

见微知著：中国股票市场的异质空间因子定价模型

戈舒怡 李少然 黎新平 苏文

(南开大学金融学院, 天津, 300000; 北京大学经济学院/北京大学金融工程实验室, 北京, 100000)

摘要

资产收益率之间的横截面相关性由公共风险因子相关性和局部相关性组成, 二者具有不同的经济学以及统计学含义。本文利用异质空间因子定价模型同时解释这两种相关性, 并在这一框架下介绍了模型的统计推断及无渐近套利理论。实证方面, 针对沪深 300 成分股和相关因子, 我们借助百万条财经新闻数据构造了资产之间的关联矩阵, 利用 2017-2021 年的日收益率估计并检验模型。结果表明: 新闻数据有助于捕捉资产之间存在显著的空间相关性, 并且在剔除行业因子影响后, 空间系数仍然十分显著, 且无论在个股层面还是行业层面, 空间系数呈现出非常明显的异质性。

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

JEL 分类号: C31, C50, G12

一、引言

在传统因子模型 (如: Sharpe, 1964 [1]; Ross, 1976 [2]) 的框架下, 同期资产收益率之间的同步移动 (comovement) 是由一些外生的公共风险因子所驱动的。近年来, 众多研究表明, 在控制了因子风险之后, 资产间仍存在某种由于局部关联所导致的依赖性 (如: Gabaix, 2011 [3]; Kou et al., 2018 [4]; 张玉华等, 2016 [5]; 张自力等, 2020 [6])。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为空间效应、网络效应或者群同效应¹。表 1 显示, 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, 这与石泽龙等 (2022) [7] 得到的结果一致, 且即使我们利用主成分分析挑选多达十个有强解释力的因子后, 因子模型仍未能完全刻画资产之间的所有横截面相关性, 残差中的空间效应仍就非常显著。

大 n (资产数量大) 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

作者简介: 戈舒怡, 助理教授, 南开大学金融学院, E-mail: sg751_shuyige@outlook.com.

李少然 (通讯作者), 助理教授, 北京大学经济学院, E-mail: lishaoran@pku.edu.cn.

黎新平, 博士, 北京大学经济学院, E-mail: xinpinglei@pku.edu.cn.

苏文, 硕士, 北京大学金融工程实验室, E-mail: suwen@stu.pku.edu.cn.

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。

Table 1: 沪深 300 成分股因子模型残差 Moran's I 检验

样本股	模型	平均 Z 值	显著个数	显著比例
沪深 300	CAPM	19.394	1192	97.95%
沪深 300	CH3	17.170	1190	97.78%
沪深 300	CH4	17.111	1191	97.86%
沪深 300	PCA-5	11.348	1152	94.66%
沪深 300	PCA-7	8.977	1109	91.13%
沪深 300	PCA-10	7.211	1025	84.22%

¹ 该表汇总了以沪深 300 成分股 2017-2021 日频收益率为研究对象，因子模型残差的 Moran's I 检验结果。该检验的目的是考察去除风险因子的影响后，股票收益率残差之间的空间相关性的强弱。若检验显著，则说明股票收益率残差之间存在空间相关性。

² CH3、CH4 分别指 Liu et al. (2019) [21] 构造的中国三因子和四因子模型。

³ PCA-5 指通过对样本股进行主成分分析，提取前 5 个主成分作为因子构造的因子模型。PCA-7，PCA-10 同理。

⁴ 2017-2021 总计 1217 个日收益率，平均 Z 值汇总了 1217 个 Moran's I 检验 Z 值的平均值，另外两列指 0.05 水平下显著个数与比例。

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，空间效应刻画的是由于局部关联所导致的内生的异质性风险传播，即一个资产受到与之存在关联的资产的风险溢出的影响。由此可见，对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，根据 Chudik et al. (2011) [8] 和 Bailey et al. (2016) [9]，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。在研究资产定价问题中，忽略资产间局部依赖性的代价是巨大的。一方面，物联时代，万物相连，金融资产间更是存在着千丝万缕的联系。如果不能准确地辨别风险的来源，将不利于我们对系统性金融风险进行识别与预防。因此，区分资产收益率横截面相关性中由共同因子暴露驱动的强相关部分与空间效应导致的弱（局部）相关部分有助于帮助我们识别不同的风险，因而可以更有效维护金融体系稳定。另一方面，资产间异质性风险的局部传播，也就是空间效应，会影响风险分散的效果以及因子模型中风险暴露系数的估计 (Abafaz, 2019 [10]; Giglio and Xiu, 2021 [11])，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且深刻的问题。本文提出了如下的异质空间因子定价模型来分别对风险因子驱动的强相关和空间效应导致的弱相关进行建模与分析。

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

其中 \mathbf{r}_t 代表 $n \times 1$ 资产收益率向量； \mathbf{f}_t 为 $K \times 1$ 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} 关于第 j 个因子 f_{jt} 的因子载荷； 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； $\boldsymbol{\Psi} = \text{diag}\{\psi_1, \dots, \psi_n\}$ 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强横截面相关性，而空间项 $\boldsymbol{\Psi}W\mathbf{r}_t$ 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。我们会在文章的第三章详细介绍它们在统计学性质上的不同以及我们如何区分这两种横截面相关性。

本文介绍的异质空间因子模型有如下三个特点：

空间效应的异质性。在国内，也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 [12]；张玉华等，2016 [5]；石泽龙等，2022 [7]），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，更符合实际情况，也有助于我们分析不同资产、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，我们发现，无论在个股层面、还是行业层面，空间系数表现出了明显的异质性，且关于异质性的渐近似然比检验统计量非常显著，说明了选用异质模型的合理性。

涵盖不可交易因子。Fama and French (1993) [13] 为首的一系列因子发掘文献大多基于某些重要的公司特征，对全市场展开排序并找出相应的头部、尾部资产组合，根据头部、尾部组合收益率的差值来代理那些公共风险因子的市场表现。也正因如此，我们将这类公共风险因子称为“可交易”因子。与这些可交易因子相对应的则是那些不可交易因子，例如宏观指标（如通货膨胀率、失业率、国民生产总值等）。宏观变量直接影响国民经济以及整个金融市场，它们显然具有共同因子的特征，但我们无法通过投资组合来完全模拟它们的真实收益率或风险溢价，因此我们称其为“不可交易”因子。可交易因子与不可交易因子共同构成了影响绝大多数个股且其风险不可被分散的共同因子。不可交易因子的引入并不会给模型的估计带来改变，但是它会极大的增加相应资产定价模型无渐进套利检验的难度，本文将为此问题介绍相应的解决方案。与国外研究相比，国内对于共同风险因子的研究和发掘方兴未艾，本文在一定程度上填补了这个空缺。

基于财经新闻构造空间关联矩阵。空间因子定价模型当中的一大难点，就是如何构建资产之间的网络关系（即空间模型中的空间权重矩阵）。已有的空间回归模型文献多是将空间权重矩阵与地理距离挂钩。然而金融资产间的相关性错综复杂、莫衷一是，挖掘并验证可以较好展现资产间局部依赖性的数据尤为重要。本文在大数据背景下，对新闻文本数据展开文本共现 (news co-mention) 分析，见微知著，发掘资产之间的空间关联性，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，则他们很可能存在某种形式的经济关联。并且，他们的新闻共现频率能够反应他们的关联强弱。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、兼并等等，它们可以较为全面的捕捉资产间的关联；其二，新闻还具有时效性，更新较快，有助于及时观测到资产间关系的变化 (Secherbina and Schlusche, 2015 [14]；Schwenkler and Zheng, 2019 [15])。北京大学金融工程实验室汇集了锐思财经文本智能分析平台和聚源数据库从 1999 年 9 月 22 日开始的数百万条新闻文本数据，并筛选了提及上市公司名称的百万余条财经新闻文本数据。本文利用此海量新闻文本数据，基于新闻共现构建了资产间的空间权重矩阵。实证结果表明：在不同稳健性检验设定下，都有 75% 以上的样本股票具有显著的局部依赖性。并且基于新闻共现的空间权重矩阵的加入能够大幅减小资产定价模型中定价误差。

本文的后续安排如下：我们将在第二章整理并总结一些国内外与本文研究相关的已有文献；第三章则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；在第四章中，我们利用沪深 300 成分股 2017-2021 的日收益率数据以及北京大学金融工

程实验室 2012-2021 年的新闻共现数据展开实证；第五章对全文进行了总结，并提出了一些对未来研究方向的展望。

二、文献综述

Fama and French (1993) [13] 的风险、市值与账面市值比三因子模型可以较好地解释美国上市公司股票收益率，其在中国市场的可推广性也早已被验证（如范龙振和余世典，2002 [16]；吴世农和许年行，2004 [17]；廖理和沈红波，2008 [18]）。但 Fama-French 的模型并不是普适的，如田利辉等 (2014) [19] 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。针对公募基金，李斌和雷印如 (2022) [20] 发现在中国市场上，无论是用 CAPM、Fama-French 三因子还是 Liu et al. (2019) [21] 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。刘维奇和张燕 (2020) [22] 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。Liu et al. (2019) [21] 则考虑到中美市场差异性，借助并改进了 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。

