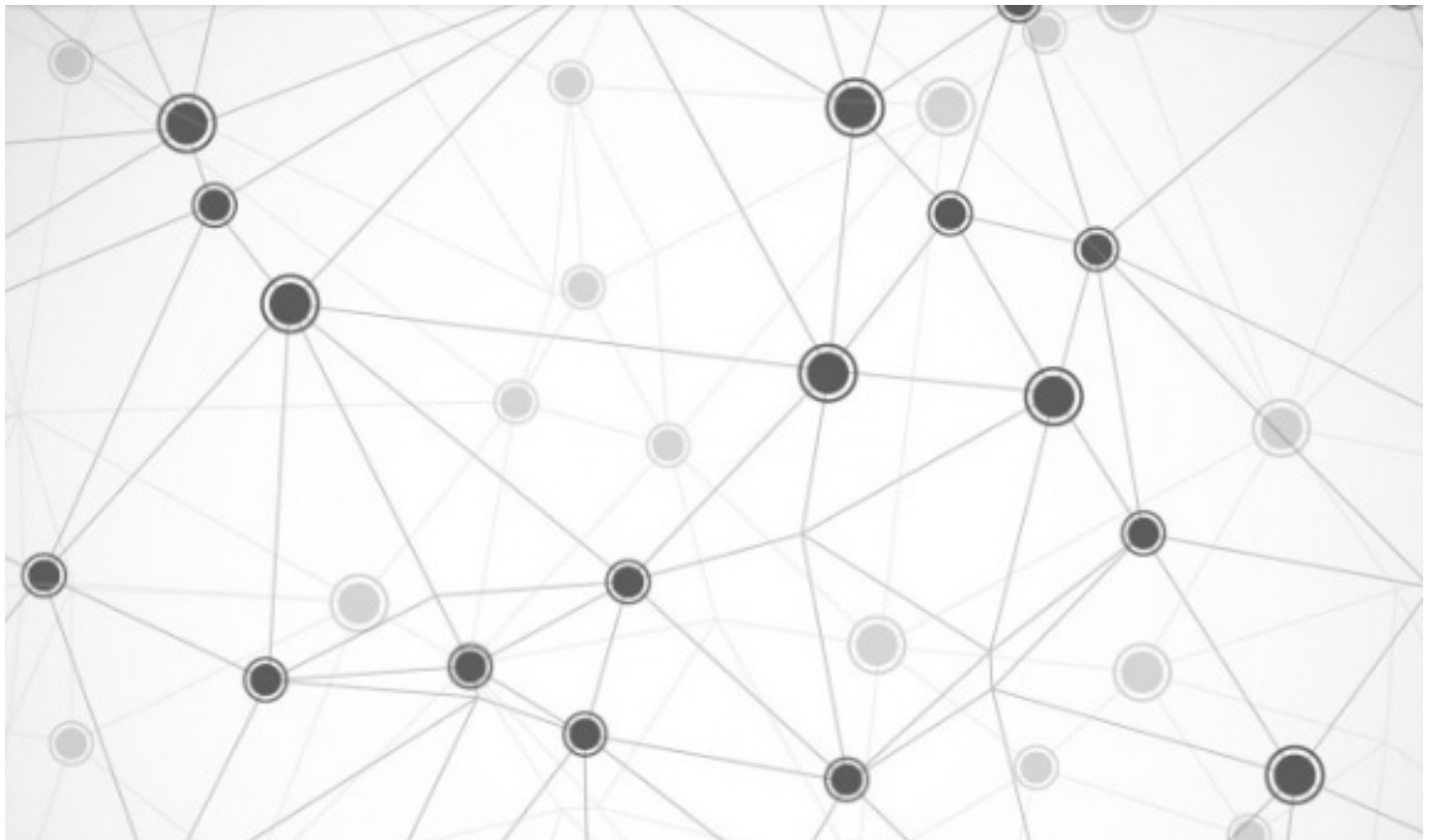


知乎

首发于
川流不息

获取 α 的新思路：科技关联度



石川

量化交易 话题的优秀回答者

已关注

「已注销」等 106 人赞同了该文章

摘要

在寻找 α 的努力中，最重要的是获得新的数据或者是对已有数据的创新用法。与数据相比，算法先进性的作用相对有限。本文介绍的 Lee et al. (2018) 提出了获取 α 的新视角。

1 引言

今天给大家介绍一篇新鲜出炉的文章，题为 Technological links and predictable returns (Lee et al. 2018)，它即将刊发于 Journal of Financial Economics，领衔作者是斯坦福大学的 Charles M. C. Lee 教授和北京大学的张然教授，这两位均长期从事基本面量化投资的研究。

该文提出了一个获取 α 的新思路：**科技关联度 (technological links)**。虽然它是以美股为研究背景（使用该因子构建的多空投资组合每个月可以获得 1.17% 的超额收益），但相信对投资 A 的小伙伴也会有很大的启发。更重要的是，该文的行文逻辑堪称因子研究的典范。相信读过本介绍后，你就能够认可这种说法。



