

证券研究报告



分析师：

郑兆磊

zhengzhaolei@xyzq.com.cn

S0190520080006

西学东渐--海外文献推荐系列之一百五十九

2023 年 11 月 23 日

报告关键点

高频交易时间的接近度可能包含着显著的选股信号。在本文中，作者提出了一种根据每笔交易与市场上其他交易的接近度对每笔交易进行分类的方法，并将其分为五种类型。通过构建与各类交易相关的归一化订单失衡指标（称之为条件订单失衡（COI）），作者研究了分解交易流对股票价格的影响。作者的实证研究结果表明：同期收益率与 COIs 存在显著的正向相关性。在可预测性方面，作者发现如果交易的 COIs 与其自身以外的股票交易相隔离，那么其与未来收益的关联性为正，反之则为负。此外，作者基于三年数据，运用 COIs 开发的交易策略对 457 只股票进行了广泛的实验，回测结果表明该交易策略实现了显著的收益和夏普比率。

相关报告

《西学东渐--海外文献推荐系列之一百五十八》2023-11-09
《西学东渐--海外文献推荐系列之一百五十七》2023-10-26
《西学东渐--海外文献推荐系列之一百五十六》2023-10-12

投资要点

- “西学东渐”作为以往学术传播的方式之一，其在如今仍具有重要的现实意义。作为 A 股市场上以量化投资为研究方向的卖方金融工程团队，我们通过大量的材料阅读，去粗取精，将认为最有价值的海外文献呈现在您的面前。
- 作为西学东渐——海外文献系列报告第一百五十九篇，本文推荐了 Yutong Lu、Gesine Reinert 和 Mihai Cucuringu 于 2022 年发表的论文《Trade Co-occurrence, Trade Flow Decomposition, and Conditional Order Imbalance in Equity Markets》。
- 高频交易时间的接近度可能包含着显著的选股信号。在本文中，作者提出了一种根据每笔交易与市场上其他交易的接近度对每笔交易进行分类的方法，并将其分为五种类型。通过构建与各类交易相关的归一化订单失衡指标（称之为条件订单失衡（COI）），作者研究了分解交易流对股票价格的影响。作者的实证研究结果表明：同期收益率与 COIs 存在显著的正向相关性。在可预测性方面，作者发现如果交易的 COIs 与其自身以外的股票交易相隔离，那么其与未来收益的关联性为正，反之则为负。此外，作者基于三年数据，运用 COIs 开发的交易策略对 457 只股票进行了广泛的实验，回测结果表明该交易策略实现了显著的收益和夏普比率。

风险提示：文献中的结果均由相应作者通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。

请务必阅读正文之后的信息披露和重要声明

目 录

1、 介绍	- 4 -
2、 相关文献	- 6 -
3、 交易与流动分解的共现性	- 7 -
3.1、交易共现	- 7 -
3.2、交易流分解	- 8 -
3.3、条件订单失衡	- 8 -
4、 探索性数据分析	- 9 -
4.1、交易共现	- 9 -
4.2、交易汇总统计	- 10 -
4.3、共现概率：零概率与经验概率	- 12 -
4.4、订单失衡的描述性统计	- 13 -
5、 条件性订单失衡对同期价格的影响	- 15 -
6、 失衡对未来回报率的预测能力	- 16 -
6.1、时间序列回归	- 16 -
6.2、基于“失衡”的投资组合排序	- 18 -
7、 条件订单失衡的投资价值	- 20 -
7.1、多-空投资组合的构建和评估	- 20 -
7.2、盈利能力	- 21 -
8、 稳健性分析	- 23 -
8.1、领域规模效应	- 23 -
8.2、日内时间效应	- 23 -
8.3、成交量测度的 COI	- 24 -
9、 结论与展望	- 24 -
参考文献	- 25 -
附录 A 领域规模效应	- 28 -
附录 B 日内时间效应	- 29 -
附录 C 成交量测度的 COI	- 30 -

图表目录

图表 1、交易共现图解	- 8 -
图表 2、以共现为条件的交易类型说明。	- 9 -
图表 3、所有交易组的汇总统计数据	- 11 -
图表 4、各类交易数量的日内分布	- 11 -
图表 5、每种交易流的零概率和经验概率	- 13 -
图表 6、所有交易组和订单失衡的汇总统计	- 14 -
图表 7、订单失衡的皮尔逊相关性	- 14 -
图表 8、同期回归结果	- 15 -
图表 9、同期回归 R^2	- 16 -
图表 10、预测性回归	- 17 -
图表 11、预测性回归的 R^2	- 18 -
图表 12、单一排序投资组合汇总	- 19 -
图表 13、双排序投资组合的年收益率	- 20 -
图表 14、多空投资组合的盈利能力	- 21 -
图表 15、多空投资组合的异常回报	- 22 -
图表 16、投资组合的累计回报	- 22 -
图表 17、不同 δ 下的实证结果	- 29 -
图表 18、按时间段分列的 COIs	- 30 -
图表 19、成交量测度的 COIs	- 31 -

报告正文

股票市场的交易共现、交易流分解和条件订单失衡

文献来源：

Yutong Lu and Gesine Reinert and Mihai Cucuringu. "Trade Co-occurrence, Trade Flow Decomposition, and Conditional Order Imbalance in Equity Markets". In: *arXiv* (2022), arXiv:2209.10334

推荐理由：

高频交易时间的接近度可能包含着显著的选股信号。在本文中，作者提出了一种根据每笔交易与市场其他交易的接近度对每笔交易进行分类的方法，并将其分为五种类型。通过构建与各类交易相关的归一化订单失衡指标（称之为条件订单失衡（COI）），作者研究了分解交易流对股票价格的影响。作者的实证研究结果表明：同期收益率与 COIs 存在显著的正向相关性。在可预测性方面，作者发现如果交易的 COIs 与其自身以外的股票交易相隔离，那么其与未来收益的关联性为正，反之则为负。此外，作者基于三年数据，运用 COIs 开发的交易策略对 457 只股票进行了广泛的实验，回测结果表明该交易策略实现了显著的收益和夏普比率。

关键词：

市场微观结构、共现分析、订单失衡、交易策略、定量金融

1、介绍

现如今，证券交易所向电子交易的转型在极大程度上重塑了市场的微观结构。由于现如今延迟时间已被降低到纳秒级（O'Hara（2015），Hirschey（2021）），市场参与者的交易策略达到了前所未有的盈利水平。市场中的代理人可以直接提交和取消限价订单。当现有限价订单由市场订单/可市场化限价订单执行时，交易即进行结算。携带不同信息的交易对相关股票的价格变化产生不同的影响，学者和从业人员因此将交易分为不同类型，并分别进行研究。例如，按交易方向分组，Chordia、Goyal 和 Jegadeesh（2016）研究了买卖双方发起的交易流程，将其分为激进买入和激进卖出。Kraus 和 Stoll（1972）以及 Lee 等人（2004）将机构与个人投资者交易分开。与上述完全基于单个交易特征的分类不同，我们感兴趣的是根据交易的投放时间相对于整个市场上其他交易的到达时间进行分类，既包括同一资产内的交易，也包括横截面上的交易。

市场参与者可以通过观察市场上的交易流向来做出交易决策，这就是我们研究的动机所在。以前的研究（Kyle（1985）、Kyle、Ou-Yang，和 Wei（2011））对高频价格的形成进行建模，并指出交易者将大订单拆分成许多小订单以掩盖其真实目的，其他市场参与者则会监控市场中的订单流以便做出交易决策。高性能交易系统的发展导致了高频交易（HFT）的惊人增长，并提高了交易策略的多样性（Hagstromer and Norden（2013））。在这个世界上，反应时间起着重要作用，因为如果不在微秒甚至纳秒内采取行动，机会就会瞬息万变。高频交易策略包括

预测交易流（Hirschey（2021））和掠夺其他市场参与者（Van Kervel 和 Menkveld（2019））。我们有兴趣探讨的问题是：某些交易是否会与其他交易发生相互作用？交易是否包含有用信息？它们是如何促进股价变动的？这些问题有助于我们揭示短期和长期的价格形成机制。

我们首先提出第 3.1 节中定义的“交易共现”概念，其是可以根据交易与其他交易之间的互动关系对交易进行识别和分组的一种工具。对于每笔给定的交易，如果两笔交易发生的时间很近，我们就认为这两笔交易同时发生并相互影响。为了定义和量化“发生时间的邻近度”，我们预先定义了邻域大小 δ 。如果两笔交易的时间差小于 δ ，则两笔交易邻近且同时发生。注意阈值 δ 是一个重要参数，它决定了同时发生的交易集合。然而，并没有严格的规则来设定这个值。直观地说，考虑到 HFT 以机构交易者作为猎物并根据机构市场订单进行交易的情况，我们目标是捕捉这些互动，并将此类交易归类为积极互动交易等类别。考虑到这一点，适当的选择应大于往返延迟加上 HFT 检测和做出交易决策的时间，而这一时间通常是不公开的。因此，我们试验了多个 δ 值，并对比了相应的结果。需要注意的是 δ 值也不宜过大，因为过大的邻域很可能将市场中不相关的交易纳入其中。在本文中，我们将 δ 视为一个超参数；为简单起见，我们仅在正文中展示了 $\delta=500\mu s$ 的经验结果，并在附录 A 中对不同的 δ 值选择进行了比较。

利用交易共现性，我们通过将所有股票的所有交易分为子组。对于某笔交易，我们通过以下两个问题来确定它属于哪一组：它是否与其他交易互动？如果是，它是否只与自己相同的股票互动，还是只与自己不同的股票互动？还是两种都有？根据答案，交易将被归入一个或两个类别，具体规则将在第 3.2 节中说明。标记所有交易后，我们研究收益与交易子组之间的关系。

我们使用订单失衡作为连接交易流和股票收益的桥梁，这在金融文献中已有深入研究。库存范式（Stoll（1978）、Spiegel 和 Subrahmanyam（1995）、Chordia、Roll 和 Subrahmanyam（2002））认为在中介市场中，买方发起的交易与卖方发起的交易之间的差额，即所谓的不平衡，会对做市商的库存造成压力。作为回应，做市商会调整存货以维持其市场风险，从而推动价格向一个方向变动。

接下来，我们在每日层面上研究了正常交易期间各类交易的总订单失衡属性及其与个股收益率的关系。数据调查表明，所有类别的条件性和无条件订单失衡都是正自相关的。条件订单失衡（COIs）都与原始订单失衡有很强的正相关性。但是条件性和无条件订单失衡之间不一定高度相关。

我们的实证结果主要集中在失衡与收益的关系上。通过回归分析，我们发现订单失衡与当日价格变化之间存在显著的正相关关系。此外，与标准回归分析相比，多元回归设置中分解订单流可显著提高 adjusted R²，这说明对当日股票开盘到收盘的回报有更加的解释力。为了利用可预测性，我们使用相同的回归分析将订单失衡与未来一天的收益进行拟合。与同期结果不同的是，统计意义上的显著关系只出现在一小部分股票中。尽管有意义的回归系数比例较低，但我们观察到，

