

技术相似性对股票收益的预测能力

——“学海拾珠”系列之一百一十七

报告日期：2022-11-23

主要观点：

本篇是“学海拾珠”系列第一百一十七篇，作者主要研究了**基于技术相关联的公司对股票收益进行预测的可行性**。研究发现，技术关联是对现有风险因子的有效拓展，目标公司的股票收益可通过技术相关联的公司回报进行预测。基于技术关联情况构造的交易策略，即做多前一个月技术关联公司回报率高的股票，做空回报率低的股票，每月超额收益高达 1.28%。

回到 A 股市场，可将公司的财报作为数据源，基于文本分析建立企业间技术层面的联系，并通过技术关联公司的历史回报和财务数据来预测目标公司的未来收益。

● 企业间技术关联的信息会反映到股价上

由于技术相似性的**专业性、不确定性和内容的特定性**，投资者将难以处理有关技术相似性的信息，因此企业间技术关联的信息反映到股价上的过程不是**立即且充分的**。在**处理能力有限或注意力不集中**的情况下，投资者不能立即充分考虑相关信息，公司股价在对技术关联的同行的新消息做出反应时会有一定滞后。

● 技术关联是对现有风险因子的有效拓展

实证结果表明，高（低）共同分析师覆盖率的技术关联公司的投资组合确实表现出比低（高）共同分析师覆盖率的投资组合具有更强（较弱）的同步性和更弱（更强）的回报可预测性。根据公司间技术关联的强度来区分，与高相似度关联相比，低相似度关联的股票回报率同步性更弱，可预测性更强，因为**投资者对低相似度关联的关注度更低**。

● 基于技术相似公司交叉预测的零成本交易策略产生显著回报

本文构建的一种零成本交易策略，即做多前一个月技术关联公司回报率高的股票，做空同行回报率低的股票，该策略**每月超额收益高达 1.28%**，且在很大程度上无法由已知的因子定价模型所解释。

● 风险提示

文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

分析师：严佳炜

执业证书号：S0010520070001

邮箱：yanjw@hazq.com

分析师：吴正宇

执业证书号：S0010522090001

邮箱：wuzhy@hazq.com

相关报告

- 1.《被动投资对共同基金管理能力和市场效率的影响——“学海拾珠”系列之一百零九》
- 2.《共同资金流 Beta 与因子定价——“学海拾珠”系列之一百一十》
- 3.《分析师的反应不足和动量策略——“学海拾珠”系列之一百一十一》
- 4.《股票短线交易与收益异象——“学海拾珠”系列之一百一十二》
- 5.《明星分析师能否在糟糕的信息环境中做出更好的覆盖决策？——“学海拾珠”系列之一百一十三》
- 6.《基金经理能选出好的“投机性”股票吗？——“学海拾珠”系列之一百一十四》
- 7.《BAB 增强版：与包含定价噪音的 Beta 为敌——“学海拾珠”系列之一百一十五》
- 8.《ETF 的资金流动是否蕴含独特信息？——“学海拾珠”系列之一百一十六》

正文目录

1 引言	4
2 数据和统计方法	6
2.1 专利数据	6
2.2.产品相似性数据	7
2.3.股票回报、公司财务和分析师预测数据	7
3 实证分析.....	8
3.1 技术相似性指标领先于产品市场关联指标	9
3.2 技术相似性的实际经济效应	10
3.3 同期协同变动	11
3.4 回报交叉可预测性	13
3.5 产品和技术相似性之间的区别	13
3.6.基于文本和传统技术相似性度量之间的区别	13
4 零成本交易策略	15
5 结论	16
风险提示:	16

图表目录

图表 1 样本描述性统计	8
图表 2 技术相似性的持续性分析	9
图表 3 技术关联经济效应	10
图表 4 股票收益的同期协同变动	12
图表 5 股票收益的可预测性	12
图表 6 非重叠组合的可预测性及共同变动	14
图表 7 替代策略的表现	15
图表 8 零成本交易策略收益	16

用户1683078417于2023-09-27日下载，仅供本人内部使用，不可传播与转载

1 引言

无论是在整个经济层面还是在行业和个别公司层面，技术创新是经济增长、生产力和盈利能力提高的重要来源。Trajtenberg, Hall 及 Hirshleifer 等学者认为创新相关的信息对决定公司价值起重要作用。然而，最近的实证研究表明，在市场中创新可能会被错误估值，Hirshleifer 等人将这种错误估值与投资者有限的处理能力联系起来。本文研究了企业自身创新活动以及技术相似公司的创新活动对资产定价的影响。具体来说，作者研究(i)企业间技术关联的信息是否反映到股价；(ii)这类信息在股价上反映是否及时和充分。

之前的研究证明了产品市场空间中重要的企业间经济联系对资产定价的影响，例如行业、供应链以及关联动量等。创新关联代表了一种新的经济联系，它也许比产品市场联系更复杂，但仍然很重要。一个企业的微观经济特异性冲击会通过企业间的直接合作或更广泛的技术投入产出渠道影响从事相关技术领域的公司。创新溢出效应使企业能够在技术同行创造的知识基础上开发自己的新技术，在生产过程中使用其他企业的创新来降低成本和提高生产率，或者扩大业务以应对技术进步带来的需求。

为了识别技术上相似的公司，作者利用专利描述中公开的信息。首先，结合大数据科学的最新进展，对来自 770 万份专利描述的约 400 千兆字节的文本数据进行自动分析，创建了“专利宇宙”：一个独特的专利与专利之间的相似性空间，在这个空间中，局部相似的专利相互连接。然后在公司层面汇总单个专利，以描述一个公司的创新空间。

创新空间的概念在经济学、金融学和管理学中早已被广泛应用。然而，目前公司技术地位的衡量主要基于引用或美国专利和商标局 (USPTO) 分配的专利分类代码，这些衡量标准因专利分类不当、主观分类以及公司引用等问题导致的偏差而受到批判。作者对技术相似性的自动化文本测量旨在减少这些偏差，因为它不依赖于人工输入且使用了对创新的完整描述，而不仅仅是专利分类或引用中包含的少量信息。

Bloom 等人认为技术空间不同于产品市场空间。作者发现技术和产品市场同行之间只有 8% 的重叠。Teece 及 Alexopoulos 认为技术创新先于产品开发，决定了产品商业化的战略方向。谷歌自 21 世纪初以来对自动驾驶汽车技术的追求提供了一个示范性的例子，本文通过文本分析发现，谷歌与汽车制造商福特、本田和 Oshkosh 卡车公司分别在 2004 年、2005 年和 2006 年存在技术相似性。相比之下，产品市场的相似性不能用来检测这些联系。虽然技术相似性的传统衡量标准--引文和专利分类可以用来识别一些技术联系，但其中很多会被忽略。例如，引文重叠法在 2014 年之前没有发现谷歌和 Oshkosh Truck 之间的任何联系。