2 科技关联度

在知识经济时代，科技实力已经成为一个公司短期盈利和长期生存的重要因素。世界上的科技巨头，如亚马逊、谷歌、苹果、英特尔等公司，它们的产品可能截然不同，但在科技层面却有着千丝万缕的联系。这些科技上的关联超越了传统的行业界限，却通常不易从公司的财务报告中辨别出来。

Technological links and predictable returns 一文研究了公司之间的科技关联度和公司股票未来预期收益率之间的关系。它背后的逻辑是公司的科研并不是独立的，一项科技进步的溢出效应将会影响科技关联度高的一系列公司，而这种影响将改变这些公司基本面，并最终（先后）反映到公司的股价中。




基于此，该文揭示了一个令人惊讶的实证关系，即**目标公司的股票收益率和与其科技关联度相近的公司前期的收益率之间有一种滞后-领先关系**。换句话说，对于任何一个目标公司，使用某种代理指标来计算它和其他公司的科技关联度，然后以该关联度为权重和其他公司的当期收益率就可以计算出一个加权收益率，该收益率对目标公司下一期的收益率有一定的预测性。因此，**以科技关联度为权重的加权收益率是一个优秀的 α 因子。这种领先-滞后关系也可以被称作“科技动量”**。

下面就来介绍如何构建这个因子。首先需要计算公司之间的科技关联度。为此，Lee et al. (2018) 使用**两个公司专利分布之间的 uncentered correlation**（就是计算相关系数的时候省去减均值的步骤，类似的做法也被 Jaffe 1986 和 Bloom et al. 2013 采用）计算科技关联度：

$$\text{TECH}_{ijt} = \frac{T_{it} T'_{jt}}{(T_{it} T'_{it})^{1/2} (T_{jt} T'_{jt})^{1/2}}$$

上式左侧 TECH_{ijt} 代表了公司 i 和 j 在第 t 期的科技关联度；上式右侧的 T_{it} 是一个 427 维的横向量，即 $T_{it} = [T_{it1}, T_{it2}, \dots, T_{it427}]$ 。427 这个数字源自美国专利商标局（United States Patent and Trademark Office）定义的 427 个科技大类。因此，一个公司的专利分布由它过去五年获取的全部科技专利在这 427 类中的比例决定。举例来说，假如从当前时点计算，某公司过去五年获得了 100 个科技专利，分别属于编号为 1（10 个），100（60 个）和 303（30 个）的三大类，则该公司的专利分布中， $T_{it1} = 10/100 = 0.1$ ，

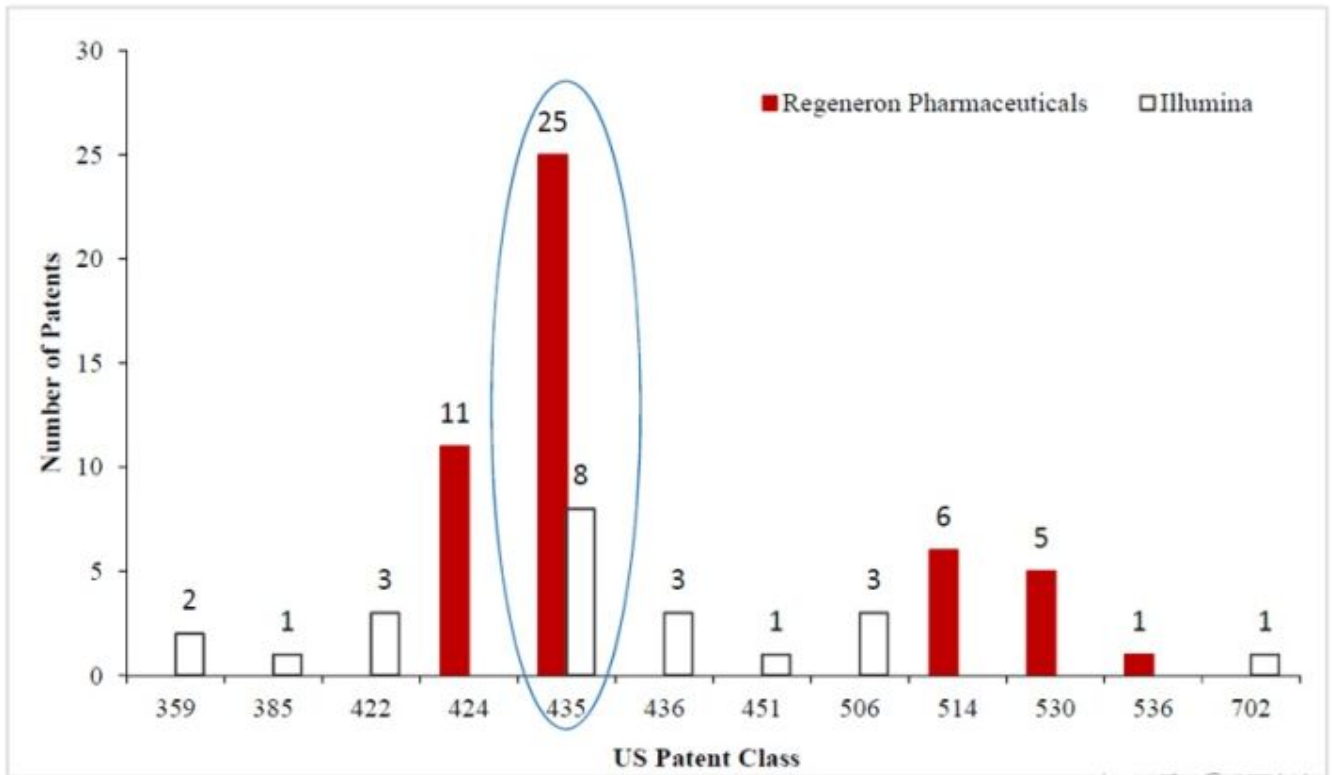
$T_{it100} = 60/100 = 0.6$ ， $T_{it303} = 30/100 = 0.3$ ，而其他大类 k 对应的 T_{itk} 值为 0，这些 T_{itk} 就构成了 t 期该公司的专利分布向量，由此就可以计算不同公司之间的科技关联性 TECH_{ijt} 。