对于那些与市场其他部分有较高互动的交易，订单失衡似乎与未来收益率呈负相关。相反，对于与市场其他股票隔离的交易，其订单失衡与未来收益率显示出微弱的正相关。

这些关联在我们随后的投资组合分析中得到了进一步证实，具体如下。我们利用这些失衡来构建交易策略。为了评估交易流分解方法的经济价值，我们以 COIs 为信号构建了信号排序投资组合。特别是，如果我们根据在预测回归中观察到的模式做出多空决策，我们将在投资组合中获得收益，最高的年化夏普比率达到 2.38。作为基准，我们建立了投资于未分解订单不平衡的投资组合，其夏普比率为负值。

本文的其余部分安排如下。第 2 节概述了我们对金融文献的贡献。在第 3 节中，我们介绍了交易共现、交易流分解和 COIs 的定义。在第 4 节中，我们以描述数据源并进行探索性分析开始我们的实证研究。随后，我们在第 5、6 节和第 7 节中分别揭示了 COIs 与同期收益率之间的关系和 COIs 的经济价值。第 8 节提供了稳健性分析和其他实证结果。最后，在第 9 节中，我们总结了研究结果，并讨论了局限性和未来的研究方向。

2、相关文献

本文在三方面对文献有所贡献。首先，我们的研究利用了一种新的共现分析的金融应用，这种统计方法被证明在空间模式分析中非常强大，并广泛应用于生物学 (Gotelli (2000)、MacKenzie, Bailey 和 Nichols (2004)、Ara'ujo et al. (2011)、自然语言处理 (NLP) (Dagan, Lee 和 Pereira (1999)、Kolesnikova (2016))、计算机视觉 (Galleguillos, Rabinovich 和 Belongie (2008)、Aaron, Taylor 和 Chew (2018)) 等领域 (Appel 和 Holden (1998)、Ye 等人 (2017))。到目前为止，共现分析在金融文献中的应用主要集中在研究新闻报道中的股票共现。Ma、Pant 和 Sheng (2011) 从在线新闻中的公司共现构建网络，并使用机器学习模型来识别公司之间的竞争关系。最近的研究，包括 Guo 等人 (2018)；Tang、Zhou 和 Hong (2019)；Wu 等人 (2019) 利用新闻中的股票共现构建网络，并将其用于股票回报率预测和投资组合分配等任务。我们的贡献在于提出了交易共现的概念。通过直接应用股票交易的共现性，我们发现这种技术有利于探索和洞察金融市场的微观结构。

其次，我们的研究为市场中交易活动之间相互作用添了新的内容。在 Kyle (1985) 的模型中，做市商通过观察市场中知情交易者和流动性交易者的总订单流来调整自己的交易策略。更激进的是，HFT 交易者可以发现机构等知情交易者 (Van Kervel 和 Menkveld (2019))，并预测其他人的交易流向 (Hirschey (2021))。各种理论模型 (Grossman 和 Miller (1988)、Brunnermeier 和 Pedersen (2005)、Yang 和 Zhu (2020) 和 Zhu (2020)) 就高频交易者和机构交易者之间的相互作用提出了各种理论模型。Van Kervel 和 Menkveld (2019) 对瑞典股市进行了实证

研究，发现当机构交易者开始拆分大单时，HFT 参与者打算顺风交易，最终与机构同向交易。

我们提出“交易共现性”这一概念，并提供了实证证据，证明股票交易的共现性并非巧合。我们不是研究交易者之间的互动，而是创新性地将交易共现作为一种工具，用于分析单个交易层面的互动性。我们对共现条件的 COIs 的研究表明，细粒度的交易互动传递了价格形成的有用信息。

最后，本文对有关订单失衡和价格形成的文献有所贡献。根据开创性的研究，订单失衡的持续性可以通过两种方式产生。首先，正如 Kyle (1985) 的模型所指出的，交易者有意在一段时间内拆分大额订单，以尽量减少其对市场的影响，这就导致了自相关失衡。Scharfstein 和 Stein (1990) 指出，订单失衡的另一个来源是羊群效应。为了探讨订单失衡如何影响价格变化，Chordia 和 Subrahmanyam (2004) 提出了一个理论模型，解释订单失衡与同期股票回报之间的正相关关系，这种关系是由做市商动态地适应订单失衡而产生的。此外，全权委托交易商跨日优化拆分订单使订单失衡具有很强的正自相关性和对未来回报的预测能力。他们使用纽约证券交易所 (NYSE) 上市股票 1988 年至 1998 年 10 年间的日度数据进行了实证研究，证实了他们的理论结果，并表现出使用订单失衡作为交易信号的盈利能力。然而，在可预测性方面还存在争议。例如，Shenoy 和 Zhang (2007) 和 Lee 等人 (2004) 发现订单失衡对未来回报没有显著的预测能力。

虽然 Chordia 和 Subrahmanyam (2004) 没有对交易流进行区分，但随后的研究表明，由不同代理人在不同时间下达的具有不同属性的有序订单会对价格变化产生不同的影响。他们发现 (Lee 等人 (2004)、Bailey 等人 (2009)、Zhang、Gu、和 Zhou (2019))，在中国市场交易者类型的私人识别数据是可用的，同时有迹象表明，机构交易流的订单失衡比个人交易者的失衡对价格的压力更大。在美国市场，Cox (2021) 最近对 2015 年至 2016 年期间标准普尔 500 指数股票的研究也得出同样的结果，该研究根据交易是否属于由机构采用的市场间扫盘，将交易分为二元类别 (Chakravarty 等 (2012))。

我们的研究利用 Covid-19 发生前最近一段时期的数据来补充美国市场秩序失衡的研究，并提出了一种无需额外私人数据集即可分解无条件贸易流的新方法。我们的研究表明，在不区分交易的情况下，订单失衡不再对未来回报具有预测能力，这是过去几十年市场微观结构演变的证据 (Chordia、Roll 和 Subrahmanyam (2002; Chordia 和 Subrahmanyam (2004)))。不过，用我们提出的方法分解的贸易流具有不同的信息量且其 COIs 确实具有预测能力。

3、交易与流动分解的共现性

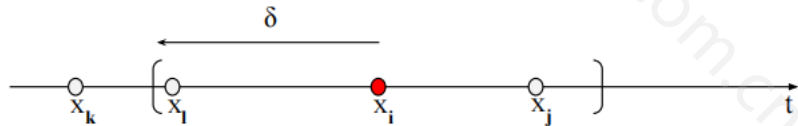
3.1、交易共现

我们首先介绍交易共现的定义。对于在时间发生的每笔交易 x_i ，在预先指定

请务必阅读正文之后的信息披露和重要声明

的 δ 下，除 x_i 本身外，在时间段 $(t_i - \delta, t_i + \delta)$ 内到达的每笔交易都被定义为与交易 x_i 共同发生。我们将阈值 δ 称为“邻域大小”，将与 x_i 同时发生的所有交易集合称为交易 x_i 的“ δ -邻域”。我们注意到，共同出现并不是等价关系。 x_i 和 x_j 完全有可能同时出现，而 x_i 和 x_l 也完全有可能同时出现，但 x_j 和 x_l 并不能同时出现。

图表 1、交易共现图解



该图可视化了交易共现的概念：给定用户定义的邻域大小 δ ，交易 x_j 到达交易 x_i 的 δ 邻域内，因此它们同时发生。相比之下，交易 x_k 位于领域外，因此这两种交易不会同时发生。交易 x_j 和 x_l 都与交易 x_i 同时发生，但它们并不同时发生。

资料来源：arXiv，兴业证券经济与金融研究院整理

3.2、交易流分解

根据共现性，我们接下来将交易分成不同的类别，这些类别以 δ 邻域为特征，其协议如图表 2。最初，我们将所有交易分为两组：隔离交易（iso）和非隔离交易（nis）。如果这笔交易不与其他交易同时出现，则标记为隔离交易。否则，交易将被标记为非隔离交易。

我们根据其 δ 邻域内的交易属性进一步分解非隔离交易。

每个非隔离交易 x_i 可分为以下三类：

1. 非自我隔离（nis-s）：交易 x_i 的 δ 邻域只包含与交易 x_i （至少一个）相同股票的交易；
2. 非交叉隔离（nis-c）：交易 x_i 的 δ 邻域只包含与交易 x_i 股票不同的股票交易；
3. 非两者隔离（nis-b）：交易 x_i 的 δ 邻域既至少包含一个相同股票的交易，也至少包含一个不同股票的交易。

这三类交易构成了非隔离交易集的分區，如图表 2 最后一行所示。我们将这种将交易分门别类的过程称为“交易流分解”。

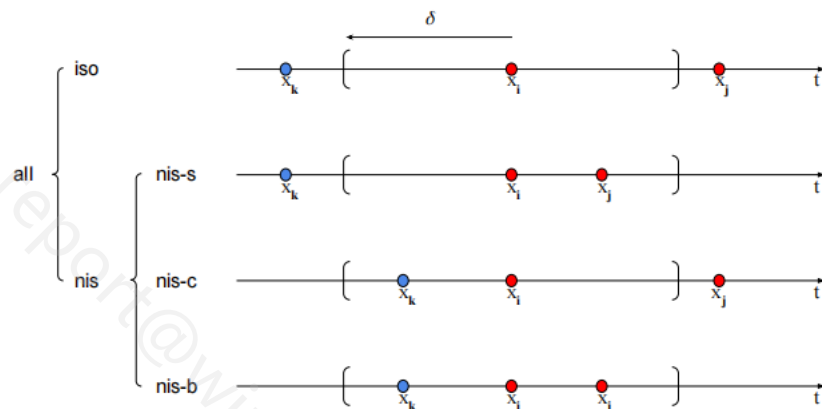
3.3、条件订单失衡

在对交易流量进行分解后，我们开始研究具有不同特征的交易对价格的影响。连接交易活动和价格变化的桥梁是订单失衡量，它被定义为买卖双方发起的交易量之间的归一化差额（Chordia 和 Subrahmanyam（2004））。对于给定的股票 i ，我们得出的条件日订单失衡量如下：

$$COI_{i,t}^{type} = \frac{N_{i,t}^{type,buy} - N_{i,t}^{type,sell}}{N_{i,t}^{type,buy} + N_{i,t}^{type,sell}} \quad (1)$$

其中， $N_{i,t}^{type,buy}$ 和 $N_{i,t}^{type,sell}$ 分别表示股票 i 在第 t 天的市场买入订单和卖出清单总数。如果分母为 0，即不存在某种类型的交易，我们将这种情况下的 COI 定义为 0。我们考虑了六种类型的 COI，上标类型的取值范围为 {all、iso、nis、nis-s、nis-c、nis-b}，表示用于计算不平衡的交易组。因此，“all”类别的 COI 即表示总的交易数量中的订单不平衡（Chordia 和 Subrahmanyam（2004））。