同时，由于因子模型只能捕捉资产间的强相关性，很多学者开始分析资产间的局部相关性和网络关系。段丙蕾 (2022) [23] 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。为研究资产间网络与收益率的关系，Fernandez (2011) [24] 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。潘荣翠等 (2012) [12] 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。张玉华等 (2016) [5] 在 S-CAPM 的基础上利用 2007-2015 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。考虑到行业间通过投入产出关系形成了复杂且紧密关联的生产网络，石泽龙等 (2022) [7] 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。

但在上述使用空间计量模型的研究中，空间回归系数常被假设为同质的，即不同资产的空间相关性的敏感程度一致，这与现实不符，且具有很大的局限性。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. (2015) [25] 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 Aquaro et al. (2021) [26] 将该模型运用在美国房价上，借助 1975 年至 2014 年期间美国 338 个大城市房价数据，发现空间依赖性显著，且各城市对空间相关性的敏感性的确存在异质性。此外，在一定正则性条件下，Ge et al. (2022) [27] 推导了异质空间因子模型下的无渐进套利条件，并提出了在资产数量较大时仍然适用的 Wald 检验。

大数据时代下，新闻、文本数据往往蕴藏着大量潜在信息。Scherbina and Schlusche (2015) [14] 和

Schwenkler and Zheng (2019) [15] 发现新闻数据揭示了公司之间的内在联系。他们利用新闻数据构造了企业之间的信息网络，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) [28] 基于企业财务文本数据构建了一个情绪指数。他们发现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) [29] 较早地利用文本数据挖掘技术，发现社交媒体中的文本蕴含对股票异常日收益率有显著解释力的有效的信息内容。此外，赵丽丽等 (2012) [30] 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。多元线性回归结果表明新闻发布对深市股票的影响力度和持续时间均强于沪市股票，且规模较小的公司的股票收益受新闻的影响较大。姜富伟等 (2021) [31] 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) [32] 和范小云等 (2022) [33] 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联动影响。沈艳等 (2019) [34] 深刻地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) [35] 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。本文则将通过新闻数据挖掘资产间的网络关系，结合异质空间因子回归模型，在我国金融市场上开展实证研究。

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，为我国金融资产定价注入新鲜血液。目前在我国仍然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，而且无论在个股层面、还是行业层面，我们确实发现空间系数表现出了明显的异质性。第二，我们以中国财经新闻大数据为辅助信息源构造了资产之间的网络关系，该方法能全面、及时捕捉公司之间的关系网络，是对中国上市公司网络以及局部风险传导研究的创新。本文的实证结果表明，加入新闻矩阵后的空间因子模型能够进一步减少定价误差。第三，我们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，在此框架下，我们还介绍了包含不可交易因子时检验无渐近套利的 Wald 检验。第四，本文中使用的新闻关联矩阵生成的数据和算法将发表在北京大学金融工程实验室，可供读者和研究者们生成任意时间段所需股票的相关数据。

三、理论基础

(一) 强相关性的刻画：因子模型

考虑如下的因子模型

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

其中 \mathbf{r}_t 代表 t 时刻 $n \times 1$ 资产收益率向量， \mathbf{f}_t 为 $K \times 1$ 风险因子， $\mathbf{B} = (\beta_{ij})_{n \times K}$ 则是因子载荷矩阵， β_{ij} 表示第 i 项资产 r_{it} 关于第 j 个因子 f_{jt} 的因子载荷。

因子模型以外生变量 \mathbf{f}_t 完全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，即这种风险无法通过简单地增加资产个数而分散化消除。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。

(二) 局部（弱）相关性的刻画：空间回归模型

考虑下述同质空间回归模型：

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

其中 \mathbf{r}_t 代表 $n \times 1$ 资产收益率向量， $W = (w_{ij})_{n \times n}$ 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 $\Omega = \text{diag}\{\boldsymbol{\sigma}^2\} = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$ 。

空间模型以内生变量 $W \mathbf{r}_t$ 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 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。

(三) 强/弱相关性的刻画：空间因子模型

本文研究如下的异质空间因子模型：

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B} \mathbf{f}_t + \boldsymbol{\Psi} W \mathbf{r}_t + \boldsymbol{\epsilon}_t, \quad (4)$$

其中 \mathbf{r}_t 代表 $n \times 1$ 资产收益率向量， \mathbf{f}_t 为 $K \times 1$ 风险因子， $\mathbf{B} = (\beta_{ij})_{n \times K}$ 则是因子载荷矩阵， β_{ij} 表示第 i 项资产 r_{it} 关于第 j 个因子 f_{jt} 的因子载荷， W 表示 $n \times n$ 维的空间权重矩阵， $\boldsymbol{\Psi} = \text{diag}\{\psi_1, \dots, \psi_n\}$ 代表空间回归系数矩阵，该对角矩阵的第 i 个对角元素为股票 i 对空间效应的敏感度 ψ_i ，当 $\psi_1 = \dots = \psi_n$ ，该模型退化为同质空间因子模型。

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。我们考虑下述四个正则性假设：

1. 零均值的残差 $\boldsymbol{\epsilon}_t$ 无自相关性且与因子 \mathbf{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 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, W 的最大行范数 $\|W\|_\infty$ 有有界, 且

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

这将保证 ΨW 的谱半径是小于 1 的, 即

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

4. 假设 3 保证了 $G(\Psi) := (I_n - \Psi W)^{-1}$ 的存在, 我们再假定 G 满足 $\|G(\Psi)\|_1$ 有界。

与 3.1 节中对因子模型的假设一致, 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 4 表明 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, 故本文不讨论滞后期收益率 r_{t-j} 的影响。

(四) 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 (4), 定义全体参数 $\boldsymbol{\theta} = (\boldsymbol{\psi}', \mathbf{b}', \boldsymbol{\sigma}^2)'$, 其中 $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_n)'$, 且 $\mathbf{b}_i = (\alpha_i, \beta_{i1}, \dots, \beta_{iK})'$, 假设 $\mathbf{S}(\Psi) = I_n - \Psi W$ 可逆, 则有

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

其中自变量 $\mathbf{x}_{it} = (1, \mathbf{f}'_t)'$, 且 $\mathbf{x}_t = (\mathbf{x}'_{1t}, \dots, \mathbf{x}'_{nt})'$, 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

$$\ell(\boldsymbol{\theta}) = -\frac{nT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^n \ln(\sigma_i^2) + T \ln |\mathbf{S}(\Psi)| - \frac{1}{2} \sum_{t=1}^T [\mathbf{S}(\Psi) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t]' \Omega^{-1} [\mathbf{S}(\Psi) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t], \quad (6)$$

其中 $\Omega = \text{diag} \{\sigma_1^2, \dots, \sigma_n^2\}$ 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

$$\sum_{t=1}^T [\mathbf{S}(\Psi) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t]' \Omega^{-1} [\mathbf{S}(\Psi) \mathbf{r}_t - \mathbf{B} \mathbf{x}_t] = \sum_{t=1}^T \sum_{i=1}^n \frac{(r_{it} - \psi_i r_{it}^W - \mathbf{x}'_{it} \mathbf{b}_i)^2}{\sigma_i^2}, \quad (7)$$

其中 $(r_{1t}^W, \dots, r_{nt}^W)'$ = $W \mathbf{r}_t = \mathbf{r}_t^W$ 。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 $\hat{\boldsymbol{\theta}}$, 如果误差项是高斯分布, 则其渐近协方差矩阵是 $\mathbf{H}^{-1}(\boldsymbol{\theta})/T$ 。如果误差项不是正态的, 那么 QMLE 的渐近协方差矩阵则由

$$V_{\hat{\boldsymbol{\theta}}} = \mathbf{H}^{-1}(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta}) \mathbf{H}^{-1}(\boldsymbol{\theta}) \quad (8)$$

给出, 其中 $\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{T} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)'$ 。

(五) 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 [5][7]，我们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：

$$H_0 : \psi_1 = \psi_2 = \cdots = \psi_n,$$

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

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

其中 $\hat{\theta}$ 是异质空间因子模型的估计，而 $\tilde{\theta}$ 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。

(六) 不可交易因子与无套利定价

无套利是资产定价中非常重要的假定。在一定前提假设下，Ge et al. (2022) [27] 给出了空间因子模型的无（渐近）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）写为超额收益形式

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

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。可以证明在无渐近套利假定下，该异质空间因子需满足 $\tilde{\alpha} \approx 0$ 。这里负号 \approx 意味着无渐近套利（即当经济系统中资产数目增加时，定价误差 $\|\tilde{\alpha}\|^2 = (\tilde{\alpha})'\tilde{\alpha}$ 是有界的。

