

基本面关联与股票收益可预测性

——基于供应链关联、地理关联与技术关联的解读

解读人：周颜锐

2023年2月16日

文献介绍

1. Market Segmentation and Cross-predictability of Returns

- author: LIOR MENZLY and OGUZHAN OZBAS
- journal: THE JOURNAL OF FINANCE
- year: 2010
- cited by 537

2. Geographic Lead-Lag Effects

- author: Christopher A. Parsons and Riccardo Sabbatucci
- journal: Review of Financial Studies
- year: 2020
- cited by 78

文献介绍

3.Technological links and predictable returns

- author: Charles M.C. Lee, Stephen Teng Sun, Rongfei Wang, Ran Zhang
- journal: Journal of Financial Economics
- year: 2019
- cited by 98

文献介绍

文献内容

- 基础研究：探究基本面关联带来的收益率交叉可预测性。
例如：企业A和企业B-基本面关联-基本面指标关联-收益率关联
- 进一步研究：可预测性的理论成因、可预测性的影响因素、基于可预测性指标投资组合的表现。

思想根源

- 对有效市场假说的拓展，即对可预测性的解释，如有限信息模型。

文献介绍

思想根源

- Case1 单个企业、两类投资者
知情投资者&普通投资者
- Case2 同行业多企业、两类投资者
知情投资者的覆盖特点
- Case3 不同行业但基本面关联企业、两类投资者

供应链关联与收益率可预测性

引言

- 有效市场假说有一个颇受质疑的前提假设，即参与市场的投资者有足够的理性，并且能够迅速对所有市场信息作出合理反应。该理论认为，在法律健全、功能良好、透明度高、竞争充分的股票市场，一切有价值的信息已经及时、准确、充分地反映在股价走势当中，其中包括企业当前和未来的价值，除非存在市场操纵，否则投资者不可能通过分析以往价格获得高于市场平均水平的超额利润。
- Hong and Stein(1999)：放松有效市场假说，认为信息能在投资者间逐步扩散，而这一假设能解释观测到的收益可预测性（过去的收益能预测未来的收益）。为检验信息渐进扩散假说，文献主要集中于研究资产对自身过去收益的滞后价格反应。

供应链关联与收益率可预测性

引言

- 在本文中，我们探索了一种与资产自身过去收益相分离的新的信息流动渠道，即对来自供应商和客户行业相关企业的冲击的延迟价格反应。
- 主要贡献：
 - 1表明投资者专业化以及由此产生的市场信息分割对价格的形成具有显著影响；
 - 2经济相关行业的信息逐渐扩散是普遍存在的；
 - 3在有限信息模型中，基本面相关企业的收益率具备交叉可预测性，而事实也是如此。

供应链关联与收益率可预测性

Section1 基本设定

A. 有限信息模型中的交叉可预测性

模型假设：

- （1）假设存在两个市场，三个日期，两类投资者；
- （2）知情投资者专注于某种市场，并能在中间阶段收到关于最终现金流的信号；
- （3）不知情投资者不会收到上述信号，他们中至少有一些人不能处理信息（要么是信息处理能力有限，要么是对其来说处理信息的成本太高）。

供应链关联与收益率可预测性

Section1 基本设定

A. 有限信息模型中的交叉可预测性

当达到中间阶段时，信息信号出现，知情投资者基于自身需求将其吸收并将信号反映到价格的变动中，但从整体的视角来看，投资者对信号的吸收是部分的。这是因为不知情投资者处理信息的能力有限或成本过高，从而其不能即时形成理性期望推断并依据理性期望来调整自身需求。结果就是，收益率呈现出正的自相关性。

供应链关联与收益率可预测性

Section1 基本设定

A. 有限信息模型中的交叉可预测性

在两个资产市场中，由于两个额外的假设，收益表现出交叉可预测性：（1）两个市场具有相关的基本面，因此一个市场中的信息信号具有关于另一个市场最终收益的信息内容；（2）两个市场是信息分割的。

由于知情投资者的专门化，源自其中一个市场的信息信号只被那些恰好专门从事该市场的投资者接收，而不被专门从事另一个市场的投资者接收。因此，具有跨市场内容的信息信号仅部分地被纳入价格，价格变化将会呈现可预测性。

