

学术前沿专栏:数智金融与资本市场高质量发展

【特约主持人】姚远:河南省特聘教授,河南省优秀青年社科专家

【主持人按语】资本市场是现代金融体系的重要组成部分,在服务实体经济、优化资源配置、防范化解金融风险 and 推动经济高质量发展中发挥着关键作用.近年来,随着我国资本市场改革持续深化,人工智能、大数据、高频交易和机器学习等新技术快速发展,正在深刻改变金融信息处理、资产价格形成与投资决策的方式.与此同时,资本市场也面临新的理论与现实问题.一方面,宏观经济、市场交易、企业盈利和投资者情绪等多源异构信息不断进入资产定价过程,如何利用机器学习方法提升资产收益预测能力,并进一步检验其经济价值,成为金融工程和资产配置研究的重要议题.另一方面,高频交易环境下市场交易行为更加复杂,异常成交与价格发现之间的关系更加隐蔽,如何基于公开高频数据识别潜在异常交易和市场操纵风险,也成为提升监管科技能力和维护市场秩序的重要问题.

在此背景下,探索数据智能赋能资本市场定价和风险治理的新方法,对于增强金融风险识别能力、推动资本市场高质量发展具有重要意义.在论文《虚假交易风险、成交量信息含量与价格发现——基于A股高频数据的经验证据》中,作者基于A股逐笔高频交易数据,构造背包式订单流闭合度指标,识别短窗口内具有高成交、高匹配但价格发现弱化特征的可疑订单流窗口,从成交量信息含量和价格发现视角为虚假交易风险的市场侧识别提供经验证据.在论文《可解释机器学习与股指收益率预测的经济价值——来自中国市场的证据》中,作者以中国股票市场代表性指数为研究对象,整合宏观经济、市场交易、企业盈利和投资者情绪等多模态信息,构建兼具预测能力和可解释性的混合机器学习框架,考察股指收益率预测的样本外表现、关键驱动因素及市场择时价值.两篇文章分别从异常交易识别和资产收益预测两个角度展开研究,既关注数据智能在投资决策和资产配置中的应用价值,也关注高频数据在市场微观结构分析和异常交易监测中的监管意义.

虚假交易风险、成交量信息含量与价格发现

——基于A股高频数据的经验证据

姚远

(河南大学 商学院,河南 开封 475004)

摘要:本文基于A股逐笔高频数据构造背包式订单流闭合度指标(KnapOFC),并通过两阶段框架识别在短窗口内同时呈现成交活跃异常、买卖订单高度匹配与价格发现弱化特征的可疑订单流窗口(NOC事件),以刻画虚假交

收稿日期:2026-05-30;修回日期:2026-06-22.

基金项目:国家社会科学基金(23BJY014).

作者简介(通信作者):姚远(1975—),女,河南开封人,河南大学教授,博士,博士生导师,研究方向为金融风险管理,
E-mail:yaoyuan@henu.edu.cn.

引用本文:姚远.虚假交易风险、成交量信息含量与价格发现——基于A股高频数据的经验证据[J].河南师范大学学报(自然科学版),2026,54(5):1-10.(Yao Yuan.Wash-trading risk, volume informativeness, and price discovery: evidence from high-frequency data in China's A-Share market[J].Journal of Henan Normal University(Natural Science Edition),2026,54(5):1-10.DOI:10.16366/j.cnki.1000-2367.2026.05.30.0001.)

易风险在公开市场数据中的间接痕迹.研究发现:第 1,A 股市场存在触发率较低但稳定可识别的 NOC 窗口,且在科创板和创业板相对更为活跃;第 2,相对于安慰剂窗口、高成交非 NOC 窗口与成交爆发匹配窗口,NOC 窗口的未来 30 min 价格影响及事件后短期已实现波动率均显著偏低,表明异常成交量未表现出与其成交规模相匹配的信息含量;第 3,NOC 窗口在订单级具有更高的 KnapOFC、更广的匹配覆盖度与更低的残余不平衡,并明显区别于高撤单、订单不平衡与尾市异动等传统异常形态.本文从成交量信息含量与价格发现视角,为高频数据条件下虚假交易风险的间接检测提供了市场侧证据,并为将订单级匹配结构纳入异常交易监测提供了可操作的指标工具.

关键词: 虚假交易风险;成交量信息含量;价格发现;背包式订单流闭合度;逐笔高频数据

中图分类号: F832.51;F830.91

文献标志码: A

文章编号: 1000-2367(2026)05-0001-10

近年来,随着高频交易快速发展,证券市场交易行为的复杂性和隐蔽性明显增强.监管层持续强调通过科技赋能监管,提升对内幕交易和市场操纵等证券期货违法违规行为的识别与处置能力.与此同时,从证监会处罚数据看,A 股市场人为操纵类违规案件在 2019 年后明显增加,并在 2020—2022 年处于较高水平^①.这表明,交易型操纵和异常交易行为仍是资本市场监管中的重点问题.由于监管处罚和违规公示通常具有事后性和滞后性,难以及时反映市场中尚未被识别的潜在风险,如何基于高频交易数据构造市场层面的间接识别指标,已成为异常交易监测中的重要问题.

已有研究基于监管处罚案例、账户级交易记录或市场监测数据识别操纵行为,并分析其市场后果^[1-3].这类研究的优势在于识别对象较为明确,但通常依赖账户身份,难以直接适用于账户身份不可见的公开高频数据环境.为缓解这一局限,后续研究开始利用价格偏离、收益回转、交易规模分布和订单流异常等市场可观测特征构造间接检测指标^[4-6].但是,现有间接识别思路多集中于价格异常或成交异常本身,对订单级买卖匹配结构及其与价格发现弱化之间关系的关注仍相对不足.

从操纵行为的经济机制看,交易型操纵并不一定依赖真实信息披露或企业基本面变化,而可能通过交易行为本身影响市场预期和价格形成.市场操纵可以划分为基于行动、基于信息和基于交易的操纵^[7];尤其在信息不对称情形下,操纵者可能通过交易误导市场并从价格偏离中获利^[8].此外,市场不确定性、账户交易行为和特定交易制度环境均可能影响操纵风险的形成与表现^[9-11].因此,在公开市场数据中识别虚假交易风险,不能只关注价格是否异常波动,还需要考察成交量、订单流和价格发现之间是否匹配.

市场微观结构理论进一步表明,成交量和订单流是价格发现的重要载体.知情交易者会通过订单流逐步将私人信息反映到价格中^[12];买卖价差和价格冲击也与信息不对称^[13]和交易信息含量^[14]密切相关;订单流信息含量研究同样表明,订单流不平衡、逆向选择风险与信息性交易之间存在紧密联系^[15-16].据此,若某一窗口内成交量异常放大主要来自信息性交易,通常应伴随较强的价格影响、后续价格调整或短期波动.相反,若成交量显著放大,但买卖方向高度对称、成交价格较为集中,同时价格影响和后续波动偏弱,则该成交量信号可能并未体现真实信息进入市场,而更可能反映低信息含量的订单流结构.

高频数据为识别短时间窗口内的异常订单流结构提供了新的可能.已有研究表明,算法交易和高频交易在一般情形下有助于提高市场流动性和价格发现效率^[17-18],但在特定市场环境或交易策略下,也可能加剧订单流短期波动和价格冲击^[19].在异常交易识别方面,逐笔成交和逐笔委托数据能够刻画订单提交、成交、撤单和价格变化的微观过程,为识别短时段可疑交易结构提供了数据基础.相关研究已利用股票日内交易网络识别操纵行为^[20].这些研究表明,虚假交易风险不仅可能表现为价格异常或成交量异常,也可能表现为买卖交易在时间、价格和数量维度上的高度可匹配性.现有闭环识别方法通常依赖账户身份数据,而公开 A 股高频数据往往只能观察订单、成交、价格和数量等市场层面信息,难以直接识别交易主体之间是否存在关联.