谷歌的例子在美国公司的样本中具有代表性。作者发现，技术相似性作为一个前瞻性的指标，预测了接下来几年的产品市场相似性。相反，产品相似性并不能预测技术相似性，因为属于同一产品领域的公司很可能已经在同一技术领域工作。

如果投资者事先考虑到关于技术关联的公开信息，那么当技术相似的同行的信息被发布到市场上时，公司的股票价格就会调整。然而，如果由于处理能力有限或注意力不集中，投资者不能立即充分考虑相关信息，公司股价在对技术同行的新消息做出反应时会有一个可预测的滞后。作者预计，由于技术相似性的专业性、不确定性和内容的特定性，投资者将难以处理有关技术相似性的信息。为了检验技术关联的资产定价回报，作者为每个公司计算基于文本的技术关联公司的投资组合回报，同时按技术相似性强度加权。

为了更好地理解作者的方法，以 2015 年 10 月的事件及其对亚马逊和技术同行股价的影响为例。2015 年 10 月 6 日至 9 日，亚马逊举办了最大规模的年度云计算会议，该会议吸引了众多与会者和名流媒体，并得到媒体广泛报道。会议上展示了许多创新性服务和工具，包括亚马逊网络服务（AWS）、物联网（IoT）、Amazon QuickSight（一种云驱动的商业智能服务）、Kinesis Firehose（一种将数据流到云的工具）和 AWS 数据库迁移。在会议后五天的窗口期内，亚马逊的累计异常回报率（CAR）达到了 5.08%。此外，在 10 月 22 日的财报中，亚马逊提供了 AWS 领域惊人增长的有力证据：销售额达到 21 亿，比上年同期增长 78%。亚马逊宣布后的五天 CAR 为 9.3%。这两个重大事件在很大程度上导致了亚马逊在整个 10 月份 14.87% 的异常回报。

亚马逊披露的好消息不仅增加了其自身股票的价值，也增加了其技术相关联的同行的股价，因为云计算的快速扩张对包括广泛公司在内的整个云生态系统都有利。例如，物联网融合了多种技术，包括无线通信、实时分析、机器学习、商品传感器和嵌入式系统，创建了一个由车辆、家用电器、可穿戴和医疗设备等物理设备以及制造业和农业中使用的各种库存物品组成的智能网络。2015 年 10 月和 11 月亚马逊技术同行的 CAR 表现显著协同变动，亚马逊所有技术关联的同行的 CAR 在公告发布当月（2015 年 10 月份）为 1.39%（ $t\text{-stat}=2.19$ ）。此外，在接下来的一个月（2015 年 11 月），所有技术关联的同行的 CAR 为 1.70%（ $t\text{-stat}=3.33$ ），显示出显著的超前-滞后效应。重要的是，行业关联不会影响亚马逊技术关联的同行的同期变动和超前-滞后效应。与亚马逊不属于同一个行业的公司的 CAR 甚至更高：10 月份为 1.55%（ $t\text{-stat}=2.05$ ），11 月份为 2.00%（ $t\text{-stat}=3.43$ ）。

亚马逊的例子代表了美国普通股票一种截面上的系统性规律。作者发现，投资者确实在股票价格中纳入了事先公开的有关技术关联的信息，并证实了基于文本分析的技术相似公司的股票回报率是同步的。但是投资者并没有立即且充分地考虑这些信息：目标公司的月度股票收益可由基于文本分析的技术相似公司的滞后组合预测。为了排除结果是由于行业动量驱动的可能，严格控制产品市场的相似性，包括两个行业组合回报：（i）市值加权的传统静态 SIC 组合和（ii）Hoberg 和 Phillips 动态的基于文本的网络行业分类（TNIC）组合，根据产品市场相似性的强度加权。此外，还控制短期反转、中期动量，各种截面特征，以及传统的技术相似性、引文重叠和美国专利商标局分类等传统方法度量关联起来的公司的股票回报。

作者研究表明，基于技术相似公司交叉预测的零成本交易策略产生了经济上和统计上的显著回报。由买入前一个月回报率高的技术相似的同行股票和做空过去回报率低的同行股票组成的交易策略产生了 1.28% 的月回报。这些策略的回报并没有被广为人知的风险因素所解释。此外，这些策略在样本期间仍然持续盈利。

基于技术关联回报的可预测性不是由技术上和产品上相似的投资组合的重叠所驱动的，并且与行业动量不同。作者构建了三个不重叠的投资组合，其中包括（i）只有基于文本的技术相似的同行，（ii）只有产品相似的同行，以及（iii）技术和产品都相似的同行。作者发现，虽然只有技术关联的公司或只有产品市场关联的公司的投资组合都能预测股票收益，但只有技术关联的投资组合的预测系数的大小要大于只有产品市场关联的投资组合的预测系数。这与有限的处理能力假说是一致的，因为关于技术关联的更细微的信息比关于产品市场关联的信息更慢地被反映到股票价格。此外，既有技术联系又有产品联系的公司的投资组合并不能预测收益，因为这类信息关联最明显，很容易被投资者识别，并立即被纳入到股票价格。

此外，作者进行了横截面测试，为有限处理能力假设提供了更直接的支持。如果

是有限的投资者注意力和处理能力驱动回报的可预测性，那么结果应该随着注意力的不集中程度而变化。因此，本文首先使用同时覆盖技术关联股票的分析师人数代表投资者对信息溢出的关注度。同时覆盖技术关联公司的分析师可能会意识到信息溢出，并将其传达给市场参与者。作者发现，高（低）共同分析师覆盖率的技术关联公司的投资组合确实表现出比低（高）共同覆盖率的投资组合具有更强（较弱）的即期同步变动和更弱（更强）的回报可预测性。第二，根据技术上相似的公司关联强度来区分。与高相似度关联相比，低相似度关联的股票回报率同步性更弱，可预测性更强，因为投资者对低相似度关联的关注度更低。

通过一系列的稳健性检验，确保结果不是由小市值股票的有限流动性或大公司和小公司之间的领先-滞后关系所驱动的。为了减少流动性不足的影响，排除价格低于 5 美元的股票，同时对样本施加额外限制的分析。首先，排除市值低于纽约证券交易所 20 百分位的公司。其次，当同行公司比预测公司大时，剔除相应关联，实证结果的主要内容仍然没有改变。为了排除结果是由客户-供应商关联驱动的可能性，控制了客户行业的投资组合收益，并从技术相似的投资组合中剔除单个客户-供应商关联。最后，将样本一分为二，发现早期（1996-2005 年）的交叉可预测性比后期（2006-2014 年）的更强。这也与处理能力有限的假说相一致，因为投资者处理大量信息的能力在早期受到计算技术的限制。研究发现 STEM（科学、技术、工程和数学）行业的结果更加有力，在这些行业中，创新信息对公司的估值至关重要。