下图来自 Lee et al. (2018)，它展示了两家公司 Regeneron Pharmaceuticals 和 Illumina 在 2002 到 2006 年间科技专利所属类别的情况（绝对数量，未经标准化）。Regeneron 是一家  公司，而 Illumina 生产生命科学工具并提供遗传分析服务。这两家公司所处完全不同的行业， 在供应链方面也没有什么联系。但是科技关联性从全新的角度揭示了它们之间的关联——这 

知乎

首发于
川流不息

验数据表明这种关联在选股方面大有可为。



出处：Lee et al. (2018)

知乎 @石川

得到 TECH_{ijt} 之后，利用它作为权重按下式计算加权收益率作为选股因子：

$$\text{TECHRET}_{it} = \frac{\sum_{j \neq i} \text{TECH}_{ijt} \times \text{RET}_{jt}}{\sum_{j \neq i} \text{TECH}_{ijt}}$$

上式左侧 TECHRET_{it} 就是公司 i 第 t 期的因子取值；右侧的 RET_{jt} 为公司 j 在 t 期的收益率，所有和 i 不同的公司 j 的收益率以 TECH_{ijt} 为权重加权在一起，构成了选股因子。由于计算科技关联度需要使用到过去五年的专利数，因此每期的候选股票池为在这段时间内至少获得了一个专利的股票（以专利的官方授予日期计算，从而避免了前视偏差）。

值得说明的是，虽然选股因子 TECHRET_{it} 的更新是月频，但是其中的科技关联度 TECH_{ijt} 更新的频率是每年一次，在每年年末使用过去五年的专利数来更新 TECH_{ijt} 。由于收益率 RET_{jt} 是月频收益率，因此最终的因子更新频率是月频。

关于数据的来源和处理方法更详尽的说明请参考 Lee et al. (2018) 中的第二节，这里不再赘述。

3 投资组合检验





首发于
川流不息

最高的第一档，做空分数最低的第十档，以此构建一个 L/S 组合，并考察该组合的收益率。L/S 组合的收益率如下表所示（该表是 Lee et al. 2018 中最重要的结果）。

Panel A: Portfolio returns						
Decile	Excess returns (%)	CAPM alpha (%)	3-Factor alpha (%)	4-Factor alpha (%)	5-Factor alpha (%)	6-Factor alpha (%)
1	0.42	-0.16	-0.33	-0.13	-0.27	-0.12
(Low)	(1.46)	(-1.04)	(-2.66)	(-1.09)	(-2.17)	(-0.99)
10	1.59	1.05	0.93	0.95	1.10	1.09
(High)	(5.38)	(5.37)	(6.41)	(6.36)	(8.11)	(7.97)
L/S	1.17	1.22	1.26	1.08	1.37	1.21
(Equal weights)	(5.47)	(5.70)	(5.88)	(4.98)	(6.49)	(5.76)
L/S	0.69	0.74	0.80	0.65	0.86	0.73
(Value-weights)	(3.19)	(3.40)	(3.62)	(2.91)	(3.81)	(3.24)

Panel B: Risk factor loadings					
	Alpha	MKT	SMB	HML	MOM
1	-0.13	1.11	0.83	-0.08	-0.21
(Low)	(-1.09)	(40.08)	(21.52)	(-1.80)	(-7.84)
10	0.95	0.95	1.07	-0.21	-0.02
(High)	(6.36)	(27.59)	(22.34)	(-3.91)	(-0.45)
L/S	1.08	-0.16	0.24	-0.13	0.20
(Equal weights)	(4.98)	(-3.14)	(3.51)	(-1.70)	(4.03)
L/S	0.65	-0.10	0.01	-0.08	0.16
(Value-weights)	(2.91)	(-2.00)	(0.16)	(-1.05)	

知乎 @石川

先来看看 Panel A。第一列是 L/S 组合相对于无风险收益率的超额收益。按等权重构建的该组合每月可以获得 1.17% 的超额收益；按照市值权重构建的投资组合每月获得 0.69% 的超额收益。Panel A 的第二到第六列汇报了考虑了其他常见的因子后，该 L/S 组合仍然能够获得的超额收益。

以第三列的 3-Factor alpha 为例，它的计算如下：**使用 TECHRET 因子的 L/S 组合收益的时间序列和 Fama-French 三因子 (Fama and French 1993) 的时间序列在时序上回归，得到的截距恰好就是截面上 L/S 无法被三因子解释的超额收益**（这是因为 Fama-French 三个因子本身是投资组合 MKT，SMB，HML 的收益率；需要这方面背景知识的小伙伴请参考[《股票多因子模型的回归检验》](#)）。