供应链关联与收益率可预测性

Section1 基本设定

B. 实证设计

- 为什么要检验供应链上的交叉可预测性效应？

基本面相关：供应商和客户企业之间可能存在相关的现金流量，因为它们之间要么直接通过交易关系进行互动。供应链上的企业也可能面临相对相似的需求或技术冲击。

- 我们为什么要关注产业所界定的经济边界？

由行业定义的经济边界也很可能代表投资者专业化引致的信息边界。事实上，股票分析师和基金经理等重要的经济代理人倾向于按照行业进行专业化分工。

供应链关联与收益率可预测性

Section1 基本设定

C. 相关文献

- 关于收益可预测性的文献，探讨了股票之间的领先滞后关系：

Hong et al. (2007): 美国的几个重要行业，如商业地产、石油、金属、零售、金融、服务等，引领着整个股市以及各项经济指标。

- 使用分析师覆盖率和机构所有权等代理变量来研究知情交易者的文献：

Hong, Lim, and Stein (2000): 表明有更多分析师关注的股票表现出更少的价格动量。这与分析师关注越多的股票对价值相关信息调整越快的观点一致。

供应链关联与收益率可预测性

Section2 实证结果

A. 数据

- CRSP and COMPUSTAT

收益率数据来自证券价格研究中心(CRSP)月度收益率数据库。样本期间为1963年7月至2005年6月。与标准做法一致，我们使用NYSE、Amex和NASDAQ股票，并剔除封闭式基金、房地产投资信托基金等。

- I/B/E/S and 13F Holdings

使用I/B/E/S数据库构建CRSP上每只股票的分析师覆盖率；使用13F Holdings数据库构建CRSP上每只股票的机构持股度量。

供应链关联与收益率可预测性

Section2 实证结果

A. 数据

- Benchmark Input-Output Surveys

使用经济分析局（BEA）的一系列基准投入产出调查来识别给定股票或行业的供应商和客户行业。BEA调查通过在表中报告商品和服务的行业间流动数量，提供了美国经济相互依赖结构的详细信息。

基于行业指派，我们接下来计算流通市值加权月度行业收益率。在计算行业组合收益后，我们进一步对每个行业形成两个单独的组合，一个由供应商行业组成，另一个由客户行业组成。为了计算这些投资组合的月回报率，我们使用商品和服务进出相关行业的流量数据作为投资组合权重。

供应链关联与收益率可预测性

Section2 实证结果

B. 初步证据

供应链上的企业或行业是否具有相关的基本面：

$$ROA_{i,t} = \alpha_i + \theta^m ROA_t^m + \theta^s ROA_{i,t}^s + \theta^c ROA_{i,t}^c + e_{i,t}, (1)$$

使用经双重聚类调整的稳健标准误 (Thompson (2009), Petersen (2009))，按公司和年份划分为公司层面，按行业 and 年份划分为行业层面。因此，剩余的独立性来源是不同年份的不同企业或行业之间。

供应链关联与收益率可预测性

Section2 实证结果

B. 初步证据

Table 1 Supply Chain Fundamentals

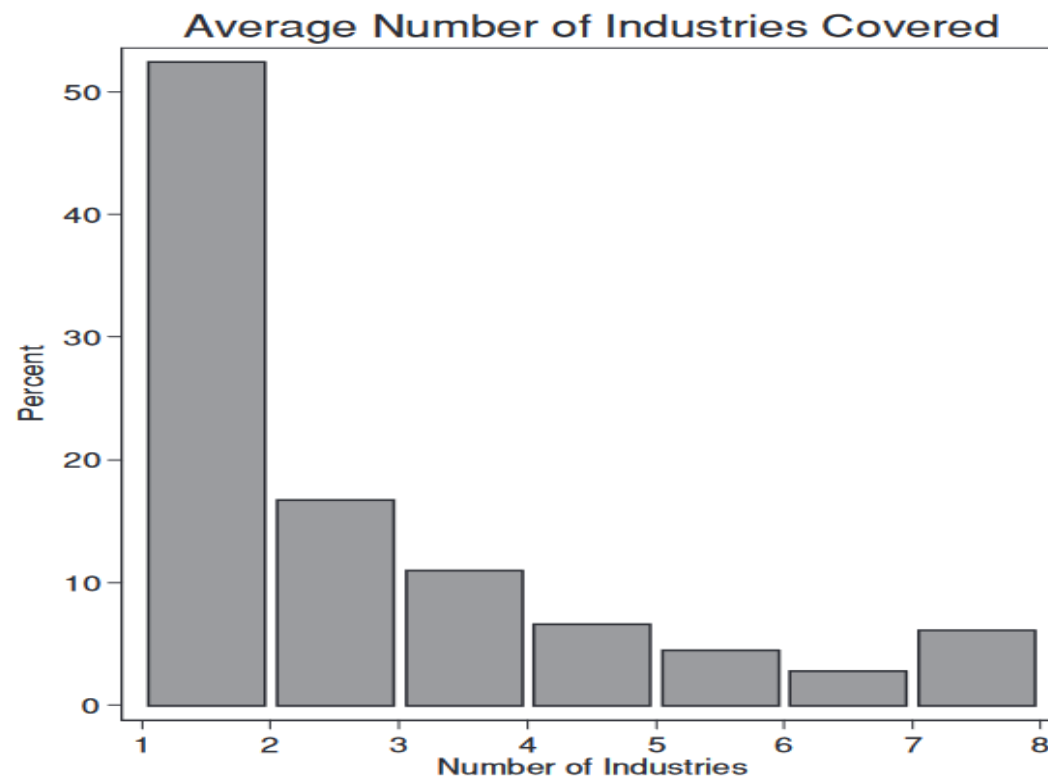
<i>Dependent Variable: ROA</i>	(1)	(2)
ROA_{market}	0.725*** (5.16)	0.172*** (2.60)
ROA_{supplier}	0.471*** (2.68)	0.398*** (4.53)
ROA_{customer}	0.465*** (2.98)	0.164* (1.90)
Fixed Effects	Yes	Yes
Clustered Standard Errors	Yes	Yes
R^2	0.592	0.394
N obs	275,291	3,002

