		<i>h</i> (HML 系数)					<i>t</i> ( <i>h</i> )				
Small	-0.37***	-0.23***	-0.07***	0.10***	0.31***	-7.69	-8.21	-2.60	3.42	9.15	
2	-0.60***	-0.34***	-0.09***	0.04	0.37***	-15.13	-12.65	-3.73	1.39	11.70	
3	-0.61***	-0.39***	-0.16***	0.12***	0.42***	-16.24	-14.58	-5.35	4.16	12.53	
4	-0.73***	-0.44***	-0.17***	0.10***	0.41***	-23.09	-12.43	-5.06	2.94	10.90	
Big	-0.82***	-0.36***	-0.01	0.31***	0.51***	-20.96	-9.11	-0.19	8.85	18.50	
		<i>n</i> (NET 系数)					<i>t</i> ( <i>n</i> )				
Small	0.03	-0.03	0.06	0.27***	0.18***	0.33	-0.55	1.14	5.13	2.87	
2	-0.10	-0.12**	-0.03	0.10*	0.04	-1.34	-2.40	-0.67	1.87	0.71	
3	-0.02	0.17***	0.09	0.05	0.15**	-0.34	3.57	1.62	0.97	2.53	
4	0.01	0.08	0.13**	0.19***	0.17**	0.11	1.28	2.08	3.04	2.50	
Big	0.02	0.03	0.19***	-0.07	-0.11**	0.26	0.36	2.85	-1.15	-2.18	
		adj. <i>R</i> <sup>2</sup>					Durbin-Watson				
Small	0.933	0.975	0.971	0.962	0.941	2.252	1.883	2.078	1.996	1.905	
2	0.955	0.976	0.974	0.958	0.945	1.983	1.889	1.818	2.270	1.943	
3	0.956	0.971	0.962	0.957	0.936	2.162	1.994	1.834	2.047	1.876	
4	0.964	0.940	0.944	0.931	0.909	1.734	1.768	2.274	1.894	1.718	
Big	0.927	0.895	0.888	0.929	0.933	1.799	1.659	1.676	1.924	2.172	

注: 括号内的为标准误差; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ 。

首先来看截距项, 在 25 个回归组合中, 三因子模型中有 5 个组合的截距项系数显著不为 0, 其中 3 个组合的 P 值小于 0.05, 1 个组合的 P 值小于 0.01, 分别在 5% 和 1% 的显著性水平下拒绝原假设。在四因子模型中, 仅有 4 个组合的截距项系数显著, 其中只有 1 个组合的 P 值小于 0.05, 另 1 个组合的 P 值小于 0.01。此外, 还有一个显著项的系数值接近 0 (0.01)。由此可以看出, 四因子模型对收益率的解释能力较三因子模型有了略微的提升。

其次, 在观察两模型的市场因子(MKT)系数时发现, 不论是三因子模型还是四因子模型, 25 个组合

下的 MKT 系数均在 1% 的显著性水平下显著。此外, 四因子模型与三因子模型的市场因子系数范围分布在 0.8~1.2 的范围内, 两者没有较大的差别。因此, 市场因子对模型的解释力度在两种模型中都表现出较强的稳定性。

接下来看规模因子(SMB)系数, 在 25 个组合中, 两个模型的 SMB 系数均有 24 个组合的系数显著, 大多数显著性水平小于 1%。这说明规模因子在模型中是一个强有力的因子。进一步观察发现, 规模因子系数  $s$  与其对应的  $t$  值与规模值成反比, 表明存在着股票的明显规模效应, 且规模较小的股票收益率更高。

与 SMB 系数类似, HML 系数也在两模型中均有 24 个组合系数显著, 并且多数以显著性水平小于 1% 通过  $t$  检验。系数的值随着账面市值比的不断升高而变得越来越大, 因此股票收益与账面市值比呈正向相关关系。

最后看网络因子(NET)系数, 25 个组合中有 11 个组合系数显著, 其余 14 个组合系数不显著且系数较小。以度值作为依据的网络因子虽然对模型有一定程度的解释, 但贡献程度远不如传统的三因子。