此外，当因子 \mathbf{f} 中包含一部分不可交易因子时，即 $\mathbf{f} = (\mathbf{f}'_1, \mathbf{f}'_2)$ ，其中 \mathbf{f}_1 是 $K_1 \times 1$ 可交易因子， \mathbf{f}_2 是 $K_2 \times 1$ 去均值后的不可交易因子，并有 $K_1 + K_2 = K$ 。我们可以将空间因子模型（4）写为

$$\tilde{r} = \underline{\alpha} + \mathbf{B}_1 \tilde{z}_1 + \mathbf{B}_2 \mathbf{f}_2 + \Psi W \tilde{r} + \epsilon, \quad (10)$$

其中 \mathbf{B}_1 和 \mathbf{B}_2 分别是系数矩阵 \mathbf{B} 的前 K_1 列和后 K_2 列， \tilde{z}_1 是可交易因子的超额收益。可以证明，在无渐近套利假定下，存在 λ_2 ，使得 $\underline{\alpha} \approx \mathbf{B}_2 \lambda_2$ ，这里 λ_2 对应的是不可交易因子的风险溢价。此外，若 $\mathbf{B}_2 \neq \mathbf{0}$ ，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：

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

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

$$W = T \frac{(\hat{\alpha}^*)' \hat{\Sigma}_{\epsilon^*}^{-1} \hat{\alpha}^*}{1 + \hat{\mu}'_z \hat{\Sigma}_z^{-1} \hat{\mu}_z}, \quad (12)$$

其中

$$\hat{\alpha}^* = (I_n - \hat{\Psi}W)^{-1} \hat{\alpha}, \quad \hat{\Sigma}_{\epsilon^*}^{-1} = (I_n - \hat{\Psi}W)' \hat{\Omega}^{-1} (I_n - \hat{\Psi}W), \quad \hat{\mu}_z = \frac{1}{T} \sum_{t=1}^T \mathbf{z}_t,$$

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：

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

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 λ_2 ，使得 $\alpha = \mathbf{B}_2 \lambda_2$ ”，我们将模型改写为

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

其中 $\alpha^* = G\alpha = (I_n - \Psi W)^{-1} \alpha$ ， $\mathbf{B}_1^* = G\mathbf{B}_1$ ， $\mathbf{B}_2^* = G\mathbf{B}_2$ ， $\epsilon_t^* = G\epsilon_t$ 。则原假设等价于“存在 λ_2 ，使得 $\alpha^* = \mathbf{B}_2^* \lambda_2$ ”，定义

$$\theta_s = ((\alpha^*)', \text{vec}(\mathbf{B}_2^*))', \quad \mathbf{g}(\theta_s) = M_{\mathbf{B}_2^*} \alpha^* = (I_n - \mathbf{B}_2^* ((\mathbf{B}_2^*)' \mathbf{B}_2^*)^{-1} (\mathbf{B}_2^*)') \alpha^*, \quad (15)$$

在原假设下，有 $\sqrt{T} \mathbf{g}(\hat{\theta}_s) \xrightarrow{d} N(0, \Omega_g)$ ，渐近协方差矩阵由 $\Omega_g = \mathbf{J}(\theta_s) V_{\theta_s} \mathbf{J}(\theta_s)'$ 给出，其中 V_{θ_s} 是 $\hat{\theta}^*$ 的渐近协方差矩阵 V_{θ^*} 中 θ_s 对应的部分，

$$\mathbf{J}(\theta_s) = \left(M_{\mathbf{B}_2^*}, - \left((\alpha^*)' \mathbf{B}_2^* ((\mathbf{B}_2^*)' \mathbf{B}_2^*)^{-1} \otimes M_{\mathbf{B}_2^*} \right) \right) + \left(M_{\mathbf{B}_2^*}, - \left((\alpha^*)' M_{\mathbf{B}_2^*} \otimes \mathbf{B}_2^* ((\mathbf{B}_2^*)' \mathbf{B}_2^*)^{-1} \right) \mathcal{K}_{n, K_2} \right)$$

则是对应的雅可比矩阵， \mathcal{K}_{n, K_2} 是 $nK_2 \times nK_2$ 转换矩阵 (Commutator Matrix)。我们可以用 $\hat{\Omega}_g = \mathbf{J}(\hat{\theta}_s) \hat{V}_{\theta_s} \mathbf{J}(\hat{\theta}_s)'$ 作为 Ω_g 的估计，但由于雅可比矩阵 \mathbf{J} 的存在， $\hat{\Omega}_g$ 不可逆。因为 $\hat{\Omega}_g$ 是半正定矩阵，一定存在正交矩阵 \hat{Q} 使得 $\hat{\Omega}_g = \hat{Q} \hat{\Lambda} \hat{Q}'$ ，其中 $\hat{\Lambda} = \text{diag}\{\hat{\lambda}_1, \dots, \hat{\lambda}_n\}$ 是 $\hat{\Omega}_g$ 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 $\tilde{\Lambda}^+ = \text{diag}\{\tilde{\lambda}_1^{-1}, \dots, \tilde{\lambda}_{n-K_2}^{-1}, 0, \dots, 0\}$ 以及

$$\hat{\Omega}_g^+ = \hat{Q} \tilde{\Lambda}^+ \hat{Q}', \quad (16)$$

这满足广义 Wald 检验的条件 (Andrews, 1987 [36])，因此

$$W = T \mathbf{g}(\hat{\theta}_s)' \hat{\Omega}_g^+ \mathbf{g}(\hat{\theta}_s) \xrightarrow{d} \chi^2(n - K_2), \quad (17)$$

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 越大意味着我们越有信心拒绝原假设。

在空间因子定价模型的基础上，我们介绍了如何检验新挖掘的可交易与不可交易因子在无渐进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。

四、实证分析

(一) 数据说明

为验证新闻信息辅助的异质空间因子模型在我国市场上的有效性,我们收集了沪深 300 成分股 2017-2021 年的日收益率数据,剔除其中缺失数据超过 100 日的股票,剩余有效资产数量为 225 支,缺失数据(未超过 100 日)用最邻近法补齐。此外,我们收集了主要行业因子的 2012-2021 年间日收益率,数据的描述性统计见表 2,收益率数据来源于 RiceQuant 平台。

Table 2: 数据的描述性统计

变量名称	均值	标准差	最小值	最大值	观测值
股票收益率	0.145%	2.834%	-19.996%	44.070%	225×1217
市场因子	0.028%	1.157%	-8.090%	5.530%	1217
价值因子	0.051%	0.673%	-2.540%	3.060%	1217
规模因子	-0.027%	0.723%	-3.630%	3.090%	1217
情绪因子	0.020%	0.640%	-2.560%	3.200%	1217
中证工业指数	0.019%	1.198%	-7.128%	5.205%	1217
中证金融地产	0.012%	1.297%	-8.085%	8.929%	1217
中证可选消费	0.010%	1.303%	-9.088%	4.855%	1217
中证能源指数	0.007%	1.529%	-8.277%	6.178%	1217

北京大学金融工程实验室汇集了数百万条从 1999 年 9 月 22 日开始的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 2012 年开始的新闻数据基本信息。可以看出,2012 年新闻数据较为稀疏,日均只有 84.29 条,但近十年(2013 往后)的数据都非常密集,2021 年日均新闻数据更是在 521.84 条,峰值达到 2590 条。同时,我们发现新闻数据在每年的分布并不均匀,比起均值,标准差(方差)非常之大。此外,附录汇总了样本股各年提及次数及关联情况的相关信息。

Table 3: 新闻数据汇总

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
日均条数	84.29	243.08	336.04	236.23	197.29	126.35	156.37	136.68	334.87	521.84	117.62
方差	2720.91	34486.47	45193.22	25200.87	17296.58	7977.94	11090.13	8939.84	98400.56	153748.28	250181.00
最大	183	657	699	589	493	359	467	464	1685	2590	1818
最小	1	1	1	1	1	1	1	1	1	1	9

¹ 目前的 2022 年数据截止于 2022 年 10 月 31 日。

大量的新闻数据潜藏了资产之间的网络结构。利用样本中第 s 年的新闻文本数据，我们构造新闻共现矩阵 $C_s = (c_{ij}^{(s)})$ ，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 C_1, C_2, \dots, C_M 提供的信息，我们可以合理地构造一个新闻关联矩阵 W 。由于动态异质空间因子模型

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\Psi}W_t\mathbf{r}_t + \boldsymbol{\epsilon}_t \quad (18)$$

的统计推断与假设检验暂缺理论支持，因此实证中只能同时综合过去 M 年的新闻共现矩阵，²令

$$\widetilde{W} = (\tilde{w}_{ij}) = \sum_{s=1}^M C_s, \quad (19)$$

再令 $W = (w_{ij}) = \left(\tilde{w}_{ij} / \sum_{k=1}^n \tilde{w}_{ik} \right)$ ，即归一化以满足权重矩阵的要求。理论上来说，为充分挖掘资产之间的空间关联程度，新闻关联矩阵的时期应比收益率时期更长。因此在本文中，我们利用 2012-2021 年总计十年的新闻数据构建了沪深 300 成分股的关联矩阵 W ，其中 $\text{rank}(W) = 216$ ，即仅有 9 支个股十年来从未与其他样本股在新闻中共现。图 1 选取 225 支样本股中的拥有最高频次新闻数据的 25 支股票并汇报其在 2012-2021 年的共现次数。可以看出，图 1 中涵盖的大多数为金融股，这符合认知，金融股确实会在新闻中与其他股票同时出现。其中共现次数最多是中信证券和华泰证券，共计 1002 次。除此以外，还有一些有显著关联的企业共现次数较多，如：五粮液和贵州茅台，共现 664 次，相互是其关联股中的第一，它们同属白酒行业；五粮液和国泰君安，共现 66 次，是五粮液关联股中的第三，而五粮液于 1998 年 4 月 27 日上市，承销商为国泰君安前身的君安证券。

