

中国股票市场的异质空间因子定价模型

戈舒怡¹, 李少然^{2,3}, 黎新平^{2,3}, 苏文⁴

(1. 南开大学金融学院, 天津 300350; 2. 北京大学经济学院, 北京 100871; 3. 北京大学金融工程实验室, 北京 100871; 4. 牛津大学数学学院, 牛津 OX2 6GG, 英国)

摘要 本文利用异质空间因子模型同时刻画 A 股市场中由共同因子驱动的全局相关和由空间关联驱动的局部相关, 并通过文本分析挖掘百万条财经新闻中隐含的股票间关联, 用以构造模型中衡量股票间关联强弱的空间矩阵。我们介绍了模型框架下的统计推断方法及渐进无套利理论, 并将渐进无套利理论推广到包含不可交易因子的情景。实证方面, 我们利用 2012–2021 年的 A 股股票和相关因子的周收益率对模型进行分析, 并将新闻关联网络与其他常用关联网络展开全方位比较。结果表明: A 股市场局部相关性的影响具有明显的异质性, 与其他关联网络相比, 新闻关联在多个维度拥有最佳表现, 这得益于新闻文本涵盖最多的行业外关联信息, 并且信息更新速度快。此外, 不可交易因子在 A 股市场上具有不可忽略的影响。

关键词 局部相关性; 文本分析; 空间计量; 新闻关联矩阵; 不可交易因子

Heterogeneous Spatial Factors Pricing Model in the Chinese Stock Market

GE Shuyi¹, LI Shaoran^{2,3}, LI Xinping^{2,3}, SU Wen⁴

(1. School of Finance, Nankai University, Tianjin 300350, China; 2. School of Economics, Peking University, Beijing 100871, China; 3. FinLab of Peking University, Beijing 100871, China; 4. Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK)

Abstract This paper employs a heterogeneous spatial factor model to accommodate both strong and local dependencies simultaneously. The spatial matrix is approxi-

收稿日期: 2025-04-02

基金项目: 国家自然科学基金 (72303119, 72495155, 72203007); 东湖高新区国家智能社会治理实验基地项目

Supported by National Natural Science Foundation of China (72303119, 72495155, 72203007); National Intelligent Society and Governance Foundation of Wuhan East Lake High-Tech Development Zone

作者简介: 戈舒怡, 博士, 副教授, 研究方向: 金融计量, 实证资产定价, E-mail: sg751_shuyige@outlook.com; 通信作者: 李少然, 博士, 助理教授, 博士生导师, 研究方向: 金融计量, 资产定价, E-mail: lishaoran@pku.edu.cn; 黎新平, 研究员, 北京大学金融工程实验室负责人, 研究方向: 量化投资, 资产定价, E-mail: xinpinglei@pku.edu.cn; 苏文, 博士研究生, 研究方向: 金融数学, 金融计量, E-mail: wen.su@maths.ox.ac.uk.

mated by a news co-mention network constructed through textual analysis of business news reports. We introduce relevant methods for the estimation and inference of asymptotic no-arbitrage theory under this framework, applicable to both tradable and non-tradable factors. Empirically, we apply the proposed model and methodology to the listed stocks in the A-share market from 2012 to 2021 and compare the performance of the news co-mention network with other competing asset linkages. Results show that the assets in the A-share market exhibit significant heterogeneity in local dependency, and the news co-mention network outperforms in many aspects due to its quick updating and comprehensive non-industrial connectivity information. Additionally, non-tradable factors have non-negligible effects.

Keywords local dependency; text analysis; spatial econometrics; news linkage matrix; non-tradable factors

1 引言

在 Sharpe (1964) 和 Ross (1976) 的传统因子模型的框架下, 同期资产收益率之间的同步移动 (comovement) 是由一些公共风险因子所驱动的. 近年来, 众多研究表明, 在控制了共同风险因子之后, 资产收益率间仍存在某些由于局部关联所导致的相关性 (Gabaix, 2011; Kou et al., 2018; 张玉华等, 2016; 张自力等, 2020). 这种刻画了异质性风险 (idiosyncratic risk) 在个体 (资产) 间传播的局部依赖性在文献中被称作空间效应、网络效应或者群同效应¹. 石泽龙等 (2022) 发现, 在中国市场上, 主流的因子模型回归后的残差存在显著的空间效应, 这说明因子模型未能完全刻画资产之间的横截面相关性.

个体数量较大的面板数据在横截面维度的相关性 (cross-sectional dependence) 往往较为复杂, 因此我们需要对不同来源与性质的横截面相关性分别进行建模与分析. 首先, 所有资产都在不同程度上受到市场风险以及其他公共风险的影响, 因此因子模型刻画的是一种无法被分散的系统性风险. 与之不同的是, 空间效应刻画的是局部关联所导致的资产收益率协同变化, 即一个资产受到与之存在关联的资产的影响. 由此可见风险因子的共同暴露与空间效应所造成的资产收益率同步移动具有不同的经济学含义. 此外, 风险因子与空间效应所导致的资产收益率同步移动分别属于强横截面相关和弱 (局部) 横截面相关, 两者具有不同的统计学性质, 因而需要被分别讨论 (Chudik et al., 2011; Bailey et al., 2016). 资产间异质性风险的局部传播会影响风险分散的效果以及因子模型中风险暴露系数与定价误差的估计 (Zareei, 2019; Giglio et al., 2021), 导致对相关资产定价模型的实证表现产生误判. 因此在模型中识别不同来源的横截面相关性是当前资产定价研究的一个重要的问题. 本文提出了如下的异质空间因子定价模型, 分别对风险因子驱动的相关和空间效应导致的弱相关进行建模与分析:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \Psi\mathbf{W}\mathbf{r}_t + \boldsymbol{\epsilon}_t, \quad (1)$$

其中 $\mathbf{r}_t = (r_{1t}, r_{2t}, \dots, r_{nt})^T$ 代表 $n \times 1$ 维的资产收益率向量; $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{Kt})^T$ 为 $K \times 1$ 维的风险因子向量; $\mathbf{B} = (\beta_{ij})_{n \times K}$ 则是因子载荷矩阵, β_{ij} 表示第 i 项资产 r_{it} 关于

¹在本文中, 空间效应、网络效应或者群同效应的含义是等价的.

第 j 个因子 f_{jt} 的因子载荷; $\mathbf{W}$ 表示 $n \times n$ 维的空间权重矩阵, 用来刻画资产间的两两关联; $\Psi = \text{diag}\{\psi_1, \dots, \psi_n\}$ 为对角线元素是异质性空间系数的对角矩阵, 表示的是各资产对于空间相关性的敏感程度. 当 $\psi_1 = \dots = \psi_n$, 该模型退化为同质空间因子模型. 在该模型中, $\mathbf{B}\mathbf{f}_t$ 代表的是由共同风险因子驱动的同质横截面相关性, 而空间项 $\Psi\mathbf{W}\mathbf{r}_t$ 则表示同期资产收益率相互影响产生的弱 (局部) 横截面相关性. 本文会在第 3 节详细介绍它们在统计学性质上的不同以及如何区分这两种横截面相关性.

本文首先广泛考察了 A 股市场中常用关联网络在空间因子模型当中的实证表现. 不同于传统的地理空间模型, 在金融市场中, 模型 (1) 中空间关联矩阵 $\mathbf{W}$ 是无法通过地理距离定义的, 但是大数据可以让研究者通过其他途径来获取股票之间可能存在的关联信息, 例如属于同行业的股票彼此之间存在行业关联, 处在供应链上下游的股票也会产生联动. 本文在实证部分总结了当前中外文献中常用的几种衡量 A 股上市公司关联性的方法, 并详细讨论了这些方法在空间因子模型中的应用表现.

其次, 本文通过可靠的统计推断来验证空间效应的异质性. 在国内, 已经逐渐有学者发现因子模型回归后的残差项具有显著的空间相关性, 并开始应用空间因子模型展开中国市场上的实证研究 (潘荣翠等, 2012; 张玉华等, 2016; 石泽龙等, 2022), 但几乎都假设空间系数是同质的, 即限定所有资产对空间相关性的敏感程度一致, 有一定的局限性. 与之相比, 本文的异质空间因子定价模型更加灵活, 允许每个资产的空间系数有差异, 这更符合实际情况, 也有助于分析不同资产、不同行业之间的局部依赖性的差异. 本文实证结果也印证了使用异质性模型的必要性.

接着, 本文在大数据背景下, 对新闻文本数据展开灵活的文本共现 (news co-mention) 分析, 发掘资产之间的内在关联性, 构建对应的新闻关联矩阵. 如果两家公司同时出现在同一篇新闻报道中, 则它们很可能存在某种形式的经济关联, 并且新闻共现频率能够反映它们关联的强弱. 新闻信息源的优势有二: 其一, 新闻内容包罗万象、不一而足, 例如同行业、竞争、生产链、合作、兼并等, 它们可以较为全面地捕捉资产间的关联; 其二, 新闻具有时效性, 更新较快, 有助于及时观测到资产间关系的变化 (Scherbina et al., 2015; Schwenkler et al., 2019).

进一步地, 本文比较不同关联网络实证表现的优劣并分析原因. 我们通过多个维度来衡量不同关联网络在实证中的表现, 例如定价误差 α 显著比例和样本内 R^2 的比较等. 根据大量的实证证据, 本文发现新闻关联矩阵在众多维度都有最优的表现. 通过比较不同关联网络涵盖非行业信息的丰富程度并计算不同关联网络每周的更新频率, 我们论证了新闻关联网络相较于其他网络的优势, 即最多的行业外信息和最快的更新频率. 本文中对比股票间关联网络质量的方法可以推广到其他股票间关联强弱的比较. Ali et al. (2020) 论述了分析师共同提及关联在美国市场具有最优的表现, 但是在 A 股市场却不具备这种效应, 原因是我国分析师主要针对特定行业进行研究, 其中能够包含的行业外信息比较局限. 从激励机制角度出发, 赵留彦和宁可 (2020) 发现新财富评选取消之后, 证券行业分析师的工作勤勉程度和荐股能力都有了显著的下降. 此外, 从市场环境的角度分析, 美股市场更成熟, 交易量大、套利机制活跃, 因此交易型信号更快被捕捉并反映在分析师报告中. 如 Qiao (2024) 也发现中国卖方分析师的预测与基于会计的信息保持一致, 而不是基于交易的信息.

最后, 本文将模型推广到涵盖不可交易因子的应用场景. 不可交易因子的引入并不会给模型的估计带来改变, 但是它会极大地增加相应资产定价模型渐进无套利检验的难度, 本文将在空间因子模型的框架下为此问题引入相应的解决方案并给出实证案例. 国内外对不可交易因子的讨论方兴未艾, 本文仅给出在考虑强弱相关性的基础上如何检验不可交易因子的定价能力的方案, 为后续的研究提供一些理论基础与实证操作的参考.

本文的后续安排如下: 我们将在第 2 节整理并总结一些国内外与本文研究相关的已有文献; 第 3 节介绍本文所需的理论基础, 包括因子模型、空间回归模型、异质空间因子模型以及对应的异质性检验和无套利检验; 在第 4 节中, 本文收集 A 股上市公司 2012–2021 的周频收益率, 结合同期北京大学金融工程实验室的新闻共现数据展开实证分析, 并将新闻网络与其他常见网络进行比较; 第 5 节对全文进行了总结并提出了一些对未来研究方向的展望.

2 文献综述

股票收益率当中的协同变化可以分为全局性与局部性两种类型. 全局性协同变化可以通过影响市场绝大多数股票收益率的共同因子来进行横截面维度的刻画. Sharpe (1964) 的 CAPM 模型给出了风险溢价的来源是个股相较于市场所承担的风险, 而 Ross (1976) 的 APT 模型则是通过不同维度的风险来源将资产定价模型扩展到多因子模型的框架. 虽然 Ross 并没有给出这些“因子”的具体含义, 甚至来源于 CAPM 的市场因子也没有被 APT 明确纳入, 但是市场因子无疑是因子定价模型中当之无愧的“最强因子”. 因子的强弱可以被量化, 进而细分为强因子、半强因子、半弱因子以及弱因子 (Connor et al., 2024), 但是因子能够成为一个全局协同变化的因素的本质是, 随着股市中上市公司的扩张, 因子对于个股影响的平均值不能趋于零. Fama-French 模型 (Fama et al., 1993) 实现了 APT 模型中因子的“可视化”, 他们基于某些公司特征 (市值以及账面市值比) 构造的多空投资组合的收益可以较好地解释美国股市无法被 CAPM 模型解释的投资组合收益率“异象”, 其在中国市场的可推广性也早已被验证 (范龙振和余世典, 2002; 吴世农和许年行, 2004). 由于这类因子可以通过股票投资组合获得, 所以被称为“可交易因子”. 但 Fama-French 的模型并不是普适的, 如田利辉等 (2014) 认为在中国股市上, 账面市值比并不是一个能显著解释收益率的因子. Liu et al. (2019) 则考虑到中美市场差异性以及中国股市存在的“壳效应”, 改进了 Fama-French 三因子模型的构造方式, 利用中国市场 2003–2019 年的数据构造了更适合中国的规模和价值因子. 针对公募基金, 李斌和雷印如 (2022) 发现, 在中国市场上, 无论是用 CAPM、Fama-French 三因子还是中国市场的规模和价值因子, 他们构造的异象投资指标均与模型回归后的异常收益 α 成正比, 反映基金经理的投资有效性的同时也说明了因子模型的解释力度不足. 因此, 不断有学者尝试进一步发掘其他能够有效解释资产收益率的因子或使用新的方法来改进经典的因子模型. 此外, 刘维奇和张燕 (2020) 在 Fama-French 三因子和五因子模型的基础上, 加入劳动成本占比因子, 发现劳动成本是我国股票收益率的重要解释成分.