供应链关联与收益率可预测性

Section2 实证结果

B. 初步证据

通过提供股权分析师专业化的证据来解决知情投资者的专业化假设



供应链关联与收益率可预测性

Section2 实证结果

C. 股票和行业层面的交叉预测效应

现在测试基于供应商和客户行业的滞后回报，股票层面的回报是否具有交叉可预测性。为此，我们进行Fama-MacBeth(1973)回归：

$$r_{i,t} = \alpha_t + \lambda_t^s r_{i,t-1}^s + \lambda_t^c r_{i,t-1}^c + \Lambda_t Z_{i,t-1} + e_{i,t}, (2)$$

其中 $Z_{i,t}$ 是滞后控制变量，如股票和行业层面的短期反转与中期动量，以及FamaFrench三因子。

供应链关联与收益率可预测性

Section2 实证结果

Table 2 Cross-predictability Effects in Stock and Industry Returns

	(1)	(2)	(3)	(4)
Constant	0.005** (2.01)	0.006** (2.33)	0.006** (2.45)	0.006** (2.35)
$r_{\text{supplier},t-1}$	0.114*** (5.03)	0.105*** (4.37)	0.117*** (5.04)	0.113*** (4.88)
$r_{\text{customer},t-1}$	0.071*** (4.11)	0.058*** (3.26)	0.059*** (3.46)	0.075*** (4.27)
$r_{\text{stock},t-1}$	-0.062*** (14.44)	-0.047*** (10.41)	-0.067*** (13.84)	
$r_{\text{stock},t-2:t-12} \times \frac{1}{11}$	0.072*** (4.07)	0.099*** (5.01)	0.099*** (5.37)	
$r_{\text{industry},t-1}$	0.134*** (15.52)	0.117*** (12.33)	0.137*** (15.29)	0.032*** (2.83)
R^2	0.028	0.051	0.034	0.091
T	492	492	492	503
Sample excludes observations:	—	<20 th NYSE percentile	No closing price at end of month $t - 1$	

供应链关联与收益率可预测性

Section2 实证结果

D. 知情投资者效应

- 有限信息模型的一个明确的预测是，在有更多的知情投资者的情况下，交叉可预测性的大小应该更低。我们现在转向测试这个预测。
- 为了代理分析师对某只股票的覆盖程度，我们计算了两种不同的衡量指标，即每月主动关注该只股票的分析师数量：
 - （1）如果分析师在过去12个月内对股票做出了预测，那么分析师是活跃的。
 - （2）要求分析师在给定的一个月内对该股票进行预测，以被认为该月是活跃的。

供应链关联与收益率可预测性

Section2 实证结果

D. 知情投资者效应

为了检验供应商和客户行业基于滞后收益的交叉可预测性效应是否对分析师覆盖率较高的股票更弱，我们使用Fama-Macbeth(1973)回归：

$$r_{i,t} = \alpha_t + \sum_{j=1}^5 \lambda_t^j A_{t-1}^j r_{i,t-1}^{related} + e_{i,t}, (3)$$

如果股票*i*在第*t* - 1个月的分析师关注水平将股票放在第*j*个分位数上，则 $A_{j,t-1}$ 等于1，否则为0。 $r^{related}$ 是供应行业收益率与客户行业收益率的均值。

供应链关联与收益率可预测性

Section2 实证结果

D. 知情投资者效应

Table 3 Analyst Coverage, Institutional Ownership, and Cross-predictability Effects