(二) 实证结果：空间因子模型

Liu et al. (2019) [21] 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。本文将以其因子模型为基准，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：

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 z_{\text{MRT},t} + \boldsymbol{\Psi}W\mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (20)$$

- (模型 2) 空间 CH3 模型：

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 z_{\text{MRT},t} + \boldsymbol{\beta}_2 z_{\text{VMG},t} + \boldsymbol{\beta}_3 z_{\text{SMB},t} + \boldsymbol{\Psi}W\mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (21)$$

- (模型 3) 空间 CH4 模型：

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 z_{\text{MRT},t} + \boldsymbol{\beta}_2 z_{\text{VMG},t} + \boldsymbol{\beta}_3 z_{\text{SMB},t} + \boldsymbol{\beta}_4 z_{\text{PMO},t} + \boldsymbol{\Psi}W\mathbf{r}_t + \boldsymbol{\epsilon}_t. \quad (22)$$

²基于原始新闻共现矩阵构造新闻关联矩阵 W 的方法有多种，如取对数处理、阈值筛选等，皆对估计结果影响不大。

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

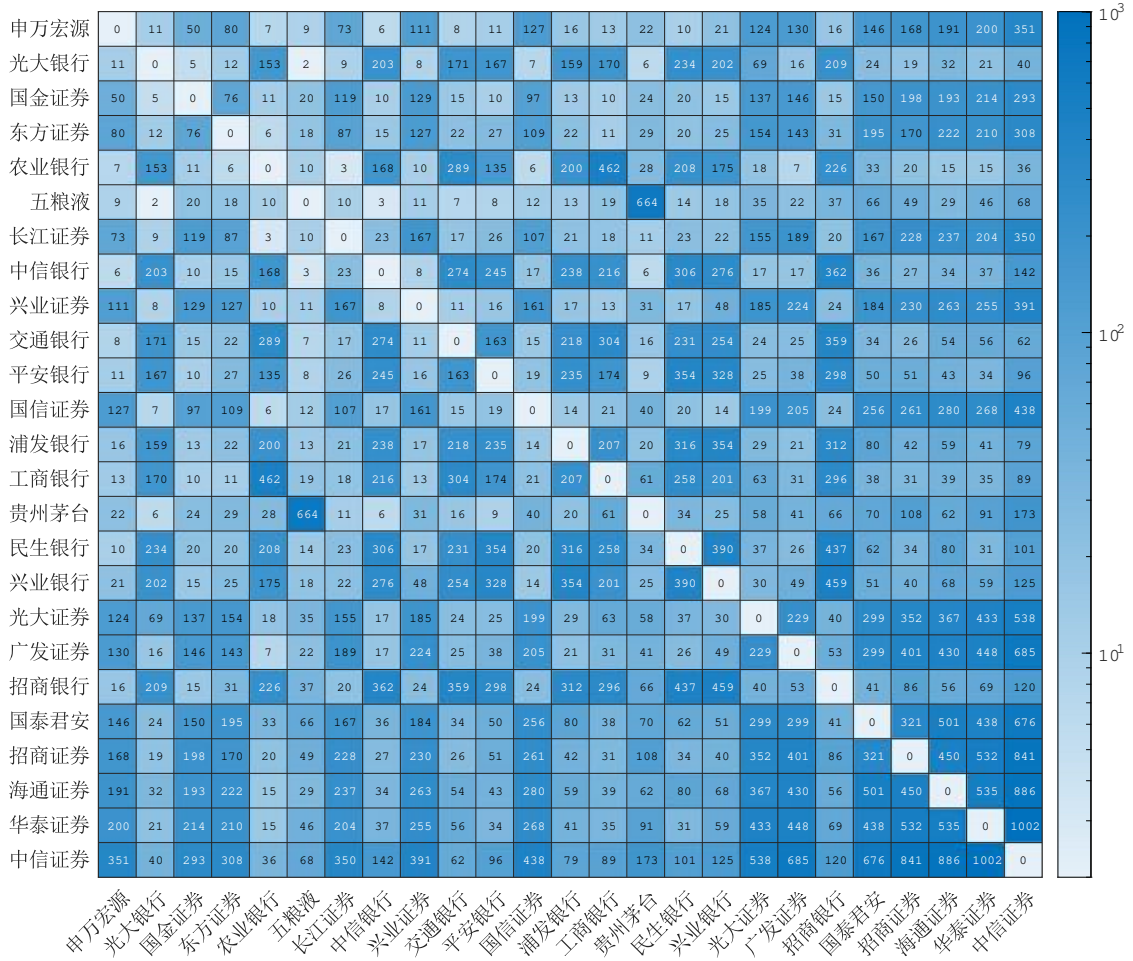


Figure 1: 新闻共现数据热力图

除此以外，我们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：

$$\begin{aligned}
 \mathbf{r}_t = & \boldsymbol{\alpha} + \beta_1 z_{MRT,t} + \beta_2 z_{VMG,t} + \beta_3 z_{SMB,t} + \beta_4 z_{PMO,t} \\
 & + \beta_5 f_{SHI,t} + \beta_6 f_{CPI,t} + \beta_7 f_{PPI,t} + \beta_8 f_{M,t} + \Psi W \mathbf{r}_t + \boldsymbol{\epsilon}_t.
 \end{aligned} \tag{23}$$

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 ），均为金融中常用的宏观变量。很可惜的是，在中国市场上，除了 SHIBOR 具有日度数据外，其余变量均只有月度数据，我们将月数据填充为日数据、去均值并缩放到标准差为 0.0125（相当于年化波动率 20%），在单位根检验通过的前提下，将变量加入估计方程。

我们的异质性空间因子模型有 $n(K+3)$ 个参数，其中 K 为因子的数目。因为参数量庞大，我们在表 4 汇总了传统因子模型和空间因子模型平均参数估计值（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。我们先看三个不含宏观因子的模型（式 20-22），在空间 CAPM 模型中，82.7% 的空间系数显著，在空间 CH3 和空间 CH4 模型中，也有 77.8% 和 77.3% 的空间系数显著，这说明沪深 300 成分股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，平均市场因子载荷略高于 1，在 CAPM 模型中是 1.0396，在 CH3 和 CH4 模型中则分

Table 4: 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	0.0011	1.0396									0.0229
std.	(0.0030)	(0.0029)									(0.0000)
% Sig	28.0%	100.0%									100.0%
● 空间 CAPM											
MG	0.0008	0.3032								0.1825	0.0217
std.	(0.0066)	(0.0060)								(0.0021)	(0.0000)
% Sig	20.0%	71.1%								82.7%	100.0%
● CH3 模型											
MG	0.0010	1.0501	0.0227	-0.1168							0.0222
std.	(0.0080)	(0.0067)	(0.0082)	(0.0058)							(0.0000)
% Sig	40.0%	100.0%	81.8%	70.7%							100.0%
● 空间 CH3 模型											
MG	0.0009	0.2910	-0.1622	-0.0303						0.1792	0.0212
std.	(0.0093)	(0.0076)	(0.0094)	(0.0068)						(0.0022)	(0.0000)
% Sig	29.8%	65.8%	80.4%	65.3%						77.8%	100.0%
● CH4 模型											
MG	0.0010	1.0531	0.0180	-0.1200	0.0237						0.0222
std.	(0.0080)	(0.0065)	(0.0075)	(0.0074)	(0.0055)						(0.0000)
% Sig	40.9%	100.0%	80.9%	69.3%	44.0%						100.0%
● 空间 CH4 模型											
MG	0.0009	0.2856	-0.1532	-0.0244	-0.0445					0.1793	0.0212
std.	(0.0093)	(0.0075)	(0.0085)	(0.0085)	(0.0065)					(0.0023)	(0.0000)
% Sig	28.4%	66.2%	78.2%	65.3%	33.8%					77.3%	100.0%
● CH4 + 宏观因子											
MG	0.0010	1.0539	0.0180	-0.1217	0.0224	0.0123	0.0016	0.0151	0.0014		0.0221
std.	(0.0074)	(0.0065)	(0.0058)	(0.0057)	(0.0075)	(0.0053)	(0.0074)	(0.0048)	(0.0049)		(0.0000)
% Sig	41.3%	100.0%	81.8%	69.8%	46.7%	8.4%	12.4%	11.6%	7.1%		100.0%
● 空间 CH4 + 宏观因子											
MG	0.0009	0.2863	-0.1527	-0.0262	-0.0454	0.0175	0.0010	0.0135	0.0061	0.1794	0.0211
std.	(0.0085)	(0.0069)	(0.0064)	(0.0061)	(0.0081)	(0.0058)	(0.0081)	(0.0052)	(0.0058)	(0.0022)	(0.0000)
% Sig	29.3%	68.0%	78.7%	65.3%	34.2%	8.4%	11.6%	14.2%	7.1%	78.2%	100.0%