然而, 股票收益率之间的局部相关性 (弱相关性) 则在较晚的时候才开始被学界和业界的研究者关注. 弱相关是一些有内在关联的股票之间的小范围协同变化. 如果这种协同变化是同时发生的, 那么文献中多使用空间模型, 也就是模型 (1) 进行诠释. 在初期的空间计量模型中, 空间回归系数常被假设为同质的, 即不同资产对空间相关性的敏感程度一致, 具有较大

的局限性. Fernandez (2011) 较早地将空间计量模型引入金融资产定价中, 利用 S-CAPM 模型对 126 家拉美公司展开实证, 研究了它们之间的空间依赖性. 张玉华等 (2016) 在 S-CAPM 的基础上利用 2007-2015 年的季度数据, 对影响股票收益率的宏微观因素进行实证研究, 发现中国股票市场存在着显著的空间依赖性. 在本文中, 我们将采用更为一般的异质空间因子模型, 它允许不同资产对空间局部相关性有不同的敏感程度, 以便于我们研究不同资产、不同行业之间的差异. Aquaro et al. (2021) 推导了异质空间因子模型的拟最大似然估计 (QMLE) 并给出了统计推断过程, 随后借助 1975 年至 2014 年期间美国 338 个大城市房价数据, 将该模型运用在美国房价上, 发现空间依赖性显著, 且各城市对空间相关性的敏感程度的确存在异质性. 此外, 在一定正则性条件下, Ge et al. (2023) 推导了异质空间因子模型下的渐进无套利条件, 并提出了在资产数量较大时仍然适用的 Wald 检验.

另一方面, 空间模型中的关联矩阵, 即模型 (1) 中的 $\mathbf{W}$ 对于研究者是未知的. 但是学者们已经关注到个体间的关联可以通过其他多种信息源进行推断, 而这些推断结果可以作为空间关联度量矩阵 $\mathbf{W}$ 的“代理矩阵”. 潘荣翠等 (2012) 基于全球 41 个地区的数据, 利用华尔街公布的经济自由度指数构建了空间权重矩阵, 发现各国股票指数之间具有显著空间集聚因素. 石泽龙等 (2022) 基于生产网络信息构造中国金融市场上各行业之间的关联矩阵, 实证结果表明他们的空间因子模型比起因子模型更符合中国行业资产定价的实际. 段丙蕾等 (2022) 全面地分析了我国 A 股市场上股票的经济关联, 包含行业关联、供应链关联、科技关联、区域关联等, 发现网络关系是预测股票收益率不可忽视的因素. Ge et al. (2024) 也发现利用股票间的关联信息可以有效地增强高维协方差矩阵的估计.

大数据时代下, 财经新闻等文本数据往往蕴藏着大量个体间潜在关联信息. Scherbina et al. (2015) 和 Schwenkler et al. (2019) 发现新闻数据揭示了公司之间的内在联系, 并且发现新闻数据中的信息可以帮助预测股票回报率. Jiang et al. (2019) 基于企业财务文本数据构建了一个情绪指数, 且该情绪指数是预测未来收益率的一个重要因素. 在国内学术界, 张永杰等 (2011) 利用文本数据挖掘技术并发现社交媒体中的文本信息对股票异常日收益率有显著的解释力. 此外, 赵丽丽等 (2012) 用支持向量回归方法将财经新闻内容量化为股市波动的一个影响因子. 姜富伟等 (2021) 构建了一个全面的中文金融情感词典, 以此计算了我国财经媒体文本情绪指标, 发现媒体文本情绪可以更准确地衡量我国股市投资者情绪的变化, 对我国股票回报有显著的样本内和样本外预测能力. 陈张杭健等 (2021) 和范小云等 (2022) 则发现股吧中的信息交互对股票价格和宏观经济变量有着影响. 沈艳等 (2019) 分析并总结了文本大数据在中国金融市场的运用方法, 并指出文本数据运用的未来方向, 包括从国内主流新闻数据提取隐含信息以及文本信息与资本市场、宏观经济、微观个体等之间的关系. 此外, 范小云等 (2022) 利用新闻共现方法挖掘我国银行业之间的网络关系, 并结合隐马尔可夫模型, 发现该方法能较好地识别和预测系统性风险. 本文将通过空间因子模型来验证新闻共现关联在 A 股市场的实证表现, 并详细分析其中的机制. 值得注意的是, 上述传统方法对新闻信息的运用往往集中于事件研究法或直接构造辅助信息相关回归变量, 而本文利用新闻信息构造空间矩阵则是一种不同的方法, 这有助于拓宽我国市场资产定价的视角.

为了让本文构建的模型更具普适性, 本文将相关方法与结论拓展到包含不可交易因子的情景中. 一些宏观经济变量已经被众多文献证明具有对股票收益率的预测能力, 例如 Fama et al. (1989); Aïtsahali et al. (2001); Gu et al. (2018); Chen et al. (2024); Dong et al. (2024) 等都发现了在模型中纳入能反映经济状况和周期的时间序列可以更好地预测相关资产未来的表现, 于是我们自然希望研究这些变量是否可以用来对股票资产进行定价.

本文边际贡献主要有以下四点: 第一, 引入异质空间因子回归模型进行资产定价分析, 为我国金融资产定价注入新鲜血液. 目前在我国仍然只有较少文献利用空间因子模型刻画资产收益率, 仅有的几篇文献也都从同质空间因子模型出发, 具有一定的局限性. 本文的实证结果也的确发现空间系数表现出了明显的异质性, 说明不同个股对于空间效应的敏感程度是不同的. 第二, 将灵活生成的 A 股新闻关联矩阵引入到空间因子模型当中, 并通过对比周频更新速度和行业外关联涵盖度来论证其应用价值. 第三, 全面对比 A 股市场常用关联网络在空间因子模型当中的定价效果, 并通过多重指标进行分析, 比较结果在一定程度上说明了新闻数据可以更全面有效地揭示资产之间关联网络结构. 第四, 在空间因子模型中讨论了具有中国市场特点的不可交易因子. 国内学术界对多因子模型的研究大多基于可交易因子, 对于如宏观因子等不可交易因子的探讨尚且存在空缺, 而我国金融市场受宏观调控影响较大, 对不可交易因子的研究有其学术与应用价值; 此外, 在空间因子模型的框架下, 本文还介绍并应用了包含不可交易因子时检验渐近无套利的 Wald 检验.

3 理论基础

3.1 强相关性的刻画: 因子模型

考虑如下的因子模型:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t. \quad (2)$$

因子模型完全以变量 $\mathbf{f}_t$ 刻画了各资产之间系统性风险的结构. 在因子模型中, 如果我们选取 $\mathbf{f}_t$ 仅为一维的市场收益率 r_{mt} , 则式 (2) 代表了 CAPM 回归方程. 如果我们选取 $\mathbf{f}_t = (r_{mt}, \text{SMB}_t, \text{HML}_t)$, 则式 (2) 代表了 Fama-French 3 因子模型.

定义 $\|\mathbf{B}\|_1 := \max_{1 \leq k \leq K} \sum_{i=1}^n |\beta_{ik}|$ 为矩阵 $\mathbf{B}$ 的最大列 1-范数. 如果当经济系统中的资产数量 n 增加, 有各列 1-范数趋于无穷且 $\|\mathbf{B}\|_1 = \mathcal{O}(n)$, 则每个因子对经济系统的影响都是持续存在的, 即这种风险无法简单地通过增加资产个数而分散、消除. 在这一假定下, 所有资产都受到共同风险因子的影响, 因此该模型刻画的是不可被分散的系统性风险, 即强横截面相关性.

3.2 弱 (局部) 相关性的刻画: 空间回归模型

考虑下述同质空间回归模型:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \psi \mathbf{W} \mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (3)$$

同式 (2), 其中 $\mathbf{r}_t$ 代表 $n \times 1$ 资产收益率向量, $\mathbf{W} = (w_{ij})_{n \times n}$ 表示 $n \times n$ 维的空间权重矩阵, ψ 代表同质空间回归系数, 误差项的协方差矩阵为 $\boldsymbol{\Omega} = \text{diag}\{\boldsymbol{\sigma}^2\} = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$.

空间模型以内生变量 $\mathbf{W}\mathbf{r}_t$ 作为回归自变量, 常被用于刻画弱 (局部) 横截面相关性, 如果 $w_{ij} \neq 0$, 实际上表明资产 i 收益率受到资产 j 同期收益率的影响. 空间权重矩阵 $\mathbf{W}$ 常被行归一化, 即满足每行之和 $\sum_{j=1}^n w_{ij} = 1$, 而空间效应的强弱则由参数 ψ 表示.

3.3 强/弱相关性的刻画: 空间因子模型

式 (1) 表达的异质空间因子模型的好处在于它区分了资产之间的强弱相关性, 从而更好地刻画了资产间收益率的关系. 如果 f_t 有变动, 则由于 $\mathbf{B}$ 的存在, $\mathbf{r}_t$ 中的大多数都会随之变动, 这是公共因子体现的强相关性. 而如果 r_{jt} 变动, 那么该变动只会带动 $\{r_{it} : w_{ij} \neq 0\}$ 的收益率变动. 如果 $\mathbf{W}$ 是稀疏的, 那么发生变动的资产只是少数, 这是空间模型体现的弱相关性. 我们考虑下述四个正则性假设:

1) 零均值的残差 ϵ_t 无自相关性且与因子 f_t 不相关, 此外, 残差的方差 $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ 有一致的上下界, 即存在常数 $\sigma_l^2 > 0$ 和 σ_u^2 , 使得 $\sigma_l^2 \leq \sigma_i^2 \leq \sigma_u^2, i = 1, 2, \dots, n$.

2) 对 $k = 1, 2, \dots, K$, 均有 $\sum_{i=1}^n \beta_{ik} \rightarrow \infty$, 且 $\|\mathbf{B}\|_1 = \mathcal{O}(n)$; 此外, $\mathbf{B}^T \mathbf{B}$ 的最小特征值随着 n 的增大而趋于无穷.

3) 空间权重矩阵 $\mathbf{W}$ 满足 $w_{11} = w_{22} = \dots = w_{nn} = 0$, 且每行之和 $\sum_{j=1}^n w_{ij} = 1$; 此外, $\mathbf{W}$ 的最大行范数 $\|\mathbf{W}\|_\infty$ 有界, 且:

$$\max\{|\psi_1|, |\psi_2|, \dots, |\psi_n|\} < \frac{1}{\|\mathbf{W}\|_\infty}.$$

4) 假设 $\mathbf{G}(\Psi) = (\mathbf{I}_n - \Psi\mathbf{W})^{-1}$ 满足 $\|\mathbf{G}(\Psi)\|_1$ 有界.

假设 1 是计量研究中的常见假设, 保证了统计推断的有效性. 其次, 在假设 2 下, 由因子驱动的相关性是一种强的横截面相关性, 它无法通过单纯地增加资产的数量而被分散. 对这些风险因子的共同暴露会导致资产全局的同步移动. 假设 3 则保证了 $\Psi\mathbf{W}$ 的谱半径是小于 1 的, 即:

$$\rho(\Psi\mathbf{W}) \leq \|\Psi\mathbf{W}\|_\infty \leq \|\Psi\|_\infty \|\mathbf{W}\|_\infty = \max_{1 \leq i \leq n} |\psi_i| \cdot \|\mathbf{W}\|_\infty < 1. \quad (4)$$

这同时保证了假设 4 中 $\mathbf{G}(\Psi) = (\mathbf{I}_n - \Psi\mathbf{W})^{-1}$ 的良好性. 实际操作中, 由于 $\mathbf{W}$ 的每行元素之和被设定为 1, 因此通过限制 ψ_i 位于 $(-1, 1)$ 之间便能满足假设 3. 假设 4 是空间计量的常见假设, 它同时也表明 $\mathbf{G}$ 的列范数并不像因子载荷矩阵 $\mathbf{B}$ 那样趋于无穷, 而是有界的, 这说明了空间项的影响是局部的, 刻画的是一种弱相关性.

需要特别注意的是, 基于套利定价理论构造的资产定价模型是同期的, 故本文不讨论滞后收益率 $\mathbf{r}_{t-j}$ 的影响.