即使技术关联企业的回报可预测性在很大程度上无法用众所周知的风险因素来解释，但另一种风险解释总是潜伏在背后，即具有相似技术的企业可能暴露一些类似的未观察到的风险因素，而这正是驱动领先-滞后效应的原因。然而，与有限处理能力假说相一致的横截面规律，很难与基于风险的解释相协调。尽管如此，为了解决这些问题，本文展示了技术相似的公司之间存在真实的经济联系，这是技术关联对资产定价影响的一个重要前提条件。作者证明了一个公司的盈利能力，用资产回报率（ROA）来衡量，与有技术关联同行的同期和滞后加权平均 ROA 相关。此外，以研发强度衡量的企业创新活动与关联企业的加权平均研发强度相关，并具有超前-滞后关系。

本文贡献如下：首先研究表明，不仅公司自身的创新，其技术关联的同行的创新也为其估值提供了重要信息。市场将有关技术相似性的信息反映到股价中，但反映过程不是立即且充分的。研究发现补充了对投资者有限认知处理能力作用的讨论，这表明投资者处理无形技术相似性信息比处理更明显的产品相似性信息更困难。

其次，对文本分析研究做出了贡献。本文对技术相似性的自动化文本测量旨在解决传统测量的偏差，并允许发现新的关联。例如，当行业由 TNIC 分类定义时，传统的基于 USPTO 类别的关联并不能捕捉到以往记录的行业动量之外的回报预测性。本文不仅表明基于文本的技术关联的同行预期股票收益超出了行业动量，而且较小规模的股票可以预测较大规模的股票。

2 数据和统计方法

2.1 专利数据

从 USPTO 存储系统中检索 USPTO 从 1976 年到 2015 年（含）颁发的美国专利的整个存储库。为尽可能实现全面性，在数据可得的情况下包括专利申请数据。因此，除了检索已颁发的专利外，作者还检索 2001 年至 2015 年（包括 2001 年）美国专利局批量数据存储系统上的专利申请。为了避免数据重复，删除在已发布专利

数据中重复出现的专利申请。此外，使用 Levenshtein 距离解决受让人姓名中的拼写错误和差异，并使用类似于 Bekkerman 和 Gavish 的方法来标准化受让人姓名。在将每项专利与标准化的受让人名称联系起来之后，作者将美国专利商标局数据库中的所有专利与证券价格研究中心 (CRSP) 数据库中报告的回报率的公司进行匹配。美国专利商标局颁发专利的时间在《官方公报》中被记录为授予日期，将专利申请公开的日期称为发布日期。本文将专利的发布日期定义为数据中列出的发布日期，否则定义为其授权日期。

作者将每个专利表示为专业术语的集合，如果两个专利（已发布或正在申请）有共同的专业术语，就认为它们是相似的。专业术语的最主要来源是维基百科，首先从维基百科页面的标题和字幕中收集术语，为弥补维基百科中的专业词汇不足，从在线专业词典中添加术语。

使用两个 MapReduce 评估专利间的相似性。第一个 MapReduce 任务查找专利中的候选专业术语 (Map 阶段)，对于每个检测到的术语，构建一个包含该术语的专利列表 (Reduce 阶段)。然后过滤太长的列表，因为它们所对应的术语太常见，无法在专利之间形成有意义的关联。第二个 MapReduce 任务在构建的列表中每一对专利 (Map 阶段) 间进行，计算两个专利之间的 L1 距离 (Reduce 阶段)，表示为由 TFIDF (术语频率-反向文档频率) 加权的共同术语向量。与专利文本中其他地方出现的术语相比，“专利要求”中出现的术语被赋予更大的权重。作者提出的方法克服了为大型文档库构建语义相似性空间的主要障碍：计算问题。因为只评估拥有共同专业术语的专利对，其方法显著减少了计算时间。

从两家公司拥有的专利的相似性中得出两家公司之间基于专利文本的技术相似性。如果 A1 拥有的专利 n1 与企业 A2 拥有的专利 n2 相似，则企业 A1 和 A2 被认为是相似的。A1 和 A2 之间的相似性是从两个专利 n1 和 n2 中较新的专利的发表年份开始检测。如果 A1 和 A2 拥有不止一对相似的专利，那么 A1 和 A2 之间的相似性就从最早的一对开始评估。对 A1 和 A2 公司的连续相似性测量是其专利之间对数加权的余弦相似性：对相似专利总量取对数并根据专利年龄（每对中较新的专利）进行折算，然后通过每个公司的专利总数的乘积取对数进行归一化。这种衡量方法考虑了一个特定公司的整个专利存量，以反映其创新历史；然而，它给较新的专利分配了更大的权重，以强调较近期创新活动的更大相关性。

2.2. 产品相似性数据

使用 Hoberg 和 Phillips (2010, 2016) 开发和使用的动态 TNIC 数据来控制公司之间的产品相似性。当检验回报协同变动和交叉可预测性时使用上一年的数据来识别企业的技术和产品相似性 (TNIC 分类)。

2.3. 股票回报、公司财务和分析师预测数据

股票回报数据来自 CRSP 月度回报数据库，从合并后的 CRSP-COMPUSTAT 数据库中获取所有证券代码为 10 或 11 的 NYSE/AMEX/NASDAQ 公司的财务报表数据和其他公司信息，剔除财年末价格低于 5 美元的股票，最终样本为在会计年度结束时账面净值 (BE) 和市值 (MV) 不含缺失值的公司。

对于每个公司 i ，计算基于文本的技术关联投资组合的回报， $r_{i,t}^{TS-text}$ ，作为属于公司 i 的基于文本的技术关联公司的平均月回报，其中根据其与公司 i 基于文本的相似性的强度进行加权。类似地，计算基于类别 (引用) 的技术关联投资组合的回报， $r_{i,t}^{TS-class}$ ($r_{i,t}^{TS-cite}$)，作为属于企业 i 的基于类 (引用) 的技术空间的关联企业的平均月回报，同样由相似性强度加权。计算产品相似投资组合的回报， $r_{i,t}^{PS}$ ，作为公司

i 动态 TNIC 行业股票的产品相似性加权平均月回报。在技术和产品相似性的所有衡量指标中，使用每年年底 t 公司间联系的公开信息，然后将其映射到 t+1 年 7 月至 t+2 年 6 月的回报数据，以计算适当的权重。

从 I/B/E/S 数据库中获得样本中每只股票的分析师覆盖率，使用历史 CUSIP 代码连接这两个数据库。股票被分析师覆盖的数量定义为在该日历年内至少为该股票发布一次 EPS 预测的分析师数量。

