

市场微观结构系列研究（一）

——从分钟频到 Level2 的探索与发现

核心观点

本文汇集了我们过去四年以来的高频因子开发经验。在高频数据开发速度加快的背景下，我们系统性整理研究成果，以期对高频因子策略的开发与研究有所帮助。核心内容主要包括两个部分：分钟频的因子挖掘实例，以及 Level2 数据的应用探讨。市场微观结构的变化常会影响历史统计规律的一致性，深入探究高频数据背后反映的规律，方才是挖掘“Alpha”的重要法门。

高频数据深度研究汇总

分钟频高频因子挖掘技巧

因子策略的本质是通过挖掘数据特征来映射标签（通常是未来收益率）的分布，根据现有的方法论，求解路径主要有三类：

- 挖掘数据的分布规律：借助表达式的持续优化，强化指标自身的截面排序效能。
- 捕捉变量的时序特征：股票市场汇聚了各类性质迥异的参与群体，不同群体的行为模式在多维度均存在显著差异，其中交易时间维度的分化，一方面源于市场制度约束，另一方面则由投资者自身交易属性决定。
- 逻辑驱动的多指标联动：量价的 Alpha 并非总是通过单一指标呈现，多指标间的关联往往还能传递部分额外信息。

基于 Level2 数据的案例分析

- 识别投资者的交易意图：订单流是市场中所有买卖订单的实时动态集合，包括订单的价格、买卖方向、交易数量和执行状态，不仅是股票价格变动的根本驱动力，更是投资者交易意图的直接反映。
- 探究不同类型投资者的信息：在股票市场中，不同类型投资者的行为特征与决策逻辑存在显著差异。清晰认知这些投资者类型，不仅是把握市场交易动态的关键，更能为投资策略制定提供重要参考。
- 适应算法的变化改进策略或是识别机构特征：由于机构在逐渐采用算法交易执行买卖订单，我们在识别机构订单的难度在不断上升；甚至最简单的“大单化小”的策略，亦能够使得 A 股市场中的大单资金流的定义标准出现偏差。尽管如此，我们仍然可以利用高频数据工具对市场的真实情况进行逆向还原。

金融工程研究

姚紫薇

yaoziwei@csc.com.cn

SAC 编号:S1440524040001

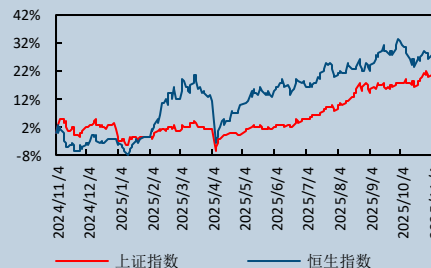
苏良

suliang@csc.com.cn

SAC 编号:S1440525090002

发布日期：2025 年 12 月 14 日

市场表现



相关研究报告

- | | |
|----------|---|
| 25.10.28 | 【中信建投金融工程及基金研究】：高频流动性与波动率因子再构建 |
| 25.09.08 | 【中信建投金融工程及基金研究】：资产配置理论、演进及国内实践 |
| 25.09.06 | 【中信建投金融工程及基金研究】：“逐鹿”Alpha 专题报告（二十八）——分钟因子模型 |
| 25.08.26 | 【中信建投金融工程及基金研究】：筹码分布因子系统构建 |
| 25.08.12 | 【中信建投金融工程及基金研究】：如何一键优化主动权益基金的基准偏离 |

目录

一、适配高频数据的“取”与“舍”	1
二、上篇：高频因子挖掘技巧三则	2
2.1 高频特征的分布信息挖掘	2
2.2 解析时序变量的选股规律	5
2.3 探究多指标关联的底层逻辑	15
三、下篇：基于 Level 2 数据的多元视角探究微观世界	23
3.1 探究真实的交易意图	23
3.2 关于投资者性质的猜想	31
3.3 与时俱进的算法交易	39
四、附录	47
4.1 MOD 修正法	47
4.2 CNIR 因子计算公式	47
风险分析	48