3.4 拟最大似然估计

假设样本收益率数据 $\{\mathbf{r}_t, t = 1, 2, \dots, T\}$ 和因子数据 $\{\mathbf{f}_t, t = 1, 2, \dots, T\}$ 满足式模型 (1), 定义全体参数 $\theta = (\psi^T, \mathbf{b}^T, \sigma^{2T})^T$, 其中 $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$, 且 $\mathbf{b}_i = (\alpha_i, \beta_{i1}, \dots, \beta_{iK})^T$, 假设 $\mathbf{S}(\Psi) = \mathbf{I}_n - \Psi\mathbf{W}$ 可逆, 则有:

$$\mathbf{r}_t = (\mathbf{I}_n - \Psi\mathbf{W})^{-1}(\mathbf{B}\mathbf{x}_t + \epsilon_t), \quad t = 1, 2, \dots, T, \quad (5)$$

其中自变量 $\mathbf{x}_{it} = (1, \mathbf{f}_t^T)^T$, 且 $\mathbf{x}_t = (\mathbf{x}_{1t}^T, \dots, \mathbf{x}_{nt}^T)^T$, 系数矩阵 $\mathbf{B} = \text{diag}\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. 拟对数似然函数可以写为:

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & -\frac{nT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^n \ln(\sigma_i^2) + T \ln |\mathbf{S}(\boldsymbol{\Psi})| - \\ & \frac{1}{2} \sum_{t=1}^T [\mathbf{S}(\boldsymbol{\Psi}) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t]^T \boldsymbol{\Omega}^{-1} [\mathbf{S}(\boldsymbol{\Psi}) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t], \end{aligned} \quad (6)$$

其中 $\boldsymbol{\Omega} = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$ 是误差项的协方差矩阵. 令对数似然函数 (6) 的梯度为 0, 方程的解是空间因子模型的拟最大似然估计 (QMLE), 如果误差项服从正态分布, 则该解就是其最大似然估计 MLE. 为易于理解, 对数似然函数的最后一部分也可以写为以下代数形式:

$$\sum_{t=1}^T [\mathbf{S}(\boldsymbol{\Psi}) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t]^T \boldsymbol{\Omega}^{-1} [\mathbf{S}(\boldsymbol{\Psi}) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t] = \sum_{t=1}^T \sum_{i=1}^n \frac{(r_{it} - \psi_i r_{it}^W - \mathbf{x}_{it}^T \mathbf{b}_i)^2}{\sigma_i^2}, \quad (7)$$

其中 $(r_{1t}^W, \dots, r_{nt}^W)^T = \mathbf{W} \mathbf{r}_t = \mathbf{r}_t^W$. 附录给出了对数似然函数的梯度与 Hessian 矩阵 $\mathbf{H}(\boldsymbol{\theta})$, 通过数值算法, 可以求得空间因子模型的拟最大似然估计 $\hat{\boldsymbol{\theta}}$.

QMLE 的渐近分布为:

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}_{\boldsymbol{\theta}}), \quad \mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta}) \mathbf{H}^{-1}(\boldsymbol{\theta}), \quad (8)$$

其中 $\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{T} (\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}) (\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^T$. 特别地, 如果误差项是高斯分布, 则有 $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{H}^{-1}(\boldsymbol{\theta}))$. 证明的过程可以参考 Aquaro et al. (2021).

3.5 异质性检验

在我国, 现有的空间计量文献大多从同质空间因子模型出发 (张玉华等, 2016; 石泽龙等, 2022), 本文则采用了更灵活的异质空间因子模型, 允许不同资产对空间依赖性的敏感程度不同. 为进一步探讨空间系数是否存在异质性, 我们考虑假设检验问题:

$$H_0: \psi_1 = \psi_2 = \dots = \psi_n.$$

备择假设是其对立. 在误差 ϵ 是高斯分布的假设下, 有渐近 LRT 统计量,

$$\chi^2 = 2(\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})) \xrightarrow{d} \chi^2(n-1),$$

其中 $\hat{\boldsymbol{\theta}}$ 是异质空间因子模型的估计, 而 $\tilde{\boldsymbol{\theta}}$ 是同质空间因子模型的估计. 由于资产数目 n 较大, 我们将此渐近卡方统计量标准化, 有:

$$Z = \frac{\chi^2 - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{d} N(0, 1), \quad \text{当 } n \rightarrow \infty.$$

该 Z 统计量越大, 我们越有信心拒绝原假设, 即异质性越明显.

3.6 (不) 可交易因子与渐近无套利检验

无套利是资产定价中非常重要的假设. 在一定前提假设下, Ge et al. (2023) 给出了空间因子模型的 (渐近) 无套利理论. 考虑存在一个无风险利率 r_f , 如果所有因子都是可交易的, 我们可以将空间因子模型 (4) 写为如下超额收益形式:

$$\tilde{r} = \tilde{\alpha} + B\tilde{z} + \Psi W\tilde{r} + \epsilon, \quad (9)$$

其中 $\tilde{r}$ 是资产的超额收益, $\tilde{z}$ 是可交易因子的超额收益, $\tilde{\alpha} = \alpha - r_f(I_n - \Psi W)\iota_n + r_f B\iota_K$ 是剔除无风险利率后的截距项, 这里 ι_m 表示一个全为 1 的 $m \times 1$ 向量. 可以证明在渐近无套利假定下, 该异质空间因子需满足 $\tilde{\alpha} \approx \mathbf{0}$. 这里符号 $\approx$ 意味着渐近无套利 (即当经济系统中资产数目增加时, 定价误差 $\|\tilde{\alpha}\|^2 = (\tilde{\alpha})^T \tilde{\alpha}$ 是有界的).

此外, 当因子 f_t 中包含一部分不可交易因子时, 即 $f = (f_1^T, f_2^T)^T$, 其中 f_1 是 $K_1 \times 1$ 可交易因子, f_2 是 $K_2 \times 1$ 去均值后的不可交易因子, 并有 $K_1 + K_2 = K$. 我们的模型 (4) 是:

$$r = \alpha + B_1 f_1 + B_2 f_2 + \Psi W r + \epsilon,$$

其中 B_1 和 B_2 分别是系数矩阵 B 的前 K_1 列和后 K_2 列. 如果记 $\tilde{r}$ 是资产的超额收益, $\tilde{z}_1$ 是可交易因子的超额收益, 则模型可以改写为:

$$\tilde{r} = \underline{\alpha} + B_1 \tilde{z}_1 + B_2 f_2 + \Psi W \tilde{r} + \epsilon, \quad (10)$$

其中 $\tilde{z}_1$ 是可交易因子的超额收益, $\underline{\alpha} = \alpha - r_f(I_n - \Psi W)\iota_n + r_f B_1 \iota_{K_1}$ 是剔除无风险利率后的截距项. 可以证明, 在渐近无套利假定下, 存在 λ_2 , 使得 $\underline{\alpha} \approx B_2 \lambda_2$, 这里 λ_2 对应的是不可交易因子的风险溢价. 此外, 若 $B_2 \neq \mathbf{0}$, 则此时不可交易因子被定价, 证明此不可交易因子应进入资产定价模型.

下面我们介绍空间因子模型的无套利检验. 为简洁, 我们省略变量的上下标注, 并将超额收益形式的空间因子模型 (9) 写成下述形式:

$$r_t = \alpha + B z_t + \Psi W r_t + \epsilon_t, \quad (11)$$

无套利检验即 $H_0: \alpha = \mathbf{0}$, Wald 统计量为:

$$W = T \frac{(\hat{\alpha}^*)^T \hat{\Sigma}_{\epsilon^*}^{-1} \hat{\alpha}^*}{1 + \hat{\mu}_z^T \hat{\Sigma}_z^{-1} \hat{\mu}_z}, \quad (12)$$

其中

$$\hat{\alpha}^* = (I_n - \hat{\Psi} W)^{-1} \hat{\alpha}, \quad \hat{\Sigma}_{\epsilon^*}^{-1} = (I_n - \hat{\Psi} W)^T \hat{\Omega}^{-1} (I_n - \hat{\Psi} W), \quad \hat{\mu}_z = \frac{1}{T} \sum_{t=1}^T z_t,$$

可以证明, 在原假设 $H_0: \alpha = \mathbf{0}$ 下, $W \xrightarrow{d} \chi^2(n)$. 如果资产数量也很大, 我们考虑标准化 Wald 统计量 $W_S = \frac{W-n}{\sqrt{2n}} \xrightarrow{d} N(0, 1)$, 显然 W_S 越大意味着越有信心拒绝原假设.

如果模型中含有不可交易因子, 同样为了简洁, 我们省略变量的上下标注, 并将模型 (10) 写成下述形式:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}_1 \mathbf{z}_{1,t} + \mathbf{B}_2 \mathbf{f}_{2,t} + \boldsymbol{\Psi} \mathbf{W} \mathbf{r}_t + \boldsymbol{\epsilon}_t, \quad (13)$$

其中 $\mathbf{f}_{2,t}$ 是零均值的不可交易因子, 为检验原假设 “存在 $\boldsymbol{\lambda}_2$, 使得 $\boldsymbol{\alpha} = \mathbf{B}_2 \boldsymbol{\lambda}_2$ ”, 我们将模型改写为:

$$\mathbf{r}_t = \boldsymbol{\alpha}^* + \mathbf{B}_1^* \mathbf{z}_{1,t} + \mathbf{B}_2^* \mathbf{f}_{2,t} + \boldsymbol{\epsilon}_t^*, \quad (14)$$

其中 $\boldsymbol{\alpha}^* = \mathbf{G}\boldsymbol{\alpha} = (\mathbf{I}_n - \boldsymbol{\Psi}\mathbf{W})^{-1}\boldsymbol{\alpha}$, $\mathbf{B}_1^* = \mathbf{G}\mathbf{B}_1$, $\mathbf{B}_2^* = \mathbf{G}\mathbf{B}_2$, $\boldsymbol{\epsilon}_t^* = \mathbf{G}\boldsymbol{\epsilon}_t$. 则原假设等价于 “存在 $\boldsymbol{\lambda}_2$, 使得 $\boldsymbol{\alpha}^* = \mathbf{B}_2^* \boldsymbol{\lambda}_2$ ”, 定义,

$$\boldsymbol{\theta}_s = ((\boldsymbol{\alpha}^*)^T, \text{vec}(\mathbf{B}_2^*)^T)^T, \quad \mathbf{g}(\boldsymbol{\theta}_s) = \mathbf{M}_{\mathbf{B}_2^*} \boldsymbol{\alpha}^* = (\mathbf{I}_n - \mathbf{B}_2^* ((\mathbf{B}_2^*)^T \mathbf{B}_2^*)^{-1} (\mathbf{B}_2^*)^T) \boldsymbol{\alpha}^*. \quad (15)$$

在原假设下, 有 $\sqrt{T} \mathbf{g}(\hat{\boldsymbol{\theta}}_s) \xrightarrow{d} N(0, \boldsymbol{\Omega}_g)$, 渐近协方差矩阵由 $\boldsymbol{\Omega}_g = \mathbf{J}(\boldsymbol{\theta}_s) \mathbf{V}_{\boldsymbol{\theta}_s} \mathbf{J}(\boldsymbol{\theta}_s)^T$ 给出, 其中 $\mathbf{V}_{\boldsymbol{\theta}_s}$ 是 $\hat{\boldsymbol{\theta}}^*$ 的渐近协方差矩阵 $\mathbf{V}_{\boldsymbol{\theta}^*}$ 中 $\boldsymbol{\theta}_s$ 对应的部分,

$$\begin{aligned} \mathbf{J}(\boldsymbol{\theta}_s) = & (\mathbf{M}_{\mathbf{B}_2^*}, -((\boldsymbol{\alpha}^*)^T \mathbf{B}_2^* ((\mathbf{B}_2^*)^T \mathbf{B}_2^*)^{-1} \otimes \mathbf{M}_{\mathbf{B}_2^*})) + \\ & (\mathbf{M}_{\mathbf{B}_2^*}, -((\boldsymbol{\alpha}^*)^T \mathbf{M}_{\mathbf{B}_2^*} \otimes \mathbf{B}_2^* ((\mathbf{B}_2^*)^T \mathbf{B}_2^*)^{-1}) \boldsymbol{\kappa}_{n, K_2}). \end{aligned}$$

则是对应的雅可比矩阵, $\boldsymbol{\kappa}_{n, K_2}$ 是 $nK_2 \times nK_2$ 转换矩阵 (commutator matrix). 我们可以用 $\hat{\boldsymbol{\Omega}}_g = \mathbf{J}(\hat{\boldsymbol{\theta}}_s) \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}_s} \mathbf{J}(\hat{\boldsymbol{\theta}}_s)^T$ 作为 $\boldsymbol{\Omega}_g$ 的估计, 但由于雅可比矩阵 $\mathbf{J}$ 的存在, $\hat{\boldsymbol{\Omega}}_g$ 不可逆. 因为 $\hat{\boldsymbol{\Omega}}_g$ 是半正定矩阵, 一定存在正交矩阵 $\hat{\mathbf{Q}}$ 使得 $\hat{\boldsymbol{\Omega}}_g = \hat{\mathbf{Q}} \hat{\boldsymbol{\Lambda}} \hat{\mathbf{Q}}^T$, 其中 $\hat{\boldsymbol{\Lambda}} = \text{diag}\{\hat{\lambda}_1, \dots, \hat{\lambda}_n\}$ 是 $\hat{\boldsymbol{\Omega}}_g$ 由大到小排序的特征值对角矩阵, 我们再定义,

$$\tilde{\lambda}_j = \hat{\lambda}_j 1_{\{\hat{\lambda}_j > 0\}} + \delta_T 1_{\{\hat{\lambda}_j = 0\}}, \quad j = 1, 2, \dots, n - K_2,$$

其中 δ_T 满足 $\delta_T + T^{-1/2} \delta_T^{-1} \rightarrow 0$, 令 $\tilde{\boldsymbol{\Lambda}}^+ = \text{diag}\{\tilde{\lambda}_1^{-1}, \dots, \tilde{\lambda}_{n-K_2}^{-1}, 0, \dots, 0\}$ 以及

$$\hat{\boldsymbol{\Omega}}_g^+ = \hat{\mathbf{Q}} \tilde{\boldsymbol{\Lambda}}^+ \hat{\mathbf{Q}}^T. \quad (16)$$

这满足广义 Wald 检验的条件 (Andrews, 1987), 因此,

$$W = T \mathbf{g}(\hat{\boldsymbol{\theta}}_s) \hat{\boldsymbol{\Omega}}_g^+ \mathbf{g}(\hat{\boldsymbol{\theta}}_s) \xrightarrow{d} \chi^2(n - K_2). \quad (17)$$

如果资产数量很大, 我们考虑标准化 Wald 统计量 $W_S = \frac{W - (n - K_2)}{\sqrt{2(n - K_2)}} \xrightarrow{d} N(0, 1)$, 同理, W_S 越大越有理由拒绝原假设.