综上所述, 通过对比三因子模型和加入网络因子的四因子模型的回归结果, 可以发现整体上四因子模型的回归结果比三因子模型更好。其中, 市场因子是所有因子中最显著、对模型解释力度最大的因子; 规模因子也表明公司规模较小的股票收益率更高; 而账面市值比因子系数的值随着账面市值比的增加而上升; 以度值为依据的网络因子虽不如传统三因子, 但对模型优化有一定程度的贡献。

### 5.3. 基于网络因子的选股策略研究

#### 1) 选股策略设计

通过四因子模型回归结果中各组合网络因子的回归系数进行选股策略研究。选取 2013 年 7 月至 2023 年 6 月期间的中国 A 股股票进行选股策略研究。考虑到中国 A 股市场的特殊性, 不便于直接进行做空操作, 因此本文只考虑做多的情况。选取一个月作为一次调整周期, 具体的选股步骤如下(王超, 2018) [18]:

第一步, 将上述已经分好的 25 组回归组合的股票, 按照网络指标的大小再次进行分组。以度值为例, 在每个按照市值 - 账面市值比分好的 25 组中, 再依次按照度值的大小分为两组, 共计 50 组。

第二步, 根据每一期分好的 50 个组合, 以对应的回归系数为基准, 构建选股矩阵。具体构建方法为: 若组合对应的回归系数为正, 则买入网络指标较高的一组; 若组合对应的回归系数为负, 则买入网络指标较低的一组。针对不同网络指标分别构建不同的选股池, 并在每一期结束后清仓后进行调仓。

第三步, 每个选股池初期总净值均设为 1, 并按等权重方法分配资金。随后, 分别计算各选股组合的收益率数据, 并与上证 A 股指数收益进行对比。最终, 从最终净值、年化收益率、年化波动率、最大回撤率和夏普比率这五个维度进行评价。

#### 2) 策略结果分析

**Table 13.** Back-testing results of different stock pool and Shanghai A-share index

**表 13.** 不同选股池与上证 A 股指数的回测结果

	最终净值	年化收益率	年化波动率	最大回撤率	夏普比率
上证 A 股	1.62	4.95%	19.82%	45.92%	0.25
度	2.08	7.60%	19.00%	45.56%	0.40
介数	2.22	8.31%	19.15%	44.32%	0.42
聚集系数	2.47	9.47%	18.49%	43.54%	0.51
接近中心性	2.37	9.00%	18.67%	42.54%	0.48

结果显示(表 13), 不同网络因子构建的选股池与上证 A 股指数的净值走势在样本期内呈现出相似的波动趋势, 但波动强度各不相同。值得注意的是, 以网络指标为依据构建的股票池收益始终高于上证 A 股指数。从最终净值和年化收益率数据来看, 网络因子策略的投资组合表现出更高的收益水平。而年化波动率和最大回撤率的数据则显示, 网络因子策略的投资组合风险相对较低。此外, 通过夏普比率的比较, 我们可以看出, 网络因子策略相对于上证 A 股指数具有更好的风险调整后收益。因此, 基于网络因子的投资策略在风险和收益之间达到了更好的平衡, 并具有明显的优势。

## 6. 总结与启示

### 6.1. 总结

1) 本文利用 2013 年 7 月至 2023 年 6 月的中国 A 股市场数据, 以股票为节点, 利用股票间相关系数构建连边, 采用阈值法建立了 120 个股市网络。

2) 计算每个样本周期内股票的度、介数、聚集系数和接近中心性等四个网络指标, 采用类似计算规模因子与账面市值比因子的方法, 计算出相应的网络因子, 并结合传统的三因子模型, 构建了包含网络因子的四因子模型。

3) 通过 GRS 检验发现, 加入四种网络因子均能提高因子模型的适用度, 其中以度值作为网络因子时, 模型的解释能力最优。对比三因子模型和以度值作为第四因子的四因子模型的回归结果, 发现四因子模型整体上能更好地解释收益。

4) 基于四因子回归结果构建网络因子选股策略进行选股并调仓, 通过回测数据发现, 选股策略结果在收益率、波动率、最大回撤率和夏普比率方面均优于上证 A 股指数。

综上所述, 本文通过引入复杂网络分析, 构建了包含网络因子的四因子模型, 提高了模型对收益的解释能力, 并在实证回测中展示了其在投资策略中的优越表现。这一研究不仅丰富了因子模型的理论框架, 也为投资者提供了更有效的选股工具, 具有重要的理论和实践意义。