图表 2、以共现为条件的交易类型说明。



我们展示了交易 x_i 的不同分类标签。颜色表示交易对应的股票。因此， x_j 表示与 x_i 相同的股票，而 x_k 表示不同的股票。第一行： x_i 是隔离 (iso) 交易， δ 邻域为空；第二行： x_i 是非隔离 (nis) 交易， δ 邻域非空；第三行： x_i 是非自我隔离 ('nis-s') 交易，其 δ 邻域中只有相同股票的其他交易； x_i 是非交叉隔离 ('nis-c') 交易，其 δ 邻域中只有不同股票的其他交易；最后一行： x_i 是非两者隔离 ('nis-b') 交易，其 δ 邻域中既有相同股票的其他交易，也有不同股票的其他交易。

资料来源：arXiv，兴业证券经济与金融研究院整理

4、探索性数据分析

在本节中，我们将简要介绍实证研究中使用的数据。通过探索性分析，我们发现了交易共现的显著模式。此外，我们还发现，分解后的交易流订单失衡彼此之间存在弱相关性，这表明我们提出的交易分解是有意义的。

4.1、交易共现

我们的研究基于 2017 年 1 月 3 日至 2019 年 12 月 10 日期间的 457 只美国股票。所选股票是标准普尔 (S&P) 500 指数中的公司，这些公司在整个样本期间都有订单数据和价格数据。

4.1.1、限价订单簿

我们从 LOBSTER 数据库 (Huang 和 Polak (2011)) 中获取限价订单簿数据，该数据库提供了在纳斯达克交易所交易的所有股票限价订单的详细情况，包括限价订单的提交、取消和已执行交易，且所有记录以时间为索引，精度可达纳秒。对于每个交易日的每只股票记录包含限价订单事件的时间戳、事件类型（提交/取消/执行）、方向（买入/卖出）、大小和价格。通过过滤限价订单执行并逆转其方向，我们就能推断出买方和卖方发起的交易，例如，执行限价买入指令意味着市场发出“卖出指令/适销限价卖出指令”。我们注意到大量市场订单会同时消耗多个现有限价订单，鉴于 LOBSTER 的高时间分辨率，我们假设不同的交易不可能有完全相同的时间戳，从而合并订单推断出具有相同时间戳的交易。

4.1.2、价格和收益率

我们从证券价格研究中心 (CRSP) 数据库中获取所研究股票范围的每日价格数据，并计算每日开盘到收盘的对数回报率为：

$$R_{i,t} = \log \frac{P_{i,t}^{Close}}{P_{i,t}^{Open}} \quad (2)$$

其中 $P_{i,t}^{Open}$ 开盘价和 $P_{i,t}^{Close}$ 是股票 i 在第 t 天的每日开盘价和收盘价。为减轻市场因素的影响，本研究还考虑了市场超额收益表示为 $r_{i,t}$ ，计算公式如下：

$$r_{i,t} = R_{i,t} - R_{SPY,t} \quad (3)$$

其中， $R_{SPY,t}$ 是跟踪标准普尔 500 指数的 SPY ETF 的日收益率。为简单起见，这里我们假设所有股票的市场贝塔系数都等于 1。

4.2、交易汇总统计

建立交易数据集后，我们为每笔交易贴上相应类型的标签。图表 3 列出了不同类型交易的数据摘要，选择的共现领域大小为 500 微秒 ($\delta = 500\mu s$)。面板 A 显示了原始数据的描述性统计，其中每个数字都是通过计算每日时间序列的平均，然后考虑所有股票的横截面平均值、中位数或标准差。平均值而言，隔离交易占交易总数的 38.29%，而大多数交易都是属于三种定义 (nis-s、nis-c、nis-b) 中的非隔离交易。约一半的非隔离交易 (占全部交易的 29.83%) 是非自我隔离交易 (nis-s)，非交叉隔离交易 (nis-c) 和非两者隔离交易 (nis-b) 的比例相似，平均值分别为 15.58% 和 16.30%。交易数量的标准差较大，这表明该群体具有异质性。面板 B 显示了不同组别交易量的百分比，其与 A 组报告百分比非常相似。有鉴于此，将交易次数作为衡量流动性的指标是合理的。

图表 3、所有交易组的汇总统计数据

Panel A: Average daily statistics of trade counts			
	Mean	Median	Std. dev.
Number of trades	3378.06	3022.07	1651.50
Percentage of iso trades	38.29	38.14	7.27
Percentage of nis trades	61.71	61.86	7.27
Percentage of nis-s trades	29.83	29.49	7.02
Percentage of nis-c trades	15.58	14.70	5.66
Percentage of nis-b trades	16.30	15.85	5.38

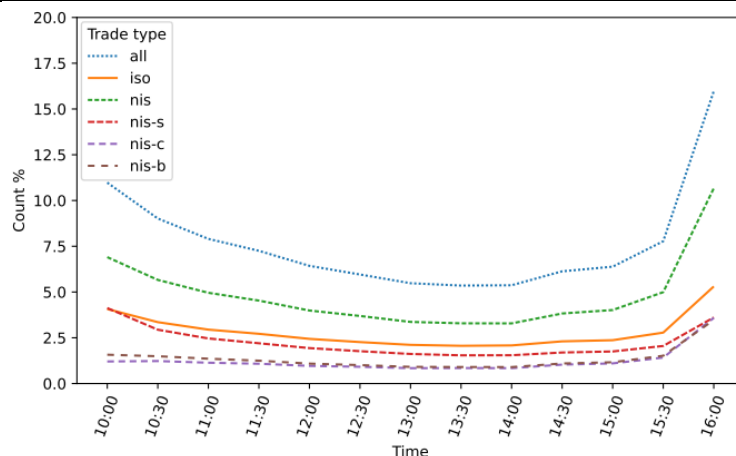
Panel B: Average daily statistics of trade volumes			
	Mean	Median	Std. dev.
Volume of trades	522301.02	458738.54	275668.69
Percentage of iso volumes	38.81	38.52	7.29
Percentage of nis volumes	61.19	61.48	7.29
Percentage of nis-s volumes	29.75	29.50	6.81
Percentage of nis-c volumes	16.00	15.01	5.87
Percentage of nis-b volumes	15.44	15.09	5.11

该表记录了每种类型交易每日横断面统计数据的时间序列平均值。我们的数据包括 457 只股票在正常交易时间内的交易记录，时间跨度为 2017 年 1 月 3 日至 2019 年 12 月 10 日。

资料来源：arXiv，兴业证券经济与金融研究院整理

图表 4 强调了在交易日开始和结束时交易活动较高的经验事实，绘制了交易的日内分布，揭示了不同交易类型略有不同的时间行为。图中显示了每半小时各类交易的数量，Y 轴表示占总交易量的百分比。我们发现所有类型的交易都在最后半小时内急剧增加。值得注意的是，经过分解后，隔离交易的流动比非隔离交易的流动更平滑，在最后半小时的爬升中斜率较低。通过进一步分离非隔离交易的子类型，我们发现非自我隔离交易（nis-s）在交易日开始的贡献更大，而其他两类交易在交易日结束外的时段都是平缓的。

图表 4、各类交易数量的日内分布



我们计算了 2017-01-03 至 2019-12-10 期间所有股票在 9:30 至 16:00 正常交易时间段内不重叠的 30 分钟间隔的 COIs。本图绘制了盘中 30 分钟 COIs 和在时间序列和横截面上的平均值。

资料来源：arXiv，兴业证券经济与金融研究院整理

4.3、共现概率：零概率与经验概率

有了订单数据，我们首先要回答以下基本问题：交易真的同时发生？还是它们的到达只是随机的、相互独立的？我们的交易流分解是否捕捉到了信号？在这一小节，我们将在完全随机订单到达的假设条件下建立一个空模型，通过比较零模型下的理论共现概率（Donges 等人（2016））和从数据中得出的经验值，我们证实股票交易之间存在 0.5 毫秒级别的共现，支持了整体交易量具有很强的跨资产互动成分的观点。从经济学角度来看，这也许是意料之中的，因为如今市场上存在大量的指数套利交易者，他们同时进行指数 ETF 与一篮子成分股的交易。

我们建立的零模型的假设条件是，对于股票 i ，在长度为 T 的时间间隔内，交易的到达遵循具有相同强度 λ_T 的独立泊松过程。在 $N_i = n_i$ 的条件下， n_i 个交易的到达时间是独立的，并且遵循 $[0, T]$ 上的均匀分布。因此，对每笔交易而言，在 T 时间段内另一笔交易落在其 δ 邻域的概率为：

$$p = \frac{2\delta}{T} \quad (4)$$

接下来，我们推导出不同类型交易流的概率如下：

$$\begin{aligned} \mathbb{P}_i(iso) &= (1-p)^{(N_i+N_{-i}-1)}, \\ \mathbb{P}_i(nis) &= 1 - (1-p)^{(N_i+N_{-i}-1)}, \\ \mathbb{P}_i(nis-s) &= [1 - (1-p)^{N_{i-1}}](1-p)^{N_{-i}}, \\ \mathbb{P}_i(nis-c) &= (1-p)^{N_{i-1}}[1 - (1-p)^{N_{-i}}], \\ \mathbb{P}_i(nis-b) &= [1 - (1-p)^{N_{i-1}}][1 - (1-p)^{N_{-i}}] \end{aligned} \quad (5)$$

其中 N_{-i} 表示除股票 i 之外市场上所有股票的交易次数。下一步是计算每只股票的概率。如图 3 所示，盘中强度并不恒定，因此我们计算每半小时（ $T=30$ 分钟）间隔的概率，并将其平均值（按强度加权）作为最终的每日概率。

图表 5 展示了按时间和股票平均的空概率和经验日概率。考虑到极小的邻域大小（ $\delta = 500\mu s$ ），如果没有共现，92.67% 的交易应该是被隔离的，然而市场上只有 38.29% 的交易是隔离交易。

图表 5、每种交易流的零概率和经验概率

	Null	Empirical
iso	92.67	38.29
nis	7.33	61.71
nis-s	0.02	29.83
nis-c	7.31	15.58
nis-b	0.00	16.30

该表显示了在零模型和真实数据下每类交易在天数和股票数中所占的平均百分比。

资料来源：arXiv，兴业证券经济与金融研究院整理

总之，有经验证据表明，交易共同发生的概念捕捉到了信号，这就促使我们对分解交易流并分别进行研究。

4.4、订单失衡的描述性统计

根据共现类型标记的交易，我们计算了每日订单失衡情况，并在图表 6 中报告了描述性统计数据。面板 A 记录了每类订单失衡的平均时间和股票。总体而言，平均无条件订单失衡为正值。在分解后，隔离和非自我隔离的订单失衡往往为正，其均值和方差均高于无条件订单失衡。与此相反，非交叉隔离和非两者隔离失衡的均值为负，但方差更大。因此，我们的研究本质上是以交易共现作为设定条件来构建具有不同行为的特征。然而，标准差远大于平均值，因此从统计学角度看，平均值与零并无显著差异。因此，均值只能作为潜在信号的一个非常微弱的指示。

面板 B 显示了各类订单失衡的平均偏自相关性。可以看出，所有订单失衡都是正的自相关。COIs 滞后 1 期的自相关性很大。在条件失衡中，非交叉隔离的订单失衡与市场上其他股票的交易密切相关，其自相关性相对较高。相比之下，非自我隔离交易订单失衡的自相关性相对较低。随着滞后期的增加，这些偏自相关性急剧下降。