样本信息如图表 1 所示，A 部分显示描述性统计数据。平均每年有 903 家公司，平均每年有 82 项专利申请。该样本涵盖了 CRSP-COMPUSTAT 普通股中市值的近 50%。平均而言，一家公司有 70 个基于文本的技术相似同行，57 个在 TNIC 分类定义的相同产品空间中运营同行，92 个通过引用关联同行，116 个通过 USPTO 分类关联同行。B 部分提供了代表样本 1% 以上的不同行业的描述性统计数据，样本显示了不同的行业代表性。其中电子设备、制药和计算机软件是三大代表行业，因为在这些领域拥有大量专利。

图表 1 样本描述性统计

Panel A: Firm characteristics									
	Mean	STD	P10	P25	P50	P75	P90		
# of firms per year	903.11	120.78	774	801	882	1027	1085		
MV per year % of COMPUSTAT	0.49	0.02	0.45	0.47	0.49	0.50	0.51		
# of patents per firm-year	81.97	398.75	2	4	11	39	136		
N_links TS-text	69.85	88.65	12	20	38	79	164		
N_links TS-cite	92.11	118.69	12	25	51	107	211		
N_links TS-class	115.68	75.07	34	60	102	156	212		
N_links PS	57.46	64.10	4	11	32	88	150		
Text-based similarity	0.13	0.09	0.05	0.06	0.11	0.17	0.26		
Class-based similarity	0.20	0.20	0.06	0.07	0.12	0.25	0.49		
Citation-based similarity	0.13	0.09	0.04	0.06	0.10	0.17	0.25		
Log MV	7.08	2.03	4.55	5.64	6.96	8.44	9.81		
Log BM	0.36	0.27	0.11	0.21	0.33	0.48	0.66		
ROA	-0.03	0.30	-0.27	-0.03	0.04	0.08	0.13		
Panel B: Descriptive statistics by industry									
	N obs. per year	%	Log MV	Log BM	ROA	N_links TS-text	N_links PS	N_links TS-cite	N_links TS-class
Electronic Equip.	123.89	14%	6.60	0.42	-0.04	108.53	81	71	105
Pharmaceutical	117.39	13%	6.60	0.23	-0.22	43.34	64	135	57
Computer Software	80.61	9%	6.94	0.31	-0.05	205.65	68	64	82
Machinery	63.17	7%	6.85	0.40	0.04	49.50	55	23	102
Medical Equip.	55.89	6%	6.37	0.31	-0.04	36.28	43	58	69
Computers	48.00	5%	6.97	0.37	-0.03	137.48	96	63	127
Measuring	41.28	5%	6.43	0.40	-0.01	33.01	63	45	96
Chemicals	34.83	4%	7.49	0.34	0.03	67.09	106	22	136
Business Service	28.17	3%	6.47	0.31	-0.06	22.67	45	28	48
Petroleum and Gas	23.11	3%	8.96	0.40	0.07	86.34	102	45	125
Consumer Goods	22.94	3%	7.40	0.32	0.03	96.62	75	13	108
Automobiles	22.11	2%	7.21	0.39	0.03	125.29	94	14	153
Business Supplies	20.22	2%	7.59	0.38	0.04	81.68	81	11	136
Construction Mat.	19.83	2%	7.07	0.44	0.04	87.87	54	18	90
Electrical Equip.	19.67	2%	6.48	0.37	-0.02	33.85	60	20	103
Food Products	15.56	2%	8.81	0.29	0.07	15.36	59	15	63
Communication	13.78	2%	8.52	0.30	-0.04	53.31	87	35	105

资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

3 实证分析

3.1 技术相似性指标领先于产品市场关联指标

为了进一步证实技术和产品的相似性反映了不同的经济联系，并验证先前关于创新效应先于产品商业化的说法，作者分析了技术相似性是否能预测未来的产品相似性，反之亦然。

图表 2 报告了回归结果，其中第 1 列的因变量是一个虚拟变量，如果给定的一对公司在 $t+1$ 年是基于文本的技术相似性，则等于 1；第 2 列的因变量是一个虚拟变量，如果给定的一对公司 $t+1$ 年属于同一 TNIC 行业，则等于 1；自变量是在 t 年测量的相同的虚拟变量。例如，当 $t+1$ 年的产品相似性 $PS(t+1)$ 为因变量时， t 年产品相似性的系数估计值 $PS(t)$ 等于 2.583（第 1 列）。当第 $t+1$ 年的技术相似性，即 $TS-text(t+1)$ 为因变量时，第 t 年基于文本的技术相似的系数估计值 $TS-text(t)$ 等于 2.850（第 2 列）。技术相似性预测了产品相似性：在第 t 年技术相似的企业更有可能在下一年成为产品市场的同行，第 1 列中 $TS-text(t)$ 的正数和统计学上的显著系数 0.034 证明了这点。相反，在第 t 年属于同一产品空间的企业在下一年成为技术相似的可能性较小，第 2 列中报告的第 t 年技术相似性虚拟变量 $TS-text(t)$ 的负系数估计值 -0.418 证明了这一点。这些结果表明，技术发展阶段先于产品商业化：产品市场的同行不太可能成为技术同行，因为他们已经在同一技术领域运作。

图表 2 技术相似性的持续性分析

	(1)	(2)
	<i>Product similar $t+1$</i>	<i>Technologically similar $t+1$</i>
PS_t	2.583*** (0.003)	-0.418*** (0.002)
$TS-text_t$	0.034*** (0.003)	2.850*** (0.002)
<i>Intercept</i>	-2.105*** (0.003)	-1.887*** (0.002)
Pseudo R^2	34.15%	40.81%

资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

对于每个表，步骤(1)、(2)和(4)在回归中同时使用了 NFLOWS、NSR 和 NEIO。这些因子对于市场因子都是高度显著的(表 3)。Mkt-RF 与 NSR 之间 adjusted-R² 最低，为 4.9%，Mkt-RF 与 NEIO 之间 adjusted-R² 的最高，为 32.0%。这可能说明 $mkt - rf$ 与 NFLOWS、 $mkt - rf$ 与 NSR 之间的正相关关系主要来自 NEIO。考虑到 NEIO 和 NSR 是显著正相关，相关值为 0.32，这是一个合理的解释。假设资金流动是不知情的、非基本面的投资决策，正系数表明，在从债券基金流向股票基金(从股票基金流向债券基金)的时期，股票被高估(低估)，市场价格暂时偏离基本价值。此外，步骤(5)表明，在从股票基金流向债券基金的净流入为负数期间，这种关系略强。这种错误定价随后被纠正。表 3 还给出了 Mkt-RF 作为滞后 NEIO 和 NSR 因变量的时间序列回归系数。与之前的研究学者 Fant LF 一致，作者主要关注 NSR 和 NEIO。步骤(3)、(6)和(7)分别基于 NSR、NEIO 和其滞后项。作者发现 Wald 统计量的 p 值低于 1%。反转延迟了一个月，可以用 ICI 延迟发布资金流数据来解释。据推测，投资者只会在一个月后才了解到新的资金流信号。然而，如果包括第一个滞后项，结果是相似的。作者发现 NEIO (NSR)与 $mkt - rf$ 的关系非常显著(不显著)，步骤(6)的