图目录

图 1: 单笔金额分位数受到股价、流动性、大单交易的影响	2
图 2: 单笔金额分位数_QUA 因子分组单调性	3
图 3: 因子与主要的风格因子暴露程度	3
图 4: 日内分时段涨跌幅因子的 IC 分布	5
图 5: 分钟涨跌幅呈现明显的“尖峰厚尾”特点	6
图 6: 分钟涨跌幅的统计指标因子 IC 测算结果	6
图 7: 累计样本的偏度因子 IC 结果	6
图 8: 累计样本的偏度因子的多空对冲收益	6
图 9: 修正收益率偏度_CSKEW 因子的分组单调性	7
图 10: 修正收益率偏度_CSKEW 因子 IC 均值为-0.053	7
图 11: 修正收益率偏度_CSKEW 的风格收益较高	7
图 12: 修正收益率偏度_CSKEW 因子优选组合超额净值	7
图 13: 分钟涨跌幅的反转效应变化（讨论日内反转效应的变化）	8
图 14: 早盘成交量占比近年来随着交易拥挤而抬升	9
图 15: 尾盘成交量占比 2021 年以来保持相对稳定	9
图 16: 盘中成交占比与股票换手率负相关	10
图 17: 早盘、盘中、尾盘成交占比因子横向对比	10
图 18: 日内分时成交量的特征提取步骤（时间重心算子在分钟成交量上的应用）	10
图 19: 衡量尾盘聚集效应的指标 A 与指标 B 测试效果	11
图 20: 成交重心与成交占比在信息上存在互补关系	11
图 21: 描述涨幅与跌幅在时间轴上分布位置的表示方法（时间重心算子在分钟涨跌幅上的应用）	11
图 22: 短线交易拥挤度_STC 因子的分组单调性	12
图 23: 短线交易拥挤度_STC 因子的 IC 均值为-0.052	12

图 24: 短线交易拥挤度_STC 因子与波动、流动性相关.....	12
图 25: 短线交易拥挤度_STC 因子优选组合超额净值.....	12
图 26: TGD 因子的分组单调性.....	13
图 27: TGD 因子的 IC 分布图 (测算 IC 和 Rank IC)	13
图 28: TGD 与风格因子的超额相似度 (暴露程度)	13
图 29: TGD 单因子增强收益表现	13
图 30: 多维度指标关联的因子体系化构建方案 (强相关指标)	15
图 31: 分钟指标相关性矩阵 (资金流相关性较高)	17
图 32: “指标对”因子 ICIR 与指标间相关系数正相关.....	17
图 33: 主力交易强度_MTS 因子的分组单调性.....	18
图 34: MTS 因子的 IC 均值为 0.06, Rank IC 均值为 0.07.....	18
图 35: MTS 因子与流动性的超额贡献重合度为 0.45.....	18
图 36: MTS 单因子增强收益表现	18
图 37: 多维度指标关联的因子体系化构建方案 (弱相关指标)	20
图 38: 基于 AMP 样本分类结果: 低波动量、高波反转.....	20
图 39: SR 因子的分组单调性.....	21
图 40: SR 因子的 IC 分布图 (测算 IC 和 Rank IC)	21
图 41: SR 与风格因子的超额相似度 (暴露程度)	21
图 42: SR 单因子增强收益表现	21
图 43: Level2 数据基本结构	23
图 44: 提交成功的委托订单有四种可能的退出方式	24
图 45: 从订单流中探寻投资者交易意图的蛛丝马迹	24
图 46: 撤单比例长期保持稳定水平	24
图 47: 撤单与成交量的时间分布规律基本一致	24
图 48: 小盘股相比大盘股流动性偏差, 撤单比例相对较高.....	25
图 49: 买卖委托的撤单比例随股价涨跌幅度变化.....	25
图 50: 大额部分撤回委托的因子构造效果一般	26
图 51: 核心指数成分股范围内彩票委托的占比	26
图 52: 集合竞价阶段“买彩票”行为较为活跃	26
图 53: 不同档位价格的差异 (区分买单、卖单)	27
图 54: 废单的价格的分布情况 (15:00 时刻订单簿快照)	27
图 55: LOTTERY 因子的分组单调性	28
图 56: LOTTERY 因子的 IC 分布图.....	28
图 57: LOTTERY 与风格因子的超额相似度 (暴露程度)	28
图 58: LOTTERY 单因子增强收益表现.....	28
图 59: PTO 因子的分组单调性	29
图 60: PTO 因子的 IC 分布图.....	29
图 61: PTO 与风格因子的超额相似度 (暴露程度)	29
图 62: PTO 单因子增强收益表现.....	29
图 63: A 股的投资者结构示意图	31
图 64: 广义主力资金净流入率_CNIR 因子 2012 年以来的 IC 变化情况.....	32