	(1)	(2)	(3)
Constant	0.008* (1.95)	0.008* (1.89)	0.008** (1.97)
$r_{\text{composite},t-1} \times \text{Rank}_{t-1}(\text{1}^{\text{st}} \text{ Quintile} - \text{Low})$	0.293*** (4.98)	0.300*** (4.95)	0.380*** (5.89)
$r_{\text{composite},t-1} \times \text{Rank}_{t-1}(\text{2}^{\text{nd}} \text{ Quintile})$	0.299*** (5.12)	0.259*** (4.65)	0.317*** (5.55)
$r_{\text{composite},t-1} \times \text{Rank}_{t-1}(\text{3}^{\text{rd}} \text{ Quintile})$	0.185*** (3.27)	0.201*** (3.71)	0.244*** (4.50)
$r_{\text{composite},t-1} \times \text{Rank}_{t-1}(\text{4}^{\text{th}} \text{ Quintile})$	0.143** (2.37)	0.157** (2.61)	0.177*** (3.16)
$r_{\text{composite},t-1} \times \text{Rank}_{t-1}(\text{5}^{\text{th}} \text{ Quintile} - \text{High})$	0.027 (0.43)	0.034 (0.55)	0.067 (1.12)
R^2	0.014	0.012	0.012
T	281	281	303

供应链关联与收益率可预测性

Section3 自融资交易策略

- 在这一节中，我们分析了基于上一节记录的交叉可预测性效应的自融资交易策略的盈利能力。由于从我们的角度分析的一个重要目的是评估交叉可预测性的经济意义，我们在这里只关注涉及行业买卖的交易策略，而不是涉及个股买卖的交易策略。
- 我们使用行业构建自融资交易策略的方式是相当标准的。在每个月初，我们根据上个月供应商和客户行业的回报将行业分为五个仓。我们将前一个月相关行业回报率处于下五分位的行业分配到第一分位，前一个月相关行业回报率处于第二分位的行业分配到第二分位，依此类推。

供应链关联与收益率可预测性

Section3 自融资交易策略

Table 5 Self-financing Trading Strategies(figures are annualized)

	Low (1)	(2)	(3)	(4)	High (5)	H – L
Panel A: Industries Sorted on $r_{\text{supplier},t-1}$						
Mean return	0.028	0.054	0.056	0.072	0.100	0.073
Standard deviation	0.159	0.175	0.178	0.178	0.162	0.110
Sharpe ratio	0.173	0.306	0.313	0.406	0.617	0.660
Panel B: Industries Sorted on $r_{\text{customer},t-1}$						
Mean return	0.016	0.060	0.052	0.065	0.085	0.070
Standard deviation	0.176	0.166	0.157	0.165	0.184	0.134
Sharpe ratio	0.090	0.361	0.333	0.395	0.463	0.520
Panel C: Industries Sorted on $r_{\text{composite},t-1}$						
Mean return	0.013	0.036	0.074	0.071	0.100	0.087
Standard deviation	0.169	0.165	0.172	0.174	0.172	0.132
Sharpe ratio	0.077	0.220	0.430	0.409	0.582	0.658

供应链关联与收益率可预测性

Section3 自融资交易策略

A.Return Factor Exposure

Table 6 Return Factor Exposure

Trading Strategy Based on	$r_{\text{supplier},t-1}$	$r_{\text{customer},t-1}$	$r_{\text{composite},t-1}$
Alpha	0.005*** (3.42)	0.005*** (2.61)	0.006*** (3.22)
$R_{\text{market}} - R_f$	0.017 (0.48)	0.051 (1.08)	0.038 (0.89)
<i>SMB</i>	0.016 (0.35)	-0.065 (1.15)	0.004 (0.07)
<i>HML</i>	-0.013 (0.24)	0.054 (0.82)	0.015 (0.23)
<i>MOM</i>	0.102*** (2.89)	0.068 (1.59)	0.144*** (3.41)
R^2	0.019	0.012	0.024
<i>N obs</i>	503	503	503

供应链关联与收益率可预测性

Section3 自融资交易策略

B.Evidence from Long/Short Equity Hedge Funds

接下来，我们检验了由最有可能利用可预测性效应的复杂市场参与者组成的多/空股票对冲基金的收益序列是否与上述三种(供应商、客户和复合)交易策略的收益序列相关。如果收益交叉可预测性是一个经济上重要的现象，人们会期望这些对冲基金利用它，他们的收益序列显示了我们的交易策略的痕迹。

供应链关联与收益率可预测性

Section3 自融资交易策略

B.Evidence from Long/Short Equity Hedge Funds

Table 7 Credit Suisse/Tremont Long/Short Equity Hedge Fund Index

	(1)	(2)	(3)	(4)
<i>Alpha</i>	0.001 (1.17)	0.001 (1.01)	0.001 (0.78)	0.001 (0.92)
<i>R_{market} - R_f</i>	0.488*** (15.26)	0.489*** (15.59)	0.497*** (15.96)	0.487*** (15.99)
<i>SMB</i>	0.216*** (6.62)	0.216*** (6.74)	0.220*** (6.95)	0.215*** (6.89)
<i>HML</i>	-0.027 (0.64)	-0.014 (0.33)	-0.021 (0.52)	-0.019 (0.48)
<i>MOM</i>	0.219*** (9.71)	0.212*** (9.52)	0.222*** (10.14)	0.205*** (9.40)
<i>Supplier Strategy</i>		0.071** (2.52)		
<i>Customer Strategy</i>			0.060*** (3.08)	
<i>Composite Strategy</i>				0.080*** (3.78)
<i>R²</i>	0.806	0.815	0.819	0.825
<i>N obs</i>	138	138	138	138