据此,本文关注的核心问题是:在账户身份不可见的 A 股逐笔高频数据条件下,能否从市场交易结构中识别出一类具有虚假交易风险特征的可疑订单流窗口?进一步地,这类窗口是否具有高成交、高匹配但价格发现弱化的独立特征?为回答这一问题,本文借鉴洗售交易识别中关于买卖交易可匹配性的思想^[21],将账户级闭环交易识别逻辑拓展至账户身份不可见的公开高频数据场景,构造背包式订单流闭合度指标(Knap-

① 数据来源于 Wind.

sack-based order-flow closure, KnapOFC), 用以衡量特定时间窗口内主动买入与主动卖出订单在时间、价格和数量维度上的可匹配程度。考虑到账户身份和实际控制关系在公开高频数据中不可直接观察, 本文将同时具有高成交、高匹配和低价发现特征的交易窗口界定为具有虚假交易风险特征的可疑订单流窗口, 并以此作为虚假交易风险的市场层面间接度量, 为异常交易监测提供间接证据。

在实证设计上, 本文采用两阶段识别框架。第 1 阶段基于分钟级高频交易数据, 在全市场范围内筛选成交量显著放大、买卖方向较为均衡、成交价格较为集中且价格影响较弱的候选窗口; 第 2 阶段在逐笔成交层面, 计算 KnapOFC 及其匹配质量指标, 检验候选窗口中的买卖订单是否能够在时间、价格和数量维度上形成较强匹配。基于 2025 年 9 月至 2026 年 4 月 A 股逐笔高频交易数据, 本文发现, A 股市场中存在一类的可疑订单流窗口。与随机安慰剂窗口、普通高成交窗口及成交爆发程度相近的窗口相比, 该类窗口虽然成交量显著放大, 但当期价格影响、未来价格冲击和事件后短期已实现波动率均未同步上升; 进一步从逐笔订单结构看, 该类窗口具有更高的 KnapOFC、更广的匹配覆盖度和更低的残余不平衡, 并区别于高撤单率^[22]、高订单不平衡^[23]和尾市异动等传统异常交易形态^[11, 24-25]。

本文的边际贡献主要体现在 3 个方面。第 1, 在识别方法上, 将账户级洗售交易的买卖可匹配思想拓展到账户身份不可见的公开高频数据场景, 构造 KnapOFC 指标, 为虚假交易风险的市场层面间接识别提供新的工具。第 2, 在研究视角上, 本文进一步考察异常成交量是否转化为价格发现, 从成交量信息含量角度拓展了异常交易研究。第 3, 在实践应用上, 本文识别出一类高成交、高匹配但价格发现弱化的可疑订单流结构, 可为异常交易监测提供参考。

1 指标构造与研究假设

从交易行为看, 虚假交易风险可能表现为: 短时间内成交量异常放大, 但买卖方向高度匹配, 成交价格较为集中, 且价格影响较弱。若某一交易窗口内成交量显著放大, 但并未伴随相应的方向性订单流、价格冲击和后续价格调整, 则该窗口中的成交量更可能反映低信息含量的交易活动, 而非真实信息进入市场所引发的价格发现过程。

基于这一逻辑, 本文关注在短时间窗口内同时表现出成交量异常放大、买卖订单高度可匹配和价格发现功能不足的订单流。具体而言, 第 1, 在分钟级高频数据中识别成交量异常放大的候选窗口; 第 2, 在逐笔订单层面考察买卖订单是否具有较强的时间、价格和数量匹配关系; 第 3, 进一步检验该类高成交窗口是否缺乏相应的价格影响和后续价格波动。只有同时满足高成交、高匹配和低价发现特征的窗口, 才被识别为具有虚假交易风险特征的可疑订单流窗口。

1.1 背包式订单流闭合度指标构造

为刻画订单级买卖匹配结构, 本文构造背包式订单流闭合度指标 (KnapOFC)。该指标用于衡量特定时间窗口内, 主动买入和主动卖出订单在时间、价格和数量维度上的可匹配程度。

本文借鉴 Cao 等^[21]关于洗售交易检测的图结构与动态规划思想, 虚假交易在订单层面通常具有 3 个特征: 买卖成交发生时间较近、成交价格差异较小、成交数量可以在买卖两侧形成较高比例的相互抵消。考虑到 A 股逐笔数据不包含账户身份, 本文保留其买卖交易可匹配的核心思想, 将其迁移到账户身份不可见的订单级高频数据场景, 具体见附录图 S1。

在窗口层面, 本文将可匹配成交量占买卖总成交量的比例定义为 KnapOFC:

$$\text{KnapOFC}_{i,d,t} = \frac{2 \times V(M)_{i,d,t}}{V(B)_{i,d,t} + V(S)_{i,d,t}}, \quad (1)$$

其中, $V(M)_{i,d,t}$ 表示窗口内买卖方向能够在时间、价格和数量约束下形成匹配的成交量; $V(B)_{i,d,t}$ 和 $V(S)_{i,d,t}$ 分别表示该窗口内主动买入和主动卖出成交量。KnapOFC 取值越高, 说明窗口内买卖订单在逐笔层面的闭合程度越强。与简单的买卖数量对称度相比, KnapOFC 不仅要求买卖总量接近, 还进一步要求成交在时间、价格和数量维度上能够形成有效匹配, 因此能够更严格地刻画订单级买卖闭合结构。具体参数设置、背包匹配流程和数值示例见附录 A。

此外,本文还计算匹配覆盖度、平均匹配时间间隔、平均匹配价格距离、残余不平衡和一对多匹配比例等分解指标,用于进一步刻画可疑订单流窗口的微观结构。

1.2 NOC 事件的两阶段识别框架

在构造 KnapOFC 的基础上,本文进一步识别非信息性订单流闭合事件.为便于表述,本文将成交量异常放大、买卖订单高度可匹配且价格发现功能不足的交易窗口称为 NOC 事件窗口。

本文采用两阶段识别框架.第一阶段为分钟级候选窗口筛选.考虑到全市场逐笔订单数据规模较大,直接对所有股票、所有交易日和所有分钟窗口进行订单级背包匹配计算成本较高.因此,本文首先基于分钟级高频面板构造低成本筛选指标,在全市场范围内识别成交量显著放大、买卖方向较为均衡、成交价格较为集中且单位成交量价格影响较弱的候选窗口.该阶段的目标是从海量交易窗口中筛选出更可能具有可疑订单流特征的样本。

第二阶段为订单级精确匹配验证.对于第一阶段识别出的候选窗口,本文回到逐笔成交数据层面,计算 KnapOFC 及其匹配质量分解指标,检验这些窗口中的买卖订单是否确实具有较强的时间、价格和数量匹配关系.通过这一设计,本文既能保证全市场识别的覆盖范围,又能在候选窗口层面保留订单级识别的微观严谨性。

在具体测度上,本文将成交爆发、订单级可匹配度、价格集中度和低价格影响 4 类信息综合为 NOC 事件识别指标:

$$\text{NOC Score}_{\text{Knap},i,d,t} = \frac{B_z + \text{KnapOFC}_z + C_z + L_z}{K}, \quad (2)$$

其中, B_z 衡量成交量相对于历史同类窗口的异常放大程度; KnapOFC_z 衡量订单级买卖可匹配度的异常程度; C_z 衡量成交价格是否集中; L_z 衡量单位成交量对应的价格影响是否显著偏低.各分量均基于历史滚动窗口进行标准化处理,避免使用未来信息。