其他列考虑的不同因子模型之后获得的超额收益可以类似的解释。其中，4 factor model 是 Fama-French 三因子 + Carhart (1997) 的动量因子，5 factor model 是 Fama and French (2015) 提出的五因子模型，而 6 factor model 是该五因子加上动量因子。结果显示，市场上其他主流因子均无法解释 L/S 获得的超额收益。

知乎

首发于
川流不息

露。这意味着该策略在市场下行、以及小市值和动量股表现好的时候额外有效。

4 回归检验

除了 portfolio tests 之外，Lee et al. (2018) 还使用 Fama and MacBeth (1973) 进行了截面回归检验，其目的是为了在控制住其他变量后考察 TECHRET 因子对于股票截面收益差异的解释程度。

Fama-MacBeth 回归是在每个时点在截面上用因子载荷和个股的收益率进行回归，从而得到每期各因子的收益率，然后在时序上取平均就得到因子的预期收益率（需要进一步了解 Fama-MacBeth 回归的朋友请参考《股票多因子模型的回归检验》）。此外，Lee et al. (2018) 通过 Newey-West 调整求出了因子收益率的 standard error，从而计算出了 Fama-MacBeth t-statistics。下表给出了实证结果。

Dep. variable	(1)	(2)	(3)	(4)
$\times 100$	RET_t	RET_t	RET_t	$RET_t - INDRET_t$
$TECHRET_{t-1}$	0.629*** (4.10)	0.583*** (5.92)	0.735*** (6.54)	0.654*** (6.06)
$INDRET_{t-1}$			0.552*** (5.20)	0.046 (0.39)
SIZE		-0.834*** (-3.38)	-0.793*** (-3.19)	-0.805*** (-3.41)
BM		0.619*** (3.95)	0.584*** (3.53)	0.525*** (3.32)
GP		0.557*** (4.26)	0.442*** (3.46)	0.410*** (3.19)
AG		-0.453*** (-5.38)	-0.430*** (-4.84)	-0.466*** (-5.53)
RD		0.443* (1.67)	0.374 (1.50)	0.401* (1.91)
RET_{t-1}		-2.291*** (-12.79)	-2.164*** (-12.51)	-2.181*** (-12.49)
MOM		0.376* (1.83)	0.428** (1.99)	0.408** (2.10)
INTERCEPT	0.984** (1.98)	1.571** (2.45)	1.556*** (3.81)	1.012*** (3.48)
Industry Fixed Effect	Yes	Yes	No	No
N	540,895	540,895	540,895	540,895
Average R^2	0.091	0.142	0.076	0.059

在 (1) 到 (3) 列中，被解释变量是股票的收益率 RET。在截面回归中，解释变量是因子载荷。在这方面的处理上，作者并没有通过时序回归求解因子载荷，而是将股票在各个因子上的取值

知乎

首发于
川流不息

在上表中，除了那些我们熟悉的因子外，其他的因子包括 Gross Profitability (GP)、Asset Growth (AG)、R&D intensity (RD) 以及 INDRET，它是目标公司所在行业的市值加权收益率。结果表明，当控制了这些变量后，TECHRET 因子的预期收益率依然显著大于零（t-statistic 在 4 以上）。

在上表的第（4）列中，作者从 RET 中减去了 INDRET 作为解释变量，从而直接排除行业动量造成的潜在影响。即便如此，Fama-MacBeth 回归结果仍然表明，TECHRET 因子的预期收益率显著大于零，其 t-statistic 高达 6.06。

除了上述控制变量外，Lee et al. (2018) 还考虑了市场中存在的其他可能造成 lead-lag 收益率效应的关联，这其中包括 customer-supplier links (Menzly and Ozbas 2010) 以及 standalone-conglomerate firm links (Cohen and Lou 2012)。结果表明，这些已有关联并不能解释新发现的科技关联度。

另一方面，Burt and Hrdlicka (2016) 指出，存在某种关联的公司可能会在一些共同因子上有近似的暴露，导致在评价新因子时出现偏差。为了排除这个影响，Lee et al. (2018) 也在构建 TECHRET 因子时使用了股票的特异性收益率（即收益率减去 4 factor model 解释的部分）。使用根据特异性收益率计算的 TECHRET 因子选股，所构建的 L/S 组合仍然能够获得主流因子无法解释的超额收益（下图）。这也再次说明 TECHRET 和这些主流的因子之间在很大程度上是正交的。