供应链关联与收益率可预测性

Section4 总结

在本文中，我们发现了基于供应商和客户行业的滞后收益率与企业 and 行业层面的收益率是交叉可预测的。这种效应在经济上似乎是显著的：利用它们的交易策略每年产生高达8.7%的溢价。此外，我们发现对于分析师关注和机构持股水平较高的股票，交叉预测效应较弱。我们还发现，机构投资者的交易行为符合知情投资者从交叉可预测收益中获利的交易行为。所有这些发现都符合有限信息模型。

地理关联与收益率可预测性

1 引言

在本文中，我们重新考虑了分析师关注与股票收益领先-滞后效应之间的关系，但至关重要的一点是，要区分分析师关注在单个公司层面的衡量，以及在两两公司之间的衡量——即不同公司被一个或多个共同分析师关注的程度（分析师数量削弱可预测性的途径之二）。

为寻求那些即使其个体成员暴露于共同的基本面变化，但通常不被同一组分析师所覆盖的公司群体，本文找到了地理关联这一刻画指标。

地理关联与收益率可预测性

1 引言

与地理上的领先滞后一个自然的比较是行业领先滞后，它涉及同一行业的企业（且其总部设在不同的城市）。行业关联指的是受到共同基本面冲击影响的同行企业之间存在的关联性，且同行业公司一般会受到共同分析师的关注。事实上，第一篇文献也得到过分析师按行业进行专业化的结论，可沿用此观点。

地理关联与收益率可预测性

2 数据与描述性统计

- 企业位置：

我们的分析涉及总部设在或接近美国最大的20个城市中心的股票。为了构建我们的样本，我们首先从1970 - 2013年期间在纽约证券交易所、纳斯达克、美国证券交易所交易的国内普通股(标码10和11)的范围开始。然后，我们根据COMPUSTAT数据库中总部所在地对应的邮政编码(ZIP)为每个企业分配一个位置变量。

总部所在地按照经济分析局(Bureau of Economic Analysis)定义的经济区域(Economic Area, EA)进行分组。EAs旨在捕获经济活动的本地节点，通常涉及一个主要的大都市地区。

地理关联与收益率可预测性

2 数据与描述性统计

- 行业分类：

除了按照总部所在地对企业进行分类外，我们还按照企业所属行业进行分组。每个月，我们将每个企业链接到一个单一的Fama-French12行业，该行业通过SIC（标准产业分类体系）指定对企业进行分组。

- 其他数据（控制变量）：

遵循Harvey et al.(2016)和Novy-Marx(2013)的做法，我们使用以下主要控制变量：公司自身滞后1个月的收益、个股12个月的动量、公司规模、账面市值比、交易量、总盈利能力、资产增长和机构持有。

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

模型假设：

- 存在 $t=0, 1, 2$ 三个日期， $i \in \{1, 2, 3\}$ 三家企业。
- 利率为0，投资者风险中性。
- 每个企业在 $t=2$ 时进行一次分红。分红的多寡取决于三个因素：行业因素 I ，地理因素 L ，企业特有因素 e 。
- 存在A、B两个行业，X、Y两个地理区域。企业1、2位于同一行业A，因此共享行业冲击 I_A ；企业2、3的总部都在Y地区因此其共享地理冲击 L_Y 。

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

结合上述假设，可得各个企业在 $t=2$ 时的分红为：

$$\pi_1 = I_A + L_X + e_1$$

$$\pi_2 = I_A + L_Y + e_2$$

$$\pi_3 = I_B + L_Y + e_3$$

各类冲击均服从独立的二维正态分布，不同冲击之间的协方差为0。

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

时间流程：

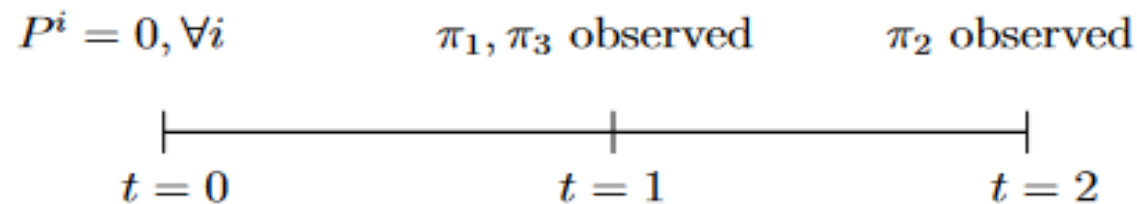


Figure 1: Timeline

最初，在 $t = 0$ 时刻，每个企业的预期红利为零，因此价格为零。在 $t = 1$ 时，公司1和3都公布了盈余信息，这些信息可以用来更新公司2的股价。当观察到 π_2 时，模型在 $t = 2$ 时结束。