图 65: 依据不同参数下因子测试效果来确定划分大单的最佳阈值	33
图 66: 2021 年以来, 绝对阈值方法下标准值大幅度漂移	33
图 67: 相对阈值方法的结论与绝对阈值方法基本一致	33
图 68: 时长在较短 (5 秒以内) 的撤回订单较多	35
图 69: 高频订单通常集中在盘口附近, 快速流转	35
图 70: 挂单时间分布规律来源于交易所的 3 秒一次的报价规则	36
图 71: TOX 因子的分组单调性	37
图 72: TOX 因子的 IC 分布图 (测算 IC 和 Rank IC)	37
图 73: TOX 与风格因子的超额相似度 (暴露程度)	37
图 74: TOX 单因子增强收益表现	37
图 75: 高频撤单占比的时序变化	38
图 76: 相比早期, 2021 年以来的连续订单金额衰减的现象更明显	39
图 77: 机构算法拆单 (大单化小) 的具体示例	40
图 78: 滞后多阶的两笔委托, 在交易方向上同样存在关联	41
图 79: 交易方向的关联性在小市值上显著性更高	41
图 80: 自相关系数随滞后阶的变化符合幂律分布的规律, 取滞后阶的对数值方便线性化处理	41
图 81: 靠近盘口的委托更容易同向连续出现	42
图 82: 小额委托的交易方向连续性更强	42
图 83: LMS 因子的分组单调性	43
图 84: LMS 因子的 IC 分布图 (测算 IC 和 Rank IC)	43
图 85: LMS 与风格因子的超额相似度 (暴露程度)	43
图 86: LMS 单因子增强收益表现	43
图 87: OST 因子的分组单调性	44
图 88: OST 因子的 IC 分布图 (测算 IC 和 Rank IC)	44
图 89: OST 与风格因子的超额相似度 (暴露程度)	44
图 90: OST 单因子增强收益表现	44
图 91: MEMO 因子的分组单调性	45
图 92: MEMO 因子的 IC 分布图 (测算 IC 和 Rank IC)	45
图 93: MEMO 与风格因子的超额相似度 (暴露程度)	45
图 94: MEMO 单因子增强收益表现	45

表目录

表 1: “量化大侦探”系列高频因子清单	1
表 2: 单笔金额分位数_QUA 因子的计算步骤	3
表 3: 修正收益率偏度_CSKEW 因子的计算步骤	7
表 4: 中证 1000 指数成分股优选月度超额收益表现 (CSKEW 因子)	8
表 5: 短线交易拥挤度_STC 因子的计算步骤	12
表 6: 时间重心偏离_TGD 因子的计算步骤	13
表 7: 中证 1000 指数成分股优选月度超额收益表现 (STC 因子)	14
表 8: 中证 1000 指数成分股优选月度超额收益表现 (TGD 因子)	14