adjusted-R2 比步骤(3)高得多。当使用两个滞后资金流作为回归变量(标准(7))时, NEIO 滞后项占主导地位。步骤(8)表明, 观察到的关系主要是由于资金从股票基金流向债券基金之后的反转效应, 因此, 这代表了个人投资者悲观的时期。鉴于投资者的相对风险厌恶程度随着财富的增加而降低, 市场压力导致将风险资产出售给更富有的投资者。考虑到风险资产投资者在市场压力时期之后有望获得可观的正回报, 这可能被解读为个人投资者的非理性行为。

3.2 技术相似性的实际经济效应

技术相似的公司之间回报率的一致性和交叉可预测性的一个重要前提是这些公司有真正的经济联系。

首先研究技术关联企业的盈利能力是否相关, 是否能进行交叉预测。用资产回报率 $ROA_{i,t}$ 来衡量目标企业*i*的盈利能力。计算每个目标企业*i*的基于文本的技术相似同行的 $ROA_{i,t}^{TS-text}$, 即所有与企业*i*有技术关联的企业的平均投资回报率, 按其技术相似性的强弱进行加权。

图表 3 技术关联经济效应

Panel A: Correlation and predictability of ROA		
	(1)	(2)
	$ROA_{i,t}$	$ROA_{i,t+1}$
$ROA_{i,t}^{TS-text}$	0.273 (3.59)	0.148 (3.43)
$ROA_{i,t-1}$	0.515 (9.28)	0.495 (5.74)
$LogBM_{i,t-1}$	0.018 (1.24)	-0.034 (-2.63)
$LogMV_{i,t-1}$	0.027 (8.91)	0.017 (5.77)
Industry and Year Fixed effects	Yes	Yes
N obs	17,046	16,198
Adj. R ²	42.02%	35.42%
Panel B: Correlation and predictability of R&D intensity		
	(1)	(2)
	$RD_{i,t}$	$RD_{i,t+1}$
$RD_{i,t}^{TS-text}$	0.195 (7.44)	0.204 (7.11)
$RD_{i,t-1}$	0.864 (72.93)	0.765 (40.02)
$ROA_{i,t-1}$	0.017 (2.66)	0.023 (3.66)
$LogBM_{i,t-1}$	-0.015 (-6.71)	0.001 (0.45)
$LogMV_{i,t-1}$	-0.002 (-6.83)	-0.002 (-7.47)
Industry and Year Fixed effects	Yes	Yes
N obs	17,048	16,198
Adj. R ²	88.14%	35.42%

资料来源: 《Technological similarity and stock return cross-predictability: Evidence from patents' big data》, 华安证券研究所

面板回归结果如图表 3 所示。在第 1 列中, 因变量是目标企业*i*在*t*年的盈利能力 $ROA_{i,t}$ 。第 2 列中, 因变量是企业*i*在*t+1*年的 ROA。主要自变量是技术关联同行的同期 ROA, 即 $ROA_{i,t}^{TS-text}$ 。此外, 作者还控制了企业*i*自身的滞后盈利能力 $ROA_{i,t-1}$, 其规模 $LogMV_{i,t-1}$ (公司*i*的市场价值的自然对数, 即流通股与*t-1*财年末股票价格的乘积), 以及账面市值比率 $LogBM_{i,t-1}$ (一加账面市值比的自然对数)。

为了控制经济周期和行业利润率，加入行业 and 年份固定效应，同时按公司和年份对标准误差进行了调整

A 组结果表明，一个公司的投资回报率与技术相关的同行的同期投资回报率是相关的。此外，技术关联同行的投资回报率可以预测企业下一年度的盈利能力。

接下来验证与创新相关活动在技术环节上的相关性和可预测性。使用研发强度（RD），即财政年度末研发支出与总资产的比率作为企业创新活动的代表。B 组结果显示，企业研发强度与上一年技术关联同行的研发强度相关，而且可以被技术关联同行的研发强度所预测，这证实了技术关联对企业创新活动确实存在影响。

3.3 同期协同变动

本节检验投资者是否将技术关联方面的信息纳入股票价格。如果投资者考虑到这些公开的信息，那么一个目标公司的股票价格应该调整以反映其技术同行的信息。因此，技术上相似的公司的股票价格应该是同步变化的。作者通过进行 Fama 和 MacBeth 回归来进行分析，其形式如下方程（1）：

$$r_{i,t} = \alpha_t + \theta_t^{TS-text} r_{i,t}^{TS-text} + \theta_t^{PS} r_{i,t}^{PS} + \theta_t^{TS-cite} r_{i,t}^{TS-cite} + \theta_t^{TS-class} r_{i,t}^{TS-class} + \theta_t r_{i,t}^{Ind} + K_t Z_{i,t} + \varepsilon_{i,t} \quad (1)$$

其中， $r_{i,t}$ 是公司 i 在 t 月份的股票收益率， $r_{i,t}^{TS-text}$ 是公司 i 在 t 月份的技术同行的组合收益率，为公司 i 基于文本分析的技术同行的相似度强弱加权平均收益率。为了确保结果不是由公司 i 的具有相似产品市场的股票收益率引起的，控制了 $r_{i,t}^{PS}$ ，即属于公司 i 动态 TNIC 行业股票按产品相似度强弱加权的组合收益率，以及 $r_{i,t}^{Ind}$ ，即与企业 i 同属 SIC 行业的企业加权平均组合回报。此外，引入 $r_{i,t}^{TS-cite}$ 以及 $r_{i,t}^{TS-class}$ 控制与公司 i 有引文重叠和美国专利商标局类别重叠的公司的加权平均收益。 $Z_{i,t}$ 包括股票收益的短期反转和中期动量以及公司的规模和账面市值比，其中使用 $r_{i,t-1}$ 来控制短期反转，使用 $r_{i,t-12}$ 、 $t-2$ 控制中期动量。为了便于比较和解释这些系数估计值的大小，将所有代表股票收益的自变量归一化，使其标准差等于 1，标准误差经过 Newey-West 调整（最多 12 个滞后期）。