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

分析师报告：

分析师在股票价格的决定中扮演着重要的角色。具体而言，存在一组以 $n \in \{1, 2, 3, \dots, N\}$ 为索引的覆盖企业2的分析师。投资者阅读分析师报告，将公司2的价格设定为 π_2 的期望，条件是分析师提供的信息。将分析师 n 出具的报告记为 r_n ：

$$P_{t=1}^2 = E[\pi_2 | (r_1, r_2, \dots, r_N)]$$

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

分析师报告：

分析师和投资者从各种各样的来源收集和分析信息，其中许多信息针对被关注公司，但在本文中也认为其可针对其余的基本面关联企业，因此分析师的报告可分为四类 $\{\pi_1, \pi_3\}$, $\{\pi_1\}$, $\{\pi_3\}$, $\{\}$ 。

由于公司2的投资者阅读了所有可获得的报告，他们利用分析师社区产生的所有信息的联合形成预期。因此，对于形成预期而言，是否所有的信息都来自于一个分析师(例如， $r_1 = \{\pi_1, \pi_3\}$ ， $r_2 = \{\}$)，或者这些信息是否分散在分析师(例如， $r_1 = \{\pi_1\}$ ， $r_2 = \{\pi_3\}$)之间都没有区别。

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

价格与收益率：

$$P_{t=1}^2 = \begin{cases} 0 & \text{if neither } \pi_1 \text{ nor } \pi_3 \text{ reported} \\ \pi_1 \left(\frac{\sigma_I^2}{\sigma_I^2 + \sigma_L^2 + \sigma_e^2} \right) & \text{if only } \pi_1 \text{ reported} \\ \pi_3 \left(\frac{\sigma_L^2}{\sigma_I^2 + \sigma_L^2 + \sigma_e^2} \right) & \text{if only } \pi_3 \text{ reported} \\ \pi_1 \left(\frac{\sigma_I^2}{\sigma_I^2 + \sigma_L^2 + \sigma_e^2} \right) + \pi_3 \left(\frac{\sigma_L^2}{\sigma_I^2 + \sigma_L^2 + \sigma_e^2} \right) & \text{if both } \pi_1 \text{ and } \pi_3 \text{ reported} \end{cases}$$

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

Industry momentum occurs when $cov(P_{t=1}^1 - P_{t=0}^1, P_{t=2}^2 - P_{t=1}^2) = cov(\pi_1, \pi_2 - P_{t=1}^2) > 0$.

Expanding this using the factor structure given in Equation 1, we have

$$\begin{aligned} cov(\pi_1 - 0, \pi_2 - P_{t=1}^2) &= cov(\pi_1, \pi_2) - cov(\pi_1, P_{t=1}^2) \\ &= \sigma_I^2 - cov\left(\pi_1, \frac{\sigma_I^2}{\sigma_I^2 + \sigma_L^2 + \sigma_\epsilon^2} p_1(N) \pi_1 + \frac{\sigma_L^2}{\sigma_I^2 + \sigma_L^2 + \sigma_\epsilon^2} p_3(N) \pi_3\right) \\ &= \sigma_I^2(1 - p_1(N)). \end{aligned}$$

Regional momentum takes a similar form:

$$\begin{aligned} cov(\pi_3 - 0, \pi_2 - P_{t=1}^2) &= cov(\pi_3, \pi_2) - cov(\pi_3, P_{t=1}^2) \\ &= \sigma_L^2 - cov\left(\pi_3, \frac{\sigma_I^2}{\sigma_I^2 + \sigma_L^2 + \sigma_\epsilon^2} p_1(N) \pi_1 + \frac{\sigma_L^2}{\sigma_I^2 + \sigma_L^2 + \sigma_\epsilon^2} p_3(N) \pi_3\right) \\ &= \sigma_L^2(1 - p_3(N)). \end{aligned}$$

地理关联与收益率可预测性

3 理论依据与制度背景

3.1 行业-地理动量模型

Sensitivity to analyst coverage. To this point, we have taken $p_1(N)$ and $p_3(N)$ as given, so as to simplify the return predictability expressions. We now attempt to be more explicit about their relationship with the number of analysts (N) covering firm 2.

Recall that a report may take on four possible values: $\{\pi_1, \pi_3\}, \{\pi_1\}, \{\pi_3\}, \{\}$. Denote the probability of each, respectively, as x, y, z , and $1 - x - y - z$. Let us assume that reports are written independently. Then, with N reports, the aggregate probability that π_1 is reported by at least one analyst $p_1(N)$, is equal to $1 - (1 - x - y)^N$. Likewise, the analogous expression for π_3 is $1 - (1 - x - z)^N = p_3(N)$.