知乎

首发于
川流不息

1	0.56	-0.01	-0.22	0.02	-0.18	0.01
(Low)	(2.00)	(-0.03)	(-1.94)	(0.19)	(-1.51)	(0.13)
2	0.72	0.17	-0.01	0.19	-0.01	0.15
	(2.74)	(1.30)	(-0.13)	(2.13)	(-0.11)	(1.70)
3	0.73	0.18	0.00	0.13	-0.03	0.08
	(2.87)	(1.58)	(0.03)	(1.65)	(-0.36)	(0.98)
4	0.84	0.32	0.13	0.27	0.10	0.22
	(3.45)	(2.84)	(1.69)	(3.70)	(1.31)	(3.01)
5	0.89	0.36	0.17	0.30	0.14	0.25
	(3.64)	(3.24)	(2.24)	(4.12)	(1.83)	(3.44)
6	1.01	0.48	0.30	0.42	0.31	0.40
	(4.12)	(4.44)	(4.05)	(5.68)	(4.04)	(5.43)
7	1.09	0.56	0.40	0.48	0.46	0.52
	(4.44)	(5.09)	(5.75)	(6.88)	(6.59)	(7.50)
8	1.14	0.60	0.47	0.53	0.56	0.61
	(4.48)	(4.86)	(5.76)	(6.49)	(7.11)	(7.62)
9	1.47	0.92	0.82	0.91	0.97	1.03
	(5.40)	(6.26)	(8.18)	(9.00)	(10.35)	(10.95)
10	1.54	0.98	0.87	0.88	1.07	1.06
(High)	(5.09)	(5.01)	(6.32)	(6.22)	(8.68)	(8.45)
L/S	0.98	0.99	1.09	0.86	1.25	1.05
(Equal weights)	(5.16)	(5.19)	(5.84)	(4.62)	(6.94)	(5.97)
L/S	0.67	0.68	0.80	0.59	0.93	0.74
(Value weights)	(3.26)	(3.32)	(3.88)	(2.84)	(4.46)	(3.60)

知乎 @石川

除了本小节介绍的这些检验之外，Lee et al. (2018) 中还包括了更多的 robustness tests，由于篇幅的问题就不逐一介绍了，感兴趣的小伙伴请阅读原文。

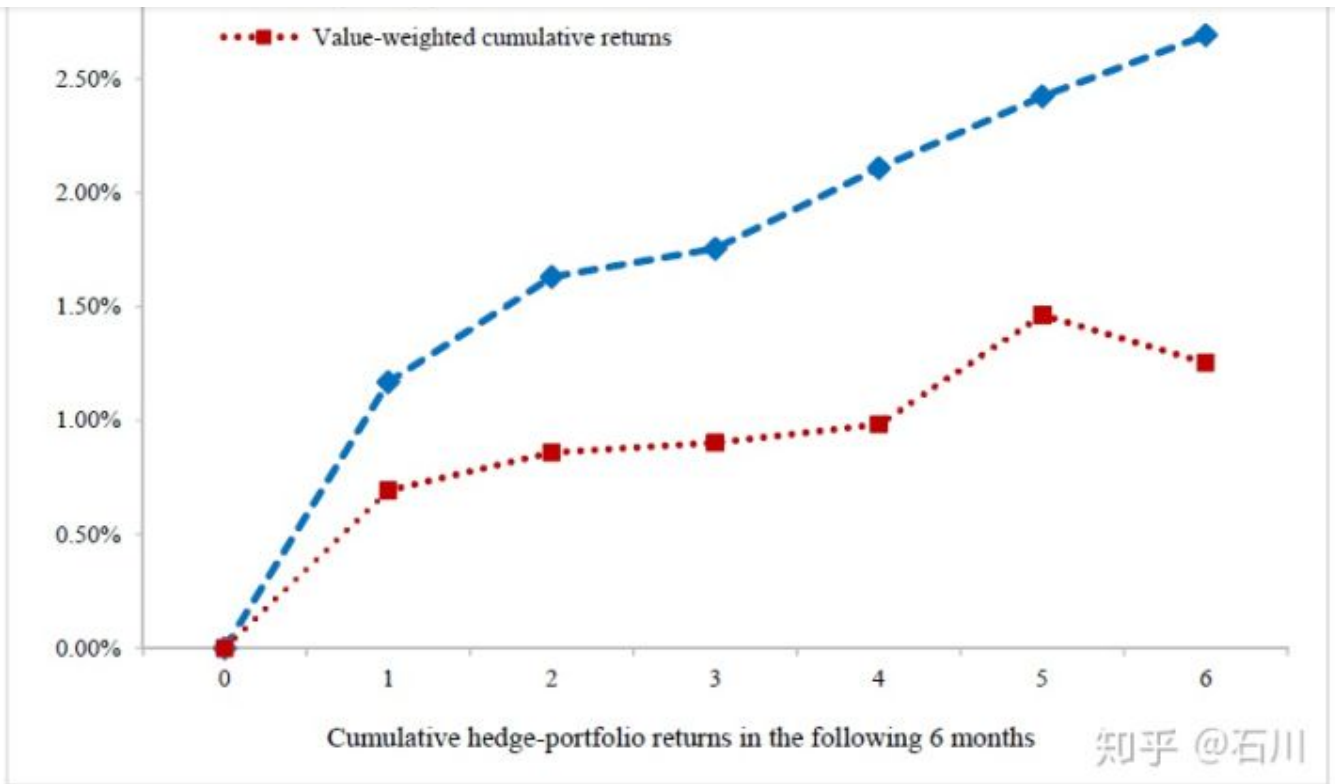
5 内在有效机制

前文的结果说明 TECHRET 因子在选股方面确实有效，而且它获得的超额收益不能被市场中常见的其他因子解释。本节就来看看它为何有效。

对于超额收益，学术界和业界主流的两种解释是**错误定价**和**风险补偿**。搞清楚 TECHRET 背后的机制至关重要：错误定价意味着投资者可以通过合理的策略获得潜在的超额收益；而风险补偿则意味着投资者获得的收益是以承担额外风险为代价的。本节和下一节分别考察错误定价和风险补偿这两种解释。