需要说明的是,本文在事件识别中纳入 L_z ,目的是将成交量异常放大但单位成交价格反应偏弱的窗口从普通高成交窗口中区分出来.因此,后文关于当期价格影响的结果更多体现为事件定义的内部一致性.为避免循环论证,本文将未来 30 min 价格影响以及事件后 5、15 和 30 min 已实现波动率作为更关键的外部验证变量.上述变量均发生在事件窗口之后,未进入 NOC 事件触发条件,因而能够检验异常成交量是否在事后继续转化为价格发现和信息吸收过程。

1.3 研究假设

基于前述指标构造和识别框架,本文提出以下研究假设。

如果低信息含量成交安排或具有虚假交易风险特征的订单流结构会在市场数据中留下可观测痕迹,那么在 A 股逐笔高频数据中应当能够观察到一类成交量异常放大、买卖订单高度匹配但价格发现功能不足的交易窗口.这类窗口未必频繁发生,但应在全市场样本中稳定存在,并可能在不同板块之间表现出一定差异.据此,本文提出假设 H1。

H1: A 股市场中存在成交量异常放大、买卖订单高度匹配但价格发现功能不足的可疑订单流窗口。

市场微观结构理论认为,若成交量异常放大主要来自信息性交易,则其应伴随方向性订单流、显著价格影响和后续价格调整.相反,若成交量异常放大主要来自买卖方向之间的高度匹配或低信息含量成交,则其价格影响应弱于普通高成交窗口.因此,相较于普通高成交窗口或成交爆发程度相近的窗口,NOE 事件窗口虽然成交量较高,但当期价格影响、未来价格影响和后续价格波动不应同步上升.据此提出假设 H2。

H2: 相较于普通高成交窗口,NOE 事件窗口的当期价格影响和后续价格调整更弱。

如果 NOE 事件捕捉的并非普通成交放大,而是具有虚假交易风险特征的订单流结构,则其应在逐笔订单层面表现出更强的买卖可匹配性,并区别于已有研究和监管实践中较为常见的异常交易形态.具体而言,NOE 窗口应具有更高的 KnapOFC 和匹配覆盖度、更低的未匹配残余不平衡,以及更小的时间和价格匹配距离.与此同时,高撤单率异常更多对应虚假申报或幌骗交易,订单不平衡通常反映方向性信息冲击或单边交易压力,尾市异动则更接近价格偏离型操纵.若 NOE 事件具有独立的微观结构特征,则其不应只是上述异常交易形态的简单替代表达,而应呈现高成交、高匹配和低价格发现的组合特征.因此,提出假设 H3。

H3: NOE 事件窗口具有更强的订单级买卖可匹配性,并显著区别于高撤单率、订单不平衡和尾市异动

等传统异常交易形态。

2 实证检验

2.1 数据来源与样本构造

本文使用 A 股全市场逐笔高频交易数据, 样本覆盖沪深主板、创业板和科创板, 样本期间从 2025 年 9 月至 2026 年 4 月, 数据包括逐笔成交、逐笔委托和成交回报记录, 能够观察每笔成交的时间、价格、数量、买卖方向以及订单生命周期。相较于日度数据或分钟级行情数据, 逐笔高频数据的优势在于能够刻画订单提交、成交和撤单的微观过程, 从而为识别短时间窗口内的买卖订单匹配结构提供基础。

本文的数据处理包含 3 个步骤, 第 1, 将原始逐笔成交数据按分钟聚合, 构造分钟级订单流面板, 并计算成交量、成交额、成交笔数、买卖方向成交量、成交价格集中度和单位成交量价格影响等变量。第 2, 对每只股票在每个分钟位置上, 利用过去 W 个交易日历史窗口构造事前标准化指标, 历史滚动指标均严格不包含当期, 以避免未来信息进入事件识别。第 3, 先基于分钟级低成本指标筛选候选窗口, 再对候选窗口回读逐笔成交数据, 计算背包式订单流闭合度指标 KnapOFC 及其匹配质量分解指标。

此外, 本文剔除集合竞价和非连续竞价时段, 仅保留连续竞价交易窗口; 剔除涨跌停锁单窗口、新股上市初期、ST 股票、停牌复牌首日以及交易停牌日观测。为避免极低成交窗口产生少数几笔成交即可完全匹配的伪闭合现象, 主样本进一步要求分钟成交笔数不低于 20 笔, 并在稳健性检验中加入成交金额门槛。

表 1 报告了 NOC 事件触发率, 以 $W=20$ 为基准, 本文共识别出 1 123 个 NOC 事件窗口, 分布于 1 086 个股票—交易日中; 按股票—交易日口径计算, 触发率为 0.166%。在 $W=60$ 的稳健窗口下, 识别出 498 个 NOC 事件的股票—交易日, 触发率为 0.076%。这表明, 本文识别的可疑订单流结构并非普遍存在的常规交易现象, 而是一类低频但稳定可观察的异常订单流窗口。

分板块看, NOC 事件触发率呈现明显差异, 以 $W=20$ 为例, 主板、创业板和科创板的事件触发率分别为 0.143%、0.195% 和 0.225%; 以 $W=60$ 为例, 三类板块触发率分别为 0.068%、0.079% 和 0.116%。总体来看, 科创板和创业板的 NOC 事件触发率高于主板, 说明该类可疑订单流结构在中小市值股票占比较高、交易活跃度更强的板块中更为明显。后文将进一步控制市值、换手率和交易活跃度等因素, 对板块差异进行稳健性检验。

总的来说, 表 1 的结果表明, A 股市场中确实存在一类稳定但低频的可疑订单流结构, 支持假设 H1。

表 1 NOC 事件触发率
Tab. 1 NOC event activation rate

Panel A 总样本 NOC 事件触发率					
W	股票数	股票—交易日观测数	含 NOC 事件的股票—交易日	NOC 事件时间/min	事件触发率/%
20	5 203	653 652	1 086	1 123	0.166
60	5 203	653 652	498	516	0.076
Panel B 分板块 NOC 事件触发率					
W	板块	股票数	股票—交易日观测数	含 NOC 事件的股票—交易日	事件触发率/%
20	主板	3 199	402 332	574	0.143
20	创业板	1 398	176 118	343	0.195
20	科创板	606	75 202	169	0.225
60	主板	3 199	402 332	272	0.068
60	创业板	1 398	176 118	139	0.079
60	科创板	606	75 202	87	0.116

2.2 描述性特征

为检验 NOC 窗口是否区别于一般高成交窗口, 本文构造多类对照组进行比较。随机安慰剂窗口为在相同股票和交易日内随机抽取的普通分钟窗口; 高成交非 NOC 窗口为成交活跃度较高但未被识别为 NOC 的窗口; 成交活跃度匹配窗口为在成交金额或成交笔数上与 NOC 窗口相近、但未满足 NOC 识别条件的窗口; 成交活跃度极端窗口为全样本中成交活跃度最高的分钟窗口。核心变量定义和更完整的变量说明放入附录,

以节约正文篇幅。

表 2 报告不同窗口的分钟级交易特征均值,结果显示,NOC 窗口不仅成交活跃度异常(Burst)显著高于随机安慰剂窗口,而且同时表现出较高的买卖数量对称度(Balance)、成交价格集中度(Concentration)和低价格冲击程度(LowImpact).相比之下,高成交非 NOC 窗口和成交活跃度匹配窗口虽然也具有明显的 Burst 特征,但其 Balance 较低,Concentration 并不明显,LowImpact 也不显著.成交活跃度极端窗口虽然 Burst 远高于其他组,但 Balance 更低,Concentration 和 LowImpact 也未明显上升.这说明 NOC 窗口不同于普通高成交窗口,而是同时表现出成交活跃异常、买卖数量对称、成交价格集中和价格反应弱化的组合特征。

表 2 不同窗口的分钟级交易特征均值(W=20)

Tab. 2 Minute-level Characteristics of Different Windows(W=20)