One of our key assumptions is that analysts are unlikely to cover firms operating in fundamentally different sectors, consistent with the patterns observed in Figure 3. Applied to the probabilities above, this implies that $x \approx z \approx 0$, which in turn implies that $p_1(N) \approx 1 - (1 - y)^N$ and that $p_3(N) \approx 0$.

地理关联与收益率可预测性

3 理论依据与制度背景

3.2 描述模型的关键假设

- 证实公司的基本面表现的确在行业与地理的层面上存在关联性

所选指标：每股收益变化量、销售额、雇员数、外部资本。

- 证实分析师的确按行业进行分布

PANEL: Portfolio Overlap				
	<i>full sample</i>	<i>firms with 1-3 analysts</i>	<i>firms with 4-9 analysts</i>	<i>firms with 10+ analysts</i>
avg # of analysts	4.81	1.83	5.84	13.81
Overlaps between geographic peers				
% firm-years with zero overlap	57.93%	68.87%	49.37%	38.28%
# of peers connected via at least one overlap	2.08	0.84	2.61	5.55
Overlaps between industry peers				
% firm-years with zero overlap	7.39%	12.67%	2.32%	0.49%
# of peers connected via at least one overlap	15.73	5.81	19.05	45.95

地理关联与收益率可预测性

4 地理与行业上的Lead-Lag关系

4.1 Fama-French回归（企业-月度层面）

在预测企业的月度收益率时，引入两个解释变量：（1）滞后的同行不同地企业组合收益率；（2）滞后的不同行同地企业组合收益率。

FM回归的形式如下：

$$r_{i,c,j,t+1} = \alpha + \beta_1 r_{c,\neq j,t} + \beta_2 r_{\neq c,j,t} + \beta_3 X_t + e_{i,c,j,t+1}$$

- $r_{i,c,j,t+1}$ 表示位于城市c，行业j，的企业i在t+1月的超额收益率
- X_t 表示一组控制变量，选取见数据说明部分

地理关联与收益率可预测性

4 地理与行业上的Lead-Lag关系

4.1 Fama-French回归（企业-月度层面）

PANEL A: 1-month predictors			
	<i>full sample</i>	<i>1970-1990</i>	<i>1991-2013</i>
$r_{city,t-1}$	0.049*** (4.12)	0.056*** (3.26)	0.044*** (2.63)
$r_{industry,t-1}$	0.174*** (9.85)	0.169*** (7.49)	0.179*** (6.72)
<i>Avg R²</i>	6.44%	6.48%	6.39%
<i>Observations</i>	1,458,783	481,729	977,054
<i># time clusters</i>	516	240	276
PANEL D: 12-months cumulative predictors			
	<i>full sample</i>	<i>1970-1990</i>	<i>1991-2013</i>
$r_{city,t-12:t-1}$	0.008** (2.50)	0.014*** (2.84)	0.003 (0.73)
$r_{industry,t-12:t-1}$	0.025*** (6.64)	0.024*** (4.54)	0.026*** (4.81)
<i>Avg R²</i>	6.43%	6.50%	6.37%
<i>Observations</i>	1,458,783	481,729	977,054
<i># time clusters</i>	516	240	276

地理关联与收益率可预测性

4 地理与行业上的Lead-Lag关系

4.2 交易策略盈利

上述结果提供了一种利用地理邻近的交叉序列相关收益的交易策略。每个月，我们对每个企业*i*进行排名，不是根据其自身的滞后收益率，而是通过总部位于同一地区的企业的平均滞后收益（但在不同行业运营）。

Momentum Trading Strategy - Quintiles										
	returns									
	Mean (%)	CAPM α	<i>t</i> -stat	FF-3 α	<i>t</i> -stat	FF-3+MOM α	<i>t</i> -stat	FF-5 α	<i>t</i> -stat	Sharpe Ratio
lowest city return	0.735	-0.258	-3.228	-0.244	-3.011	-0.223	-2.713	-0.199	-2.333	0.207
	0.876	-0.077	-1.112	-0.038	-0.520	0.002	0.021	-0.016	-0.191	0.318
	1.027	0.109	1.432	0.127	1.583	0.178	2.309	0.132	1.553	0.452
	0.949	-0.009	-0.102	-0.002	-0.022	-0.024	-0.288	0.013	0.122	0.366
highest city return	1.158	0.212	3.196	0.211	3.029	0.195	2.629	0.217	2.714	0.526
5-1 spread	0.423 [3.65]	0.471	[4.16]	0.455	[3.99]	0.419	[3.40]	0.417	[3.319]	