在空间因子定价模型的基础上, 本文介绍了如何检验新挖掘的可交易与不可交易因子在渐进无套利设定下是否可以减少定价误差, 为之后的相关研究提供了新的思路.

4 实证分析

4.1 数据说明与模型设定

4.1.1 收益率数据

为验证基于股票间关联信息的异质空间因子模型在我国市场上的实证表现, 本文以 2021 年 12 月 A 股全部上市公司为总样本, 并使用相应的周频收益率为研究对象. 周频收益率作为一个折中的选择, 既减缓了日频收益率所受到的噪音和市场微观结构的干扰, 也避免了月频收益率所导致的样本数量少的问题, 因而在 A 股市场的研究中更加常用 (He et al., 2021; 段丙蕾等, 2022; Ge et al., 2025).

在全样本股票的基础上, 我们首先筛选出 2012 年以来收益数据完整且股票状态保持正常的股票, 此外, 从中剔除上市不足一年的股票和 ST 股, 剩余 1353 只, 最后, 由于金融地产股的特殊性, 其在文献中经常被单独讨论, 我们在主要结果中将 108 只金融地产股剔除, 但在稳健性检验中给出包含金融地产股的结果. 关于因子收益率, 我们应用 Liu et al. (2019) 提出的四因子模型 (CH4 模型), 即: 市场因子、价值因子、规模因子和情绪因子. CH4 模型是考虑到中国市场的实际情况而构造的因子模型, 实证表明其对中国金融资产收益率解释能力优于传统的 Fama-French 三因子和五因子模型. 因子数据收益率的描述性统计见表 1.

表 1 因子数据的描述性统计

变量名称	均值	标准差	最小值	中位数	最大值	观测值
无风险利率	0.04%	0.02%	0.01%	0.03%	0.10%	513
市场因子	0.17%	2.98%	-14.17%	0.34%	10.13%	513
价值因子	0.21%	1.82%	-8.03%	0.11%	12.03%	513
规模因子	0.09%	2.03%	-11.38%	0.22%	10.31%	513
情绪因子	0.16%	4.67%	-39.86%	-0.35%	59.82%	513

4.1.2 股票间关联数据

在空间因子定价模型中, 空间关联矩阵 $\mathbf{W}$ 衡量了样本公司股票收益率之间的关联. A 股市场上, 常被研究的关联网络有: 新闻共现、分析师关联、概念关联、行业关联、城市关联、省份关联、专利相似关联、统计关联、供应链关联、残差关联等. 表 2 汇总了这些常用的关联网络, 包括网络名称、网络描述以及对应的文献.

上述关联大致可分为两类 (Ge et al., 2024): 第一类是判断性关联, 即关联是否存在, 量化后的该关联是一个 0-1 变量, 如行业关联、城市关联等; 另一类是可以通过一个连续变量来衡量关联强弱, 例如分析师共同提及关联等. 此外, 需要补充说明的有以下几点: 概念股关联曾被 Du et al. (2022) 进行过详细的讨论, 作者论述了该关联的存在以及基于该关联的投资组合可以产生很强的“动量溢出”效应. 然而遗憾的是, 该文章的部分数据通过网络爬虫获取, 我们暂时没有找到完整数据, 只能通过其他最常用的概念股数据代替. 由于没有强制披露政策, A 股供应链数据质量较差, 因此本文的结果不能代表该网络实际的价值. 最后是过往非系统性关联数据, 为了避免使用样本期内股票收益率关联信息而造成内生性的问题, 本文使用样本期前半年的收益率数据来计算股票间过往收益率相关性产生的关联网络.

表 2 常用关联网络

网络名称	网络简介	参考文献
Sentence	2012–2021 年财经新闻文本中两家公司同句共现次数	
Passage	2012–2021 年财经新闻文本中两家公司同篇共现次数	
Analyst	利用中国研究数据服务平台 (CNRDS) 中分析师报告的相关数据剔除空值并消除重复项后对两家公司共享分析师的关联性进行识别. 矩阵相应元素为 2012 年 1 月至 2020 年 12 月两家公司共享分析师团队的数量.	Ali et al. (2020)
Concept	利用 RESSET 数据库关于概念和成分股的相关数据, 把矩阵相应元素设为两家公司间共同概念的数量. 由于数据可用性, 概念动量分析的样本期从 2016 年 8 月至 2020 年 12 月, 在此期间共识别出 336 个概念.	Du et al. (2022)
Industry	利用 RESSET 数据库行业分类的历史数据, 按年将两家申万一级行业分类相同公司的矩阵中相应元素设为 1, 否则为 0, 最终矩阵由 2012–2021 年的矩阵加总形成.	
City	利用 CSMAR 的公司总部位置的历史数据, 按年将总部在相同城市 (共有 399 个城市包括 4 个直辖市) 的两家公司的矩阵相应元素设为 1, 否则为 0, 最终矩阵由 2005–2020 年的矩阵加总形成.	Parsons et al. (2020)
Province	与 City 类似, 按年将总部在相同省份的两家公司的矩阵相应元素设为 1, 否则为 0, 最终矩阵由 2005–2020 年的矩阵加总形成.	Parsons et al. (2020)
Tech	利用 CCER 数据库的专利数据 (包括国际专利分类 (IPC) 专利类别和专利授权日期等信息), 把矩阵相应元素设为由 2012–2020 年内各个公司在 IPC 中每个类别的专利的比例向量计算得到的技术相似性.	Lee et al. (2019)
Statistical	把在五个公司层面特征: 收盘价 (P)、公司规模 (SIZE)、账面市值比 (Bm)、净资产收益率 (ROE) 和资产增长率 (AG) 上具有最近欧几里得距离 (最小的 10%) 的股票之间的矩阵相应元素设为 1, 否则为 0, 样本期为 2012–2020 年.	He et al. (2021)
Customer	利用上市公司年报披露的关于其前五大客户的信息 (样本期为 2010 年 1 月至 2020 年 12 月), 按年将矩阵中相应元素设为一家公司在另一家公司销售额中所占比例 (%), 否则为 0, 最终矩阵由 2010–2020 年的矩阵加总形成. 由于证监会并未强制要求披露其前几大客户的真实名称, 而且一些重要客户并非上市公司, 所识别的客户-供应商关系的数量远小于其他关系.	Cohen et al. (2008)
Resid	2011 年 6 月至 2011 年 12 月的股票收益数据在 CH4 因子模型中的残差的相关系数矩阵取绝对值形成.	

4.1.3 财经新闻关联数据

本文重点考察基于财经新闻的文本关联在 A 股市场空间因子模型的效力. 北京大学金融工程实验室汇集了从 1999 年 9 月 22 日开始的数百万条锐思财经文本智能分析平台和聚源数据库的新闻文本数据, 并从中筛选 1138247 条包含上市资产的新闻文本, 表 3 汇总了从 2014 年开始的新闻数据基本信息. 可以看出, 新闻数据在每年的分布并不均匀, 比起均值, 标准差 (方差) 非常大.

图 1 选取中证 800 成分股中拥有较高频次新闻数据的 25 支股票并展示其在 2012–2021 年的共现次数. 有一些有显著关联的企业共现次数较多, 如: 五粮液 (000858) 和贵州茅台 (600519), 共现 133 次, 相互是其关联股中的第一, 它们同属白酒行业, 说明新闻数据捕捉

表 3 新闻数据汇总

	2014	2015	2016	2017	2018	2019	2020	2021	2022
日均条数	336.04	236.23	197.29	126.35	156.37	136.68	334.87	521.84	117.62
方差	45193.22	25200.87	17296.58	7977.94	11090.13	8939.84	98400.56	153748.28	250181.00
最大	699	589	493	359	467	464	1685	2590	1818
最小	1	1	1	1	1	1	1	1	9

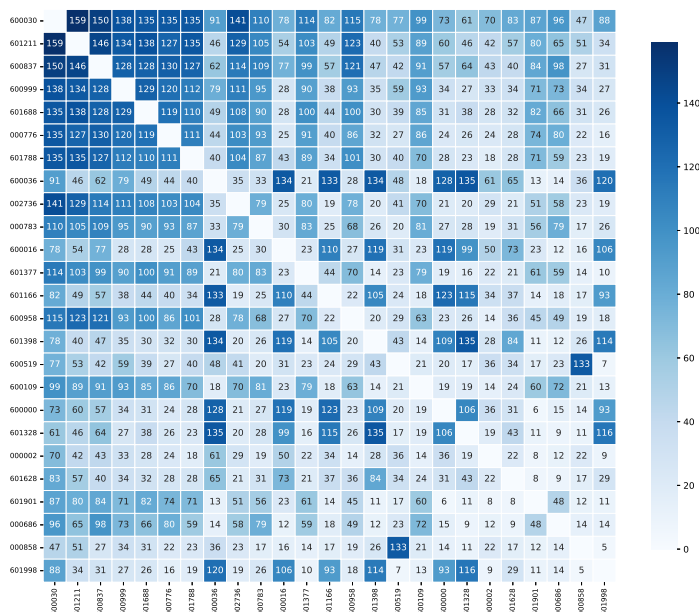


图 1 新闻共现数据热力图

到了行业间关联。此外, 金融股拥有较高的共现次数, 如: 中信证券 (600030)、国泰海通 (601211) 等。值得注意的是, 五粮液和国泰海通的共现次数要多于其和其他金融股的共现次数, 而五粮液 1998 年在 A 股上市时的主承销商是国泰海通的前身君安证券, 这说明新闻数据可以发掘一些潜在关联。还有一些其他个股有较高的共现次数, 如沃尔核材 (002130) 和长园集团 (600525), 这与它们之间长达四年之久的控股权之争有关。

4.1.4 关联数据描述性统计与空间矩阵标准化

本文所用到的关联网络数据的描述性统计如表 4 所示。值得注意的是, 原始网络信息并不满足空间因子模型中的 $\sum_{j=1}^n w_{ij} = 1$ 这一条件, 需要进行预处理。

如新闻共现网络, 我们记样本中第 s 年的新闻共现矩阵为 $\mathbf{C}_s = (c_{ij}^{(s)})$, 其中 $c_{ij}^{(s)}$ 指的是资产 i 与资产 j 在第 s 年在同一篇新闻中共同出现的次数, 其中 $s = 1, 2, \dots, M$ 。通过 $\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_M$ 提供的信息, 我们可以合理地构造一个新闻关联矩阵 $\mathbf{W}$ 。令

$$\tilde{\mathbf{W}} = (\tilde{w}_{ij}) = \sum_{s=1}^M \mathbf{C}_s. \quad (18)$$

表 4 关联网数据的描述性统计

	非零股票对	均值	标准差	最小值	中位数	最大值	Rank ($\mathbf{W}$)
Sentence	40279	2.01	7.80	1	1	727	1225
Passage	87544	1.64	4.52	1	1	664	1231
Analyst	109223	5.72	15.65	1	1	437	1091
Concept	440680	2.35	1.07	1	2	15	1241
Industry	44492	8.87	2.36	1	10	10	1245
City	23583	13.13	3.61	1	14	16	1134
Province	57786	13.34	3.04	1	14	16	1245
Tech	140644	0.13	0.19	0.00	0.06	1.00	605
Statistical	292079	2.68	1.83	1	2	9	1245
Customer	156	10.93	23.50	0.19	4.63	210.72	148
Resid	774390	0.08	0.06	0.00	0.07	0.93	1245