窗口	成交额(10 ⁶ 元)	成交笔数	Burst	Balance	Concentration	LowImpact
NOC 窗口	16.6	511.0	5.73	0.898	2.97	3.05
随机安慰剂窗口	2.29	144.6	0.88	0.550	−0.07	−0.01
高成交非 NOC 窗口	6.77	347.7	5.63	0.488	−0.57	−0.26
成交活跃度匹配窗口	17.7	577.5	5.45	0.536	−0.08	−0.05
成交活跃度极端窗口	126	3 968.5	506.93	0.354	−0.52	−0.26

注:成交活跃度异常衡量成交量、成交金额和成交笔数相对于历史同分钟位置的异常程度;买卖数量对称度取值越高,表示主动买入与主动卖出成交量越接近;成交价格集中度和低价格冲击程度为标准化指标,数值越高分别表示成交价格越集中、单位成交量引起的价格变动越小。

2.3 NOC 事件与价格发现

假设 H2 认为,若 NOC 事件主要反映低信息含量成交或虚假交易风险信号,则其成交量异常放大不应伴随相应的价格发现.本文以多类对照窗口为参照,检验 NOC 事件窗口的未来价格影响和事件后短期波动差异.基准回归模型为:

$$Y_{i,d,t} = \alpha + \beta \text{NOC}_{i,d,t} + \gamma X_{i,d,t} + \mu_i + \delta_d + \theta_t + \varepsilon_{i,d,t}, \tag{3}$$

其中, $Y_{i,d,t}$ 表示未来 30 min 绝对收益率(R)、当期价格冲击(PI)、窗口内价格振幅(Range)和未来 30 min 价格冲击(FuturePI30)等价格发现指标; $\text{NOC}_{i,d,t}$ 为 NOC 事件窗口指示变量,在 NOC 窗口中取 1,在相应对照窗口中取 0; $X_{i,d,t}$ 为成交额、成交笔数、撤单率和订单不平衡等控制变量; μ_i 、 δ_d 和 θ_t 分别表示股票固定效应、交易日固定效应和分钟固定效应,标准误在股票层面聚类。

表 3 报告了 NOC 事件窗口相对于不同对照窗口的当期及未来价格影响.Panel A-C 显示,NOC 窗口的分钟绝对收益率、单位成交量价格冲击和窗口内价格振幅均显著低于对照窗口,说明本文识别的 NOC 窗口具有高成交但弱价格反应的特征.需要说明的是,由于事件识别过程中已纳入低价格影响条件,Panel A-C 的结果主要体现事件定义的内部一致性.为避免循环论证,本文进一步考察事件窗口之后的 FuturePI30.Panel D 显示,在安慰剂窗口、普通高成交非 NOC 窗口和成交爆发匹配窗口三类对照下,NOC 事件后的未来 30 min 价格影响均显著为负,且在 W=20 和 W=60 两类历史窗口下保持一致.这表明,NOC 窗口中的异常成交量并未在事后转化为更强的价格调整或信息吸收过程.总体而言,表 3 的结果支持成交量异常放大但价格发现功能弱化的核心判断。

2.4 NOC 事件与已实现波动率

为进一步检验成交量信息含量,本文增加后续已实现波动率检验.若异常放大的成交量主要反映信息进入市场,则事件发生后价格应继续吸收信息,表现为较高的短期已实现波动率;相反,若 NOC 事件主要反映低信息含量订单流,则其后续已实现波动率不应显著上升.本文分别计算事件后 5 min、15 min 和 30 min 已实现波动率:

$$RV_{i,d,t}^{(h)} = \sqrt{\sum_{i,d,t+\tau}^h} \quad h \in \{5, 15, 30\}, \tag{4}$$

并以 $RV^{(5)}$ 、 $RV^{(15)}$ 和 $RV^{(30)}$ 作为被解释变量进行回归.回归设定与式(3)保持一致,仍控制成交额、成交笔数、撤单率和订单不平衡,并加入股票、交易日和分钟固定效应。

附录表 S1 结果显示,在不同对照组和两类历史窗口设定下,NOC 事件在事件后 5、15 和 30 min 已实现波动率上的系数均显著为负,无论相对于安慰剂窗口、普通高成交非 NOC 窗口,还是成交爆发程度相近窗口,NOC 事件后的短期波动均更低,说明该结果并非由一般高成交或成交爆发强度差异驱动.结合表 S1 中未来价格影响较弱的结果,NOC 窗口虽然伴随成交量异常放大,但事件后并未表现出持续的信息吸收和价格发现过程.需要说明的是,较低的后续波动并不意味着 NOC 事件改善了市场质量,而是表明该类成交量异常未体现出与其交易规模相匹配的信息冲击.

表 3 NOC 事件与当期及未来价格影响
Tab. 3 NOC Events and Contemporaneous and Future Price Impacts

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:被解释变量为 R						
NOC 事件	−0.001 9*** (−28.07)	−0.002 2*** (−16.15)	−0.002 3*** (−22.74)	−0.003 2*** (−13.91)	−0.002 4*** (−10.21)	−0.003 7*** (−5.91)
观测值	12 352	5 676	12 353	5 675	2 246	1 032
R ²	0.152	0.160	0.208	0.199	0.285	0.270
Panel B:被解释变量为 PI						
NOC 事件	−0.000 1*** (−28.81)	−0.000 1*** (−16.09)	−0.000 1*** (−24.13)	−0.000 2*** (−14.35)	−0.000 1*** (−11.04)	−0.000 2*** (−5.98)
观测值	12 352	5 676	12 353	5 675	2 246	1 032
R ²	0.131	0.142	0.194	0.186	0.315	0.280
Panel C:被解释变量为 Range						
NOC 事件	−0.002 2*** (−23.25)	−0.002 5*** (−16.85)	−0.003 0*** (−24.17)	−0.004 1*** (−14.80)	−0.002 8*** (−11.80)	−0.004 7*** (−8.09)
观测值	12 353	5 676	12 353	5 676	2 246	1 032
R ²	0.230	0.277	0.309	0.300	0.312	0.366
Panel D:被解释变量为 FuturePI30						
NOC 事件	−0.000 2*** (−9.41)	−0.000 2*** (−5.97)	−0.000 2*** (−10.50)	−0.000 3*** (−6.89)	−0.000 2*** (−4.75)	−0.000 3*** (−4.85)
观测值	10 604	4 880	10 644	4 871	1 908	896
R ²	0.034	0.040	0.054	0.056	0.053	0.149
对照组	安慰剂窗口	安慰剂窗口	普通高成交 非 NOC 窗口	普通高成交 非 NOC 窗口	成交爆发 匹配窗口	成交爆发 匹配窗口
历史窗口	20	60	20	60	20	60
控制变量	Y	Y	Y	Y	Y	Y
μ_i	Y	Y	Y	Y	Y	Y
δ_d	Y	Y	Y	Y	Y	Y
θ_t	Y	Y	Y	Y	Y	Y

注:括号内为 t 统计量.控制变量包括 Log(成交额)、成交笔数、撤单率和订单不平衡.***、**、* 分别表示在 1%、5%、10% 水平上显著,下同.

2.5 NOC 事件的订单级买卖匹配结构

根据 H3,如果 NOC 事件并非普通成交放大,而是具有虚假交易风险特征的可疑订单流结构,则其不仅应表现为成交量异常放大和价格发现弱化,还应在逐笔成交层面呈现更强的买卖可匹配性,并区别于高撤单率、订单不平衡和尾市异动等传统异常交易形态.

为检验这一判断,本文比较 NOC 窗口与多类对照窗口的订单级匹配质量分解指标.表 4 显示,NOC 窗口的 KnapOFC 为 0.554 0,高于各类对照窗口;其匹配覆盖度为(Coverage)0.614 0,也高于全部对照组,说明 NOC 窗口的买卖匹配存在于逐笔成交中.与此同时,NOC 窗口的残余不平衡(Residual Imbalance)仅为

0.102 0,明显低于各类对照窗口,表明经过背包匹配后,买卖剩余量的方向性较弱,具有更明显的成交闭合特征.价格匹配方面,NOC 窗口的平均价格距离为(Price Gap)0.387 0 tick,低于多数对照窗口,说明其匹配成交在价格上也较为接近.