地理关联与收益率可预测性

4 地理与行业上的Lead-Lag关系

4.3 分析师覆盖的作用机制

在本文中，分析师对Lead-Lag关系的缓解作用可归纳为两条：（1）分析师越多，价格对信息的吸收效率更高；（2）分析师重叠程度越大，提前吸收的相关信息更多。

是否有办法来区分两类作用？如果看传统的行业关联，则难以区分，因为分析师往往是按行业进行分化的。但是地理关联的企业之间一般不存在分析师的重叠效应。

地理关联与收益率可预测性

4 地理与行业上的Lead-Lag关系

4.3 分析师覆盖的作用机制

PANEL A: Number of analysts					
	(0)	(1-4)	(5-9)	(10+)	Δ High/Low
$r_{city,t-1}$	0.052*** (2.86)	0.046** (2.51)	0.083*** (3.82)	0.054** (2.49)	[0.5957]
$r_{industry,t-1}$	0.230*** (9.21)	0.183*** (7.63)	0.116*** (4.34)	0.051** (2.05)	[0.00***]
$Avg R^2$	6.39%	7.27%	10.73%	13.96%	
$Observations$	503,536	317,030	167,289	170,333	
$\# time clusters$	336	336	336	336	

地理关联与收益率可预测性

4 地理与行业上的Lead-Lag关系

4.3 分析师覆盖的作用机制

行业层面的领先滞后效应总体上大于地区层面的领先滞后效应，但主要局限于小企业，以及分析师覆盖率和交易量较低的企业。地区层面的领先滞后虽然平均而言更小，但似乎同样适用于不同规模、交易量和分析师覆盖率的企业。

一个小疑问：

第一篇文献中的供应链上下游企业之间的可预测性会随着分析师覆盖数的上升而下降，但是又假设分析师不会吸收本行业以外的信息，这样就与此处地理关联的情形相矛盾。

科技关联与收益率可预测性

分组检验

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)

科技关联与收益率可预测性

回归检验

回归设定：

- 时间：1963–2012
- 方法：Fama-Macbeth回归
- 自变量：企业收益率RET、基于行业收益率的超额收益率RET-INDRET
- 因变量：滞后一期的TECHRET
- 控制变量：滞后的INDRET、企业规模、账面市值比、总盈利、资产增长、研发强度、滞后RET、动量
- 特殊处理：把每个月解释变量映射到0至1的十分位上

科技关联与收益率可预测性

回归检验

Dep. variable $\times 100$	(1) RET_t	(2) RET_t	(3) RET_t	(4) $RET_t - INDRET_t$
$TECHRET_{t-1}$	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	0.065

科技关联与收益率可预测性

中国数据实证结果

表 6: TechRet 回归系数

Variables	<i>Dependent variable: Return(%)</i>	
	(1)	(2)
$TechRet_{t-1}$	0.1496*** (24.348)	0.0739*** (3.449)
$ln(Msmvttl)_{t-1}$	-0.6415*** (-19.948)	-0.2044*** (-6.528)
HHI_{t-1}	0.1298 (0.489)	0.4807 (0.939)
BM_{t-1}	0.5014*** (5.414)	0.2565*** (2.600)
$MOM12_{t-1}$	-0.0021** (-2.417)	0.0028*** (3.055)
Ret_{t-1}	-0.0376*** (-8.665)	-0.0440 *** (-11.439)
Constant	15.5312*** (21.181)	23.8181*** (20.602)
Observations	155670	155670
Fixed effects	No	Yes

Notes:

[1]*p<0.1; **p<0.05; ***p<0.01

[2]Standard Errors are heteroscedasticity and autocorrelation robust (HAC)

科技关联与收益率可预测性

中国数据实证结果

表 7: 基于 TechRet 的三分位分组

Panel A:Equal-weighted portfolios	Low	Mid	High	High-Low
Excess Return(%)	0.84 (1.1)	0.92 (1.14)	1.11 (1.45)	0.27 (1.48)
CAPM alpha	0.36 (1.09)	0.42 (1.37)	0.61** (1.97)	0.26 (1.39)
Fama-French five-factor alpha	-0.19 (-1.3)	-0.09 (-0.66)	0.04 (0.33)	0.24 (1.3)
Panel B:Value-weighted portfolios	Low	Mid	High	High-Low
Excess Return(%)	1.04 (1.58)	1.27 (1.61)	1.88** (2.49)	0.84*** (2.88)
CAPM alpha	0.61*** (2.7)	0.79*** (4.31)	1.4*** (5.66)	0.79*** (2.95)
Fama-French five-factor alpha	0.43** (2.2)	0.85*** (4.34)	1.21*** (5.49)	0.78*** (2.92)

Notes:

[1]* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

[2]Standard Errors are heteroscedasticity and autocorrelation robust (HAC)