图表 6、所有交易组和订单失衡的汇总统计

Panel A: Statistics of daily order imbalances			
	Mean	Median	Std. dev.
all	0.0024	0.0026	0.1163
iso	0.0108	0.0105	0.1493
nis	-0.0037	-0.0028	0.1256
nis-s	0.0090	0.0087	0.1471
nis-c	-0.0265	-0.0225	0.1847
nis-b	-0.0163	0.0128	0.1985

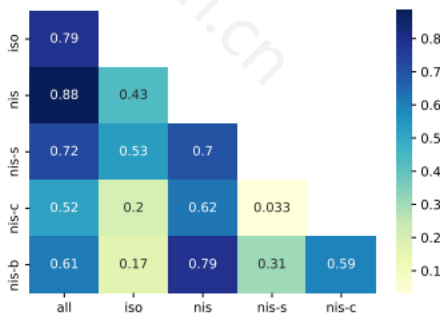
Panel B: Average partial autocorrelation of order imbalances			
lag	1	2	3
all	0.270	0.089	0.040
iso	0.259	0.094	0.045
nis	0.261	0.092	0.044
nis-s	0.213	0.084	0.045
nis-c	0.316	0.124	0.068
nis-b	0.243	0.096	0.059

本表显示了所选 457 只股票在 2017-01-03 至 2019-12-10 期间的 COIs 统计摘要。面板 A 记录了 COIs 的均值、中位数和标准差。面板 B 显示了 COIs 的部分自相关性，取所有股票平均值。

资料来源：arXiv，兴业证券经济与金融研究院整理

图表 7 展示了在 $\delta = 500$ 微秒下 COIs 的皮尔逊相关性(所有股票的平均值)。所有类型的订单失衡都相互正相关，但强弱不同，可能相当低。无条件订单失衡是一个例外，它与其他所有类型的订单失衡都有很强的相关性。隔离失衡与非隔离失衡及其子类型之间的相关性很低。

图表 7、订单失衡的皮尔逊相关性



对于每种类型的订单失衡，我们首先考虑 2017-01-03 至 2019-12-10 期间的日值向量，然后计算相关矩阵，最后求出所有股票的平均值。

资料来源：arXiv，兴业证券经济与金融研究院整理

正如预期的那样，隔离和非隔离条件会产生不同的特征。此外，通过分解非隔离交易得到的三个订单失衡与非隔离订单失衡的总和也有很强的相关性，但相互之间的相关性较弱。更详细地探讨它们之间的关系后，我们发现来源于未与市场其他股票共同交易的非自我隔离交易的订单失衡与隔离订单失衡的相关性相对更高。相比之下，与市场联系更紧密的非交叉隔离交易和非两者隔离交易的订单失衡与隔离订单失衡的相关性较低。因此，我们确信分解后的订单失衡是可区

分的特征，所有成对相关系数都小于 0.6，并且它们可以揭示股票市场的结构特性，而这些特性是无法通过观察总订单流量推断出的。

5、条件性订单失衡对同期价格的影响

为了评估各类订单失衡对未来回报的同期影响，我们对每只股票 i 采用以下时间序列回归方法：

$$r_{i,t} = \alpha_i + \beta_i COI_{i,t}^* + \epsilon_{i,t}, \quad (6)$$

其中， $r_{i,t}$ 是股票 i 在时间 t 的收益率；假定 $\epsilon_{i,t}$ 为均值为零的正态分布，随机变量 $\{i,t\}$ 假定为独立变量。在推断时，我们采用 Newey-West 估计法（Newey 和 West（1994））修正残差项 $\epsilon_{i,t}$ 中的异方差性和自相关性。我们采用显著性水平为 5% 的双尾 t 检验。

在图表 8 中，我们报告了同期回归的结果。与之前的研究一致，几乎所有股票的无条件订单失衡都与收益率显著正相关。此外，我们的条件订单失衡（COI）对当日同期收益率也有显著的正向影响，尤其是隔离的 COI。值得注意的是，分解非隔离交易产生的三种订单失衡类型（nis-s、nis-c 和 nis-b）对系数平均值和显著系数比例的影响相对较弱。

图表 8、同期回归结果

	Average coefficient	Percentage positive	Percentage positive and significant	Percentage negative and significant	Average R^2 (%)
all	2.56	100.00	97.16	0.00	7.16
iso	2.00	99.34	96.06	0.00	7.15
nis	1.80	97.81	91.03	0.22	4.55
nis-s	1.11	88.62	69.80	2.64	2.83
nis-c	0.82	93.87	76.37	0.87	2.48
nis-b	0.84	95.62	81.62	0.66	2.61

本表总结了使用公式(6)进行 457 次回归的结果。‘Average coefficient’表示所有回归系数的平均值。‘Percentage’表示所选股票的比例。‘Significant’表示使用双尾 t 检验在 5% 的显著性水平上具有统计学意义。‘Average R^2 ’表示所有股票回归 R^2 的平均值。

资料来源：arXiv，兴业证券经济与金融研究院整理

重点关注由同期收益率回归所解释的方差百分比，我们发现以所有交易计算的“all”组 COI 产生的 R^2 最高为 7.16%。对交易流进行分解后，“iso”组 COI 虽然只计算了 38.29% 的交易，但其解释的方差与“all”组 COI 相当。与“all”组或“iso”组 COI 相比，用“nis”组 COI 对回报率进行回归所得到的 R^2 较低。因此价格影响并不与数量成正比，而似乎是由交易类型驱动的。这表明，与市场上其他交易有明显共现关系的交易所产生的价格压力是不均匀的且值得单独研究。

除了研究个体条件订单不平衡对收益的显著影响外，我们还对分解总订单失

衡所获得的额外信息感兴趣。为此，我们拟合了以下三个回归方法：

$$r_{i,t} = \alpha_i + \beta_i^{all} COI_{i,t}^{all} + \epsilon_{i,t}$$

$$r_{i,t} = \alpha_i + \beta_i^{iso} COI_{i,t}^{iso} + \beta_i^{nis} COI_{i,t}^{nis} + \epsilon_{i,t}$$

$$r_{i,t} = \alpha_i + \beta_i^{iso} COI_{i,t}^{iso} + \beta_i^{nis-s} COI_{i,t}^{nis-s} + \beta_i^{nis-c} COI_{i,t}^{nis-c} + \beta_i^{nis-b} COI_{i,t}^{nis-b} + \epsilon_{i,t}, (7)$$

其中， β_i^* 为回归系数， $\epsilon_{i,t}$ 为假定为高斯噪声的残差项。每次回归都会输入一组从整个交易流分区中得出的 COIs。我们使用 R^2 来评估其性能。考虑到特征数量的影响，我们还使用 *adjusted* R^2 作为评价指标。

从图表 9 中我们可以看到，考虑交易类型后， R^2 和 *adjusted* R^2 都有明显改善。以无条件模型为基准，将市场订单分为隔离订单和非隔离订单后，总方差的解释率增加了 0.92%。通过进一步分解非隔离交易， R^2 增加了 1.05%。这两项改进在统计上都很显著。综上，我们成功地将具有不同同期价格影响的交易从整个交易流中分离出来，且这样的分解有助于解释每日价格变化。

图表 9、同期回归 R^2

	Average coefficient	Percentage positive	Percentage positive and significant	Percentage negative and significant	Average R^2 (%)
all	2.56	100.00	97.16	0.00	7.16
iso	2.00	99.34	96.06	0.00	7.15
nis	1.80	97.81	91.03	0.22	4.55
nis-s	1.11	88.62	69.80	2.64	2.83
nis-c	0.82	93.87	76.37	0.87	2.48
nis-b	0.84	95.62	81.62	0.66	2.61

我们对选定的 457 只股票中的每只股票进行方程(7)中的三次回归，并报告 R^2 和 *adjusted* R^2 的平均值。我们在 ‘*Adj. R²*’ 一栏中记录了每次回归的 *adjusted* R^2 与前一行回归相比的增长情况。上标***表示使用单尾配对 t 检验在 1% 的范围内显著，最后一列是对应的 t 值。

资料来源：arXiv，兴业证券经济与金融研究院整理

6、失衡对未来回报率的预测能力

结合订单失衡的同期效应，研究其预测能力也很重要。在本节中，我们发现 iso 和 nis-s 订单失衡与未来收益率呈正相关，而 nis、nis-c 和 nis-b 的 COIs 与未来收益率呈负相关。此外，我们还发现，分解交易流以及同时使用多个 COIs 包含了预测次日回报的信号。我们利用回归和投资组合排序方法进行举证。

6.1、时间序列回归

我们重复上一节的回归分析思路，探讨 COIs 与前一天市场超额收益之间的关系。首先，我们将未来收益率与当前 COIs 分别进行回归，模型为：

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} COI_{i,t}^* + \epsilon_{i,t+1}, \quad (8)$$

其中， $\beta_{i,t}$ 为回归系数， $\epsilon_{i,t}$ 为残差项，假定残差项服从均值为零的独立同分布正态分布。

图表 10 列出了回归结果。正如预期，与同期影响不同，显著系数的大小和百分比都很低，无条件订单失衡的平均系数约等于零。在我们的研究期间，我们无法论证 Chordia 和 Subrahmanyam (2004) 提出的理论模型，即在没有未来订单失衡的情况下，失衡与前一天收益之间会产生正相关关系。然而，通过对交易进行分类，我们可以稍微加强上述信号。我们的研究结果表明，隔离订单和非自我隔离订单执行所产生的价格压力显示出适度的预测能力。此外，非隔离 (nis)、非交叉隔离 (nis-c) 和非两者隔离 (nis-b) 的交易失衡与未来价格变化呈负相关。其中，“nis”COI 的平均系数最大，10.5% 的个股估计系数为负且显著，而只有 1.31% 的系数为正且显著。就 R^2 而言，所有分解的交易流的 COI 均优于未分解（即总的）交易流的 COI。

图表 10、预测性回归

	Average coefficient	Percentage positive	Percentage positive and significant	Percentage negative and significant	Average R^2 (%)
all	-0.04	47.48	2.84	5.25	0.15
iso	0.05	56.89	5.47	2.84	0.17
nis	-0.09	42.89	1.97	6.13	0.15
nis-s	0.03	54.92	4.81	3.28	0.16
nis-c	-0.10	33.26	1.31	10.50	0.20
nis-b	-0.08	34.79	1.31	6.13	0.15

本表总结了对选定的 457 只股票每只都使用公式(8)进行线性回归的结果。‘Average Coefficient’表示所有回归系数的横截面平均值。‘Percentage’表示所选股票的比例。‘Significant’表示使用双尾 t 检验在 5% 的显著性水平上具有统计意义。‘Average R^2 ’表示所有股票回归 R^2 的平均值。