注: 1245 只股票的网络可视作 774390 个股票对, 其中关联不为 0 的视为“非零股票对”。
Rank($\mathbf{W}$) 为归一化后空间矩阵的秩。

再令 $\mathbf{W} = (w_{ij}) = (\tilde{w}_{ij} / \sum_{k=1}^n \tilde{w}_{ik})$, 即归一化以满足权重矩阵的要求, 其他网络也作此归一化处理²。

在本篇文章中, 我们建立的模型是异质性空间因子模型, 其中空间矩阵 $\mathbf{W}$ 是一个固定的常矩, 不随时间 t 变化。理论上来说, 随时间变化的空间矩阵的假设将更广泛、更真实。但是由于时变网络下资产的收益率成为了非平稳时间序列, 现有的计量理论中, 时变空间矩阵的异质性空间因子模型的渐近性质尚未被建立。根据 Aquaro et al. (2021) 和 Ge et al. (2023), 在研究不涉及预测问题的情况下, 使用全样本期的网络信息构造空间矩阵是一种折中的方法。

4.1.5 回归模型设定

本文以 CAPM 模型和 Liu et al. (2019) 的 CH3、CH4 模型为基准, 加入空间回归项, 逐一研究下述几个空间因子模型, 其中资产收益率和可交易因子收益率均为超额收益:

- (模型 1) 空间 CAPM 模型:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \beta_1 z_{\text{MRT},t} + \boldsymbol{\Psi} \mathbf{W} \mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (19)$$

- (模型 2) 空间 CH3 模型:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \beta_1 z_{\text{MRT},t} + \beta_2 z_{\text{VMG},t} + \beta_3 z_{\text{SMB},t} + \boldsymbol{\Psi} \mathbf{W} \mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (20)$$

- (模型 3) 空间 CH4 模型:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \beta_1 z_{\text{MRT},t} + \beta_2 z_{\text{VMG},t} + \beta_3 z_{\text{SMB},t} + \beta_4 z_{\text{PMO},t} + \boldsymbol{\Psi} \mathbf{W} \mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (21)$$

²如果原始矩阵 $\tilde{\mathbf{W}}$ 的某 i_0 行全为 0, 那么归一化后的 $\mathbf{W}$ 中该行也全为 0, 是不满足空间权重矩阵的要求的, 但是实际上因为第 i_0 样本股与其他股无空间关联性, 则在似然函数中可以被因式分解出来, 不影响到空间因子模型回归的结果。因此, $\mathbf{W}$ 的每行要么和为 1, 要么全为 0。

CH3 和 CH4 模型是 Liu et al. (2019) 考虑到中国市场的实际情况, 构造了适合中国市场的价值、规模和情绪因子³, 即 VMG (value-minus-growth)、SMB (small-minus-big) 和 PMO (pessimistic-minus-optimistic) 因子, 实证表明其对中国上市公司收益率的解释能力优于传统的 Fama-French 三因子和五因子模型. 除此以外, 本节还将讨论如下的混合了可交易因子与不可交易因子的异质空间因子模型:

- (模型 4) 空间 CH4 + 宏观因子模型:

$$r_t = \alpha + \beta_1 z_{MRT,t} + \beta_2 z_{VMG,t} + \beta_3 z_{SMB,t} + \beta_4 z_{PMO,t} + \beta_5 f_{SHI,t} + \beta_6 f_{CPI,t} + \beta_7 f_{PPI,t} + \beta_8 f_{M,t} + \Psi W r_t + \epsilon_t, \quad (22)$$

其中我们选用的不可交易因子分别是 SHIBOR (上海银行间同业拆放利率)、CPI (居民消费价格指数)、PPI (工业生产者价格指数)、M (货币供应量 M_0), 均为可以获得周频数据的常用宏观变量, 数据来源为聚源数据库.

4.2 异质性检验

在中国股票市场上, 针对空间因子模型, 大多数学者使用同质空间因子模型进行实证检验 (潘荣翠等, 2012; 张玉华等, 2016; 石泽龙等, 2022), 本文应是首篇使用异质空间因子模型在中国股票市场上展开实证的文章, 因此检验模型的异质性是否显著是有必要的. 在表 5 中, 以 Sentence 网络的 CAPM 模型为例, 同质空间系数的估计值 $\psi_{\text{homo}} = 0.735$, 异质空间系数的平均估计值 $\psi_{\text{MG}} = 0.748$, 非常接近, 但其异质性 LRT 检验非常显著, 这说明了各个个股对空间效应的敏感程度是不同的. 在其他网络或因子模型中, 也出现了类似的情况. 同时, 异质性系数下的模型 R^2 也显著高于同质性模型, 引入异质性系数对 R^2 的提升有时甚至强于增加因子. 如在 Passage 网络的 CH-3 模型中, 同质模型 $R^2_{\text{homo}} = 0.397$, 增加了一个因子后的 CH-4 模型的同质 $R^2_{\text{homo}} = 0.398$, 仅仅略微增加. 但在 CH-3 模型中考虑异质性系数, 模型的平均 R^2 可以提升到 0.407. 上述结果说明我们应该使用异质空间因子模型进行定价.

此外, City、Tech 和 Customer 网络中的空间系数显著比例较低, 说明这三个网络并不合适作为空间矩阵, 或说这三个网络没有体现出资产之间的关联性.

4.3 空间因子定价模型的无套利检验

在用因子模型进行无套利资产定价时, 人们常用一些截距项相关指标来检验模型是否满足无套利假定 (详见 Fama et al. (2015) 中表 5): 如模型截距项的绝对值 ($|\alpha|$) 在所有样本股中的均值、模型截距项的绝对值均值占样本股平均收益率绝对值均值的比重等.

表 6 汇总了不同网络下 CAPM 模型、CH-3 模型和 CH-4 模型的无套利检验结果. 其中, Sentence 网络和 Passage 网络都有助于提升因子模型的解释力度, 显著的 α 比例从原始 CAPM 模型的 2.73% 降低为 1.12% (Sentence 网络) 和 1.37% (Passage 网络), 模型 R^2 则从 0.274 升至 0.388 和 0.384, 说明新闻数据隐含了资产之间的关联信息. 除了新闻关联外, Analyst 网络和 Industry 网络也显著提升了模型的解释力度, 尤其在 CH-3 和 CH-4 模型中,

³因子数据可于 Robert F. Stambaugh 教授主页 <https://finance.wharton.upenn.edu/~stambaugh/> 下载.

表 5 不同关联网络下的模型异质性检验

	Sentence	Passage	Analyst	Concept	Industry	City	Province	Tech	Statistical	Customer	Resid
• CAPM 模型											
% Sig	93.0%	95.3%	85.5%	93.3%	94.6%	66.8%	87.1%	45.1%	95.4%	11.5%	93.6%
ψ_{MG}	0.748	0.846	0.732	0.801	0.698	0.373	0.608	0.368	0.823	0.027	0.775
ψ_{homo}	0.735	0.930	0.869	0.936	0.688	0.209	0.533	0.824	0.902	0.171	0.878
R^2_{hete}	0.388	0.384	0.370	0.355	0.381	0.317	0.340	0.313	0.362	0.279	0.348
R^2_{homo}	0.373	0.375	0.364	0.342	0.368	0.293	0.319	0.304	0.355	0.276	0.330
p -value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
• CH-3 模型											
% Sig	71.2%	71.5%	71.2%	63.4%	71.2%	29.9%	39.9%	33.0%	62.0%	7.3%	32.4%
ψ_{MG}	0.554	0.662	0.574	0.741	0.525	0.167	0.283	0.301	0.695	0.016	0.621
ψ_{homo}	0.570	0.819	0.748	0.858	0.580	0.056	0.236	0.736	0.852	0.103	0.794
R^2_{hete}	0.413	0.407	0.406	0.373	0.403	0.362	0.366	0.370	0.374	0.358	0.365
R^2_{homo}	0.396	0.397	0.399	0.370	0.394	0.356	0.359	0.366	0.372	0.356	0.363
p -value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
• CH-4 模型											
% Sig	71.1%	71.8%	71.1%	63.3%	71.4%	29.6%	39.8%	32.9%	61.8%	7.5%	32.3%
ψ_{MG}	0.554	0.661	0.574	0.740	0.525	0.168	0.283	0.301	0.694	0.016	0.620
ψ_{homo}	0.569	0.819	0.747	0.857	0.580	0.056	0.236	0.736	0.851	0.103	0.793
R^2_{hete}	0.415	0.408	0.407	0.374	0.404	0.363	0.368	0.371	0.376	0.359	0.366
R^2_{homo}	0.398	0.398	0.400	0.371	0.395	0.358	0.361	0.368	0.374	0.357	0.364
p -value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

注: % Sig 为异质模型空间系数项在 0.05 显著性水平下显著的比例; ψ_{MG} 为异质模型空间系数项估计值的平均数 (n 只个股的平均); ψ_{homo} 为同质模型空间系数项的估计值 (只有一个); R^2_{hete} 和 R^2_{homo} 分别为异质模型和同质模型的 R^2 平均数 (n 只个股的平均); p 代表异质性 LRT 检验的 p 值.

它们与 Passage 网络的回归结果非常接近. 以 CH-4 模型和指标 $\frac{A|\alpha|}{A|r|}$ 为例, 原始 CH-4 模型中 $\frac{A|\alpha|}{A|r|} = 45.9\%$, 加入 Sentence 网络将其降至 36.8%, 而 Industry 网络则将其降至 39.3%, 略不如 Sentence 网络. 和前文类似, 加入 CH-3 和 CH-4 因子虽然提升了 R^2 , 但不一定能降低个股的平均异常收益 ($|\alpha|$). 在所有网络中, Sentence 网络表现最优. 在 Sentence 网络下, 模型的平均异常收益最低, 显著的 α 个数最少, 平均 R^2 也最高. 此外, 加入 Customer 网络则几乎不能提升因子模型的解释力度. Resid 网络在 CAPM 模型中表现尚可, 但一旦我们增加因子模型中的因子个数, Resid 网络就不能提供太多的其余信息了, 这也正说明了我们需要寻找其他外部辅助信息来帮助我们识别资产之间的关联结构.

4.4 空间因子模型回归结果

根据前文的结果, 对比其他网络信息, Sentence 网络在中国股票市场上有较强的解释力, 我们在表 7 汇总了传统因子模型和 Sentence 网络下空间因子模型的平均参数估计值 (mean group estimate) 及其标准误差, 并且计算了每 n 个异质性参数在 0.05 显著性水平下参数显著的比例.

表 6 不同关联网络下的模型估计结果

	NONE	Sentence	Passage	Analyst	Concept	Industry	City	Province	Tech	Statistical	Customer	Resid
• CAPM 模型												
% Sig	2.73%	1.12%	1.37%	1.77%	2.49%	2.01%	2.81%	3.05%	2.57%	1.37%	2.65%	2.49%
$A \alpha $	0.16%	0.11%	0.12%	0.13%	0.13%	0.13%	0.14%	0.14%	0.14%	0.12%	0.16%	0.13%
$\frac{A \alpha }{A r }$	36.8%	26.7%	27.3%	30.1%	30.2%	29.6%	33.4%	32.0%	33.3%	27.5%	36.5%	31.3%
MG	0.12%	0.02%	0.00%	0.02%	0.01%	0.04%	0.08%	0.05%	0.08%	0.02%	0.12%	0.03%
R^2	0.274	0.388	0.384	0.370	0.355	0.381	0.317	0.340	0.313	0.362	0.279	0.348
• CH-3 模型												
% Sig	7.63%	4.02%	4.58%	4.50%	6.43%	5.46%	7.39%	7.31%	7.39%	4.26%	7.47%	8.35%
$A \alpha $	0.20%	0.16%	0.16%	0.17%	0.18%	0.17%	0.19%	0.19%	0.19%	0.16%	0.19%	0.19%
$\frac{A \alpha }{A r }$	45.5%	36.4%	37.9%	38.8%	42.7%	38.9%	44.8%	44.5%	44.1%	38.4%	45.2%	45.4%
MG	0.09%	0.02%	0.01%	0.03%	0.00%	0.04%	0.07%	0.07%	0.06%	0.03%	0.09%	0.03%
R^2	0.355	0.413	0.407	0.406	0.373	0.403	0.362	0.366	0.370	0.374	0.358	0.365
• CH-4 模型												
% Sig	8.11%	4.10%	4.74%	4.98%	6.51%	5.54%	7.79%	7.31%	7.71%	4.26%	7.95%	8.67%
$A \alpha $	0.20%	0.16%	0.16%	0.17%	0.19%	0.17%	0.19%	0.19%	0.19%	0.17%	0.20%	0.20%
$\frac{A \alpha }{A r }$	45.9%	36.8%	38.4%	39.2%	43.2%	39.3%	45.1%	44.8%	44.5%	38.8%	45.6%	45.9%
MG	0.09%	0.02%	0.01%	0.03%	0.00%	0.04%	0.07%	0.06%	0.06%	0.03%	0.09%	0.03%
R^2	0.357	0.415	0.408	0.407	0.374	0.404	0.363	0.368	0.371	0.376	0.359	0.366