在考察该因子获取的超额收益的可持续性上，作者发现 L/S 投资组合在未来几个月内都可以持续获得收益（下图）；表明科技动量是一个价格发现的过程，随着投资者逐渐意识到科技关联公司带来的新息，股价也随之反映完全。**这或许说明价格对于与科技有关的基本面消息的吸收是缓慢的，导致了错误定价。**

知乎

首发于
川流不息

为验证上述猜想，Lee et al. (2018) 研究了以下三个方面：

1. 科技相关新息 (innovation) 的性质；
2. 投资者对这类新息的有限注意力 (limited attention) ；
3. 投资者的套利成本。

在第一方面，实证结果表明，TECHRET 因子的强度和目标公司的 technology intensity 以及 technology specificity (强度和专度) 有关。举例来说，在强度方面，R&D 开销大的公司获得的 TECHRET 因子收益更高；在专度方面，该文以专利集中度作为衡量专度的指标并发现专度高的公司获得的 TECHRET 因子收益更高。

Lee et al. (2018) 指出，对于行业应用集中度高的专利类别来说，科技新息被价格反映的速度更慢一些。而上述的结果与这个说法一致。对于科技专度更高的公司，TECHRET 因子包含了更多的该公司的估值信息。此外投资者对于科技专度高的公司的估值变化反应不足。这两种原因导致了较慢的信息扩散过程。

在第二方面，为了检验投资者的 limited attention，作者的猜想是关注度低的公司——特征是市值小、分析师报告和媒体报道更少、机构投资者占比低——可以获得更高的 TECHRET 因子收益率。为此，作者分别构建了 dummy 指标，并进行了回归分析，结果证实了上述猜想。

在最后一方面，作者的假设是那些套利成本高的公司能够获得更高的 TECHRET 因子收益率。1 特异性波动率 (Baker and Wurgler 2006, 2007) 以及负面新闻 (Hong et al. 2000) 作为套成本的代理指标，Lee et al. (2018) 的分析结果和上述猜想一致。对于 limited attention 和套成本方面的回归分析结果如下表所示。

知乎

首发于
川流不息

$TECHRET_{t-1}$	0.839*** (5.58)	0.829*** (5.07)	0.897*** (4.75)	0.838*** (4.41)	0.293*** (3.33)	0.544*** (3.41)
$TECHRET_{t-1} \times$	-0.592***					
$Size > median$	(-3.05)					
$TECHRET_{t-1} \times$		-0.434**				
$Analyst > median$		(-2.50)				
$TECHRET_{t-1} \times$			-0.519***			
$InstitOwn > median$			(-2.72)			
$TECHRET_{t-1} \times$				-0.804***		
$News > median$				(-3.00)		
$TECHRET_{t-1} \times$	前四个 dummy 变量是大于均值时取值为 1。它们的回归结果是负数说明小于均值的公司能获得更高的超额收益。				0.509*** (3.35)	
$IdioVol > median$						0.933** (2.15)
$TECHRET_{t-1} \times$						
$Bad\ News$						
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	540,895	462,323	412,786	165,300	540,895	540,895
Average R^2	0.146	0.119	0.119	0.108	0.147	0.145

上述结果从科技新息性质、投资者的有限注意力以及套利成本方面证实了价格对与科技类基本面消息的吸收是缓慢的，从而造成了错误定价。

6 风险解释站不住脚

除了错误定价这种解释外，另一种常见的解释是从风险补偿的角度，即因子之所以获得超额收益是因为它暴露于某种未知的风险。然而 Lee et al. (2018) 的分析说明，这种解释并不成立。

由于“未知”，我们不可能罗列所有潜在的风险然后考察 TECHRET 因子在上面的暴露如何。取而代之，Lee et al. (2018) 从另外四个角度来分析风险补偿说。本文着重介绍其中的两个。

第一个是考察股票在盈余公告期的收益情况，这是一种被学术界普遍认可的方法。它背后的逻辑是，如果某个异象和错误定价有关，则该因子在盈余公告期内应该比其他时间内获得更高的收益，这是因为最新的盈余报告有助于修正投资者之前对该股票的估值错误。而反过来，如果该异象是源自风险补偿，我们将不会观察到上述现象，换句话说，该因子在不同时期（无论是否盈余公告期内）的收益率应该大致相当。

回归结果（下表）说明，在考虑了一系列必要的控制变量后，盈余公告期内 TECHRET 因子能获得非盈余公告期内 4 倍以上的收益率，这是风险补偿说完全无法解释的。

<i>Dep. variable</i>	(1)	(2)
$\times 100$	<i>DLYRET</i>	<i>DLYRET</i>
<i>TECHRET</i>	0.291 [*]	0.398 ^{**}
	(1.69)	(2.25)
<i>TECHRET</i> \times <i>EDAY</i>	1.610 ^{***}	1.660 ^{***}
	(4.85)	(4.97)
<i>EDAY</i>	0.264 ^{***}	0.272 ^{***}
	(14.02)	(14.41)
<i>Lagged Controls</i>	No	Yes
<i>Day Fixed Effects</i>	Yes	Yes
<i>N</i>	10,522,482	10,522,112
<i>Average R²</i>	0.105	0.112