表 9: 日内分钟指标梳理.....	16
表 10: 主力交易强度_MTS 因子的计算步骤.....	18
表 11: 中证 1000 指数成分股优选月度超额收益表现 (MTS 因子)	19
表 12: 强反转_SR 因子的计算步骤.....	21
表 13: 中证 1000 指数成分股优选月度超额收益表现 (SR 因子)	22
表 14: 订单类型在主要板块中应用的分布	25
表 15: 彩票委托_LOTTERY 因子的计算步骤.....	28
表 16: 收盘参与度_PTO 因子的计算步骤.....	29
表 17: 中证 1000 指数成分股优选分月度超额收益表现 (LOTTERY 因子)	30
表 18: 中证 1000 指数成分股优选分月度超额收益表现 (PTO 因子)	30
表 19: 投资者类型性质.....	31
表 20: 买入订单委托金额分位数统计结果	34
表 21: 卖出订单委托金额分位数统计结果	34
表 22: 不同订单规模组下投资者群体订单占比	34
表 23: 沪深交易所接受的市价委托类型.....	36
表 24: 毒流动性_TOX 因子的计算步骤.....	37
表 25: 中证 1000 指数成分股优选月度超额收益表现 (TOX 因子)	38
表 26: 长期记忆强度_LMS 因子的计算步骤.....	43
表 27: 分拆痕迹_OST 因子的计算步骤.....	44
表 28: 高维记忆_MEMO 因子的计算步骤.....	45
表 29: 中证 1000 指数成分股优选分月度超额收益表现 (LMS 因子)	46
表 30: 中证 1000 指数成分股优选分月度超额收益表现 (OST 因子)	46
表 31: 中证 1000 指数成分股优选分月度超额收益表现 (MEMO 因子)	46
表 32: MOD 修正法处理资金流的冲击的计算步骤.....	47
表 33: 广义主力净流入率_CNIR 因子的计算步骤.....	47

一、适配高频数据的“取”与“舍”

基于高频数据的因子研究，从本质上而言始终伴随着取舍的权衡。相较于中低频价量数据聚焦趋势性、方向性信息的特点，高频数据凭借其毫秒级的颗粒度与高密度的观测维度，能穿透中长期数据难以覆盖的微观交易细节，为 Alpha 挖掘提供更丰富的潜在线索，在拓展因子计算广度、捕捉瞬时市场异动信号上的优势尤为突出。但与此同时，数据量的指数级倍增带来的挑战同样不容忽视：不仅直接推高了存储、算力层面的难度，更催生了显著的信息噪音干扰，大量无效、冗余的瞬时波动数据极易掩盖核心信号，增加有效信息的筛选成本。

作为市场中较早涉足高频数据与因子研究的团队，我们已对高频因子的样本外表现开展了长期、持续的跟踪验证，完整经历多轮市场牛熊周期与风格切换。在扎实的实践积累中我们深刻洞察：一方面，市场微观结构（如买卖盘深度、订单流特征、流动性分层、交易成本变化等）会直接作用于高频因子的信号传导效率，进而显著影响其有效性边界；另一方面，高频因子的构造过程本身更易受噪音数据干扰，导致过拟合风险相较于中低频因子更为凸显，且其背后的信息来源高度分散、影响因素交织复杂，既涵盖中观的主题热点，也涉及微观资金博弈、订单执行节奏等多重维度，进一步提升了因子筛选、验证与落地的复杂度。

如何在丰富的数据中更高效地挖掘有效信息，是我们长期以来持续探索的方向。虽会遇到挑战，但“外练武功，内修心法”，正是我们探寻高频 Alpha 的核心法门。

表 1：“量化大侦探”系列高频因子清单

因子类别	因子代码	中文名称	因子构造方式（简括）
分钟频	QUA	单笔金额分位数	分钟频单笔成交金额的 10%分位值/90%分位值
	MTS	主力交易强度	分钟单笔成交金额与分钟成交额的相关系数
	CSKEW	修正收益率偏度	排除开盘后首个半小时的日内分钟涨跌幅偏度指标
	SR	强反转	利用分钟单笔成交金额对分钟涨跌幅分两组，计算高低差
	TGD	时间重心偏离	分别将跌幅和涨幅加总求时间重心，计算二者的截面残差
	STC	短线交易拥挤度	以尾盘成交半小时为例，计算成交量的时间重心
Level2 频	LOTTERY	彩票委托	涨停价卖单的主动撤回比例
	PTO	收盘参与度	计算废单（卖单）占当日自由流通股本的比例
	TOX	毒流动性	计算 5 秒以内撤单占 30 秒以内撤单的比例
	MEMO	高维记忆	以挂单方向的 1-100 阶相关系数为样本计算偏度和峰度
	LMS	长期记忆强度	以挂单方向的 1-100 阶相关系数为样本回归系数取截距项
	OST	分拆痕迹	将挂单方向数值序列转为傅立叶频谱，计算振幅序列的峰度