¹ 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。

别是 1.0501 和 1.0531, 显著比例均为 100.0%, 但加入空间因子模型之后, 平均因子载荷降低到 30% 左右及以下, 显著比例降低到了 70% 左右。在 CH3 模型中, 价值因子的因子载荷从 0.0227 变化为-0.1622, 规模因子的因子载荷则从-0.1168 变化为-0.0303, 但显著比例几乎无区别。CH4 模型中, 情绪因子的因子载荷 β_4 从 0.0237 变为-0.0445, 显著比例也有所降低。

在加入空间项后, 三个传统因子模型的异常收益 α 均得到了或多或少的解释。例如在 CAPM 模型中, 尚有 28.0% 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 20.0%。在 CH3 和 CH4

模型中，均原有 40.0% 以上的 α 显著，加入空间项使得显著的 α 降低到 30% 以下。但值得注意的是，加入规模、价值和情绪因子似乎并不有助于解释异常收益，在传统因子模型中，最简单 CAPM 模型里显著的 α 更少，在增加因子得到的 CH3 和 CH4 模型里，具有显著非零异常收益的个股反而增多。

最后，我们发现在加入宏观因子后，宏观因子本身的显著程度一般，因此整体的估计结果与 CH4 模型回归并无明显差异， ψ 显著比例仍在 80% 左右， ψ 平均值也仍然是 0.18 左右。从解释异常收益的角度来看，加入宏观因子的 CH4 模型尚且无法解释 41.3% 个股的显著异常收益，而加入空间项后，异常收益显著的公司只剩下 29.3%，提升较大。

如刘霖和秦宛顺（2004）[37]、杨子暉和李东承（2018）[38] 所言，中国的资本市场并不有效，且存在套利机会，异常收益普遍显著，新闻矩阵虽然能够解释部分因子模型无法解释的异常收益，但仍任重道远。

Table 5: 异质性检验

	异质空间系数			同质空间系数			渐近 LRT 检验		
	α	β_1	ψ	α	β_1	ψ	Z	p	
● 空间 CAPM 模型									
MG	0.0008	0.3032	0.1825	MG	0.0009	0.3041	0.2786		
std.	(0.0066)	(0.0060)	(0.0021)	std.	(0.0068)	(0.0064)	(0.0023)	422.47	0.0000
% Sig	20.0%	71.1%	82.7%	% Sig	29.3%	55.6%			
● 空间 CH3 模型									
MG	0.0009	0.2910	0.1792	MG	0.0009	0.3081	0.1848		
std.	(0.0093)	(0.0076)	(0.0022)	std.	(0.0095)	(0.0078)	(0.0023)	348.43	0.0000
% Sig	29.8%	65.8%	77.8%	% Sig	28.0%	59.1%			
● 空间 CH4 模型									
MG	0.0009	0.2856	0.1793	MG	0.0009	0.3032	0.1845		
std.	(0.0093)	(0.0075)	(0.0023)	std.	(0.0095)	(0.0077)	(0.0023)	347.72	0.0000
% Sig	28.4%	66.2%	77.3%	% Sig	28.9%	55.6%			
● 空间 CH4 + 宏观因子模型									
MG	0.0009	0.2863	0.1794	MG	0.0009	0.3041	0.1845		
std.	(0.0085)	(0.0069)	(0.0022)	std.	(0.0086)	(0.0071)	(0.0023)	347.07	0.0000
% Sig	29.3%	68.0%	78.2%	% Sig	29.3%	55.6%			

¹ 同质空间因子模型的 ψ 仅有一个估计值，故无显著比例。

在四个空间回归模型中，平均 ψ 大致都在 0.18 左右，但实际 225 支股中 ψ 的差异明显。如在空间 CH4 模型中， ψ 的最大值为 0.779，最小值为 -0.995，标准差为 0.176。表 ?? 直观地展示了运用同质和异质空间因子模型所带来的结果差异。尤其在空间 CAPM 模型中，同质模型的 ψ 估计值是 0.2786，与市场平均值相去甚远。渐近 LRT 的 Z 统计量也相当显著（见表 5），足以说明使用异质空间因子模型的合理性。在空间 CH3、CH4 和加入宏观因子的 CH4 模型中，同质模型的 ψ 估计值虽然与异质模型的市场平

均差距不大，但同质模型显然无法描绘出个股之间的异质性。

(三) 稳健性检验：行业因子

一般来说，同行业内的个股相关性更强，空间依赖性也更强。为了避免新闻关联矩阵捕捉的只是行业内的相关性，我们应在空间回归模型中加入行业因子，即：

- (模型 5) 空间 CAPM + 行业模型：

$$r_t = \alpha + \beta_1 z_{MRT,t} + \beta_I z_{IND,t} + \Psi W r_t + \epsilon_t. \quad (24)$$

- (模型 6) 空间 CH3 + 行业模型：

$$r_t = \alpha + \beta_1 z_{MRT,t} + \beta_2 z_{VMG,t} + \beta_3 z_{SMB,t} + \beta_I z_{IND,t} + \Psi W r_t + \epsilon_t. \quad (25)$$

- (模型 7) 空间 CH4 + 行业模型：

$$r_t = \alpha + \beta_1 z_{MRT,t} + \beta_2 z_{VMG,t} + \beta_3 z_{SMB,t} + \beta_4 z_{PMO,t} + \beta_I z_{IND,t} + \Psi W r_t + \epsilon_t. \quad (26)$$

- (模型 8) 空间 CH4 + 宏观因子 + 行业模型：

$$r_t = \alpha + \beta_1 z_{MRT,t} + \beta_2 z_{VMG,t} + \beta_3 z_{SMB,t} + \beta_4 z_{PMO,t} + \beta_5 f_{SHL,t} + \beta_6 f_{CPI,t} + \beta_7 f_{PPI,t} + \beta_8 f_{M,t} + \beta_I z_{IND,t} + \Psi W r_t + \epsilon_t. \quad (27)$$

Table 6: 样本股分类汇总

序号	分类	分类包含	数量	基准
1	制造业	C	121	中证全指工业指数 000988
2	金融地产	J E K	55	中证全指金融地产指数 000992
3	生活、文体和运输	F G I L M P Q R	31	中证全指可选消费指数 000989
4	能源资源类	A B D	18	中证全指能源指数 000986

根据中国证券监督管理委员会 (CSRC) 公布的上市公司行业分类结果，可以将所有上市公司分为 19 类，而我们的 225 支样本股在分类后，制造业占据了半数 (121 支)，金融业排名第二，有 41 支，但其余行业均无超过 10 支股票。因此我们合理地合并一些分类 (分类字母对应名称可参见附录表 10)，并选取合适的基准行业因子，合并后的行业分类如表 6 所示。需要注意的是，如果行业分类非常的细 (即每个行业中的股票数非常少)，那么行业因子很可能对于某些大公司的股票而言是内生的，因此我们尽量选择较大的分类以避免这个问题。

Table 7: 行业因子回归结果

	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	β_I	ψ	σ^2
● CAPM + 行业												
MG	0.0011	0.7133								0.3589		0.0229
std.	(0.0026)	(0.0025)								(0.0022)		(0.0000)
% Sig	30.2%	94.2%								68.4%		100.0%
● 空间 CAPM + 行业												
MG	0.0008	0.1322								0.2619	0.1655	0.0214
std.	(0.0058)	(0.0056)								(0.0047)	(0.0022)	(0.0000)
% Sig	20.0%	67.1%								59.1%	82.2%	100.0%
● CH3 模型												
MG	0.0012	0.6993	-0.1591	-0.1207						0.3443		0.0217
std.	(0.0075)	(0.0060)	(0.0070)	(0.0069)						(0.0051)		(0.0000)
% Sig	37.8%	80.4%	80.9%	68.4%						73.3%		100.0%
● 空间 CH3 模型												
MG	0.0009	0.1447	-0.2174	-0.0489						0.2209	0.1622	0.0210
std.	(0.0087)	(0.0070)	(0.0080)	(0.0078)						(0.0058)	(0.0023)	(0.0000)
% Sig	28.9%	64.9%	80.4%	63.1%						62.7%	76.4%	100.0%
● CH4 模型												
MG	0.0012	0.7013	-0.1582	-0.1200	-0.0050					0.3414		0.0217
std.	(0.0076)	(0.0062)	(0.0062)	(0.0069)	(0.0069)					(0.0052)		(0.0000)
% Sig	38.7%	81.8%	80.4%	68.4%	30.2%					72.4%		100.0%
● 空间 CH4 模型												
MG	0.0010	0.1411	-0.2088	-0.0419	-0.0452					0.2176	0.1624	0.0210
std.	(0.0088)	(0.0071)	(0.0069)	(0.0079)	(0.0078)					(0.0060)	(0.0023)	(0.0000)
% Sig	28.9%	64.4%	79.6%	61.8%	30.7%					62.2%	76.9%	100.0%
● CH4 + 宏观因子												
MG	0.0012	0.7024	-0.1584	-0.1215	-0.0063	0.0122	0.0016	0.0142	0.0020	0.3410		0.0216
std.	(0.0073)	(0.0063)	(0.0056)	(0.0054)	(0.0058)	(0.0070)	(0.0056)	(0.0068)	(0.0048)	(0.0046)		(0.0000)
% Sig	39.1%	81.3%	80.4%	68.4%	32.0%	8.9%	9.8%	11.1%	7.1%	71.6%		100.0%
● 空间 CH4 + 宏观因子												
MG	0.0010	0.1420	-0.2082	-0.0436	-0.0462	0.0171	0.0027	0.0140	0.0051	0.2173	0.1625	0.0209
std.	(0.0083)	(0.0066)	(0.0062)	(0.0058)	(0.0062)	(0.0076)	(0.0060)	(0.0075)	(0.0052)	(0.0051)	(0.0023)	(0.0000)
% Sig	28.9%	64.4%	79.1%	61.8%	31.6%	8.9%	11.1%	13.3%	7.1%	61.3%	77.8%	100.0%