图表 4 中报告了从每月横截面回归得到的系数时间序列的平均值 t 统计值。结果显示，一个公司的股票回报率与基于文本的技术同行的投资组合同时发生变化。

图表 4 股票收益的同期协同变动

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Intercept</i>	0.021 (3.23)	0.019 (3.77)	0.020 (3.14)	0.018 (3.68)	0.018 (3.6)	0.017 (3.95)	0.017 (3.86)
$r_{i,t}^{TS-text}$	0.036 (25.13)	0.024 (17.32)					0.014 (8.92)
$r_{i,t}^{TS-cite}$			0.030 (22.18)	0.019 (15.37)			0.009 (6.50)
$r_{i,t}^{TS-class}$					0.032 (26.18)	0.021 (17.64)	0.013 (10.64)
$r_{i,t}^{PS}$		0.027 (31.22)		0.028 (32.3)		0.027 (17.64)	0.024 (29.43)
$r_{i,t}^{Ind}$	0.022 (35.00)	0.016 (25.66)	0.022 (35.21)	0.016 (25.18)	0.022 (35.17)	0.017 (25.95)	0.015 (23.93)
$r_{i,t}$	-0.005 (-4.55)	-0.006 (-5.12)	-0.005 (-4.44)	-0.006 (-5.06)	-0.005 (-4.58)	-0.006 (-5.08)	-0.006 (-5.14)
$r_{i,t-12,t-2}$	0.000 (0.24)	-0.001 (-0.36)	0.000 (0.20)	-0.001 (-0.42)	0.000 (0.26)	0.000 (-0.29)	-0.001 (-0.48)
<i>LogBM</i>	0.002 (0.50)	0.002 (0.86)	0.002 (0.71)	0.003 (1.13)	0.003 (1.13)	0.003 (1.26)	0.004 (1.33)
<i>LogSize</i>	-0.001 (-2.26)	-0.001 (-2.20)	-0.001 (-2.27)	-0.001 (-2.24)	-0.001 (-2.19)	-0.001 (-2.12)	-0.001 (-2.15)
T	210	210	210	210	210	210	210
Adj. R-Square	8.58%	10.20%	8.38%	10.13%	8.61%	10.15%	10.58%

资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

图表 5 股票收益的可预测性

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Intercept</i>	0.018 (2.39)	0.018 (2.45)	0.019 (2.41)	0.018 (2.47)	0.018 (2.46)	0.017 (2.36)	0.016 (2.25)
$r_{i,t}^{TS-text}$	0.007 (3.61)	0.006 (3.61)					0.006 (4.12)
$r_{i,t}^{TS-cite}$			0.005 (2.99)	0.003 (2.3)			0.001 (0.64)
$r_{i,t}^{TS-class}$					0.004 (2.05)	0.002 (1.31)	0.000 (0.08)
$r_{i,t}^{PS}$		0.003 (2.56)		0.004 (3.01)		0.004 (1.31)	0.003 (2.93)
$r_{i,t}^{Ind}$	0.002 (2.1)	0.001 (1.05)	0.002 (2.35)	0.001 (1.29)	0.002 (2.37)	0.001 (1.32)	0.001 (0.93)
$r_{i,t}$	-0.002 (-1.56)	-0.002 (-1.48)	-0.002 (-1.59)	-0.002 (-1.54)	-0.002 (-1.58)	-0.002 (-1.49)	-0.002 (-1.59)
$r_{i,t-12,t-2}$	0.000 (-0.11)	-0.001 (-0.38)	0.000 (-0.07)	-0.001 (-0.34)	0.000 (-0.06)	0.000 (-0.31)	-0.001 (-0.45)
<i>LogBM</i>	0.001 (0.29)	0.001 (0.23)	0.001 (0.42)	0.001 (0.42)	0.001 (0.33)	0.001 (0.31)	0.002 (0.58)
<i>LogSize</i>	-0.001 (-2.11)	-0.001 (-2.09)	-0.001 (-2.11)	-0.001 (-2.06)	-0.001 (-2.23)	-0.001 (-2.13)	-0.001 (-2.02)
T	209	209	209	209	209	209	209
Adj. R-Square	5.22%	5.70%	5.21%	5.73%	5.36%	5.76%	6.01%

This table reports time-series averages of coefficient estimates from monthly cross-sectional predictive regression:

$$r_{i,t+1} = \alpha_t + \lambda_t^{TS-text} r_{i,t}^{TS-text} + \lambda_t^{PS} r_{i,t}^{PS} + \lambda_t^{TS-cite} r_{i,t}^{TS-cite} + \lambda_t^{TS-class} r_{i,t}^{TS-class} + \lambda_t^{Ind} r_{i,t}^{Ind} + \Lambda_t Z_{i,t} + \varepsilon_{i,t+1} \quad (2)$$

资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

3.4 回报交叉可预测性

通过以下形式的 Fama 和 MacBeth 回归测试技术相似公司股票回报的交叉可预测性，检验技术相似性信息是否完全纳入股价：

$$r_{i,t} = \alpha_t + \lambda_t^{TS-text} r_{i,t-1}^{TS-text} + \lambda_t^{PS} r_{i,t-1}^{PS} + \lambda_t^{TS-cite} r_{i,t-1}^{TS-cite} + \lambda_t^{TS-class} r_{i,t-1}^{TS-class} + \lambda_t r_{i,t-1}^{Ind} + \Lambda_t Z_{i,t-1} + \varepsilon_{i,t} \quad (2)$$

因变量是公司 i 在 t 月的股票收益，自变量与方程 (1) 中的自变量相同，只是时间为 $t-1$ 月。

图表 5 中报告了月度横截面回归得到的系数时间序列的平均值及 t 统计值。结果表明，技术同行的滞后回报交叉预测了目标公司的股票回报。例如，第七列包括所有控制变量后， $r_{i,t-1}^{TS-text}$ 的系数为 0.006，大于 $r_{i,t-1}^{PS}$ 的系数，而且系数估计值显著。SIC 行业组合的可预测性被 TNIC 产品相似性所包含。

图表 5 第 3-6 列验证传统的技术相似性度量是否能够预测股票收益。结果表明，在控制了由 3 位数 SIC (第 3 列) 和 TNIC (第 4 列) 行业分类所捕获的行业动量，基于引文的投资组合 $r_{i,t-1}^{TS-cite}$ 的系数估计该指数是正数且显著。无论是基于分类的投资组合还是基于引用重叠度的投资组合，其对股票回报的预测都无法超过基于文本衡量方法所捕获的技术动量。

3.5 产品和技术相似性之间的区别

虽然技术相似性和产品相似性是不同的经济联系，但这对相似性之间存在一些重叠。为了区分产品和技术相似性的影响，作者计算了三个非重叠组的投资组合收益：(i) 技术上与目标公司相似但产品不相似的同行 $r_{i,t-1}^{TS-text NOT PS}$ ，(ii) 产品相似但技术不相似的同行 $r_{i,t-1}^{PS NOT TS-text}$ ，以及 (iii) 产品和技术相似的同行 $r_{i,t-1}^{TS-text AND PS}$ 。