资料来源：中信建投证券

在过去一段时间内，行业内倡导的高频因子低频化使用的方式，凭借其对高频信号的适度平滑、落地门槛的显著降低，以及与中低频策略的良好兼容性，推动高频因子策略的应用普及度实现快速攀升；AI 技术在量化研究领域的深度渗透与工具化普及，彻底改变了传统因子挖掘的效率边界，大量基于高频数据构造的因子被批量生成、快速迭代。这一趋势既极大丰富了量化策略的因子储备，也带来了新的行业挑战：一方面，批量挖掘催生了明显的策略同质化现象，部分因子的逻辑重叠度偏高；另一方面，易于捕捉的线性或简单非线性信号逐渐被挖掘殆尽，传统高频因子的可拓展研究空间持续收窄，行业内对高频数据价值再挖掘的需求日益迫切。

在此背景下，笔者希望通过整理我们过往的研究成果，探究这类数据后续能够继续开发的方向。

二、上篇：高频因子挖掘技巧三则

分钟行情与日频行情的数据结构基本一致，均涵盖收盘价、成交量等，不同点在于两类数据的颗粒度，前者能够观察到更加细致的日内规律。本节所涉及的分钟数据不包含由 Level2 降频而来的衍生指标，原因是笔者想聚焦于分钟数据本身，探讨传统中低频量价规律向分钟频率迁移的适应性。依托传统的量价数据和行业深入的研究，我们不仅能够刻画 A 股市场的各类交易异象，还可借助不同时段的信息差异制定交易策略。

因子策略的本质是通过挖掘数据特征来映射标签（通常是未来收益率）的分布，根据现有的方法论，求解路径主要有三类：

（1）**挖掘数据的分布规律**：借助表达式的持续优化，强化指标自身的截面排序效能。通常，适用这类方法的指标反应了市场长期存在的收益来源，例如，动量（涨跌幅）、流动性（单笔成交金额）、波动率（振幅）等等。本小节将以分钟频率单笔成交金额这一指标为例，讨论如何再进一步提升超额收益。

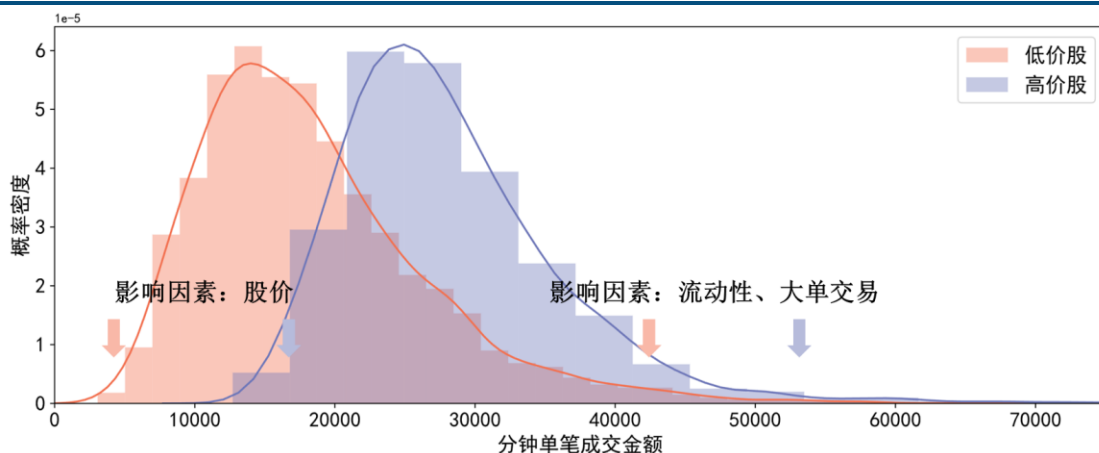
（2）**捕捉变量的时序特征**：股票市场汇聚了各类性质迥异的参与群体，不同群体的行为模式在多维维度均存在显著差异，其中交易时间维度的分化，一方面源于市场制度约束，另一方面则由投资者自身交易属性决定。基于这一研究思路，我们展示了两类不同方法，分别用以构造三个描述现象背后规律的因子。

（3）**逻辑驱动的多指标联动**：量价的 Alpha 并非总是通过单一指标呈现，多指标间的关联往往还能传递部分额外信息。利用好高频数据的丰度优势，我们可以批量测试多个指标间的相关性是否具有预示作用，还可以通过主观逻辑去寻找关键变量，对目标变量进行切割以提纯有效信息。