知乎 @石川

第二个角度是 standardized unexpected earnings (SUE, 未预期盈余)，它是一个非收益率指标，因此不会被对风险的控制不足所影响。Lee et al. (2018) 检验了 TECHRET 能否预测未来的 SUE。由于 SUE 是公司未来现金流的决定因素，如果 TECHRET 能够预测 SUE 则说明前者带来的超额收益和公司基本面的改变相关，而非风险补偿。

实证结果如下表所示，它说明 TECHRET 对 SUE 有统计上显著的预测性。此外，Panel B 的结果表明，当前季度的 TECHRET 对未来三个季度的 SUE 都有显著的预测性，且这种预测性在逐步减弱。这一结果有力的佐证了该因子可能来源于错误定价，而非风险补偿。

	SUE_t	SUE_t	SUE_t	SUE_t
$TECHRET_{t-1}$	0.203*** (8.65)	0.128*** (7.36)	0.127*** (6.51)	0.079*** (4.29)
SUE_{t-1}		1.674*** (37.54)	1.847*** (37.59)	1.905*** (26.78)
SUE_{t-2}		0.655*** (28.90)	0.751*** (29.32)	0.736*** (23.63)
SUE_{t-3}		0.427*** (21.77)	0.521*** (23.29)	0.494*** (19.09)
SUE_{t-4}		-0.683*** (-23.32)	-0.518*** (-15.85)	-0.489*** (-13.54)
<i>Firm Fixed Effect</i>	Yes	Yes	No	No
<i>Industry Fixed Effect</i>	No	No	Yes	Yes
<i>Quarter Fixed Effect</i>	Yes	Yes	Yes	No
<i>N</i>	101,922	101,922	101,922	101,922
<i>Adj/Avg. R²</i>	0.213	0.408	0.362	0.412

Panel B: Earnings predictability over longer periods

	(1) SUE_t Quarter 1	(2) SUE_{t+1} Quarter 2	(3) SUE_{t+2} Quarter 3	(4) SUE_{t+3} Quarter 4
$TECHRET_{t-1}$	0.079*** (4.29)	0.064*** (3.01)	0.032* (1.76)	0.008 (0.37)
<i>Controls</i>	Yes	Yes	Yes	Yes
<i>N</i>	101,922	99,179	96,448	93,835
<i>Average R²</i>	0.412	0.413	0.413	0.412

知乎@石川

此外，Lee et al. (2018) 还指出技术变化风险以及竞争替代风险均无法解释 TECHRET 因子。综合本节和上一节的结果，Lee et al. (2018) 认为，TECHRET 获得超额收益的原因在于人们对科技新息可造成的股价变化反应不足，而非额外的风险补偿。

7 结语

Lee et al. (2018) 是一篇研究因子的典范。

以下高度概括一下它的行文逻辑：首先它从世界经济发展造成的公司之间越来越密切的联系出发提出了科技关联度这个新视角，并选择了适当的代理指标（专利分布之间的相关系数）来计算公司之间科技关联度的强弱。为了检验这个因子在解释股票截面预期收益率差异上的作用，该文使用了业界广泛流行的 portfolio tests 和 regression tests 指出该因子确实能够获得超额收益，并通过系列更为细致的 robustness tests 来确认这一点。该文最后错误定价和风险补偿两个主流角度析了该因子有效的内在机制，并指出它背后的原理是投资者对于科技新息的反应不足。

知乎

首发于
川流不息

(专利数据、股票数据等)；其次是经过几十年来无数学者在顶级期刊上发表的丰硕成果的积淀，一个新的因子被提出后，应该进行哪些 tests、使用哪些主流因子来分析这个新因子、以及如何识别该因子是源自风险补偿还是投资者对它的反应不足等，有大量被反复验证过的文献形成一个科学的分析框架。这种积累不是一朝一夕能形成的，这实在是让人羡慕，也值得我们的学术界和业界学习。

我一直以来的观点是，在寻找 α 的努力中，最重要的是获得新的数据或者是对已有数据的创新用法。与数据比起来，算法先进性的作用相对有限。在这方面，Lee et al. (2018) 的贡献尤为突出，提出了获取 α 的新视角。

这是真正的计量经济学。

这是真正的金融工程。

这是真正为人们理解股票截面预期收益差异而做出的卓越努力。

参考文献

- Baker, M. and J. Wurgler (2006). Investor Sentiment and the Cross-section of Stock Returns. *Journal of Finance*, Vol. 61(4), 1645 – 1680.
- Baker, M. and J. Wurgler (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, Vol. 21(2), 129 – 152.
- Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, Vol. 81(4), 1347 – 1393.
- Burt, A. and C. M. Hrdlicka (2016). Understanding Network-based Measures of Information Diffusion. Unpublished working paper, University of Washington.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, Vol. 52(1), 57 – 82.
- Cohen, L. and D. Lou (2012). Complicated Firms. *Journal of Financial Economics*, Vol. 104(2), 383 – 400.
- Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *American Economic Review*, Vol. 76(5), 984 – 1001.
- Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, Vol. 81(3), 607 – 636.
- Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, Vol. 33(1), 3 – 56.
- Fama, E. F. and K. R. French (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, Vol. 116(1), 1 – 22.
- Lee, C. M. C., S. T. Sun, R. Wang, and R. Zhang (2018). Technological Links and Return