### 6.2. 启示

对于投资者来说, 首先, 鉴于网络因子能够提高多因子模型的适用度, 可以在量化投资策略中考虑网络指标的使用, 同时进一步研究不同的网络指标, 以确定最有效的因子组合。其次, 可以探索更多与网络结构相关的指标, 如节点重要性、网络连通性等, 进一步完善多因子模型, 提升投资决策的精准性。综上所述, 结合股市网络指标进行多因子模型分析具有一定的应用前景, 但仍需不断深化研究、优化策略, 并注意风险管理, 以实现长期稳健的投资回报。

对于监管者而言, 监管机构可以制定相关规定, 规范量化投资机构和基金管理人在使用网络指标时的行为。这包括对网络因子的计算方法、数据来源和使用方式进行监管, 以确保投资者利益和市场稳定。

## 基金项目

杭州市哲学社会科学项目(编号: Z23JC039); 国家社会科学基金项目(编号: 19BTJ044)。

## 参考文献

- [1] Watts, D.J. and Strogatz, S.H. (1998) Collective Dynamics of "Small-World" Networks. *Nature*, **393**, 440-442. <https://doi.org/10.1038/30918>
- [2] Barabási, A.L. and Albert, R. (1999) Emergence of Scaling in Random Networks. *Science*, **286**, 509-512. <https://doi.org/10.1126/science.286.5439.509>
- [3] 吴婕, 许忠好, 翟心彤. 基于复杂网络的新冠疫情下股票市场波动模型研究[J]. 应用概率统计, 2022, 38(4):

- 603-616.
- [4] 徐玉华, 赵玥, 高新柠, 等. 复杂网络理论下我国股市行业板块风险传染效应研究[J]. 江西科学, 2022, 40(2): 207-214.
- [5] 胡宁. 网络视角下的中国股票市场系统性风险[J]. 中国外资, 2021(18): 82-84.
- [6] Markowitz, H. (1952) Portfolio Selection. *The Journal of Finance*, **7**, 77-91.  
<https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- [7] Sharpe, W.F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, **19**, 425-442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- [8] Ross, S.A. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, **13**, 341-360.  
[https://doi.org/10.1016/0022-0531\(76\)90046-6](https://doi.org/10.1016/0022-0531(76)90046-6)
- [9] Fama, E.F. and French, K.R. (1992) The Cross-Section of Expected Stock Returns. *Journal of Finance*, **47**, 427-465.  
<https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- [10] Fama, E.F. and French, K.R. (2015) A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, **116**, 1-22.  
<https://doi.org/10.1016/j.jfineco.2014.10.010>
- [11] 李志冰, 杨光艺, 冯永昌, 景亮. Fama-French 五因子模型在中国股票市场的实证检验[J]. 金融研究, 2017(6): 191-206.
- [12] 赵胜民, 闫红蕾, 张凯. Fama-French 五因子模型比三因子模型更胜一筹吗——来自中国 A 股市场的经验证据[J]. 南开经济研究, 2016(2): 41-59.
- [13] 赵合. 基于三因子模型下投资者情绪对我国 A 股收益率的影响研究[D]: [硕士学位论文]. 昆明: 云南师范大学, 2023.
- [14] 潘映锦. ESG 因子在我国 A 股市场上的有效性研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2024.
- [15] Liu, J., Stambaugh, F.R. and Yuan, Y. (2019) Size and Value in China. *Journal of Financial Economics*, **134**, 48-69.  
<https://doi.org/10.1016/j.jfineco.2019.03.008>
- [16] 吴翎燕, 韩华, 宋宁宁. 基于相关系数和最佳阈值的股票网络模型构建[J]. 复杂系统与复杂性科学, 2013, 10(4): 49-55+85.
- [17] Gibbons, M.R., Ross, S.A. and Shanken, J. (1989) A Test of the Efficiency of a Given Portfolio. *Econometrica*, **57**, 1121-1152. <https://doi.org/10.2307/1913625>
- [18] 王超. 基于复杂网络度值的多因子模型选股策略研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2018.