注: ‘NONE’ 指将空间矩阵设置为零矩阵, 此时回归结果即为传统因子模型回归结果。% Sig 为模型截距项在 0.05 显著性水平下显著的比例; $A|\alpha|$ 为模型截距项绝对值的均值, $A|r|$ 为个股平均收益率绝对值的均值; MG 为截距项参数估计值的平均数 (mean group estimate); R^2 为模型的平均拟合优度。本表格将相关指标下最优表现结果进行加粗。

我们先看三个不含宏观因子的模型, 在空间 CAPM 模型中, 93.0% 的空间系数显著, 在空间 CH3 和空间 CH4 模型中, 也有 71.2% 和 71.1% 的空间系数显著, 这说明我国上市公司中普遍存在显著的空间依赖性, 且这一空间依赖性是由单独的因子模型无法解释的。其次, 空间回归项很明显影响因子的估计值与显著程度, 尤其是市场因子, 在不加入空间项时, 平均市场因子载荷接近 1, 在 CAPM 模型中是 1.0893, 在 CH3 和 CH4 模型中则分别是 0.9538 和 0.9557, 显著比例均为 100.0%, 但加入空间因子模型之后, 平均因子载荷明显降低。这是因为在没有加入空间关联的时候, 股票间的局部影响会被公共因子“吸收”, 当我们对局部关联建模之后, 这部分被“吸收”的影响会被分离出来。我们发现, 空间系数平均估计值与市场平均因子载荷之和基本上接近于未加入空间项之前的市场平均因子载荷。这说明空间因子模型的确将资产之间的共同移动分解为公共关联和局部关联两部分。如 CH3 模型中, 平均市场因子载荷原为 0.9538, 而加入空间项后变为 0.4238, 与此同时空间系数平均估计值为 0.5541, 其和略高于 0.9538, 这是符合认知的, 因为如果不加入空间项, 则局部关联很可能会被误识别为公共关联, 而公共关联造成了资产间的全局移动, 总变化量大于局部移动, 因此原始因子模型中的平均市场因子载荷会比空间模型中两部分之和略小。这些现象也说明了我们需要在模型中加入空间项, 否则局部关联将被误识别为公共关联, 导致额外的误差。

与市场因子类似, 在 CH3 模型中, 价值因子的因子载荷从 -0.0607 变化为 -0.0234 , 规模因子的因子载荷则从 0.7580 变化为 0.4267 。CH4 模型中, 情绪因子的因子载荷 β_4 从

表 7 空间因子模型回归结果

	因子系数								空间系数	误差方差	
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	ψ	σ^2
• CAPM											
MG	0.001	1.089									0.054
std.	(0.002)	(0.002)									(0.000)
% Sig	2.7%	100.0%									100.0%
• 空间 CAPM											
MG	0.000	0.286								0.748	0.049
std.	(0.003)	(0.003)								(0.003)	(0.000)
% Sig	1.1%	41.5%								93.0%	100.0%
• CH3 模型											
MG	0.001	0.954	-0.061	0.758							0.050
std.	(0.005)	(0.005)	(0.005)	(0.004)							(0.000)
% Sig	7.6%	100.0%	41.8%	87.8%							100.0%
• 空间 CH3 模型											
MG	0.000	0.424	-0.023	0.427						0.554	0.048
std.	(0.006)	(0.006)	(0.006)	(0.005)						(0.004)	(0.000)
% Sig	4.0%	50.1%	31.2%	60.5%						71.2%	100.0%
• CH4 模型											
MG	0.001	0.956	-0.064	0.756	0.015						0.050
std.	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)						(0.000)
% Sig	8.1%	100.0%	40.6%	86.8%	16.1%						100.0%
• 空间 CH4 模型											
MG	0.000	0.424	-0.024	0.427	0.001					0.554	0.048
std.	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)					(0.004)	(0.000)
% Sig	4.1%	49.8%	30.7%	61.0%	14.7%					71.1%	100.0%
• CH4 + 宏观因子											
MG	0.001	0.955	-0.066	0.755	0.014	-0.023	0.002	-0.010	-0.035		0.050
std.	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)		(0.000)
% Sig	9.6%	100.0%	40.8%	86.8%	15.7%	4.1%	3.7%	3.1%	4.2%		100.0%
• 空间 CH4 + 宏观因子											
MG	0.000	0.425	-0.025	0.427	0.000	-0.014	-0.006	-0.009	-0.018	0.552	0.048
std.	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.000)
% Sig	5.1%	50.1%	31.3%	60.6%	14.7%	3.6%	3.6%	2.4%	3.6%	71.2%	100.0%

注: MG 指样本股的平均估计值, std. 则指其标准误差, % Sig 表示样本股中估计值在 5% 显著性水平下显著的比例。

0.0154 变为 0.0006, 显著比例也有所降低。此外, 宏观因子的因子载荷变化并不明显, 但大多数可交易因子的因子载荷在空间因子模型中都有向 0 靠近的趋势。这主要是因为可交易因子

本身就可视作各资产的投资组合, 和空间项中的 $\mathbf{W}\mathbf{r}_t$ 存在着相互解释的关系, 空间系数的显著性也证明了这一点 (CH3 和 CH4 模型的空间系数显著性不如 CAPM 模型), 而不可交易因子则不能视作各资产的投资组合, 因此变化程度并不明显。

在加入空间项后, 三个传统因子模型的异常收益 α 得到了不同程度的解释。例如在 CAPM 模型中, 尚有 2.7% 的个股拥有显著非零的 α , 而在加入空间项后, 这一比例降低到了 1.1%; 在 CH3 和 CH4 模型中, 均原有 7.5% 以上的 α 显著, 加入空间项使得显著的 α 降低到 4% 左右。值得注意的是, 在本文的实证中, 加入规模、价值和情绪因子后, 具有显著非零异常收益的个股反而增多。这种现象在 A 股实证资产定价的文献中可谓屡见不鲜, 例如 Li et al. (2023) 中表 4、表 6 和 Hou et al. (2023) 的实证结果。究其原因, 还是要追溯到 Fama-French 多因子模型 (Fama et al., 1993), 多因子模型提出的基础是因为根据某些公司特征作为基准的多空投资组合出现了传统 CAPM 模型无法解释的显著套利空间, 而多因子模型有助于解释这些异象产生的风险溢价。CH3 和 CH4 在中国市场对这些异象产生的风险溢价有着比传统 CAPM 模型更好的解释力度 (Liu et al., 2019), 但这并非表明它们一定会更好地降低全市场显著的 α 的个数, 所以表中出现的这些“异常”也是情理之中的。

4.5 稳健性检验: 涵盖金融地产股的实证结果

在本小节中, 我们在模型中加入金融地产股重新进行回归, 并探讨模型结果是否发生变化。如表 8 所示, 当我们加入金融地产股后, Sentence 网络仍然在各网络数据中表现最优, 模型整体 R^2 比起无金融地产股时有所上升。以空间 CH-4 模型为例, Sentence 网络下显著的截距项占比达到 4.36%, 比起无金融地产股时的 4.10% 略有上升, 这可以理解为金融地产股中截距项不为 0 的个股略多于其余类型股票; 此外, 模型平均 R^2 从 0.415 上升到 0.420, 说

表 8 不同关联网下的模型估计结果 (包含金融地产)

	NONE	Sentence	Passage	Analyst	Concept	Industry	City	Province	Tech	Statistical	Customer	Resid
• CAPM 模型												
% Sig	2.73%	1.18%	1.55%	1.85%	2.96%	2.22%	2.96%	3.18%	2.59%	1.40%	2.66%	2.73%
$A \alpha $	0.15%	0.11%	0.12%	0.13%	0.13%	0.13%	0.14%	0.14%	0.14%	0.12%	0.15%	0.14%
$\frac{A \alpha }{A r }$	36.1%	26.8%	27.8%	30.0%	31.0%	29.5%	33.6%	32.4%	33.3%	27.4%	35.9%	31.8%
MG	0.12%	0.01%	0.00%	0.02%	0.01%	0.03%	0.07%	0.05%	0.07%	0.02%	0.12%	0.03%
R^2	0.279	0.391	0.382	0.371	0.357	0.389	0.321	0.343	0.314	0.363	0.283	0.350
• CH-3 模型												
% Sig	7.54%	4.29%	5.03%	4.36%	6.65%	5.40%	7.39%	7.46%	7.76%	4.36%	7.39%	8.35%
$A \alpha $	0.19%	0.16%	0.16%	0.16%	0.18%	0.16%	0.19%	0.19%	0.19%	0.16%	0.19%	0.20%
$\frac{A \alpha }{A r }$	45.2%	36.5%	38.4%	38.6%	43.3%	38.6%	44.9%	44.8%	44.3%	38.3%	44.9%	46.1%
MG	0.08%	0.02%	0.01%	0.03%	0.00%	0.04%	0.07%	0.06%	0.05%	0.02%	0.08%	0.03%
R^2	0.361	0.419	0.410	0.411	0.377	0.412	0.367	0.371	0.374	0.379	0.363	0.370
• CH-4 模型												
% Sig	8.06%	4.36%	5.10%	4.88%	6.58%	5.32%	7.98%	7.76%	7.91%	4.66%	7.91%	8.65%
$A \alpha $	0.19%	0.16%	0.17%	0.17%	0.19%	0.17%	0.19%	0.19%	0.19%	0.16%	0.19%	0.20%
$\frac{A \alpha }{A r }$	45.8%	36.9%	38.9%	39.0%	43.8%	39.0%	45.4%	45.2%	44.9%	38.8%	45.5%	46.6%
MG	0.08%	0.02%	0.01%	0.03%	0.00%	0.03%	0.06%	0.06%	0.05%	0.02%	0.08%	0.03%
R^2	0.362	0.420	0.411	0.412	0.379	0.413	0.369	0.373	0.376	0.380	0.364	0.371

明虽然显著有超额收益的个股数量增加,但加入金融地产股后模型对个股的平均解释力度也增加了.从整体来看,截距项检验的结果基本与不含金融地产股的结果保持一致.

4.6 稳健性检验: 去除行业关联

谈到上市公司资产间的关联,人们首先想到的一定是行业关联.行业关联可以说是被研究得最早,受到关注度最多的一类关联.虽然在上节实证结果中,实证结果发现分析师共同提及关联和新闻文本关联在捕捉股票间弱相关能力上都显著优于行业关联,但是我们依然好奇,这类关联在排除了行业关联后,是否仍然为挖掘股票间的局部相关性提供了帮助?本小节我们将进行检验.首先,从不同的关联中剔除行业内关联,即:

$$\mathbf{W}_{\text{no_ind}} = \mathbf{W} \otimes (\mathbf{I}_{n \times n} - \mathbf{1}^+(\mathbf{W}_{\text{ind}})),$$

其中, $\otimes$ 是矩阵 Hardmard 积, $\mathbf{I}_{n \times n}$ 为元素全为 1 的矩阵, $\mathbf{1}^+(\cdot)$ 为矩阵元素的正示性函数, 定义为:

$$\mathbf{1}^+(\mathbf{A})_{ij} = \mathbf{1}\{\mathbf{A}_{ij} > 0\},$$

即不为零的元素 (此处等价于正数) 返回为 1, 零元素返回为 0. 对于任意的关联矩阵 $\mathbf{W}$, 我们都可以通过上式进行去行业化转换, 并将转换后得到的 $\mathbf{W}_{\text{no_ind}}$ 重新代入模型得到回归结果, 如表 9 所示.

在去除了行业关联后, 虽然 City、Province 和 Resid 等网络已经不能非常有效地捕捉其余的信息, 但其余大部分关联网络下的空间因子模型仍然可以有效地减少因子模型的 α 显著

表 9 去行业关联后空间模型回归结果

	Sentence	Passage	Analyst	Concept	Industry	City	Province	Tech	Statistical	Customer	Resid
• CAPM 模型											
% Sig	1.53%	1.69%	2.01%	2.49%		2.97%	3.13%	2.57%	1.69%	2.65%	2.65%
$A \alpha $	0.12%	0.13%	0.14%	0.13%		0.14%	0.14%	0.14%	0.12%	0.16%	0.14%
$\frac{A \alpha }{A r }$	29.0%	29.8%	32.0%	30.6%		33.6%	32.2%	33.3%	28.0%	36.7%	31.6%
MG	0.03%	0.02%	0.03%	0.02%		0.08%	0.05%	0.08%	0.03%	0.12%	0.03%
R^2	0.342	0.346	0.333	0.346		0.312	0.334	0.313	0.356	0.276	0.343
• CH-3 模型											
% Sig	5.38%	5.94%	5.78%	7.31%		7.71%	7.71%	7.39%	5.38%	7.47%	8.59%
$A \alpha $	0.17%	0.18%	0.18%	0.19%		0.19%	0.20%	0.19%	0.17%	0.19%	0.20%
$\frac{A \alpha }{A r }$	40.2%	41.4%	41.6%	44.0%		45.5%	45.6%	44.1%	40.2%	45.4%	46.9%
MG	0.04%	0.04%	0.04%	0.03%		0.08%	0.07%	0.06%	0.04%	0.09%	0.06%
R^2	0.372	0.372	0.374	0.365		0.360	0.363	0.370	0.368	0.356	0.359
• CH-4 模型											
% Sig	5.62%	6.02%	5.94%	7.79%		8.19%	8.03%	7.71%	5.30%	7.95%	8.92%
$A \alpha $	0.17%	0.18%	0.18%	0.19%		0.20%	0.20%	0.19%	0.17%	0.20%	0.20%
$\frac{A \alpha }{A r }$	40.7%	41.9%	42.1%	44.4%		45.8%	45.9%	44.5%	40.5%	45.8%	47.2%
MG	0.04%	0.03%	0.04%	0.02%		0.08%	0.07%	0.06%	0.04%	0.09%	0.06%
R^2	0.374	0.373	0.375	0.366		0.361	0.364	0.371	0.370	0.358	0.361

比例 (在表 9 中, CAPM、CH-3、CH-4 模型的 α 显著比例分别是 2.73%, 7.54%, 8.06%), 这说明这些辅助数据的确与行业关联有一部分重合, 但在行业外仍然能提供资产之间的关联信息. 值得注意的是, 在去除行业前, Sentence 网络显著优于 Passage 网络, 但在去除行业关联后, Sentence 网络和 Passage 网络表现非常接近, 这也侧面说明了很大一部分的行业关联在同一句子中出现, 但在同一篇文章 (非同句) 中出现频率明显下降.