2.1 高频特征的分布信息挖掘

单笔成交金额反映成交订单的平均金额大小，若是某个时段内盘口附近单笔委托的金额整体水平提高，则在相同成交金额下，撮合成功的订单笔数更少，平均计算的成交金额会增加。

图 1：单笔金额分位数受到股价、流动性、大单交易的影响



数据来源：中信建投证券

我们认为，主力资金的挂单行为与普通散户存在显著差异，其核心特征是挂单金额更大、成交意愿更强，通常委托价格集中在盘口价附近。因此单笔成交金额偏高，往往可作为主力资金参与交易的潜在信号。不过需要注意的是，低价、小市值或流动性不足的股票，在交易相对清淡的时段，交易所撮合机制可能促成两笔大额委托的意外成交，进而导致单笔成交金额出现异常偏高的情况，这类情形并非主力资金参与的信号。

单笔金额数据的处理和构造因子过程中存在以下难点：（1）股价相关：不同股价间的单笔金额存在较为明显的差异，例如贵州茅台相比较于其他低价股，单笔成交金额平均水平相差很大；（2）低流动性股票：在流动性不足（小单委托较少）的时候，两笔较大的订单被撮合成交会使得单笔金额出现零星的异常值。我们需要构造的是能够横向对比不同股票的定量指标，考虑到数据的局限，制定了针对性的处理方法。

分位数处理：避免极端值的干扰；

标准化处理：将较大的分位数与较小的分位数相比，利用比值实现去量纲。

于是，我们构造了单笔成交金额分位数_QUA 因子来刻画上述特征，其计算方法如下：

表 2: 单笔金额分位数_QUA 因子的计算步骤

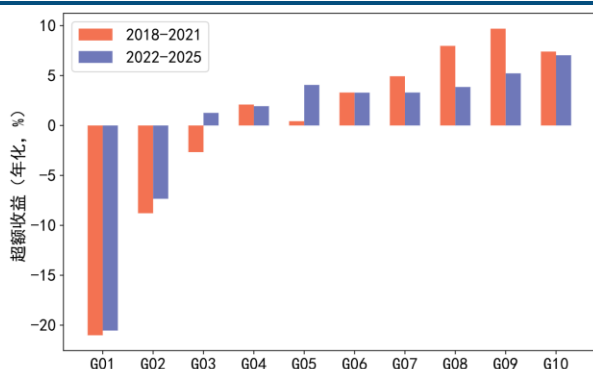
步骤	说明
第一步	选定股票，逐个交易日回顾分钟单笔成交金额序列，记为 A
第二步	将 A 从小到大排序，剔除最大的 10 个样本值（极端值）
第三步	构造指标 $S = (A_{10\%} - A_{min}) / (A_{max} - A_{min})$ ，其中 A_{max} 和 A_{min} 分别为最大值和最小值， $A_{10\%}$ 为 10% 分位值
第四步	S 指标为 QUA 因子的日频信号，例如，滚 20 个交易日计算 S 指标的均值即为月度选股因子

资料来源：中信建投证券

在测算因子表现时，我们剔除行业与市值的影响，同时排除不具备交易条件（如 ST、涨跌停）的股票，在月度（20 日）的调整频率下观察因子在全市场十分组的收益表现。单笔金额分位数_QUA 因子 IC 为-0.054，Rank IC-0.078，全样本内每年的多空收益可以达到 35.7%；在时段 1（2018-2021）与时段 2（2021-2025）内，QUA 因子的分组效果基本保持一致，均在空头端表现出异常高的负向收益。