图表 6 第 1、2 列报告了同期和预测回归的结果。产品相似的同行的股票收益率与目标公司表现出高度的同期性，但是，他们的投资组合预测能力较低。相比之下，只有技术上相似的公司股票收益率显示出较低的同期复合性和较高的可预测性。这些结果表明技术相似公司的回报率的可预测性并不是由技术和产品相似公司之间的重叠所驱动的。基于文本的技术相似性衡量方法捕捉到了产品相似性所不能反映的重要经济联系。其次，这些结果支持了投资者有限注意力驱动回报交叉可预测性的猜想。根据投资者有限注意力的假设，更细微的信息进入股票价格的速度会更慢。技术的相似性传达了重要但更复杂的经济联系，因此，与产品关联相比，技术关联的预测能力更强。

3.6 基于文本和传统技术相似性度量之间的区别

作者还研究了基于文本的度量、传统度量、引用重叠和 USPTO 分类所捕获的技术相似性的效应。对于每个公司，计算以下非重叠组的投资组合回报：(i) 基于文本的技术相似但不基于引用 (类) 相似的同行 $r_{i,t-1}^{TS-text NOT TS-cite}$ ($r_{i,t-1}^{TS-text NOT TS-class}$)，(ii) 基于引用 (类) 的相似但不基于文本的相似的对等体 $r_{i,t-1}^{TS-cite NOT TS-text}$ ($r_{i,t-1}^{TS-class NOT TS-text}$)，以及 (iii) 对等体既基于引用 (类) 又基于文本， $r_{i,t-1}^{TS-text AND TS-cite}$ ($r_{i,t-1}^{TS-text AND TS-class}$)。图表 6 第 3 列和第 4 列报告了非重叠引用和基于文本的投资组合的同期和预测回归结果，第 5 列和第 6 列报告了类别和基于文本投资组合的回归结果。

通过基于文本的和传统的相似性度量相关联公司表现出最高的联动。仅通过一个指标 (基于文本或引用 (类别) 的指标) 关联的公司表现出较低的协同性。然而，只有基于文本的技术相似的同行人才能预测股票收益，仅基于引用或类的技术同行不能预测股票收益。

基于引用和基于分类的技术关联的同期协同变动较弱，缺乏可预测性，这表明这些方法度量的是技术相似性的噪声。文本分析不仅过滤了现有度量中发现的噪声，还提取了传统度量无法捕获的重要信息，正如仅基于文本的技术相似的同行的统计和经济上显著的同期协同性和可预测性所表明的那样。

图表 6 非重叠组合的可预测性及共同变动

	(1)	(2)	(3)	(4)	(5)	(6)
	$r_{i,t}$	$r_{i,t+1}$	$r_{i,t}$	$r_{i,t+1}$	$r_{i,t}$	$r_{i,t+1}$
<i>Intercept</i>	0.019 (3.97)	0.018 (2.33)	0.017 (3.84)	0.016 (2.24)	0.018 (3.9)	0.017 (2.31)
$r_{i,t}^{TS-text NOT PS}$	0.004 (2.32)	0.006 (3.74)				
$r_{i,t}^{TS-text AND PS}$	0.014 (15.46)	0.001 (1.06)	0.003 (3.13)			
$r_{i,t}^{PS NOT TS-text}$	0.021 (23.27)	0.003 (3.13)				
$r_{i,t}^{TS-text NOT TS-cite}$			0.005 (4.35)	0.002 (2.53)		
$r_{i,t}^{TS-text AND TS-cite}$			0.011 (11.35)	0.004 (3.85)		
$r_{i,t}^{TS-cite NOT TS-text}$			0.005 (4.98)	-0.001 (-0.91)		
$r_{i,t}^{TS-text NOT TS-class}$					0.002 (1.76)	0.003 (2.93)
$r_{i,t}^{TS-text AND TS-class}$					0.011 (10.17)	0.003 (3.39)
$r_{i,t}^{TS-class NOT TS-text}$					0.008 (7.63)	0.000 (-0.19)
$r_{i,t}^{PS}$			0.024 (27.33)	0.003 (3.08)	0.025 (30.18)	0.003 (2.88)
$r_{i,t}^{TS-cite}$	0.010 (6.22)	0.001 (0.45)			0.011 (7.81)	0.001 (0.91)
$r_{i,t}^{TS-class}$	0.014 (9.18)	0.001 (0.56)	0.015 (12.11)	0.001 (0.46)		
$r_{i,t}^{Ind}$	0.016 (23.27)	0.000 (-0.07)	0.016 (24.67)	0.001 (0.91)	0.015 (23.52)	0.001 (0.78)
$r_{i,t}$	-0.007 (-5.49)	-0.003 (-1.9)	-0.006 (-4.9)	-0.002 (-1.66)	-0.006 (-5.23)	-0.002 (-0.43)
$r_{i,t-12,t-2}$	-0.001 (-0.67)	-0.001 (-0.76)	-0.001 (-0.51)	-0.001 (-0.47)	-0.001 (-0.57)	-0.001 (-0.43)
<i>LogBM</i>	0.004 (1.35)	0.002 (0.59)	0.003 (1.27)	0.002 (0.65)	0.003 (1.18)	0.002 (0.61)
<i>LogSize</i>	-0.001 (-2.14)	-0.001 (-2.2)	-0.001 (-2.32)	-0.001 (-2.32)	-0.001 (-2.22)	-0.001 (-2.03)
T	210	209	210	209	210	209
Adj. R-Square	11.448%	6.619%	10.46%	6.07%	10.57%	6.05%

资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

4 零成本交易策略

本节验证基于技术相似公司回报的零成本交易策略的盈利能力。构建如下交易策略：根据 t 年 6 月底纽约证券交易所的市值中位数将公司独立分为两组（小、S 或大、B），并根据上个月底技术相似同行的回报将其分为五组：低（1）至高（5）。持有这些投资组合一个月，并计算投资组合（S/1、S/2、S/3、S/4、S/5、B/1、B/2、B/3、B/4 和 B/5）的等权（市值加权）月度回报。然后计算五个从低到高投资组合的月度规模调整回报率，分别为 $(S/1+B/1)/2$ 、 $(S/2+B/2)/2$ 、 $(S/3+B/3)/2$ ，以及 $(S/4+B/4)/2$ 和 $(S/5+B/5)/2$ 。