4.7 涵盖不可交易因子的无套利检验

如 2.6 节所言, 当模型中涵盖了不可交易因子, 则传统的截距项绝对值均值等统计量已经无法用以对无套利假设进行检验. 为此, 本小节借助 2.6 节的 Wald 检验理论, 针对涵盖不可交易因子的空间因子定价模型展开探讨. 值得注意的是, 当模型涵盖不可交易因子, 则 Wald 统计量的时间、空间复杂度均高达 $\mathcal{O}(n^4)$, 故我们仅以沪深 300 成分股为例, 展示涵盖不可交易因子的无套利检验⁴.

我们利用 Sentence 网络在沪深 300 成分股中进行了 Wald 检验, 检验结果见表 10, 可以看出, 仅是增加可交易因子无益于降低定价误差. 如表 10 所示, CH4 模型的 W_S 大于 CH3 模型的 W_S , 且 CH3 模型的 W_S 大于 CAPM 模型的 W_S . 但是宏观因子的加入是有助于解释异常收益的, 如 CH4+ 宏观因子模型的 W_S 甚至优于 CAPM. 此外, 标准化 Wald 统计量在八个模型中均在 30 以上, 无套利检验在八个模型中均无法通过. 在 CAPM 模型中, $W_S = 36.61$, 加入新闻矩阵后, W_S 下降至 33.37, 仍然非常显著, 但异常收益或多或少得到解释. 其余几个模型也均有类似的结论, 其中提升最大的是空间 CH4 模型, 将 W_S 从 51.00 降为 44.88.

表 10 无套利 Wald 检验

	标准化 Wald	p 值
因子模型		
• CAPM	36.61	0.000
• CH3 模型	49.96	0.000
• CH4 模型	51.00	0.000
• CH4 模型 + 宏观因子	35.18	0.000
空间因子模型		
• 空间 CAPM	33.37	0.000
• 空间 CH3 模型	44.06	0.000
• 空间 CH4 模型	44.88	0.000
• 空间 CH4 模型 + 宏观因子	34.60	0.000

5 结论与展望

本文将资产之间的横截面相关性拆分成由公共因子解释的强相关性部分以及同期收益率解释的局部相关性部分, 并用新闻数据来构造同期收益率之间的关联, 通过异质空间因子

⁴如果以全部 1353 只个股作为样本股进行 Wald 检验, 则内存需求 (约 200GB) 已经远超一台标准电脑的配置.

模型在中国市场展开实证. 结果表明, 基于新闻共现的局部空间相关性是资产收益的重要解释成分, 且在剔除行业影响后, 空间相关性仍然十分显著.

本文还讨论了空间因子模型框架下的渐近无套利定价理论, 介绍了仅含有可交易因子以及同时含有可交易因子与不可交易因子这两种情形的 Wald 检验. 实证检验结果表明, 加入宏观因子有助于解释无套利.

值得说明的是, 本文所使用的异质空间因子模型是静态而非动态的, 即空间矩阵 $\mathbf{W}$ 是固定的而非动态的 $\mathbf{W}_t$, 因此理论上只能用以检验市场上是否存在空间相关性, 不能用以预测. 动态的异质空间因子模型构建起来是有难度的, 因其密度函数是时变的, 要在 $T \rightarrow \infty$ 给出合适的渐近理论就需要 $\mathbf{W}_t$ 满足一些平稳性质, 如: 存在常数 c_i , 使得当 T 趋于无穷时, 有

$$\{(\mathbf{I}_n - \Psi \mathbf{W}_T)^{-1}\}_{ii}(\mathbf{W}_T)_{ii} \rightarrow c_i, \quad i = 1, 2, \dots, n,$$

其中 $\mathbf{A}_{ii}$ 指的是矩阵 $\mathbf{A}$ 的 (i, i) 元, 但在真实网络信息数据中该条件严苛且难以验证. 尽管如此, 动态的异质空间因子模型在理论上将会更匹配资产定价的实际, 是未来的研究方向之一. 除了动态矩阵, 另外一个值得改进之处是 Aquaro et al. (2021) 提出的将空间项拆为正负两部分 $\Psi^+ \mathbf{W}^+ \mathbf{r}_t + \Psi^- \mathbf{W}^- \mathbf{r}_t$, 这一问题涉及对新闻信息进行情感正负分类, 用正向关联构造 $\mathbf{W}^+$, 而用负向关联构造 $\mathbf{W}^-$.

由于资产之间存在这样的空间效应, 仅利用市场上的公共因子可能很难完全刻画资产之间的风险关联. 为应对此, 金融机构和研究机构可以广泛利用大数据技术挖掘更多潜在的股票关联信息, 例如通过类似于新闻文本分析的方式揭示股票间的隐性关联. 此外, 监管部门应要求上市公司提高信息透明度, 特别是与供应链和客户关系相关的信息. 这将有助于捕捉个股之间的局部关联性, 进一步提升中国股票市场的研究质量和风险管理能力.

当然, 整体来看, 目前还难以在我国金融市场上找到一个完全贴切的资产定价模型, 本文的异质空间因子模型虽然很大程度地提升了因子模型的可解释性, 但仍然有不少的个股存在显著的异常收益. 挖掘更有效的因子、提出更稳健的模型, 有待业界学界研究者去展开进一步探究.

参 考 文 献

- 陈张杭健, 吴粤, 李世炳, 任飞, (2021). 股吧个体信息交互对股价联动关系的影响研究 [J]. 管理科学学报, 24(5): 47-69.
- Chen Z H J, Wu Y, Li S B, Ren F, (2021). Research on the Influence of Individual Information Interaction on Stock Price Linkage in Stock Bar[J]. Journal of Management Sciences in China, 24(5): 47-69.
- 段丙蕾, 汪荣飞, 张然, (2022). 南橘北枳: A 股市场的经济关联与股票回报 [J]. 金融研究, (2): 171-188.
- Duan B L, Wang R F, Zhang R, (2022). Tangerine in the South and Trifoliate Orange in the North: Economic Linkages and Stock Returns in A-Share Market[J]. Journal of Financial Research, (2): 171-188.
- 范龙振, 余世典, (2002). 中国股票市场的三因子模型 [J]. 系统工程学报, (6): 537-546.
- Fan L Z, Yu S D, (2002). Three-factor Model of China's Stock Market[J]. Journal of Systems Engineering, (6): 537-546.

- 范小云, 王业东, 王道平, 郭文璇, 胡焱翊, (2022). 不同来源金融文本信息含量的异质性分析——基于混合式文本情绪测度方法 [J]. 管理世界, 38(10): 78–95.
- Fan X Y, Wang Y D, Wang D P, Guo W X, Hu X Y, (2022). Heterogeneity Analysis of Information Content of Financial Texts from Different Sources — Based on Mixed Text Sentiment Measurement Method[J]. Journal of Management World, 38(10): 78–95.
- 范小云, 王业东, 王道平, (2022). 基于新闻大数据与机器学习的中国银行业系统性风险研究 [J]. 世界经济, 45(4): 3–30. Doi: 10.19985/j.cnki.cassjwe.2022.04.006.
- Fan X Y, Wang Y D, Wang D P, (2022). A Study of the Systemic Risk Measurement and Early Warning Mechanism of the Banking Industry in China Based on News and Machine Learning[J]. The Journal of World Economy, 45(4): 3–30. Doi: 10.19985/j.cnki.cassjwe.2022.04.006.
- 姜富伟, 孟令超, 唐国豪, (2021). 媒体文本情绪与股票回报预测 [J]. 经济学 (季刊), 21(4): 1323–1344.
- Jiang F W, Meng L C, Tang G H, (2021). Media Text Sentiment and Stock Return Prediction[J]. China Economic Quarterly, 21(4): 1323–1344.
- 李斌, 雷印如, (2022). 中国公募基金挖掘了股票市场异象吗?[J]. 金融研究, (9): 188–206.
- Li B, Lei Y R, (2022). Do Mutual Funds Exploit Stock Market Anomalies in China?[J]. Journal of Financial Research, (9): 188–206.
- 刘霖, 秦宛顺, (2004). 中国股票市场套利定价模型研究 [J]. 金融研究, (6): 44–55.
- Liu L, Qin W S, (2004). Research on Arbitrage Pricing Model in Chinese Stock Market[J]. Journal of Financial Research, (6): 44–55.
- 刘维奇, 张燕, (2020). 资产定价与劳动成本占比 [J]. 中国管理科学, 28(12): 1–11.
- Liu W Q, Zhang Y, (2020). Asset Pricing and the Proportion of Labor Cost[J]. Chinese Journal of Management Science, 28(12): 1–11.
- 潘荣翠, 张鑫, 韩跃红, 尚朝秋, (2012). 基于空间计量分析的全球股票价格指数宏观影响因素 [J]. 云南师范大学学报 (哲学社会科学版), 44(1): 135–144.
- Pan R C, Zhang X, Han Y H, Shang C Q, (2012). Macro Influencing Factors of Global Stock Price Index Based on Spatial Econometric Analysis[J]. Journal of Yunnan Normal University (Humanities and Social Sciences Edition), 44(1): 135–144.
- 沈艳, 陈赟, 黄卓, (2019). 文本大数据分析在经济学和金融学中的应用: 一个文献综述 [J]. 经济学 (季刊), 18(4): 1153–1186.
- Shen Y, Chen Y, Huang Z, (2019). Application of Textual Big Data Analysis in Economics and Finance: A Literature Review[J]. China Economic Quarterly, 18(4): 1153–1186.
- 石泽龙, 傅强, 李山, (2022). 生产网络与行业收益率: 基于时变网络依赖参数的空间因子模型 [J]. 系统管理学报, 31(3): 453–466.
- Shi Z L, Fu Q, Li S, (2022). Production Network and Industry Stock Returns: Spatial Factor Model Based on Time-Varying Network Dependent Parameters[J]. Journal of Systems & Management, 31(3): 453–466.
- 田利辉, 王冠英, 张伟, (2014). 三因素模型定价: 中国与美国有何不同?[J]. 国际金融研究, (7): 37–45.
- Tian L H, Wang G Y, Zhang W, (2014). Three-factor Model Pricing: How is China Different from the United States?[J]. Studies of International Finance, (7): 37–45.
- 吴世农, 许年行, (2004). 资产的理性定价模型和非理性定价模型比较研究——基于中国股市的实证分析 [J]. 经济研究, (6): 105–116.
- Wu S N, Xu N H, (2004). A Comparative Study of Rational Pricing Model and Irrational Pricing Model of Assets — An Empirical Analysis Based on China's Stock Market[J]. Economic Research