资料来源：arXiv，兴业证券经济与金融研究院整理

接下来，我们根据以下模型，将未来 1 天的股票回报率与三组 COIs 分别进行回归：

$$r_{i,t+1} = \alpha_i + \beta_i^{all} COI_{i,t}^{all} + \epsilon_{i,t}$$

$$r_{i,t+1} = \alpha_i + \beta_i^{iso} COI_{i,t}^{iso} + \beta_i^{nis} COI_{i,t}^{nis} + \epsilon_{i,t}$$

$$r_{i,t+1} = \alpha_i + \beta_i^{iso} COI_{i,t}^{iso} + \beta_i^{nis-s} COI_{i,t}^{nis-s} + \beta_i^{nis-c} COI_{i,t}^{nis-c} + \beta_i^{nis-b} COI_{i,t}^{nis-b} + \epsilon_{i,t}, \quad (9)$$

其中， β_i^* 为回归系数， $\epsilon_{i,t}$ 表示残差项，假定为独立且同分布的均值为零的高斯噪声分布。

从图表 11 中，我们可以看到 R^2 和 $adjusted R^2$ 值都有稳定的提高。从横截面来看， $adjusted R^2$ 的增长在 1% 的水平上具有统计意义。因此，我们得出结论，

以共同发生为条件的订单失衡有助于短期回报预测。我们推测，根据这种共现性对交易流进行分解可以提升对未来回报的预测能力。

图表 11、预测性回归的 R^2

	R^2	adj. R^2	Δ adj. R^2	t value
all	0.15	0.016	—	—
iso + nis	0.34	0.061	0.045***	15.92
iso + nis-s + nis-c + nis-b	0.64	0.094	0.033***	19.82

我们对选定的 457 只股票中的每只股票进行方程(9)中的三次回归，并报告 R^2 和adjusted R^2 的平均值。我们在‘ Δ adj. R^2 ’一栏中记录了每次回归的adjusted R^2 与前一行回归相比的增长情况。上标***表示使用单尾配对 t 检验在 1%的范围内显著，最后一列是对应的 t 值。

资料来源：arXiv，兴业证券经济与金融研究院整理

6.2、基于“失衡”的投资组合排序

为了证实我们对未来回报与不同类型 COI 之间正负关系的研究结果，我们采用了投资组合排序法（Cattaneo 等人（2020）、Fama 和 French（1993））将订单失衡转化为投资组合。对于每类 COI，我们根据其失衡值从低到高将股票分为 5 个五分位组合。考虑到多重特征，我们进一步为每对 COIs 创建 5×5 双重排序组合。失衡排序的投资组合权重相同，只持有股票的多头头寸，投资组合的日收益率按所有股票的平均收益率计算。在 2017-01-03 至 2019-12-10 的整个样本期内，对失衡排序投资组合进行的回溯测试强化了这一发现，即 iso 和 nis-s 失衡是动量信号，而 nis、nis-c 和 nis-b 失衡是反转信号，它们对未来回报的影响不同。在 2017 年 1 月 3 日至 2019 年 12 月 10 日的整个样本期间，对不平衡排序的投资组合进行的回溯测试强化了这一发现：iso 和 nis-s 失衡是动量信号，而 nis、nis-c 和 nis-b 失衡是反转信号，它们对未来回报的影响各不相同。

6.2.1、单一排序投资组合

图表 12 面板 A 记录了单一排序投资组合的年化收益率。我们注意到，在第一行中，无条件失衡排序投资组合的收益率为负值，且在五分位数上波动，这证实了无条件订单失衡与未来收益率之间没有明显的线性关系。然而，在分解交易流后，我们发现随着等序失衡的增加，回报率的增长是单调的，这加强了等序失衡与未来回报率的正相关性。同样，nis-s 也呈上升趋势，这也是正相关的表现。与此相反，我们观察到其他类型 COIs 的平均回报率在下降，这与我们的时间序列回归结果一致，证实了负相关关系。

面板 B 显示了每个投资组合中股票的日平均 COIs。COIs 是有符号的，表示为“Low”和“High”投资组合对应于符号相反的强信号。我们观察到，所有信号强度的分布大致对称，且以 0 为中心。在每一行中，不存在由平均 COI 值无差别的股票组成的五分位投资组合。然而，五等分法的投资组合收益既不对称，也不单调（iso 除外）。通过比较面板 A 各行的回报率，我们发现最大正回报率的幅度总是

小于最小负回报率的绝对值。因此，我们推测 COIs 对未来回报的正负影响是不对称的，负值影响更大。

此外，就对未来回报的负面影响而言，COIs 的最大值并不会导致次日的最大跌幅。例如，“iso”COI 的最低五分位数和第二五分位数投资组合的回报率相似，而对于“nis”COI 的投资组合，第四五分位数投资组合的平均回报率最低，为-3.85%，而最高的五分位数投资组合的回报率则上升至-2.63%。对于上述现象，我们解释为极端不平衡会导致次日出现强劲反转，因为一些投资者的目标是保持稳定的风险敞口水平。

图表 12、单一排序投资组合汇总

Pannel A: Annualized returns					
	Low	2	3	4	High
all	-1.57	-1.77	-1.95	-3.61	-0.49
iso	-4.25	-4.11	-2.13	-0.91	2.00
nis	0.54	-1.76	-1.69	-3.85	-2.63
nis-s	-4.09	-2.56	-2.28	-2.70	2.25
nis-c	4.67	-0.56	-2.75	-5.72	-5.03
nis-b	3.60	-1.91	-1.87	-4.55	-4.66

Pannel B: Average daily COIs					
	Low	2	3	4	High
all	-0.16	-0.05	0.00	0.06	0.17
iso	-0.20	-0.06	0.01	0.08	0.22
nis	-0.18	-0.06	0.00	0.06	0.17
nis-s	-0.20	-0.07	0.01	0.08	0.21
nis-c	-0.24	-0.10	-0.03	0.05	0.19
nis-b	-0.28	-0.11	-0.01	0.08	0.24

本表显示按 COI 排序的五分位投资组合的统计数据。每行包含五个投资组合，其构建方法是每天按照前一天 COI 的类型（由行指数表示）从低到高对所有股票进行排序，并将每只股票分配到列名所示的相应五分位投资组合中。断点分别为每日计算的各类 COI 的 20%、40%、60% 和 80%。面板 A 显示了每个投资组合的年化收益率，计算方法是将 2017-01-03 至 2019-09-10 期间的每日收益率取平均值，再乘以 252。面板 B 报告了样本期内投资组合中股票的平均每日 COI。

资料来源：arXiv，兴业证券经济与金融研究院整理

6.2.2、双排序投资组合

为了进一步研究 COIs 之间的相互作用，我们通过对分解交易流的每一对失衡信号进行独立的双重排序来建立投资组合。图表 13 列出所有投资组合的年化收益率，其中每个区块包含 25 个投资组合，根据行名和列名所示的一对信号进行排序。

在每列等值区块中，平均收益率随着隔离交易的低 COIs 值到高 COIs 值依次上升。与此相反，使用等 COI 进行控制时，收益通常会随着非隔离 COI 值变高而下降。对最强信号进行双重排序，可在区块的右上角和左下角获得最高和最低收益。最强回报的幅度为 11.37%和-14.53%，与单个信号排序的最强回报幅度相比也有所提高。除了 nis-s-nis-b，对每对动量和反转 COI 特征进行双重排序时，都会

出现相同的模式和改进。然而与未来回报率具有相同关系的成对 COIs 的模式并不明显。例如，从等 iso-nis-s 排序的区块来看，我们无法观察到沿行和沿列的单调模式。因此，我们推测，动量信号和反转信号带有不同的信息，同时纳入这两种信号可提高预测能力。

图表 13、双排序投资组合的年收益率

Low 2 3 4 High						Low 2 3 4 High					
nis						nis-s					
Low	-0.99	-3.96	-7.91	-10.97	-14.53		-4.04	-3.37	-3.54	-11.71	-4.70
2	-1.85	-6.37	-1.72	-4.35	-9.36		-8.91	-4.41	-3.37	0.18	0.84
3	1.53	1.40	-0.36	-3.42	-9.84	iso	-2.00	-2.34	-1.10	-4.17	1.85
4	6.67	-1.32	-0.78	-4.34	-2.51		5.85	-1.85	-2.93	-4.71	2.26
High	11.37	5.12	2.63	-0.14	2.13		-9.24	3.73	3.19	1.83	2.54
nis-c						nis-b					
Low	0.26	-3.51	-7.36	-6.75	-11.63		0.58	-4.41	-6.89	-8.08	-9.22
2	7.20	-1.49	-7.10	-8.76	-9.40		1.06	-3.93	-3.07	-9.25	-6.86
3	5.94	-0.05	-1.30	-5.75	-6.93	iso	5.64	-1.27	0.59	-7.17	-6.85
4	3.54	0.82	3.63	-7.87	-4.87		6.90	1.07	-0.69	-5.00	-5.66
High	17.65	8.04	-0.13	3.56	-1.16		9.39	2.96	0.78	5.28	0.27
nis-s						nis-c					
Low	-1.50	2.03	9.51	3.92	12.41		3.90	-1.35	-4.49	-4.44	-8.27
2	-8.91	-1.40	-0.68	5.32	10.06		2.94	-2.89	0.09	-7.59	-0.93
3	-6.94	-5.46	-2.40	0.48	9.35	nis	7.58	3.77	-8.48	-4.87	-7.06
4	0.15	-9.50	-5.94	-3.67	3.16		3.42	-4.54	-0.17	-7.20	-5.95
High	-19.3	1.73	-11.61	-9.62	0.44		5.49	-3.71	3.11	-2.51	-4.50
nis-b						nis-e					
Low	2.88	-4.62	-6.16	0.16	0.95		2.49	-4.20	-6.03	-8.86	-8.27
2	5.12	-3.16	-1.69	-12.87	-1.21		1.15	-1.41	-1.86	-6.83	-3.61
3	10.16	-0.26	-1.99	-7.03	2.52	nis-s	6.24	-2.66	-4.14	-4.35	-6.79
4	6.24	8.17	-1.41	-6.59	-7.73		4.95	0.67	-5.2	-5.53	-6.93
High	-7.31	13.17	-3.48	2.61	-4.79		8.40	5.82	6.08	-1.66	-2.84
nis-b						nis-b					
Low	1.21	-10.21	-8.62	-9.36	-4.83		6.42	1.68	2.42	5.58	6.17
2	1.10	-0.92	-1.23	-12.11	-3.38		2.27	-1.12	1.63	-5.22	-4.60
3	7.68	-4.22	-0.69	-5.57	-11.59	nis-c	1.50	-2.99	-1.39	-7.6	-1.43
4	5.73	2.84	-2.01	-2.46	-10.15		4.19	-5.67	-8.07	-6.04	-5.22
High	6.14	11.71	1.79	-0.02	-0.16		-9.01	-4.11	-1.08	-3.45	-6.83

本表显示基于每一对 COIs 的双排序投资组合的年化收益率。每个面板包含 5×5 双排序投资组合。为了构建投资组合，我们每天将所有股票按照前一天 COIs 的两种类型（以行指数和列名称表示）从低到高分别排序到五个五分位数。然后将两种排序的交集创建 25 个双排序投资组合。每个投资组合的年化收益率是通过取其从 2017-01-03 到 2019-09-10 的每日收益率的平均值并乘以 252 计算得出的。