需要注意的是,极端成交爆发窗口的 Price Gap 为 0.343 0 tick,略低于 NOC 窗口,但其 KnapOFC 和 Coverage 明显更低,Residual Imbalance 明显更高.这说明单一维度的近价格成交匹配并不足以构成 NOC 结构;NOC 事件的核心特征在于较高的订单级匹配度、较高的匹配覆盖度和较低的残余不平衡同时出现.总体而言,表 4 结果支持 H3,即 NOC 事件窗口在逐笔成交层面具有更强的买卖可匹配性,并不同于高撤单率、订单不平衡和尾市异动等传统异常交易形态.

表 4 NOC 窗口与对照窗口的匹配质量分解
Tab. 4 Matching quality decomposition for NOC and control windows

窗口	KnapOFC	Coverage	Residual Imbalance	Time Gap/s	Price Gap/tick
NOC 窗口	0.554 0	0.614 0	0.102 0	11.38	0.387 0
安慰剂窗口	0.385 0	0.544 0	0.450 0	10.46	0.731 0
普通高成交非 NOC 窗口	0.328 0	0.495 0	0.512 0	8.84	0.661 0
成交爆发匹配窗口	0.356 0	0.514 0	0.465 0	7.94	0.753 0
极端成交爆发窗口	0.186 0	0.312 0	0.662 0	7.72	0.343 0
高撤单率异常窗口	0.193 0	0.268 0	0.728 0	10.46	0.793 0
尾市异动窗口	0.334 0	0.472 0	0.429 0	4.36	0.587 0

注:KnapOFC 衡量买卖订单在时间、价格和数量约束下的可匹配程度;Coverage 为匹配覆盖度,衡量参与匹配的成交单元占窗口总成交单元的比例;Residual Imbalance 为残余不平衡,衡量背包匹配后未匹配剩余量的方向性偏离程度;Time Gap 为平均匹配时间间隔;Price Gap 为平均匹配价格距离,以最小报价单位 tick 计量.

2.6 稳健性检验

接下来本文从阈值选择、分时段样本和板块控制 3 个方面进行稳健性检验,详细结果见附录表 S2-S4.

NOC 事件识别依赖成交爆发阈值和买卖平衡阈值.前者刻画当前窗口成交量相对于个股自身历史基准的异常放大程度,采用相对倍数作为判据,以控制不同股票和板块之间的流动性差异;后者刻画短窗口内主动买入与主动卖出成交量的接近程度,用以排除明显由单边交易压力驱动的高成交窗口.基准设定中,成交爆发阈值取 2.5,买卖平衡阈值取 0.80;稳健性检验进一步将成交爆发阈值调整为 2.0 和 3.0,将买卖平衡阈值调整为 0.75 和 0.85,分别对应相对宽松和相对严格的筛选标准.结果显示,随着阈值收紧,识别出的事件数量逐步下降,说明阈值调整能够有效改变事件筛选强度.更重要的是,在全部阈值组合下,NOC 事件在 FuturePI30、RV⁽⁵⁾、RV⁽¹⁵⁾ 和 RV⁽³⁰⁾ 上的系数均为负,且均在 1%水平上显著,表明本文关于价格发现弱化和后续信息吸收不足的结论并不依赖 2.5 和 0.80 这一基准参数组合.

第二,考虑到样本期间为 2025 年 9 月至 2026 年 4 月,本文将样本划分为 2025 年 9 月至 12 月和 2026 年 1 月至 4 月两个子期间进行检验.附录表 S3 结果显示,两个子期间内 NOC 事件的未来价格影响和事件后已实现波动率均显著更低,表明结论并非由某一特定市场阶段驱动.

第三,针对科创板和创业板触发率较高的问题,本文进一步控制市值、换手率、日内振幅、成交额、量比以及行业和交易日固定效应.附录表 S4 结果显示,在(W=20)的基准历史窗口下,科创板和创业板系数仍显著为正;在(W=60)的扩展历史窗口下,科创板边际显著,创业板不再显著.这表明板块差异在扩大历史参照期并控制股票特征和交易活跃度后有所减弱.

总体而言,稳健性检验支持本文主要结论:NOC 事件中的异常成交量未表现出与其成交规模相匹配的价格发现和后续信息吸收.

3 结论

本文针对账户身份不可见条件下虚假交易风险的间接检测问题,基于 A 股逐笔高频数据构造背包式订单流闭合度指标 KnapOFC,并据此识别在短窗口内同时呈现成交量异常放大、买卖订单高度匹配与价格发

现弱化特征的可疑订单流窗口(NOC事件).本文关注此类风险在公开高频数据中可能留下的市场层面可观测痕迹,即异常成交量是否具有相应的信息含量,以及买卖订单是否在逐笔层面呈现高度可匹配结构.

主要结论如下:第1,A股市场存在一类稳定但触发率较低的、具有虚假交易风险特征的可疑订单流窗口.在20日基准历史窗口下,本文识别出1123个NOC窗口,分布于1086个股票—交易日,触发率为0.166%;在60日扩展历史窗口下,触发率为0.076%.科创板和创业板触发率高于主板,表明此类窗口呈现一定的板块差异.第2,NOC窗口并非普通高成交窗口.相对于安慰剂窗口、高成交非NOC窗口与成交爆发匹配窗口,NOC窗口的未来30min价格影响显著更低,事件后5、15和30min已实现波动率也显著更低,说明其成交量异常放大并未转化为持续的信息吸收和价格调整.需要说明的是,较低的后续波动并不意味着市场质量改善,而是表明该类成交量异常未体现出与其成交规模相匹配的信息冲击.第3,NOC窗口在订单级具有更强的买卖可匹配性,其KnapOFC、匹配覆盖度和残余不平衡均明显区别于高撤单、订单不平衡和尾市异动等传统异常形态.总体而言,NOC事件表现为一种高成交、高匹配、低价格发现的订单流结构,而非普通高成交或已有异常交易形态的简单替代.

在监管应用层面,本文的结果显示:第1,监管系统可在现有指标基础上增设“成交异常放大、KnapOFC较高、价格发现偏弱”的组合预警规则,对同时满足多项条件的股票—窗口给予重点关注.第2,对被预警窗口,可进一步调取参与成交的账户关系、交易席位、下单时序和成交对手信息,核查是否存在关联账户对倒、自成交或一致行动人在买卖两侧进行配合交易的情形.第3,可设置观察、预警和核查3个层级,以平衡漏报与误报;同时应明确,KnapOFC等市场侧指标仅用于提示具有虚假交易风险特征的可疑订单流,不直接构成违规认定.最终认定仍需结合账户级证据、交易目的、资金关系和监管规则进行综合判断.

后续研究可结合账户级监管数据、程序化交易标识、投资者结构和个股基本面信息,进一步检验不同类型交易主体在NOC事件形成中的作用机制.

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2026.05.30.0001).