- Journal, (6): 105–116.
- 杨子晖, 李东承, (2018). 我国银行系统性金融风险研究——基于“去一法”的应用分析 [J]. 经济研究, 53(8): 36–51.
- Yang Z H, Li D C, (2018). Research on the Systemic Financial Risk of China's Banks — Based on the Application Analysis of “Cut One Method”[J]. Economic Research Journal, 53(8): 36–51.
- 张玉华, 宋韞赟, 张元庆, (2016). 基于空间面板数据模型的股票收益率影响因素分析 [J]. 中国软科学, (5): 172–183.
- Zhang Y H, Song Y Y, Zhang Y Q, (2016). Analysis of Factors Influencing Stock Returns Based on Spatial Panel Data Model[J]. Chinese Soft Science Magazine, (5): 172–183.
- 张永杰, 张维, 金曦, 熊熊, (2011). 互联网知道的更多么? —— 网络开源信息对资产定价的影响 [J]. 系统工程理论与实践, 31(4): 577–586.
- Zhang Y J, Zhang W, Jin X, Xiong X, (2011). Does the Internet Know More? — Open Source Information and Asset Pricing[J]. Systems Engineering — Theory & Practice, 31(4): 577–586.
- 张自力, 闫红蕾, 张楠, (2020). 股票网络、系统性风险与股票定价 [J]. 经济学 (季刊), 19(1): 329–350.
- Zhang Z L, Yan H L, Zhang N, (2020). Stock Network, Systemic Risk and Stock Pricing[J]. China Economic Quarterly, 19(1): 329–350.
- 赵丽丽, 赵茜倩, 杨娟, 王铁军, 李庆, (2012). 财经新闻对中国股市影响的定量分析 [J]. 山东大学学报 (理学版), 47(7): 70–75.
- Zhao L L, Zhao X Q, Yang J, Wang T J, Li Q, (2012). The Quantitative Analysis of the Impact of Financial News on Stock Market[J]. Journal of Shandong University (Natural Science), 47(7): 70–75.
- 赵留彦, 宁可, (2020). 声誉激励与分析师行为——基于新财富最佳分析师评选的实证研究 [J]. 经济科学, (3): 73–85.
- Zhao L Y, Ning K, (2020). Reputation Incentive and Analyst Behavior: An Empirical Study Based on the New Wealth Best Analyst Appraisal[J]. Economic Science, (3): 73–85.
- 朱恩伟, 吴璟, 刘洪玉, (2019). 基于新闻文本共现性的银企关系分析——以房地产上市公司为例 [J]. 金融研究, (2): 117–135.
- Zhu E W, Wu J, Liu H Y, (2019). Bank-Firm Relationship Analysis Based on Co-occurrence in News Texts: Evidence from Listed Real Estate Companies in China[J]. Journal of Financial Research, (2): 117–135.
- Ali U, Hirshleifer D, (2020). Shared Analyst Coverage: Unifying Momentum Spillover Effects[J]. Journal of Financial Economics, 136(3): 649–675.
- Andrews D W, (1987). Asymptotic Results for Generalized Wald Tests[J]. Econometric Theory, 3(3): 348–358.
- Aquaro M, Bailey N, Pesaran M H, (2015). Quasi Maximum Likelihood Estimation of Spatial Models with Heterogeneous Coefficients[R]. USC-INET Research Paper: 15–17.
- Aquaro M, Bailey N, Pesaran M H, (2021). Estimation and Inference for Spatial Models with Heterogeneous Coefficients: An Application to US House Prices[J]. Journal of Applied Econometrics, 36(1): 18–44.
- Aït-Sahali Y, Brandt M W, (2001). Variable Selection for Portfolio Choice[J]. The Journal of Finance, 56(4): 1297–1351.
- Bailey N, Holly S, Pesaran M H, (2016). A Two-stage Approach to Spatio-temporal Analysis with Strong and Weak Cross-sectional Dependence[J]. Journal of Applied Econometrics, 31(1): 249–280.
- Chen L, Pelger M, Zhu J, (2024). Deep Learning in Asset Pricing[J]. Management Science, 70(2): 714–

- 750.
- Chudik A, Pesaran M H, Tosetti E, (2011). Weak and Strong Cross-section Dependence and Estimation of Large Panels[J]. *The Econometrics Journal*, 14(1): C45–C90.
- Cohen L, Frazzini A, (2008). Economic Links and Predictable Returns[J]. *The Journal of Finance*, 63(4): 1977–2011.
- Connor G, Korajczyk R A, (2024). Semi-strong Factors in Asset Returns[J]. *Journal of Financial Econometrics*, 22(1): 70–93.
- Dong C, Connor G, Li S, Linton O B, (2024). A Dynamic Semiparametric Characteristics-Based Model For Portfolio Selection[R/OL]. Available at SSRN: 4863571.
- Du Q, Liang D, Chen Z, Tu J, (2022). Concept Links and Return Momentum[J]. *Journal of Banking & Finance*, 134: 106329.
- Fama E F, French K R, (1989). Business Conditions and Expected Returns on Stocks and Bonds[J]. *Journal of Financial Economics*, 25(1): 23–49.
- Fama E F, French K R, (1993). Common Risk Factors in the Returns on Stocks and Bonds[J]. *Journal of Financial Economics*, 33(1): 3–56.
- Fama E F, French K R, (2015). A Five-factor Asset Pricing Model[J]. *Journal of Financial Economics*, 116(1): 1–22.
- Fernandez V, (2011). Spatial Linkages in International Financial Markets[J]. *Quantitative Finance*, 11(2): 237–245.
- Gabaix X, (2011). The Granular Origins of Aggregate Fluctuations[J]. *Econometrica*, 79(3): 733–772.
- Ge S, Li S, Linton O, (2023). News-implied Linkages and Local Dependency in the Equity Market[J]. *Journal of Econometrics*, 235(2): 779–815.
- Ge S, Li S, Zheng H, (2025). Diamond Cuts Diamond: News Co-mention Momentum Spillover Prevails in China[J]. *Journal of Banking and Finance*, 171: 107356.
- Ge S, Li S, Linton O B, Liu W, Su W, (2026). Should We Augment Large Covariance Matrix Estimation with Auxiliary Network Information?[R/OL]. Available at SSRN: 4975753.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4975753.
- Giglio S, Xiu D C, (2021). Asset Pricing with Omitted Factors[J]. *Journal of Political Economy*, 129(7): 1947–1990.
- Gu S, Kelly B, Xiu D, (2020). Empirical Asset Pricing via Machine Learning[J]. *The Review of Financial Studies*, 33(5): 2223–2273.
- He W, Wang H, Wang Y, Yu J, (2021). Similar Stocks[R/OL]. Available at SSRN: 3815595.
- Hou K, Qiao F, Zhang X, (2023). Finding Anomalies in China[R]. Fisher College of Business Working Paper.
- Jiang F, Lee J, Martin X, Zhou G, (2019). Manager Sentiment and Stock Returns[J]. *Journal of Financial Economics*, 132(1): 126–149.
- Kou S, Peng X, Zhong H, (2018). Asset Pricing with Spatial Interaction[J]. *Management Science*, 64(5): 2083–2101.
- Lee C M, Sun S T, Wang R, Zhang R, (2019). Technological Links and Predictable Returns[J]. *Journal of Financial Economics*, 132(3): 76–96.
- Li Z, Liu L X, Liu X, John Wei K C, (2024). Replicating and Digesting Anomalies in the Chinese A-share Market[J]. *Management Science*, 70(8): 5066–5090.

- Liu J, Stambaugh R F, Yuan Y, (2019). Size and Value in China[J]. Journal of Financial Economics, 134(1): 48–69.
- Parsons C A, Sabbatucci R, Titman S, (2020). Geographic Lead-lag Effects[J]. The Review of Financial Studies, 33(10): 4721–4770.
- Qiao F, (2024). Do Analysts Disseminate Anomaly Information in China?[J]. Journal of Banking & Finance, 165: 107221.
- Ross S, (1976). The Arbitrage Theory of Capital Asset Pricing[J]. Journal of Economic Theory, 13(3): 341–360.
- Scherbina A, Schlusche B, (2015). Economic Linkages Inferred From News Stories and the Predictability of Stock Returns[J]. Available at SSRN: 2363436. <http://dx.doi.org/10.2139/ssrn.2363436>.
- Schwenkler G, Zheng H, (2019). The Network of Firms Implied by the News[R]. Boston University Questrom School of Business Research Paper: 3320859. <http://dx.doi.org/10.2139/ssrn.3320859>.
- Sharpe W F, (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk[J]. The Journal of Finance, 19(3): 425–442.
- Zareei A, (2019). Network Origins of Portfolio Risk[J]. Journal of Banking & Finance, 109: 105663.

附录

对数似然函数 (6) 的梯度由下述各一阶导给出:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \psi_i} &= -T \text{tr}\{[\mathbf{S}(\boldsymbol{\Psi})]^{-1} \mathbf{I}_{ii} \mathbf{W}\} + \sum_{t=1}^T \frac{r_{it}^W (r_{it} - \psi_i r_{it}^W - \mathbf{x}_{it}^T \mathbf{b}_i)}{\sigma_i^2}, \quad i = 1, 2, \dots, n, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i} &= \sum_{t=1}^T \frac{\mathbf{x}_{it} (r_{it} - \psi_i r_{it}^W - \mathbf{x}_{it}^T \mathbf{b}_i)}{\sigma_i^2}, \quad i = 1, 2, \dots, n, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_i^2} &= -\frac{T}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} \sum_{t=1}^T (r_{it} - \psi_i r_{it}^W - \mathbf{x}_{it}^T \mathbf{b}_i)^2, \quad i = 1, 2, \dots, n,\end{aligned}$$

其中 $\mathbf{I}_{ii}$ 表示的是一个除了第 (i, i) 元为 1, 其他位置元素全为 0 的 n 阶方阵. 实际上, 通过观察对数似然函数关于 $\mathbf{b}$ 和 $\boldsymbol{\sigma}^2$ 的偏导, 可以得到 $\boldsymbol{\theta}$ 的 QMLE 的内在关系, 即:

$$\hat{\mathbf{b}}_i = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\tilde{\mathbf{r}}_i - \hat{\psi}_i \tilde{\mathbf{r}}_i^W), \quad i = 1, 2, \dots, n,$$

其中

$$\mathbf{X} = \begin{pmatrix} 1 & \mathbf{f}_1^T \\ 1 & \mathbf{f}_2^T \\ \vdots & \vdots \\ 1 & \mathbf{f}_T^T \end{pmatrix}, \quad \tilde{\mathbf{r}}_i = (r_{i1}, \dots, r_{iT})^T, \quad \tilde{\mathbf{r}}_i^W = (r_{i1}^W, \dots, r_{iT}^W)^T,$$

以及

$$\hat{\sigma}_i^2 = \frac{1}{T} \|\tilde{\mathbf{r}}_i - \hat{\psi}_i \tilde{\mathbf{r}}_i^W - \mathbf{X}_i \hat{\mathbf{b}}_i\|^2, \quad i = 1, 2, \dots, n.$$

对数似然函数的 Hessian 矩阵是:

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} \\ \cdot & \mathbf{H}_{22} & \mathbf{H}_{23} \\ \cdot & \cdot & \mathbf{H}_{33} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^T} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\psi} \partial \mathbf{b}^T} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\psi} \partial (\boldsymbol{\sigma}^2)^T} \\ \cdot & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b} \partial \mathbf{b}^T} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b} \partial (\boldsymbol{\sigma}^2)^T} \\ \cdot & \cdot & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\boldsymbol{\sigma}^2) \partial (\boldsymbol{\sigma}^2)^T} \end{bmatrix}.$$

定义 $\mathbf{G} = (g_{ij})_{n \times n} = \mathbf{W}(\mathbf{I}_n - \boldsymbol{\Psi}\mathbf{W})^{-1}$, 则 Hessian 矩阵由下述各部分给出:

$$\begin{aligned} \mathbf{H}_{11} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} = \begin{cases} -T g_{ij} g_{ji}, & i \neq j, \\ -T g_{ii}^2 - \frac{(\mathbf{r}_i^W)^T (\mathbf{r}_i^W)}{\sigma_i^2}, & i = j, \end{cases} \\ \mathbf{H}_{12} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \mathbf{b}_j^T} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \mathbf{b}_j^T} = \begin{cases} 0, & i \neq j, \\ -\frac{(\mathbf{r}_i^W)^T \mathbf{X}}{\sigma_i^2}, & i = j, \end{cases} \\ \mathbf{H}_{13} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \sigma_j^2} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \sigma_j^2} = \begin{cases} 0, & i \neq j, \\ -\frac{(\mathbf{r}_i^W)^T (\mathbf{r}_i - \psi_i \mathbf{r}_i^W - \mathbf{X} \mathbf{b}_i)}{\sigma_i^4}, & i = j, \end{cases} \\ \mathbf{H}_{22} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i \partial \mathbf{b}_j^T} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i \partial \mathbf{b}_j^T} = \begin{cases} 0, & i \neq j, \\ -\frac{\mathbf{X}^T \mathbf{X}}{\sigma_i^2}, & i = j, \end{cases} \\ \mathbf{H}_{23} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i \partial \sigma_j^2} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i \partial \sigma_j^2} = \begin{cases} 0, & i \neq j, \\ -\frac{\mathbf{X}^T (\mathbf{r}_i - \psi_i \mathbf{r}_i^W - \mathbf{X} \mathbf{b}_i)}{\sigma_i^4}, & i = j, \end{cases} \\ \mathbf{H}_{33} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma_i^2 \partial \sigma_j^2} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma_i^2 \partial \sigma_j^2} = \begin{cases} 0, & i \neq j, \\ -\frac{T}{2\sigma_i^4} + \frac{(\mathbf{r}_i - \psi_i \mathbf{r}_i^W - \mathbf{X} \mathbf{b}_i)^T (\mathbf{r}_i - \psi_i \mathbf{r}_i^W - \mathbf{X} \mathbf{b}_i)}{\sigma_i^6}, & i = j. \end{cases} \end{aligned}$$