资料来源：arXiv，兴业证券经济与金融研究院整理

7、条件订单失衡的投资价值

如前几节所述，有证据表明条件订单失衡包含解释和预测个股收益的信号。在本节中，我们利用排序形成多空投资组合，从而利用其经济价值。我们基于失衡的交易策略产生了明显的利润和显著的异常回报。高额的交易利润也为分解后的交易流的 COIs 所具有的预测能力提供了重要证据。

7.1、多-空投资组合的构建和评估

我们根据不平衡排序的五分位投资组合设计实用的交易策略。在每个交易日的上午 9:30 时，我们买入第一个（或最后一个）五分位数组组合，并卖空最后一个（或第一个）五分位数组组合，以获得相同数量的动量（或反转）信号，从而实现自我平衡。每天，我们在下午 16:00 关闭所有仓位，以避免隔夜效应。总体而言，每日收益率是多头和空头不平衡排序投资组合收益率之间的差额。

为了评估盈利能力，我们比较了投资组合的年化收益率以及年化夏普比率（Sharpe（1994）），定义为：

$$SR_p := \frac{\text{mean}(R_{p,t}) - R_f}{\text{std}(R_{p,t})} \times \sqrt{252}, \quad (10)$$

其中， $R_{p,t}$ 为投资组合的日收益率， R_f 为相关期间的日平均无风险利率， R_f 等于0.00625%。

7.2、盈利能力

我们构建了多空投资组合，并在图表 14 中报告了其盈利能力指标。面板 A 显示了年化收益率，行名和列名分别表示了基于 COIs 的单排序和双排序投资组合的对角线和非对角线数值。我们发现，纳入多个 COIs 可以提高投资组合的收益，这也证明了交易流分解技术可以创建有利可图的 COI 信号。例如，“iso-nis”双排序对应的多空策略收益率为 25.85%，比简单地分别对“iso”和“nis”COI 排序高出 19.60%和 22.67%。通过对“nis-s”和“nis”COI 进行双重排序，最高年化收益率达到 31.71%。一般来说，如果将对角线以外的值与行列上相应的对角线值进行比较，我们会发现所有 10 个双重排序的多空组合的收益率至少高于其中一个单一排序的基准值，而且 10 个组合中有 9 个组合的收益率高于两个基准值。

面板 B 中的夏普比率证实了我们对 COI 信号经济价值的发现。经风险调整后，我们的交易策略仍然有利可图，双重排序优于单一信号交易。以“iso”和“nis-c”排序的投资组合的夏普比率最高，达到 2.38，其次是以“nis-s-nis”排序的投资组合的夏普比率为 2.09。此外，80%的非对角线值至少高于一个相应的对角线值，50%的非对角线值高于两个对角线值。因此，有迹象表明，对于投资者来说，在做出交易决策时纳入多种类型的 COIs 是有经济效益的。

图表 14、多空投资组合的盈利能力

Panel A: Annualized returns					
	iso	nis	nis-s	nis-c	nis-b
iso	6.25	25.85	6.58	29.39	18.55
nis		3.18	31.71	8.41	7.65
nis-s			6.34	16.74	11.10
nis-c				9.70	13.25
nis-b					8.26
Panel B: Annualized Sharpe ratios					
	iso	nis	nis-s	nis-c	nis-b
iso	1.29	1.67	0.99	2.38	1.77
nis		0.41	2.09	1.19	1.26
nis-s			1.36	1.65	0.72
nis-c				1.62	1.78
nis-b					1.58

本表显示了按相应的行指数和列名称表示的 COIs 排序的多空投资组合的年化收益率和夏普比率。对角线值和非对角线值分别表示单排序和双排序投资组合。面板 A 列出了投资组合的年化收益率，计算方法是将 2017-01-03 至 2019-09-10 期间的每日收益率取平均值，然后乘以 252。面板 B 报告了根据公式 (10) 计算的样本期内的年化夏普比率。

资料来源：arXiv，兴业证券经济与金融研究院整理

从资产定价的角度来看，COIs 是异常收益的独特且重要的来源。我们将多空投资组合的超额收益与 Fama-French 五因子 (Fama 和 French (2015)) 进行回归，并在图表 15 中显示了它们的 α 值。除了基于“nis”COI 的单一排序外，所有投资组合都产生了统计意义上的显著异常回报，这证明利润无法用普通风险因子来解释。

图表 15、多空投资组合的异常回报

Annualized alpha					
	iso	nis	nis-s	nis-c	nis-b
iso	5.23** (2.46)	5.23*** (2.79)	5.11* (1.70)	29.11*** (4.21)	17.27*** (3.06)
nis		1.54 (0.68)	31.02*** (3.63)	7.29** (2.14)	6.27** (2.19)
nis-s			4.92** (2.38)	15.32*** (2.82)	9.72 (1.25)
nis-c				7.78*** (2.65)	12.05*** (3.10)
nis-b					6.63*** (2.65)

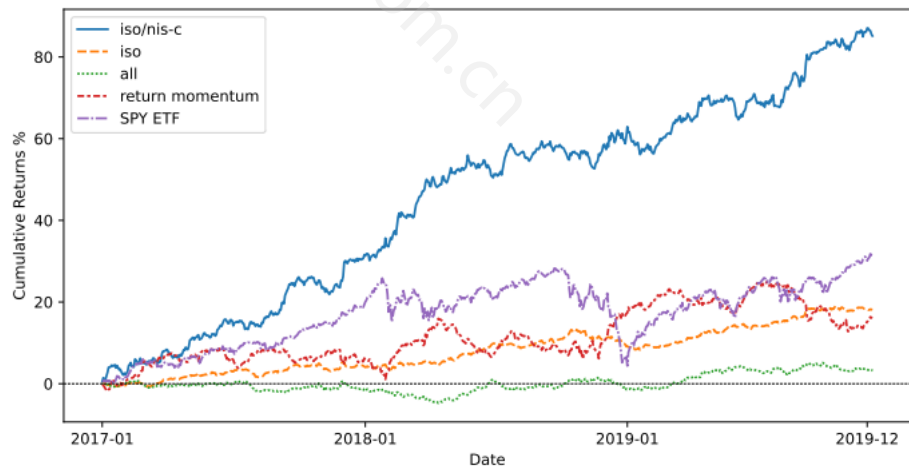
本表记录了与 Fama-French 5 因子模型 (Fama 和 French (2015)) 相关的多空投资组合的异常收益率 α 。对于每个多空投资组合，我们根据 5 个因子对投资组合的超额收益进行时间序列回归，

$$R_{p,t} - R_{f,t} = \alpha_p + b_p(R_{M,t} - R_{f,t}) + s_pSMB_t + h_pHML_t + r_pRMW_t + c_pCMA_t + e_{p,t}$$

其中， α_p 为投资组合的异常收益率，解释变量为市场、规模、价值、盈利能力和投资因素， $e_{p,t}$ 为特异性项。对角线值和非对角线的数值分别表示单排序多空组合和双排序多空组合。上标*、**和***表示 10%、5%和 1% 的统计显著性，括号内为相应的 t 值。

资料来源：arXiv，兴业证券经济与金融研究院整理

图表 16、投资组合的累计回报



本图绘制了 2017-01-03 至 2019-12-10 期间五个投资组合的累计收益率。这些组合包括：(1) “iso/nis-c”：根据“iso”和“nis-c”COIs 双重排序的多空组合；(2) “iso”：根据“iso”COI 单一排序的多空组合；(3) “all”：根据分解后的交易流量 COI 单一排序的多空组合；(4) “return momentum”：根据前几日回报单一排序的多空组合；(5) “SPY ETF”：跟踪标准普尔 500 指数的 SPDR 标准普尔 500 ETF 信托基金。

资料来源：arXiv，兴业证券经济与金融研究院整理

我们还将我们的策略与三个基准投资组合进行了比较。第一，我们构建了一个包含“all”COI 的多空投资组合，以评估分解交易流的经济价值。第二，我们根

据前几天的收益率构建了另一个基准多空投资组合。由于 COIs 与同期收益率存在显著相关性，因此有必要证明价格并不能完全揭示盈利能力。第三，我们选择 SPY 作为可交易市场组合，以整体市场表现为基准。在基于 COI 的策略中，我们选择“iso”COI 的单排序组合以及“iso”和“nis-c”COI 的双排序组合作为代表。图表 16 显示了选定的基于 COI 的多空投资组合和基准的累计回报。在测试期间，我们发现使用分解交易流的 COIs 没有获得任何利润。回报动量多空组合和 SPY 的夏普比率分别为 0.45 和 0.73。“iso”单一排序组合的年化收益率与收益动量相似，但波动性更小，夏普比率达到 1.29。显然，双排序投资组合的回报率和夏普比率都优于其他所有投资组合。

8、稳健性分析

在本节中，我们将简要评述交易共现的识别的稳健性以及条件订单失衡的构建。更多详情见附录。

8.1、领域规模效应

交易共现的定义和单个交易的分类取决于邻域大小 δ 的选择。当考虑到 $\delta = 0$ 的极端情况时，所有交易都是相互隔离的。随着 δ 的逐渐增大，隔离交易会变成非隔离交易的一个子类型。同时，非自我隔离交易和非交叉隔离交易都只能变成非两者隔离交易。最终，当 δ 足够大时，所有交易都是非隔离的；具体地说，所有交易都变成了非两者隔离的。因此，随着 δ 值的增大，隔离交易的数量会减少，非隔离交易和非两者隔离交易的数量会单调增加。因此，非自我隔离和非交叉隔离交易的数量最初会增加，达到各自的最大值后，它们开始减少。

我们对附录 A 中的 8 个 δ 值重复进行了分析。对于较小的邻域规模，同期影响和预测能力的模式是稳健的。然而，当 δ 达到 50 毫秒时，交易共现作为过滤器的性能就会下降。此外，不同类型的 COIs 在不同的 δ 值下都能达到最佳效果，这表明该方法在结合来自多个 δ 值的信号方面具有潜在优势。

8.2、日内时间效应

不同盘中时段的交易活动对价格有不同的影响。我们注意到，交易活动在每个交易日的第一个小时和最后半个小时更为密集。最近的一些研究，如 Cont、Cucuringu 和 Zhang (2021) 在计算失衡时排除了这些波动时段以确保稳健性，而其他研究 (Chu 和 Qiu (2021)) 则特别关注这些半小时时段的失衡。考虑到日内时间效应，我们分别研究了三个时间段内的 COIs，即 9:30-10:00、10:00-15:30 和 15:30-16:00，并在附录 B 中记录了我们的研究结果。

我们关于同期回报与失衡关系的研究结果在每个时期都成立。此外，我们还发现分解后的交易流的预测能力来自不同的时间段。最后一小时的“iso”和“nis-s”COIs 有助于预测未来收益。另一方面，“nis-c”COI 的预测能力来自最后半小时