参 考 文 献

- [1] 吴崇林,张小芹,刘杰,等.洗售交易、机构投资者异质性与股票流动性[J].中央财经大学学报,2022(2):41-53.
Wu C L, Zhang X Q, Liu J, et al. Wash trading, institutional investor heterogeneity and stock liquidity[J]. Journal of Central University of Finance & Economics, 2022(2): 41-53.
- [2] 刘振清,王旭,姚远.基于新型指标的股票价格操纵识别模型研究[J].江淮论坛,2020(4):119-125.
Liu Z Q, Wang X, Yao Y. Research on stock price manipulation identification based on new index[J]. Jiang-huai Tribune, 2020(4): 119-125.
- [3] Akter M, Cumming D, Ji S. Natural disasters and market manipulation[J]. Journal of Banking & Finance, 2023, 153(C): 106883.
- [4] Cong L W, Li X, Tang K, et al. Crypto wash trading[J]. Management Science, 2023, 69(11): 6427-6454.
- [5] 吴崇林,刘杰,李志冰,等.开盘价操纵与定价效率:基于高频数据的开盘价操纵识别模型[J].经济学报,2021,8(4):27-56.
Wu C L, Liu J, Li Z B, et al. Opening price manipulation and pricing efficiency: a detection model of opening price manipulation based on high frequency data[J]. China Journal of Economics, 2021, 8(4): 27-56.
- [6] 李志辉,王近,李梦雨.中国股票市场操纵对市场流动性的影响研究:基于收盘价操纵行为的识别与监测[J].金融研究,2018(2):135-152.
Li Z H, Wang J, Li M Y. A study on China's stock market manipulation's effects on market liquidity: based on closing price manipulation behavior's identification and monitoring[J]. Journal of Financial Research, 2018(2): 135-152.
- [7] Allen F, Gale D. Stock-price manipulation[J]. Review of Financial Studies, 1992, 5(3): 503-529.
- [8] Aggarwal R K, Wu G J. Stock market manipulations[J]. The Journal of Business, 2006, 79(4): 1915-1953.
- [9] Cumming D, Ji S, Peter R, et al. Market manipulation and innovation[J]. Journal of Banking & Finance, 2020, 120: 105957.
- [10] Comerton-Forde C, Putnins T J. Measuring closing price manipulation[J]. Journal of Financial Intermediation, 2011, 20(2): 135-158.
- [11] 李志辉,常心宇,魏斌.责任投资与股票市场操纵:基于ESG基金持股的研究[J].证券市场导报,2025(4):54-65.
Li Z H, Chang X Y, Wei B. Responsible investment and stock market manipulation: a study based on ESG fund holdings[J]. Securities Market Herald, 2025(4): 54-65.
- [12] Kyle A S. Continuous auctions and insider trading[J]. Econometrica, 1985, 53(6): 1315.
- [13] Glosten L R, Milgrom P R. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders[J]. Journal of Financial Economics, 1985, 14(1): 71-100.

[14] Hasbrouck J.Measuring the information content of stock trades[J].The Journal of Finance,1991,46(1):179-207.

[15] Easley D,López de Prado M M,O'Hara M.Flow toxicity and liquidity in a high-frequency world[J].Review of Financial Studies,2012,25(5):1457-1493.

[16] Easley D,López de Prado M M,O'Hara M.The microstructure of the "flash crash":Flow toxicity,liquidity crashes,and the probability of informed trading[J].The Journal of Portfolio Management,2011,37(2):118-128.

[17] Hendershott T,Jones C M,Menkveld A J.Does algorithmic trading improve liquidity [J].The Journal of Finance,2011,66(1):1-33.

[18] Brogaard J,Hendershott T,Riordan R.High-frequency trading and price discovery[J].The Review of Financial Studies,2014,27(8):2267-2306.

[19] Kirilenko A,Kyle A S,Samadi M,et al.The flash crash;high-frequency trading in an electronic market[J].The Journal of Finance,2017,72(3):967-998.

[20] 扈文秀,马丽,张建锋.基于股票日内交易加权网络的超短期股票交易型操纵识别研究[J].数据分析与知识发现,2019,3(10):118-126.

Hu W X,Ma L,Zhang J F.Identifying ultra-short-term market manipulation with stock intraday trading weighted network[J].Data Analysis and Knowledge Discovery,2019,3(10):118-126.

[21] Cao Y,Li Y H,Coleman S,et al.Detecting wash trade in financial market using digraphs and dynamic programming[J].IEEE Transactions on Neural Networks and Learning Systems,2016,27(11):2351-2363.

[22] Aitken M, Cumming D, Zhan F. High frequency trading and end-of-day price dislocation[J]. Journal of Banking & Finance, 2015, 59: 331-349.

[23] Comerton-Forde C,Putniņš T J.Stock price manipulation;prevalence and determinants[J].Review of Finance,2014,18(1):23-66.

[24] Titman S,Wei C S,Zhao B.Corporate actions and the manipulation of retail investors in China;an analysis of stock splits[J].Journal of Financial Economics,2022,145(3):762-787.

[25] Chen T,Gao Z Y,He J B,et al.Daily price limits and destructive market behavior[J].Journal of Econometrics,2019,208(1):249-264.

Wash-trading risk, volume informativeness, and price discovery:
evidence from high-frequency data in China's A-Share market

Yao Yuan

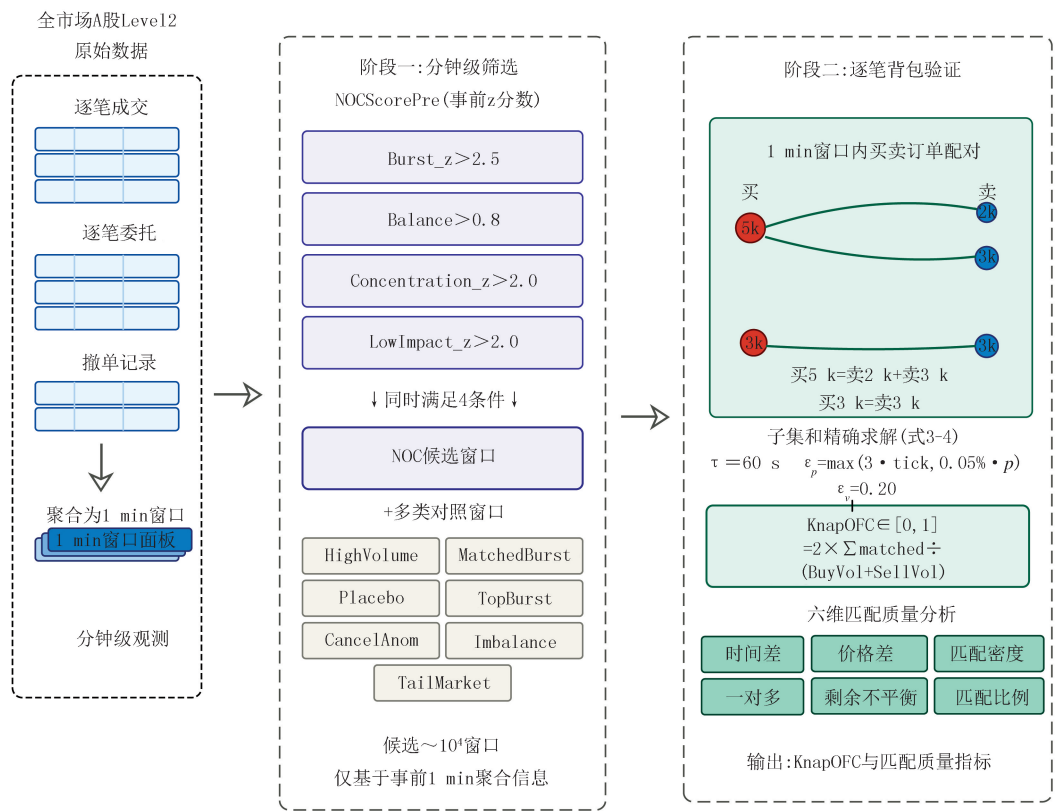
(Business School, Henan University, Kaifeng 475004, China)

Abstract: Using tick-by-tick high-frequency data from China's A-share market, this paper constructs a knapsack-based order-flow closure measure, KnapOFC, and develops a two-stage framework to identify suspicious order-flow windows that simultaneously exhibit abnormal trading activity, high buy-sell order matching, and weakened price discovery, referring to as NOC events. These events are used to capture indirect market-level traces of false-trading risk in publicly observable high-frequency data. The main findings are as follows. First, NOC windows exist in the A-share market with a low but stable occurrence rate, and are relatively more active on the STAR Market and ChiNext. Second, compared with placebo windows, high-volume non-NOC windows, and volume-burst-matched windows, NOC windows exhibit significantly lower future 30 min price impact and lower short-term realized volatility after the event, suggesting that abnormal trading volume does not carry information content commensurate with its trading scale. Third, at the order level, NOC windows display higher KnapOFC, broader matching coverage, and lower residual imbalance, and are clearly distinguishable from traditional abnormal trading patterns such as high cancellation, order imbalance, and end-of-day price movements. From the perspective of trading-volume information content and price discovery, this paper provides market-side evidence for the indirect detection of false-trading risk using high-frequency data, and offers an operational indicator for incorporating order-level matching structures into abnormal trading surveillance.