知乎

首发于
川流不息Returns. *Journal of Finance*, Vol. 65(4), 1555 – 1580.**免责声明：**文章内容不可视为投资意见。市场有风险，入市需谨慎。

原创不易，请保护版权。如需转载，请联系获得授权，并注明出处，谢谢。已委托“维权骑士”
([维权骑士_免费版权监测/版权保护/版权分发](#)) 为进行维权行动。

编辑于 2019-07-03

[多因子模型](#) [有效市场假说](#) [资产定价](#)[▲ 赞同 106](#) [▼](#) [● 18 条评论](#) [🔗 分享](#) [★ 收藏](#) [...](#)

文章被以下专栏收录

**川流不息**

北京量信投资管理有限公司是一家在中国基金业协会备案登记的专业私募基金管理人...

[关注专栏](#)

推荐阅读

**机器学习能否助力风险投资？**

石川

多因子模型水平测试题

作者：李腾 链接：
<https://zhuanlan.zhihu.com/p/20190824>
来源：知乎 著作权归作者所有，转
载请联系作者获得授权。本试题由
李腾、陈烨、邓岳、@陈志岗 整
理，欢迎补充！ 1 试卷说明 测试...

jerry...

发表于量化投资那...

**【012】A
alpha, 大**

llangli

18 条评论[⇌ 切换为时间排序](#)[写下你的评论](#)

知乎

首发于
川流不息

llanglli

11 个月前

确实是研究典范。你的总结整理也很精彩，赞一个

👍 1



石川 (作者) 回复 llanglli

11 个月前

谢谢

👍 赞



DavidCharge

11 个月前

最近几年这个field的bar提高的很快，不像以前那么容易灌水了。不过美国的专利数据随着时间有变动的，将来回头看可复制性不一定高

👍 3



石川 (作者) 回复 DavidCharge

11 个月前

密切关注下，谢谢

👍 赞



王二二

11 个月前

为啥要使用uncentered correlation呢，这个和一般的correlation有什么区别呢

👍 赞



石川 (作者) 回复 王二二

11 个月前

uncentered correlation 的定义和两个向量余弦是完全一致的，它的取值范围是 0 到 1 之间，两个向量越接近，它们的夹角越接近 0，余弦越接近 1；两个向量越相互垂直，它们的余弦越接近 0。一般的 correlation 的取值是 -1 到 1。余弦或者 uncentered correlation 是一种常见的计算两个向量相似性或“距离”的指标。Lee et al. 2018 也是承袭了最早的 Jaffe 1986 的方法。

👍 2



陈皇宇 Renco

11 个月前

昨天刚好看到这篇文章哈哈

👍 赞



石川 (作者) 回复 陈皇宇 Renco

11 个月前

之前看大神的文章和回答学了很多东西！感谢大神回复！

👍 赞



陈皇宇 Renco 回复 石川 (作者)

11 个月前

哈我就一普通人，互相学习，共同进步！



展开其他 1 条回复



雪在烧

11 个月前

这个有没有幸存者误区嫌疑？科技发展太快，落后的被快速淘汰，于是测试的数据都是幸存者？

👍 1



石川 (作者) 回复 雪在烧

11 个月前

找机会我帮你问问原作者。。。感谢留言。

👍 赞



哈德逊河畔的鹅

11 个月前

感谢大大分享，之后通读一遍。所以说alternative data真是太重要了

👍 赞



shiqiang

11 个月前

“首先是研究美股的数据非常完善（专利数据、股票数据等）；其次是经过几十年来无数学者在顶级期刊上发表的丰硕成果的积淀，一个新的因子被提出后，应该进行哪些 tests、使用哪些主流因子来分析这个新因子、以及如何识别该因子是源自风险补偿还是投资者对它的反应不足等，有大量被反复验证过的文献形成一个科学的分析框架。”

这个确实太重要了，国内这几年也在慢慢积累吧。

也同样您的观点：超额收益来自于超额的信息，或者对现有信息的加工获取到的更多新超额信息。

👍 2



浮士德2.0

11 个月前

看过您的好几篇文章。发现您无论在写文章，还是读文章，都很追求结构上的完整和逻辑的严谨精练。这点我和你很像（笑哭）

👍 赞



石川 (作者) 回复 浮士德2.0

11 个月前

谢谢鼓励！这评价很高啊，不敢当。🙏

👍 赞



stardust

9 个月前

非常同意:寻找 α 的努力中，最重要的是获得新的数据或者是对已有数据的创新用法。与数据比起来，算法先进性的作用相对有限。但是最可悲的是，很多大机构内部过于强调开发算法。

知乎



首发于
川流不息



Messiah

5 个月前

最近在看三级的equity，这篇文章很有启发

赞