以外的时段。此外，对于“nis-b”交易，只有 10:00-15:30 的 COI 才有助于预测次日开盘至收盘的市场超额收益。

8.3、成交量测度的 COI

除了纳入交易数量外，人们通常还将订单失衡定义为买方和卖方发起的交易量之间的标准化差异。我们研究了个股收益与交易量订单失衡之间的关系，并分析了相应的交易策略。更多详情见附录 C。

在交易量衡量标准下，我们的研究结果是稳健的。我们观察到与计数 COIs 相同的模式，但注意到在所有交易类型中，与交易量失衡相关的同期回归 R^2 比与计数失衡相关的回归 R^2 低约 1%。这一发现与之前的研究(Chan 和 Lakonishok(1995)、Chordia 和 Subrahmanyam (2004))一致，这些研究提供的证据表明，交易数量能更好地捕捉机构为优化执行而拆分订单所带来的价格压力。

9、结论与展望

在本文中，我们提出了“交易共现”的概念，它将时间上相互接近的交易联系起来，从而能够在细粒度上研究股票交易之间的相互作用。在与其他交易共同发生的条件下，我们将每笔交易分为五组。我们计算每种类型的分解交易流(COI)，并研究其对个股回报的同期影响、预测能力以及其经济价值。

我们的实证结果表明，分解交易流对价格的影响各不相同。“iso”交易流的 COI 就能解释当日收益率的变化，与使用未分解的总交易流的 COI 的解释能力相当，而加入交易流的 COIs 则进一步提高了解释能力。就可预测性而言，我们观察到，平均而言，未来回报与“iso”和“nis-s”的 COIs 呈正相关，而与“nis”、“nis-c”和“nis-b”的 COIs 呈负相关。此外，交易流分解具有显著的经济价值，根据以往 COIs 的发展方向构建多空投资组合会明显提高交易策略的盈利能力。

最后，我们提出了两个未来发展方向，特别是考虑到我们目前在数据可用性和计算能力方面的局限性。首先，我们通过经验证明了根据交易与其他交易的共同发生率来分解交易的重要性，但我们无法确定是谁发起了某些类型的交易。利用私人数据集区分不同类型的交易者(Tumminello 等人(2012)、Cont、Cucuringu、Glukhov 等人(2021))，并发现交易互动背后的机制，这将是一个有趣的研究方向。例如，检测机构等知情交易者是否会成功隐藏其交易目的，从而导致其交易极有可能与其他交易隔离，这将是一个令人感兴趣的问题。如果能识别出高频交易者，就值得应用共现分析来了解高频交易者如何对其他市场参与者的交易活动做出反应。其次，为了实现数据缩减的目的，我们只研究限价订单的执行情况，而非所有限价订单簿事件。过去的研究发现，限价订单的提交和取消也会对价格产生影响。将我们的想法扩展到限价订单的共同发生，并深入研究订单流失衡(Eisler、Bouchaud 和 Kockelkoren (2012)、Cont、Kukanov 和 Stoikov (2014)、Xu、Gould 和 Howison (2018)、Cont、Cucuringu 和 Zhang (2021))可能会很

有趣，这与我们的 COIs 类似。

参考文献

- [1] Jesse S Aaron, Aaron B Taylor, and Teng-Leong Chew. “Image co-localization-co-occurrence versus correlation”. In: *Journal of cell science* 131.3 (2018), jcs211847.
- [2] Anne E Appel and George W Holden. “The co-occurrence of spouse and physical child abuse: a review and appraisal.” In: *Journal of family psychology* 12.4 (1998), p. 578.
- [3] Miguel B Ara’ujo et al. “Using species co-occurrence networks to assess the impacts of climate change”. In: *Ecography* 34.6 (2011), pp. 897–908.
- [4] Warren Bailey et al. “Stock returns, order imbalances, and commonality: Evidence on individual, institutional, and proprietary investors in China”. In: *Journal of Banking & Finance* 33.1 (2009), pp. 9–19.
- [5] Markus K Brunnermeier and Lasse Heje Pedersen. “Predatory trading”. In: *The Journal of Finance* 60.4 (2005), pp. 1825–1863.
- [6] Matias D Cattaneo et al. “Characteristic-sorted portfolios: Estimation and inference”. In: *Review of Economics and Statistics* 102.3 (2020), pp. 531–551.
- [7] Sugato Chakravarty et al. “Clean sweep: Informed trading through intermarket sweep orders”. In: *Journal of Financial and Quantitative Analysis* 47.2 (2012), pp. 415–435.
- [8] Louis KC Chan and Josef Lakonishok. “The behavior of stock prices around institutional trades”. In: *The Journal of Finance* 50.4 (1995), pp. 1147–1174.
- [9] Tarun Chordia, Amit Goyal, and Narasimhan Jegadeesh. “Buyers versus sellers: who initiates trades, and when?” In: *Journal of Financial and Quantitative Analysis* 51.5 (2016), pp. 1467–1490.
- [10] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. “Order imbalance, liquidity, and market returns”. In: *Journal of Financial Economics* 65.1 (2002), pp. 111–130.
- [11] Tarun Chordia and Avanidhar Subrahmanyam. “Order imbalance and individual stock returns: Theory and evidence”. In: *Journal of Financial Economics* 72.3 (2004), pp. 485–518.
- [12] Xiaojun Chu and Jianying Qiu. “Forecasting stock returns using first half an hour order imbalance”. In: *International Journal of Finance & Economics* 26.3 (2021), pp. 3236–3245.

- [13] Rama Cont, Mihai Cucuringu, Vacslav Glukhov, et al. “Analysis and modeling of client order flow in limit order markets”. In: *Available at SSRN* (2021).
- [14] Rama Cont, Mihai Cucuringu, and Chao Zhang. “Price impact of order flow imbalance: multi-level, cross-sectional and forecasting”. In: *arXiv e-prints* (2021), arXiv–2112.
- [15] Rama Cont, Arseniy Kukanov, and Sasha Stoikov. “The price impact of order book events”. In: *Journal of financial econometrics* 12.1 (2014), pp. 47–88.
- [16] Justin Cox. “ISO order imbalances and individual stock returns”. In: *Journal of Financial Research* 44.1 (2021), pp. 5–23.
- [17] Ido Dagan, Lillian Lee, and Fernando CN Pereira. “Similarity-based models of word cooccurrence probabilities”. In: *Machine learning* 34.1 (1999), pp. 43–69.
- [18] Jonathan F Donges et al. “Event coincidence analysis for quantifying statistical interrelationships between event time series”. In: *The European Physical Journal Special Topics* 225.3 (2016), pp. 471–487.
- [19] Zoltan Eisler, Jean-Philippe Bouchaud, and Julien Kockelkoren. “The price impact of order book events: market orders, limit orders and cancellations”. In: *Quantitative Finance* 12.9 (2012), pp. 1395–1419.
- [20] Eugene F Fama and Kenneth R French. “A five-factor asset pricing model”. In: *Journal of financial economics* 116.1 (2015), pp. 1–22.
- [21] Eugene F Fama and Kenneth R French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of Financial Economics* 33.1 (1993), pp. 3–56.
- [22] Carolina Galleguillos, Andrew Rabinovich, and Serge Belongie. “Object categorization using co-occurrence, location and appearance”. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2008, pp. 1–8.
- [23] Nicholas J Gotelli. “Null model analysis of species co-occurrence patterns”. In: *Ecology* 81.9 (2000), pp. 2606–2621.
- [24] Sanford J Grossman and Merton H Miller. “Liquidity and market structure”. In: *the Journal of Finance* 43.3 (1988), pp. 617–633.
- [25] Li Guo et al. “News co-occurrence, attention spillover, and return predictability”. In: *Attention Spillover, and Return Predictability* (November 18, 2018) (2018).
- [26] Björn Hagströmer and Lars Nordén. “The diversity of high-frequency traders”. In: *Journal of Financial Markets* 16.4 (2013), pp. 741–770.
- [27] Nicholas Hirschey. “Do high-frequency traders anticipate buying and selling pressure?” In: *Management Science* 67.6 (2021), pp. 3321–3345.

- [28] Ruihong Huang and Tomas Polak. “Lobster: Limit order book reconstruction system”. In: *Available at SSRN 1977207* (2011).
- [29] Olga Kolesnikova. “Survey of word co-occurrence measures for collocation detection”. In: *Computacion y Sistemas* 20.3 (2016), pp. 327–344.
- [30] Alan Kraus and Hans R Stoll. “Parallel trading by institutional investors”. In: *Journal of Financial and Quantitative Analysis* 7.5 (1972), pp. 2107–2138.
- [31] Albert S Kyle. “Continuous auctions and insider trading”. In: *Econometrica: Journal of the Econometric Society* (1985), pp. 1315–1335.
- [32] Albert S Kyle, Hui Ou-Yang, and Bin Wei. “A model of portfolio delegation and strategic trading”. In: *The Review of Financial Studies* 24.11 (2011), pp. 3778–3812.
- [33] Yi-Tsung Lee et al. “Order imbalances and market efficiency: Evidence from the Taiwan Stock Exchange”. In: *Journal of Financial and Quantitative Analysis* 39.2 (2004), pp. 327–341.
- [34] Zhongming Ma, Gautam Pant, and Olivia RL Sheng. “Mining competitor relationships from online news: A network-based approach”. In: *Electronic Commerce Research and Applications* 10.4 (2011), pp. 418–427.
- [35] Darryl I MacKenzie, Larissa L Bailey, and James D Nichols. “Investigating species co-occurrence patterns when species are detected imperfectly”. In: *Journal of Animal Ecology* 73.3 (2004), pp. 546–555.
- [36] Whitney K Newey and Kenneth D West. “Automatic lag selection in covariance matrix estimation”. In: *The Review of Economic Studies* 61.4 (1994), pp. 631–653.
- [37] Maureen O’Hara. “High frequency market microstructure”. In: *Journal of Financial Economics* 116.2 (2015), pp. 257–270.
- [38] David S Scharfstein and Jeremy C Stein. “Herd behavior and investment”. In: *The American Economic Review* (1990), pp. 465–479.
- [39] William F Sharpe. “The sharpe ratio”. In: *Journal of Portfolio Management* 21.1 (1994), pp. 49–58.
- [40] Catherine Shenoy and Ying Jenny Zhang. “Order imbalance and stock returns: Evidence from China”. In: *The Quarterly Review of Economics and Finance* 47.5 (2007), pp. 637–650.
- [41] Matthew Spiegel and Avaniidhar Subrahmanyam. “On intraday risk premia”. In: *The Journal of Finance* 50.1 (1995), pp. 319–339.
- [42] Hans R Stoll. “The supply of dealer services in securities markets”. In: *The Journal of Finance* 33.4 (1978), pp. 1133–1151.