Keywords: wash-trading risk; volume informativeness; price discovery; knapsack-based order-flow closure; trade-by-trade high-frequency data

[责任编辑 陈留院 杨浦]

附 录



图S1 KnapOFC指标构建逻辑图
Fig.S1 Logic diagram of KnapOFC indicator construction

附录 A

本附录给出 KnapOFC(背包式订单流闭合度,Knapsack order-flow closure)指标的完整计算流程.

A.1 参数与数据设置

本附录的计算基于含主动买卖方向标识的逐笔成交数据.每个成交单元 o 记录成交时间 $T(o)$ 、成交价格 $P(o)$ 与成交数量 $V(o)$, 并标记为主动买入或主动卖出.匹配仅在同一股票、同一交易日、同一窗口内进行,不跨窗口.基准参数设置如表 A1 所示.

表 A1 KnapOFC 计算的基准参数设置

Tab. A1 Baseline parameter settings for KnapOFC computation

符号	含义	基准取值	说明
t	日内分钟窗口	1 min	仅在同一窗口内匹配,不跨窗口;窗口基于连续交易记录构造.
lot	最小交易单位	100 股	若样本保留零股成交,可按实际股数作为计量单位进行整数化处理.
$tick$	最小价格变动单位	0.01 元	普通 A 股股票基准取 0.01 元;科创板、创业板等证券品种按交易所规定的最小价格变动单位取值.
τ	时间约束	60 s	买卖两笔成交的时间差上限.
ϵ_p	价格约束	$\max(3 \times tick, 5 \times 10^{-4} \times P)$	价差上限,随价位自适应;取目标成交价.表示相对成交价的万分之 5.
ϵ_v	数量误差容限	0.20	组合成交量与目标成交量的相对误差上限.

窗口基于交易所逐笔数据中的连续交易记录构造;集合竞价阶段(如开盘、收盘集合竞价)按独立窗口处

理,或在样本构造中予以剔除,不与连续交易记录混合匹配,以避免交易制度差异对匹配结果的干扰。

A.2 候选匹配集合构造

对于股票在交易日 d 的分钟窗口 t , 设主动买入成交集合与主动卖出成交集合分别为:

$$B_{i,d,t} = \{b_1, b_2, \dots, b_m\}, S_{i,d,t} = \{s_1, s_2, \dots, s_n\}.$$

对任一买入成交 b_k 首先从卖出集合中筛选出同时满足时间接近与价格接近条件的候选匹配集合:

$$C(b_k) = \{s_l \in S_{i,d,t} : |T(b_k) - T(s_l)| \leq \tau, |P(b_k) - P(s_l)| \leq \epsilon_p(b_k)\}.$$

其中价格约束阈值随价位自适应,基准设定为: $\tau = 60$ s, $\epsilon_p(b_k) = \max(3 \times \text{tick}, 5 \times 10^{-4} \times P(b_k))$.

对任一卖出成交 s_l , 按相同规则从买入集合中构造候选匹配集合 $C(s_l)$. 该步骤保证进入背包求解的候选成交必须首先满足时间接近与价格接近条件。

A.3 单笔成交的背包匹配问题

以买入成交 b_k 为例, 令其成交数量 $V(b_k)$ 作为背包容量, 相反方向候选卖出成交集合 $C(b_k)$ 中各笔成交数量 $V(s_l)$ 作为可选择物品. 背包求解的目标是在候选卖出成交中寻找一个子集 $A_k \subseteq C(b_k)$, 使其成交数量之和尽可能接近 $V(b_k)$. 形式上, 可写为: $A_k^* = \arg \min_{A_k \subseteq C(b_k)} |V(b_k) - \sum_{s_l \in A_k} V(s_l)|$.

若最优组合满足数量误差约束: $\frac{|V(b_k) - \sum_{s_l \in A_k^*} V(s_l)|}{V(b_k)} \leq \epsilon_v, \epsilon_v = 0.20$, 则认为买入成交 b_k 与卖出成交组合 A_k^* 构成一组有效匹配. 卖出成交的匹配过程与此对称, 即以其成交数量为背包容量, 在候选买入成交集合中寻找成交数量最接近的组合。

A.4 动态规划求解

定义布尔状态 $DP(j, u) = 1$, 表示在前 j 个候选成交中存在某一子集, 其成交量之和恰为 u (单位: 手). 初始条件为: $DP(0, 0) = 1, DP(0, u) = 0, (u > 0)$.

状态变量 u 的取值上界 U_o 定义为: $U_1 = \text{ceil}(1 + \epsilon_v)Q_o, U_2 = Q_1 + \dots + Q_q, U_o = \min(U_1, U_2)$. 其中 $\text{ceil}(\cdot)$ 表示向上取整. U_1 为目标成交量允许上浮 ϵ_v 后的整数上界, U_2 为候选成交总量, U_o 取二者中的较小值. 由此无需枚举超过目标可接受上限的成交量, 从而将状态空间控制在 $O(q \times U_o)$ 之内; 时间与价格约束 (τ, ϵ_p) 进一步限制了候选数 q , 使单笔求解的开销在逐笔数据上可控。

状态转移为:

$$DP(j, u) = \begin{cases} DP(j-1, u) \vee DP(j-1, u - Q_j), & u \geq Q_j, \\ DP(j-1, u), & 0 \leq u < Q_j. \end{cases}$$

完成状态转移后, 在误差范围内寻找距 Q_o 最近的可达成成交量:

$$u^* = \arg \min_{u: DP(q, u) = 1} |Q_o - u|, \quad \text{s.t.} \quad \frac{|Q_o - u|}{Q_o} \leq \epsilon_v.$$

若存在满足条件的 u^* , 则沿状态转移路径逆向回溯(对每个 j 判断 u^* 是否由包含 c_j 的状态到达), 即可恢复对应的候选成交子集, 作为该目标成交的最优匹配组合; 若不存在满足条件的 u^* , 则该目标成交不计入有效匹配。

若同一目标成交存在多个数量误差相同的候选组合, 本文按候选成交时间升序、成交单元序号升序确定回溯所得的组合, 以保证单笔匹配结果的唯一性。

A.5 匹配顺序与去重规则

由于买入成交与卖出成交均各自作为目标参与求解, A.3-A.4 会同时产生“以买入为目标”和“以卖出为目标”两类潜在匹配组. 本附录将两类匹配组合并进入同一候选池, 再统一排序与去重, 从而避免同一成交单元被重复计入。

具体而言, 对候选池中每一组潜在匹配 g , 记录其相对数量误差与可匹配成交量, 并按如下规则依次接受: (1) 相对数量误差由小到大; (2) 误差相同时, 可匹配成交量由大到小; (3) 仍相同时, 按目标成交单元的时间戳升序; 若再相同, 按目标成交单元的序号升序。

按上述顺序逐组考察: 若某候选匹配组中的目标成交单元或其组合中的任一相反方向成交单元已被先

前接受的匹配组占用,则跳过该组;否则接受该组,并将其涉及的全部成交单元标记为已占用,以保证每一笔成交单元最多参与一次有效匹配.