¹ 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。

表 7 汇总了加入行业因子后空间因子模型的回归结果。我们首先可以看出在四个传统因子模型加入行业因子后, 市场因子的显著性会降低, 样本股平均市场因子载荷从 1 左右降低到 0.7 左右, 原因是行业因子解释了原先由市场因子解释的部分收益率。在空间因子模型中, 加入行业因子后空间系数仍然非常显著, 四个模型中显著的 ψ 的占比分别为 82.2%、76.4%、76.9% 和 77.8%, 但系数估计平均值有所下降, 从原先的 0.179-0.183 降低到了 0.162-0.166。因子模型回归结果中 β_I 平均在 0.35 左右, 显著比例在 70% 左右, 而四个空间因子模型中 β_I 在 0.22 左右 (除了空间 CAPM 中 $\beta_I = 0.2619$), 这意味着新闻矩

阵确实捕捉了一部分行业内相关性，与行业因子有重合作用，但在剔除行业内相关性的影响后，空间系数仍然非常显著。

从异常收益 α 的角度看，CAPM + 行业模型中显著的 α 比例有 30.2%，反而不如原始的 CAPM 模型，随后在加入空间项后，其显著比例下降至 20.0%，可以认为考虑行业因子后，新闻矩阵仍然有效地解释了部分异常收益。空间 CH3 + 行业和空间 CH4 + 行业模型中， α 显著比例分别在 37.8% 和 38.7%，加入新闻矩阵后，显著比例均下降到 28.9%。在空间 CH4+ 宏观因子 + 行业模型中，异常收益非零的比例有 39.1%，新闻矩阵的加入使得其下降到了 28.9%。

Table 8: 不同行业平均估计值

	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	β_I	ψ	σ^2
● 制造业 ($n_I = 121$)												
MG	0.0015	0.3033	-0.3426	-0.1064	-0.0803	0.0138	0.0071	0.0277	0.0107	0.0594	0.1750	0.0249
std.	(0.0007)	(0.1807)	(0.1313)	(0.1136)	(0.1202)	(0.0749)	(0.0629)	(0.0667)	(0.0717)	(0.1094)	(0.0361)	(0.0005)
% Sig	33.1%	66.9%	81.8%	59.5%	33.1%	8.3%	12.4%	14.0%	5.8%	55.4%	72.7%	100.0%
● 金融地产 ($n_I = 55$)												
MG	0.0000	-0.0715	0.0711	0.0764	-0.0027	0.0060	0.0057	0.0060	0.0255	0.4504	0.1525	0.0130
std.	(0.0004)	(0.0837)	(0.0821)	(0.0581)	(0.0638)	(0.0392)	(0.0329)	(0.0349)	(0.0375)	(0.0944)	(0.0195)	(0.0003)
% Sig	21.8%	54.5%	87.3%	63.6%	36.4%	3.6%	9.1%	14.5%	5.5%	78.2%	87.3%	100.0%
● 生活、文体和运输 ($n_I = 31$)												
MG	0.0009	0.1462	-0.2561	-0.0681	0.0156	0.0588	-0.0095	-0.0234	-0.0273	0.3267	0.1213	0.0213
std.	(0.0006)	(0.1967)	(0.1122)	(0.0959)	(0.1037)	(0.0642)	(0.0538)	(0.0573)	(0.0614)	(0.1305)	(0.0315)	(0.0004)
% Sig	29.0%	64.5%	61.3%	54.8%	12.9%	16.1%	12.9%	9.7%	9.7%	54.8%	80.6%	100.0%
● 资源能源类 ($n_I = 18$)												
MG	0.0006	-0.2964	-0.0759	0.0538	-0.0561	0.0008	-0.0144	0.0114	-0.0390	0.3783	0.1803	0.0173
std.	(0.0005)	(0.1120)	(0.1049)	(0.0836)	(0.0861)	(0.0521)	(0.0438)	(0.0464)	(0.0499)	(0.0526)	(0.0243)	(0.0004)
% Sig	22.2%	77.8%	66.7%	83.3%	38.9%	16.7%	5.6%	11.1%	16.7%	61.1%	77.8%	100.0%

¹ MG 指行业内的平均估计值，std. 则指其标准误差，% Sig. 表示行业内估计值在 5% 显著性水平下显著的比例。

此外，我们也计算了每个行业内的平均估计值（见表 8），并发现行业中呈现出了一些异质性。资源能源类股票和制造业股票对新闻相关性的敏感度较高，平均 ψ 分别为 0.1803 和 0.1750，但金融地产股的 ψ 显著比例更高，占到 87.3%，但四个行业分类中空间系数显著比例均在 70% 以上。再考虑因子部分，在两个主要因子（市场因子 + 行业因子）上，除了资源能源类以外，其余三个行业分类整体来看都与市场其他股或行业内其他股呈现比较显著的正相关，但是行业因子的加入同时也稀释了市场因子对收益率的解释能力。比较特殊的是资源能源类股票，它的行业因子载荷是 $\beta_I = 0.1803$ ，和其余三者相仿，但此时其则市场负平均因子载荷是 -0.2964，呈现负相关，显著比例 77.8%。我们同时也在加入行业因子后对异质性进行了检验，Z 统计量为 365.20，仍然非常显著。

在行业层面，与市场关联度较低的金融地产股的平均 β_I 为 0.4504，显著比例也相当高，为 87.3%，另外三个行业的显著比例都在 60% 左右或略微低于 60%。值得注意的是制造业股票，制造业股票与市场

关联度高，而其行业 β_I 接近于 0。从异常收益看，空间因子模型在金融地产、资源能源类股票上的解释程度较高，仅有 21.8% 和 22.2% 的比例显著，但在生活、制造业股上，其显著比例达到 29.0% 和 33.1%，略显不足。

(四) 无套利检验

考虑模型 1 - 模型 4，利用前文探讨的 Wald 统计量进行对应的无套利检验，检验结果见表 9。

Table 9: 无套利 Wald 检验

	标准化 Wald	p 值
因子模型		
• CAPM	36.61	0.0000
• CH3 模型	49.96	0.0000
• CH4 模型	51.00	0.0000
• CH4 模型 + 宏观因子	35.18	0.0000
空间因子模型		
• 空间 CAPM	33.37	0.0000
• 空间 CH3 模型	44.06	0.0000
• 空间 CH4 模型	44.88	0.0000
• 空间 CH4 模型 + 宏观因子	34.60	0.0000

可以看出，仅是增加可交易因子无益于降低定价误差。如表所示，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。

(五) 小结

横截面弱相关性是资产定价中不可忽视的部分，而新闻共现数据能够帮助我们挖掘资产间的弱相关性。实证表明：无论考虑 CAPM 模型、多因子模型、宏观因子模型，还是加入行业因子，回归结果中的空间系数项均能达到 75% 以上显著（最低的是空间 CH3+ 行业因子模型，76.4%），且比起原因子模型，异常收益显著的比例均大幅度下降，不仅印证了新闻矩阵的有效性，更说明了局部相关性能解释因子模型无法解释的部分。此外，异质性在中国市场中显然存在，无论是行业层面，还是个股层面，空间系数的估计值均有明显差异。

但由于中国市场的特殊性，很难找到一个完全贴切的定价模型，因此所有回归模型中显著的 α 均在 20% 及以上，即有五分之一以上的个股收益率显著偏离估计方程。从 Wald 检验结果也可以看出，虽然新闻矩阵的加入有效地降低了标准化 Wald 统计量的值，但是 p 值一直处于 0.0000，显著程度非常高。