图表 7 替代策略的表现

Panel A: Equal weighted portfolios						
	Excess Return (percent)	CAMP α (percent)	3-Factor α (percent)	4-Factor α (percent)	5-Factor α (percent)	6-Factor α (percent)
Low(1)	0.46 (0.86)	-0.22 (-0.78)	-0.31 (-1.19)	-0.14 (-0.58)	-0.18 (-0.66)	-0.10 (-0.4)
(2)	0.99 (2.31)	0.42 (2.08)	0.23 (1.62)	0.32 (2.47)	0.22 (1.52)	0.27 (2.01)
(3)	1.02 (2.3)	0.42 (2.11)	0.26 (1.65)	0.38 (2.62)	0.28 (1.7)	0.34 (2.25)
(4)	1.42 (3.3)	0.86 (3.95)	0.69 (4.36)	0.80 (5.53)	0.67 (4.14)	0.72 (5.01)
High(5)	1.74 (3.26)	1.13 (3.25)	0.97 (3.51)	1.10 (4.13)	1.10 (4.13)	1.16 (4.56)
H - L	1.28 (2.97)	1.34 (3.09)	1.28 (3.02)	1.24 (2.9)	1.28 (2.94)	1.27 (2.89)
Panel B: Value weighted portfolios						
	Excess Return (percent)	CAMP α (percent)	3-Factor α (percent)	4-Factor α (percent)	5-Factor α (percent)	6-Factor α (percent)
Low(1)	0.34 (0.67)	-0.31 (-1.19)	-0.35 (-1.37)	-0.22 (-0.88)	-0.20 (-0.73)	-0.14 (-0.53)
(2)	0.59 (1.47)	0.05 (0.27)	-0.06 (-0.38)	0.05 (0.33)	-0.06 (-0.37)	-0.01 (-0.06)
(3)	0.90 (2.31)	0.37 (2.26)	0.24 (1.83)	0.33 (2.78)	0.20 (1.42)	0.24 (1.98)
(4)	1.24 (3.19)	0.72 (4.2)	0.61 (4.07)	0.72 (5.2)	0.58 (3.7)	0.63 (4.49)
High(5)	1.30 (2.72)	0.72 (2.54)	0.64 (2.58)	0.75 (3.12)	0.71 (2.85)	0.76 (3.17)
H - L	0.96 (2.26)	1.03 (2.41)	0.99 (2.33)	0.96 (2.25)	0.91 (2.05)	0.90 (2.02)

资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

图表 7 的 A 部分和 B 部分分别显示了等权和市值加权投资组合的结果。平均月规模调整后的投资组合回报率与技术同行的前一个月回报率呈单调增长。具体而言，技术同行回报高（5）的等权（市值）加权投资组合的平均超额收益为 1.74%，而同行回报低的组合超额收益为 0.46%。做多技术关联公司回报高的组合和做空低的等权（市值）加权投资组合的零成本策略每月收益率为 1.28%。

为检验该异常回报是否来自系统风险，作者将收益对以下模型进行奉献调整，包括 CAPM 模型、三因子 Fama-French 模型（包括超额市场收益、SML 和 HML 因素）、四因子模型（包括前三个因素和 MOM）、五因子 Fama-French 模型（用 RMW 和 CMA 因子）以及六因子模型（将 MOM 添加到 Fama French 五因子中）。图表 7 显示，即使经过风险调整，基于过去一个月技术同行回报形成的投资组合收益也是单调增加，该投资策略的异常回报不受风险调整的影响。

总的来说，与技术同行的可预测性是对现有风险因子的拓展，可以产生可观的回报。图表 8 描述了样本期内交易策略的累积回报。

图表 8 零成本交易策略收益



资料来源：《Technological similarity and stock return cross-predictability: Evidence from patents' big data》，华安证券研究所

5 结论

本文研究了如何将技术相似性信息纳入到股价中。利用专利文本中公开的、大量的、极其丰富的数据，作者构建了一种**新的企业技术相似性度量方法**。与传统的基于 USPTO 指定的专利技术分类或常见引用法的度量方式不同，该构建方式避免了专利审查员和专利律师的自由裁量权以及公司引用倾向和实践导致的偏差。重要的是，**基于文本的技术相似性度量捕获了传统度量无法发现的重要新技术关联**。

本文证明，技术关联的相关信息被反映在股票价格中，技术相似股票的同期协同变动证明了这一点。然而，正如技术关联公司的回报可预测性所表明，技术关联的相关信息并没有立即和充分地反映在股价中。

本文提出一种零成本交易策略，即做多前一个月技术关联公司回报率高的股票，做空技术关联公司回报低的股票，每月溢价高达 1.28%。随着时间的推移，这种交易策略表现一致，并且在很大程度上无法由实证定价模型解释。尽管创新在新的知识经济中日益重要，但与产品市场关联的信息相比，创新关联的无形信息被反映到股价中的速度更慢。本文发现，**基于技术关联公司的回报可预测性强于基于产品市场关联的回报**。

横截面分析结果支持有限信息处理能力假设作为技术动量的解释。相比关联程度较弱的同行，对密切关联公司的冲击会更快地反映到股价上。此外，与共同分析师数量较少的技术同行相比，共同分析师数量较多的技术同行受到的冲击将更快地被反映到股价。

文献来源：

核心内容摘选自 Ron Bekkerman 在 The Review of Asset Pricing Studies 上的论文《Technological similarity and stock return cross-predictability: Evidence from patents' big data》。

风险提示：

文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

重要声明

分析师声明

本报告署名分析师具有 PRC 证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

免责声明

华安证券股份有限公司经 PRC 证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告中的信息均来源于合规途径，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资回报，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券及其所属关联机构可能会持股报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经华安证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。如欲引用或转载文献内容，务必联络华安证券研究所并获得许可，并需注明出处为华安证券研究所，且不得对文献进行有悖原意的引用和删改。如未经本公司授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

投资评级说明

以本报告发布之日起 6 个月内，证券（或行业指数）相对于同期沪深 300 指数的涨回撤为标准，定义如下：

行业评级体系

- 增持—未来 6 个月的投资回报领先沪深 300 指数 5%以上；
- 中性—未来 6 个月的投资回报与沪深 300 指数的变动幅度相差-5%至 5%；
- 减持—未来 6 个月的投资回报落后沪深 300 指数 5%以上；

公司评级体系

- 买入—未来 6-12 个月的投资回报领先市场基准指数 15%以上；
- 增持—未来 6-12 个月的投资回报领先市场基准指数 5%至 15%；
- 中性—未来 6-12 个月的投资回报与市场基准指数的变动幅度相差-5%至 5%；
- 减持—未来 6-12 个月的投资回报落后市场基准指数 5%至；
- 卖出—未来 6-12 个月的投资回报落后市场基准指数 15%以上；
- 无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。市场基准指数为沪深 300 指数。