A.6 窗口汇总与 KnapOFC 计算

对一组被接受的有效匹配 g , 将目标成交数量与相反方向组合成交数量中的较小值计入其可匹配成交量: $V(M)_g = \min\{V(o_g), \sum (c_j \in A_g^*)V(c_j)\}$.

窗口层面的可匹配成交量为所有被接受匹配组之和: $V(M)_{(i,d,t)} = \sum (g \in G_{(i,d,t)})V(M)_g$, 其中 $G_{i,d,t}$ 为股票 i 在交易日 d 的窗口 t 内最终被接受的有效匹配组集合. 在此基础上, KnapOFC 定义为:

$$\text{KnapOFC}_{i,d,t} = \frac{2 \times V(M)_{i,d,t}}{V(B)_{i,d,t} + V(S)_{i,d,t}}.$$

其中 $V(B)_{i,d,t}$ 与 $V(S)_{i,d,t}$ 分别为窗口内主动买入与主动卖出成交总量. KnapOFC 取值范围为 $[0, 1]$: 当买卖两侧不仅总量接近, 而且能在逐笔层面通过时间、价格与数量约束形成较多有效匹配时, 指标较高; 反之, 即使买卖总量较为对称, 只要逐笔成交难以形成有效匹配, 指标仍会较低.

A.7 典型示例

以一只股票在某 1 分钟窗口内的逐笔成交为例 ($lot=100$ 股, $tick=0.01$ 元). 该窗口共有 3 笔主动买入与 3 笔主动卖出, 如表 A.2 所示.

表 A2 窗口内逐笔成交数据
Tab.A2 Execution-level data

成交单元	方向	时刻	价格/元	数量/股	数量/手
b_1	买入	10:15:05	10.00	500	5
b_2	买入	10:15:40	10.01	200	2
b_3	买入	10:15:50	10.08	300	3
s_1	卖出	10:15:10	10.00	300	3
s_2	卖出	10:15:20	10.01	200	2
s_3	卖出	10:15:35	9.99	200	2

窗口内主动买入总量为 $V(B)=1\,000$ 股, 主动卖出总量为 $V(S)=700$ 股. 由于所有成交均发生在同一 1 min 窗口内, 时间约束满足; 但买入成交 b_3 的价格为 10.08 元, 与所有卖出成交的价差均超过 $\epsilon_p=0.03$, 因此不进入候选匹配集合.

对于 b_1 , 候选卖出成交为 s_1, s_2 和 s_3 , 对应数量分别为 3 手、2 手和 2 手. 以 b_1 的 5 手成交量为背包容量, 动态规划可找到组合 $s_1 + s_2$, 因此, 计入可匹配成交量 500 股. 对于 b_2 , 候选卖出成交中存在 2 手成交, 可形成精确匹配, 计入可匹配成交量 200 股. 若同一目标成交存在多个误差相同的候选组合, 则按照候选成交时间升序和成交单元序号升序确定唯一匹配组合.

进一步地, 本文将以买入成交和以卖出成交为目标得到的潜在匹配组合并入同一候选池, 并按照数量误差由小到大、可匹配成交量由大到小、目标成交时间由早到晚的顺序依次接受. 若某一成交单元已被此前接受的匹配组使用, 则后续包含该成交单元的匹配组被跳过, 以避免重复计入. 根据上述规则, 本例中最终被接受的有效匹配为 $b_1 \leftrightarrow (s_1, s_2)$ 和 $b_2 \leftrightarrow s_3$, 可匹配成交量为: $V(M) = 500 + 200 = 700$.

因此, 该窗口的 KnapOFC 为: $\text{KnapOFC} = \frac{2 \times 700}{1\,000 + 700} = \frac{1\,400}{1\,700} \approx 0.824$.

可见, KnapOFC 并非简单的买卖成交量对称度, 只有同时满足时间接近、价格接近和数量可组合匹配条件的成交, 才会被计入可匹配成交量; 价格偏离较大的成交即使发生在同一窗口内, 也不会被认定为有效匹配.

表 S1 NOC 事件与已实现波动率

Tab. S1 NOC events and volatility

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:被解释变量为 $RV^{(5)}$						
NOC 事件	−0.001 9*** (−14.39)	−0.002 4*** (−9.75)	−0.002 5*** (−15.75)	−0.004 1*** (−11.61)	−0.001 9*** (−8.49)	−0.003 1*** (−4.23)
观测值	12 352	5 676	12 350	5 676	2 245	1 032
R^2	0.085	0.093	0.156	0.144	0.127	0.100
Panel B:被解释变量为 $RV^{(15)}$						
NOC 事件	−0.002 9*** (−15.88)	−0.003 8*** (−9.80)	−0.003 7*** (−16.39)	−0.006 4*** (−12.94)	−0.002 7*** (−8.34)	−0.004 1*** (−4.77)
观测值	12 352	5 676	12 350	5 676	2 245	1 032
R^2	0.105	0.105	0.163	0.168	0.126	0.117
Panel C:被解释变量为 $RV^{(30)}$						
NOC 事件	−0.003 7*** (−15.95)	−0.004 7*** (−9.69)	−0.004 6*** (−16.96)	−0.007 5*** (−12.15)	−0.003 0*** (−8.01)	−0.004 5*** (−4.54)
观测值	12 352	5 676	12 350	5 676	2 245	1 032
R^2	0.105	0.098	0.156	0.169	0.120	0.118
对照组	安慰剂窗口	安慰剂窗口	普通高成交 非 NOC 窗口	普通高成交 非 NOC 窗口	成交爆发 匹配窗口	成交爆发 匹配窗口
历史窗口	20	60	20	60	20	60
控制变量	Y	Y	Y	Y	Y	Y
μ_i	Y	Y	Y	Y	Y	Y
δ_d	Y	Y	Y	Y	Y	Y
θ_t	Y	Y	Y	Y	Y	Y

注：***、**、* 分别表示在 1%、5%、10% 水平上显著，下同。

表 S2 不同阈值设置汇总表

Tab. S2 The summary table of different

W	被解释变量	NOC 系数范围	显著性
20	FuturePI30	−0.000 092 至 −0.000 070	均 1%显著
20	RV5	−0.001 783 至 −0.001 294	均 1%显著
20	RV15	−0.002 683 至 −0.001 948	均 1%显著
20	RV30	−0.003 075 至 −0.002 331	均 1%显著
60	FuturePI30	−0.000 154 至 −0.000 113	均 1%显著
60	RV5	−0.002 542 至 −0.001 890	均 1%显著
60	RV15	−0.003 466 至 −0.002 874	均 1%显著
60	RV30	−0.004 065 至 −0.003 407	均 1%显著

表 S3 分时段核心回归结果

Tab. S3 Main regression results of different samples

W	子期间	FuturePI30	RV ⁽⁵⁾	RV ⁽¹⁵⁾	RV ⁽³⁰⁾
20	2025-09 至 2025-12	-0.000 093***	-0.001 475***	-0.002 311***	-0.002 814***
20	2026-01 至 2026-04	-0.000 075***	-0.001 501***	-0.002 233***	-0.002 638***
60	2025-09 至 2025-12	-0.000 141***	-0.002 337***	-0.003 357***	-0.003 996***
60	2026-01 至 2026-04	-0.000 086***	-0.002 189***	-0.003 247***	-0.003 771***

表 S4 板块差异控制回归结果

Tab. S4 Main regression results of different stocks

W	板块变量	系数	t 统计量	显著性
20	科创板	0.000 614	2.90	***
20	创业板	0.000 638	4.43	***
60	科创板	0.000 281	1.94	*
60	创业板	0.000 096	1.02	不显著