最后，在加入行业因子后，我们发现行业因子和空间项有一定程度重合。传统因子模型中，行业因子载荷平均在 0.35 左右，但在空间因子模型中降低为 0.22 左右。同样，在未加入行业因子时，平均 ψ 在 0.18 左右，而加入行业因子后，估计值降低到 0.16 左右，但仍然非常显著。这些情况说明了新闻矩阵在一定程度上捕捉了行业关联，因而空间项的加入稀释了行业因子的影响。但基于新闻共现的空间矩阵能够较为全面地捕捉资产间的各种关联，所以在控制了行业因子的影响后，空间效应依然十分显著，说明了由空间项刻画的弱（局部）依赖性是不可忽视的定价因素。

五、结论与展望

本文将资产之间的横截面相关性 (cross-sectional dependence) 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，见微知著，通过异质空间因子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，且在剔除行业因子的影响后，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。检验结果表明，加入宏观因子有助于解释无套利，新闻矩阵也是解释异常收益的一大因素，但整体来看， α 的显著程度仍然非常高。

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，即新闻矩阵 W 是固定的而非动态的 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施行起来是有难度的，一方面很难确定新闻矩阵所用数据时期与收益率时期的匹配关系，另一方面则是似然函数是时变的，要在 $T \rightarrow \infty$ 给出合适的渐近理论就必须要求 W_t 满足一些平稳性质，如：存在常数 c ，使得当 T 趋于无穷时，有

$$\{(I_n - \Psi W_T)^{-1}\}_{ii} (W_T)_{ii} \rightarrow c_i, \quad i = 1, 2, \dots, n,$$

其中 A_{ii} 指的是矩阵 A 的 (i, i) 元，但在真实新闻数据中这可能又是十分苛刻且难以验证的。不过无论如何，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，是未来的研究话题之一。

当然，整体来看，目前还尚难在我国金融市场上找到一个完全贴切的资产定价模型，我们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，但异常收益显著的个股仍然在五分之一以上。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

- [1] Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The journal of finance, 19(3), 425-442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>

-
- [2] Ross, S. 1976 The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341–360. https://doi.org/10.1142/9789814417358_0001
- [3] Gabaix, X. 2011. The granular origins of aggregate fluctuations. *Econometrica*, 79(3), 733-772. <https://doi.org/10.3982/ECTA8769>
- [4] Kou, S., Peng, X., and Zhong, H. 2018. Asset pricing with spatial interaction. *Management Science*, 64(5), 2083-2101. <https://doi.org/10.1287/mnsc.2016.2627>
- [5] 张玉华、宋韞贇和张元庆, 2016, 《基于空间面板数据模型的股票收益率影响因素分析》, 《中国软科学》第5期, 第172-183页。
- [6] 张自力、闫红蕾和张楠, 2020, 《股票网络、系统性风险与股票定价》, 《经济学(季刊)》第1期, 第329-350页。
- [7] 石泽龙、傅强和李山, 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》, 《系统管理学报》第3期, 第453-466页。
- [8] Chudik, A., Pesaran, M. H., and Tosetti, E. 2011. Weak and strong crosssection dependence and estimation of large panels. *The Econometrics Journal*, 14(1), C45-C90. <https://doi.org/10.1111/j.1368-423X.2010.00330.x>
- [9] Bailey, N., Holly, S., and Pesaran, M. H. 2016. A twostage approach to spatiotemporal analysis with strong and weak crosssectional dependence. *Journal of Applied Econometrics*, 31(1), 249-280. <https://doi.org/10.1002/jae.2468>
- [10] Zareei, A. 2019. Network origins of portfolio risk. *Journal of Banking & Finance*, 109, 105663 <https://doi.org/10.1016/j.jbankfin.2019.105663>.
- [11] Giglio, S., and Xiu, D. C. 2021. Asset Pricing with Omitted Factors. *Journal of Political Economy*, 129(7), 1947-1990. <https://doi.org/10.1086/714090>
- [12] 潘荣翠、张鑫、韩跃红和尚朝秋, 2012, 《基于空间计量分析的全球股票价格指数宏观影响因素》, 《云南师范大学学报(哲学社会科学版)》第1期, 第135-144页。
- [13] Fama, E. F., and French, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- [14] Scherbina, A., and Schlusche, B. 2015. Economic linkages inferred from news stories and the predictability of stock returns. Available at SSRN 2363436. <http://dx.doi.org/10.2139/ssrn.2363436>

- [15] Schwenkler, G., and Zheng, H. 2019. The network of firms implied by the news. Boston University Questrom School of Business Research Paper, (3320859). <http://dx.doi.org/10.2139/ssrn.3320859>
- [16] 范龙振和余世典, 2002, 《中国股票市场的三因子模型》, 《系统工程学报》第6期, 第537-546页。
- [17] 吴世农和许年行, 2004, 《资产的理性定价模型和非理性定价模型的比较研究——基于中国股市的实证分析》, 《经济研究》第6期, 第105-116页。
- [18] 廖理和沈红波, 2008, 《Fama-French 三因子模型与股权分置改革效应研究》, 《数量经济技术经济研究》第9期, 第117-125页。
- [19] 田利辉、王冠英和张伟, 2014, 《三因素模型定价: 中国与美国有何不同?》, 《国际金融研究》第7期, 第37-45页。
- [20] 李斌和雷印如, 2022, 《中国公募基金挖掘了股票市场异象吗?》, 《金融研究》第9期, 第188-206页。
- [21] Liu, J., Stambaugh, R. F. and Yuan, Y. 2019. Size and value in China. *Journal of financial economics*, 134(1), 48-69. <https://doi.org/10.1016/j.jfineco.2019.03.008>
- [22] 刘维奇和张燕, 2020, 《资产定价与劳动成本占比》, 《中国管理科学》第12期, 第111页。doi: [10.16381/j.cnki.issn1003-207x.2020.0043](https://doi.org/10.16381/j.cnki.issn1003-207x.2020.0043).
- [23] 段丙蕾, 2022, 《南橘北枳: A股市场的经济关联与股票回报》, 《金融研究》第2期, 第171-188页。
- [24] Fernandez V. 2011. Spatial linkages in international financial markets. *Quantitative Finance*. 11(2): 237-245. <https://doi.org/10.1080/14697680903127403>
- [25] Aquaro, M., Bailey, N. and Pesaran, M. H. 2015. Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. USC-INET Research Paper, (15-17). <http://dx.doi.org/10.2139/ssrn.2623192>
- [26] Aquaro, M., Bailey, N. and Pesaran, M. H. 2021. Estimation and inference for spatial models with heterogeneous coefficients: an application to US house prices. *Journal of Applied Econometrics*, 36(1), 18-44. <https://doi.org/10.1002/jae.2792>
- [27] Ge, S., Li, S. and Linton, O. 2022. News-implied linkages and local dependency in the equity market. *Journal of Econometrics*. <https://doi.org/10.1016/j.jeconom.2022.07.004>
- [28] Jiang, F. and Lee, J. and Martin, X., and Zhou, G. 2019. Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1): 126-149. <https://doi.org/10.1016/j.jfineco.2018.10.001>

- [29] 张永杰、张维、金曦和熊熊, 2011, 《互联网知道的更多么?——网络开源信息对资产定价的影响》, 《系统工程理论与实践》第4期, 第577-586页。
- [30] 赵丽丽、赵茜倩、杨娟、王铁军和李庆, 2012, 《财经新闻对中国股市影响的定量分析》, 《山东大学学报(理学版)》第7期, 第70-75+80页。
- [31] 姜富伟、孟令超和唐国豪, 2021, 《媒体文本情绪与股票回报预测》, 《经济学(季刊)》第4期, 第1323-1344页。
- [32] 陈张杭健、吴粤、李世炳和任飞, 2021, 《股吧个体信息交互对股价联动关系的影响研究》, 《管理科学学报》第24卷第5期, 第47-69页。
- [33] 范小云、王业东、王道平、郭文璇和胡焯翊, 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》, 《管理世界》第10期, 第78-95页。
- [34] 沈艳、陈赞和黄卓, 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》, 《经济学(季刊)》第4期, 第1153-1186页。doi:10.13821/j.cnki.ceq.2019.03.01.
- [35] 范小云、王业东和王道平, 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》, 《世界经济》第4期, 第3-30页。doi:10.19985/j.cnki.cassjwe.2022.04.006.
- [36] Andrews, D. W. 1987. Asymptotic results for generalized Wald tests. *Econometric Theory*, 3(3), 348-358. <https://doi.org/10.1017/S0266466600010434>
- [37] 刘霖和秦宛顺, 2004, 《中国股票市场套利定价模型研究》, 《金融研究》第6期, 第44-55页。
- [38] 杨子晖和李东承, 2018, 《我国银行系统性金融风险研究——基于“去一法”的应用分析》, 《经济研究》第8期, 第36-51页。
- [39] 朱恩伟、吴璟和刘洪玉, 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》, 《金融研究》第2期, 第117-135页。

Heterogeneous Spatial Factors Pricing Model in China Market

Shuyi Ge Shaoran Li Xinping Li Wen Su

(School of Finance, Nankai University; School of Economics, Peking University; FinLab of Peking University)

Summary: Co-movement or cross-section correlation of high dimension assets return is caused by common risk factors dependency and local dependency. With the help of heterogeneous spatial factors pricing model, we analyse these two dependencies and introduce the statistical inference procedure as well as the asymptotic no arbitrage theory. Besides, we construct the linkage matrix by news co-mentioned information, using the data of

HS300 stocks during 2017-2021, the empirical results show that the news linkage matrix helps capture the highly significant local dependency and spatial correlation among assets return. Finally, we also find that the spatial coefficients are still significant after we add the industry factors, and the coefficients appear to be heterogeneous between industries and individuals.