- [43] Yi Tang, Yilu Zhou, and Marshall Hong. “News co-occurrences, stock return correlations, and portfolio construction implications”. In: *Journal of Risk and Financial Management* 12.1 (2019), p. 45.
- [44] Michele Tumminello et al. “Identification of clusters of investors from their real trading activity in a financial market”. In: *New Journal of Physics* 14.1 (2012), p. 013041.
- [45] Vincent Van Kervel and Albert J Menkveld. “High-frequency trading around large institutional orders”. In: *The Journal of Finance* 74.3 (2019), pp. 1091–1137.
- [46] Qiong Wu et al. “A deep learning framework for pricing financial instruments”. In: *arXiv.org* (2019).
- [47] Ke Xu, Martin D Gould, and Sam D Howison. “Multi-level order-flow imbalance in a limit order book”. In: *Market Microstructure and Liquidity* 4.03n04 (2018), p. 1950011.
- [48] Liyan Yang and Haoxiang Zhu. “Back-running: Seeking and hiding fundamental information in order flows”. In: *The Review of Financial Studies* 33.4 (2020), pp. 1484–1533.
- [49] Shujing Ye et al. “Co-occurrence and interactions of pollutants, and their impacts on soil remediation—a review”. In: *Critical reviews in environmental science and technology* 47.16 (2017), pp. 1528–1553.
- [50] Ting Zhang, Gao-Feng Gu, and Wei-Xing Zhou. “Order imbalances and market efficiency: New evidence from the Chinese stock market”. In: *Emerging Markets Review* 38 (2019), pp. 458–467.

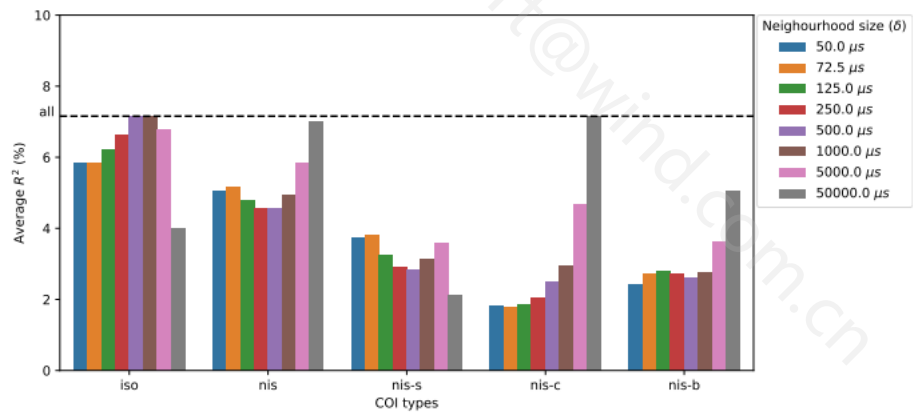
附录 A 领域规模效应

为了研究领域大小对条件订单不平衡的影响，我们对每个 $\delta \in \{0.05 \text{ ms}, 0.075 \text{ ms}, 0.125 \text{ ms}, 0.25 \text{ ms}, 0.5 \text{ ms}, 1 \text{ ms}, 5 \text{ ms}, 50 \text{ ms}\}$ 重复进行回归和组合分析，结果如图表 17 所示。

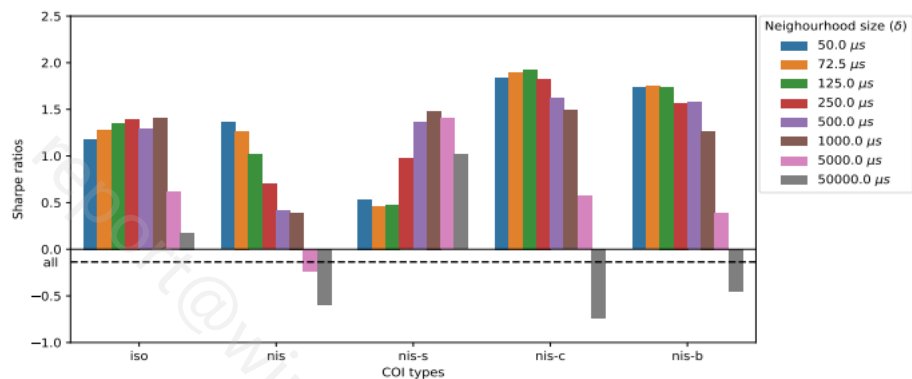
图表 17a 显示了同期回归的平均 R^2 。在 $\delta = 1$ 毫秒时，隔离订单失衡的 R^2 最高。相比之下，非隔离失衡的 R^2 直方图呈 U 型，在 $\delta = 0.5$ 毫秒时最小。对于三类非隔离订单失衡，非自我隔离和非两者隔离失衡的 R^2 随 δ 值的增加呈下降趋势。随着 δ 的增大，非交叉隔离的失衡现象能解释更多的回报差异。

图表 17b 详细显示了按 δ 排序的不同 COI 类型的多空投资组合的夏普比率。我们注意到，每种类型的订单失衡的夏普比率在不同的 δ 值上达到峰值。

图表 17、不同 δ 下的实证结果



(a) Average R^2 of regressing contemporaneous returns against COIs for different δ s.



(b) Sharpe ratios of COI-based long-short portfolios for different δ s.

(a) 图表示不同 δ 下同期收益率与 COIs 的平均回归 R^2 。

(b) 图表示不同 δ 下基于 COI 的多空投资组合的夏普比率。

资料来源：arXiv，兴业证券经济与金融研究院整理

附录 B 日内时间效应

我们研究了不同盘中时间间隔的 COIs。首先，我们将当时开盘到收盘的市场准入回报与每个 COI 分别进行回归，以评估它们对当日价格变化的影响。图表 18 面板 A 列出了所有此类回归的 R^2 。剔除最开始半个小时和最后半个小时的交易并不会明显改变我们发现的失衡与收益之间的关系。无论在哪个时段，只用“iso”交易得出的 COIs 都足以解释与使用所有交易时相近的方差。请注意，尤其是在第一小时，价格影响主要来自于隔离交易。

其次，我们通过构建单一序列的多空组合对每个 COI 进行交易，并在面板 B 中显示年化夏普比率。我们可以合理地预期，在正常交易期末期的交易活动更有助于预测 COI。我们观察到，最后一小时的“iso”和“nis-s”COIs 信号导致夏普比率分别上升 1.03 和 0.67，显著提高了投资组合的利润。相反，最后半小时的非交叉隔离 COI 并不是预测未来收益的良好信号。就非交叉隔离交易（nis-b）而言，只有在波动较小的交易时段，COI 才能预测未来收益。

图表 18、按时间段分列的 COIs

Panel A: Contemporaneous regression R^2 (%)				
	9:30 – 10:00	10:00 – 15:30	15:30 – 16:00	9:00 – 16:00
all	1.84	6.41	2.98	7.16
iso	3.23	6.21	2.06	7.15
nis	0.68	4.14	2.30	4.55
nis-s	0.61	2.48	1.15	2.83
nis-c	0.51	2.37	1.03	2.48
nis-b	0.21	2.38	1.39	2.61

Panel B: Annualized Sharpe ratios				
	9:30 – 10:00	10:00 – 15:30	15:30 – 16:00	9:00 – 16:00
all	-0.87	-0.65	0.79	-0.14
iso	-0.78	0.44	2.31	1.29
nis	0.73	0.67	-0.39	0.41
nis-s	-0.55	0.92	2.03	1.36
nis-c	1.21	1.39	0.29	1.62
nis-b	-0.18	1.57	0.33	1.57

我们计算了所选 457 只股票在 2017-01-03 至 2019-12-10 期间每天 9:30-10:00、10:00-15:30 和 15:30-16:00 的 COIs。面板 A 使用公式 (6) 对每只股票开盘至收盘的同期市场超额收益率与各类 COIs 的线性回归，并总结了所有股票的平均 R^2 。面板 B 显示了由公式 (10) 计算的基于不同盘中时间段 COIs 多空单一排序投资组合的年化夏普比率。两个面板的最后一列均以每日 COIs 为基准。

资料来源：arXiv，兴业证券经济与金融研究院整理

附录 C 成交量测度的 COI

本节中，我们不考虑交易数量，而是分析交易量订单流不平衡，定义为：

$$COI_{i,t}^{type} = \frac{V_{i,t}^{type,buy} - V_{i,t}^{type,sell}}{V_{i,t}^{type,buy} + V_{i,t}^{type,sell}} \#(11)$$

其中， $V_{i,t}^{type,buy}$ 和 $V_{i,t}^{type,sell}$ 分别表示股票 i 在第 t 天的市场买单和市场卖单总量。我们重复对成交量失衡的分析，结果见图表 19。

图表 19、成交量测度的 COIs

Panel A: Time series regression					
	Average coefficient	Percentage positive	Percentage positive and significant	Percentage negative and significant	Average R^2 (%)
all	2.14	100.00	97.59	0.00	6.27
iso	1.71	100.00	97.37	0.00	6.16
nis	1.52	98.25	91.68	0.00	3.97
nis-s	1.07	92.78	77.24	0.66	3.03
nis-c	0.58	93.43	65.65	0.66	1.22
nis-b	0.72	97.16	81.18	0.66	2.09
Panel B: Annualized Sharpe ratios					
	iso	nis	nis-s	nis-c	nis-b
iso	2.11	1.83	2.35	2.17	1.96
nis		-0.13	0.99	0.88	0.93
nis-s			1.39	1.79	1.22
nis-c				1.42	1.36
nis-b					1.18

我们由公式 (11) 计算了选定的 457 只股票在 2017-01-03 至 2019-12-10 期间以成交量衡量的 COIs。面板 A 总结了使用公式 (6) 对 457 只股票进行回归的结果。‘Average Coefficient’表示所有回归系数的平均值。‘Percentage’表示所选股票的比例。‘Significant’表示使用双尾 t 检验在 5% 的显著性水平上具有统计学意义。‘Average R^2 ’表示所有股票回归 R^2 的平均值。面板 B 显示按相应行指数和列名称表示的 COIs 排序的多空投资组合的年化夏普比率。对角线上和非对角线的数值分别代表单一排序和双重排序的投资组合。样本期内的年化夏普比率由公式 (10) 得出。

资料来源：arXiv，兴业证券经济与金融研究院整理

风险提示：文献中的结果均由相应作者通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。

定量研究专题报告

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并登记为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，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收到任何形式的补偿。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后的12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅。其中：沪深两市以沪深300指数为基准；北交所市场以北证50指数为基准；新三板市场以三板成指为基准；香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		增持	相对同期相关证券市场代表性指数涨幅在5%~15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%~5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们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，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

使用本研究报告的风险提示及法律声明

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告仅供兴业证券股份有限公司（以下简称“本公司”）的客户使用，本公司不会因接收人收到本报告而视其为客户。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约，投资者自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效，任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违反当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民（1934年美国《证券交易所》第15a-6条例定义为本「主要美国机构投资者」除外）。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本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。未经授权的转载，本公司不承担任何转载责任。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

兴业证券研究

上海	北京	深圳
地址：上海浦东新区长柳路36号兴业证券大厦15层 邮编：200135 邮箱： research@xyzq.com.cn	地址：北京市朝阳区建国门大街甲6号SK大厦32层01-08单元 邮编：100020 邮箱： research@xyzq.com.cn	地址：深圳市福田区皇岗路5001号深业上城T2座52楼 邮编：518035 邮箱： research@xyzq.com.cn