Keywords: Local Dependency, Text Analysis, Spatial Econometrics, News Linkage Matrix, Non-tradable factors

JEL Classification: C31, C50, G12

附录

附录 A: 对数似然函数的导数

对数似然函数 (6) 的梯度由下述各一阶导给出:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \psi_i} &= -T \text{tr} \{ [S(\boldsymbol{\Psi})]^{-1} I_{ii} W \} + \sum_{t=1}^T \frac{r_{it}^W (r_{it} - \psi_i r_{it}^W - \mathbf{x}'_{it} \mathbf{b}_i)}{\sigma_i^2}, \quad i = 1, 2, \dots, n, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i} &= \sum_{t=1}^T \frac{\mathbf{x}_{it} (r_{it} - \psi_i r_{it}^W - \mathbf{x}'_{it} \mathbf{b}_i)}{\sigma_i^2}, \quad i = 1, 2, \dots, n, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_i^2} &= -\frac{T}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} \sum_{t=1}^T (r_{it} - \psi_i r_{it}^W - \mathbf{x}'_{it} \mathbf{b}_i)^2, \quad i = 1, 2, \dots, n,\end{aligned}$$

其中 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 σ^2 的偏导, 可以得到 $\boldsymbol{\theta}$ 的 QMLE 的内在关系, 即

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

其中

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

以及

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

对数似然函数的 Hessian 矩阵是

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} \\ \cdot & \mathbf{H}_{22} & \mathbf{H}_{23} \\ \cdot & \cdot & \mathbf{H}_{33} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \psi'} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \mathbf{b}'} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial (\sigma^2)'} \\ \cdot & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b} \partial \mathbf{b}'} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b} \partial (\sigma^2)'} \\ \cdot & \cdot & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\sigma^2) \partial (\sigma^2)'} \end{bmatrix}.$$

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

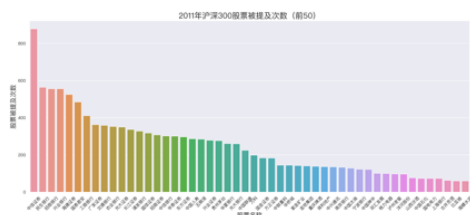
$$\begin{aligned}
 \mathbf{H}_{11} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} = \begin{cases} -T g_{ij} g_{ji}, & i \neq j, \\ -T g_{ii}^2 - \frac{(\mathbf{r}_i^W)'(\mathbf{r}_i^W)}{\sigma_i^2}, & i = j, \end{cases} \\
 \mathbf{H}_{12} &= \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \mathbf{b}'_j} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \mathbf{b}'_j} = \begin{cases} 0, & i \neq j, \\ -\frac{(\mathbf{r}_i^W)' \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)'(\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} \right), \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mathbf{b}_i \partial \mathbf{b}'_j} = \begin{cases} 0, & i \neq j, \\ -\frac{\mathbf{X}' \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}'(\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)'(\mathbf{r}_i - \psi_i \mathbf{r}_i^W - \mathbf{X} \mathbf{b}_i)}{\sigma_i^6}, & i = j, \end{cases}
 \end{aligned}$$

附录 B: 样本股按 CSRC 分类表

Table 10: 中国证券监督管理委员会行业分类: 225 支样本股

序号	门类	数量
A	农林牧渔业	3
B	采矿业	10
C	制造业	121
D	电力、热力、燃气及水生产和供应业	5
E	建筑业	7
F	批发和零售业	5
G	交通运输、仓储和邮政业	10
H	住宿和餐饮业	0
I	信息传输、软件和信息技术服务业	8
J	金融业	41
K	房地产业	7
L	租赁和商务服务业	2
M	科学研究和技术服务业	1
N	水利、环境和公共设施管理业	0
O	居民服务、修理和其他服务业	0
P	教育	1
Q	卫生和社会工作	3
R	文化、体育和娱乐业	1
S	综合	0

附录 C：样本股各年提及次数



(a) 2011



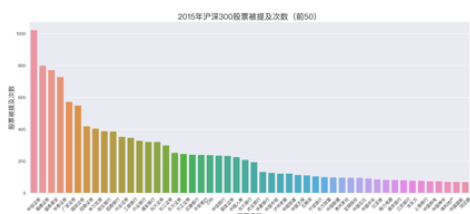
(b) 2012



(c) 2013



(d) 2014



(e) 2015



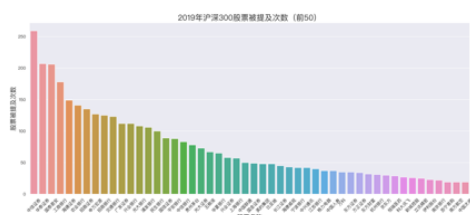
(f) 2016



(g) 2017



(h) 2018



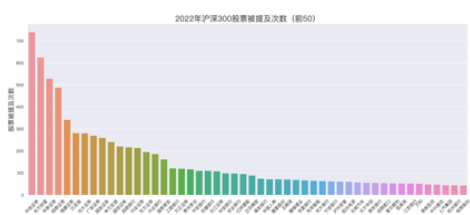
(i) 2019



(j) 2020

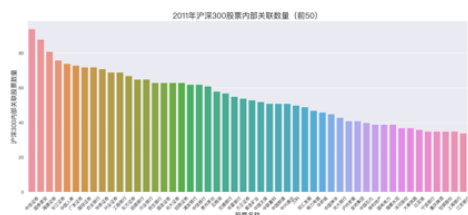


(k) 2021

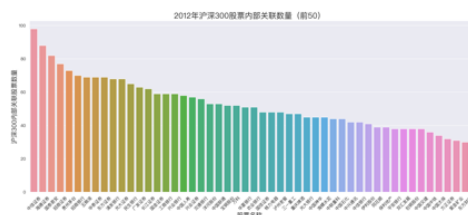


(l) 2022

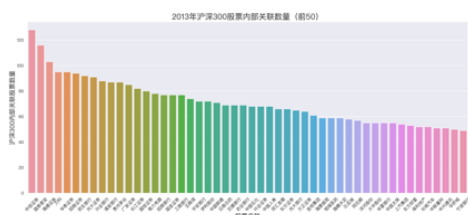
附录 D: 样本股各年关联股票数



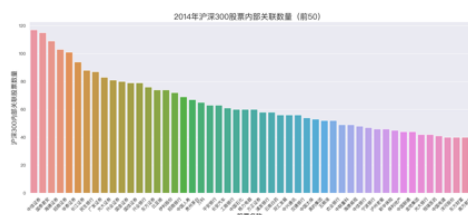
(m) 2011



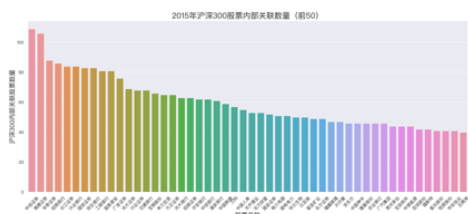
(n) 2012



(o) 2013



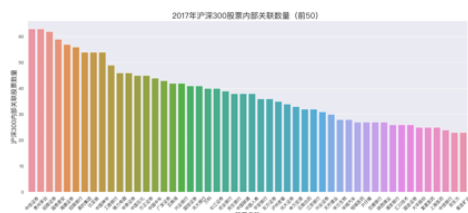
(p) 2014



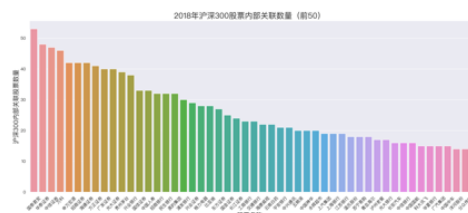
(q) 2015



(r) 2016



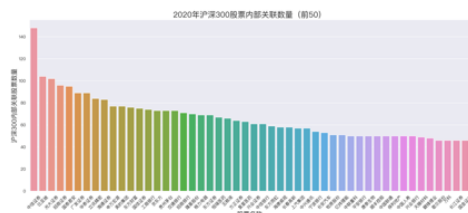
(s) 2017



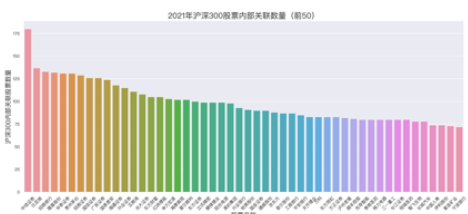
(t) 2018



(u) 2019



(v) 2020



(w) 2021



(x) 2022