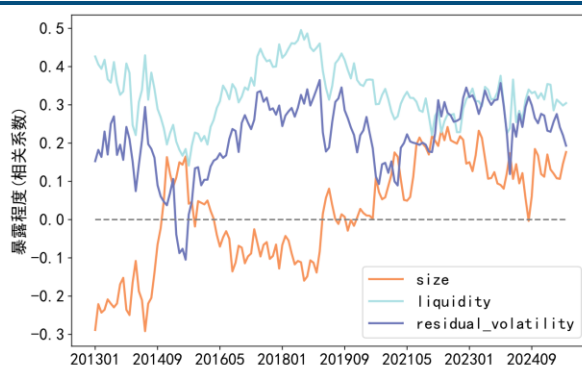
图 3 展示了 QUA 因子在流动性、市值等 Barra 风格上的暴露程度。

图 2: 单笔金额分位数_QUA 因子分组单调性



数据来源：Wind、中信建投证券，20180102-20250804

图 3: 因子与主要的风格因子暴露程度



数据来源：Wind、中信建投证券，20130104-20250818

QUA 因子在空头部分的收益尚可，但在多头部分的收益明显较弱（图 2），可实现的收益偏低，这也是一般量价因子的通病。在风格偏离上，截止 2025 年 8 月，QUA 因子在市值上的暴露程度仅为 0.18，但是在 2013 年 1 月至 2014 年 10 月和 2016 年 4 月至 2019 年 1 月两段区间内为负值；流动性风格暴露长期稳定在 0.3 附近，在扣除流动性部分收益之后，多空收益由 35.7%降低至 17.8%，IC 仅为-0.036，这说明了 QUA 本质上还是以风格收益为主导的 Half-Alpha。

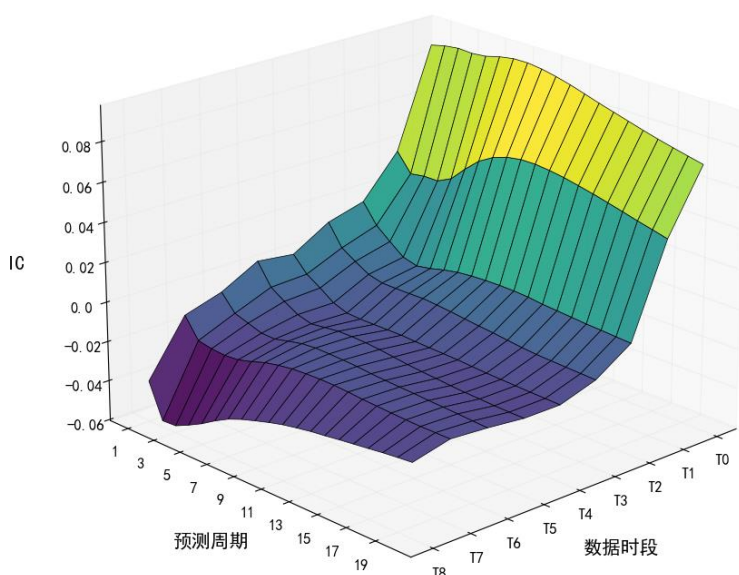
单笔成交金额与流动性关系密切：当市场流动性不足时，其最小值能够逼近最小交易单位对应的金额，而最大值往往与市场交易活跃度直接相关。尽管我们已采用比值形式构造因子，用以规避不同价格股票之间的差异，但 QUA 因子与换手率因子的 IC 序列相关系数仍高达 0.85。若对流动性暴露施加约束，该因子可能难以在组合优化过程中拓宽有效边界。事实上，我们并不排斥由风格带来的收益，而是顾虑在因子贡献的超额收益中风格收益成分占比过高，这可能会在某些极端行情下，引发组合的超额大幅回撤。

2.2 解析时序变量的选股规律

我们在分析日内量价信息时，核心关注点在于特定时间段内是否蕴含特殊规律。这一问题可通过多种讨论形式展开：探讨在上午、下午两个交易时段的反转性质差异，或计算早盘、尾盘及盘中的个股成交量占比，进而区分不同时段的交易者行为差异等。目前在这块的研究相对充分，笔者仅以较为常见的分钟因子为例介绍分析如何构造表达式，描述分时段性质差异对未来收益率的预测作用。

造成量价信息在日内分布不均的原因主要是 A 股的交易时间段不连续。出于研究目的，我们将日内交易划分为 9 个时段：T0 为隔夜时段，T1 至 T8 则自 9:30 开盘后，间隔半小时划分一组；选取过去 20 个交易日作为时间窗口，分别计算各个对应时段涨跌幅的平均值，测算其与不同周期的未来收益率的相关系数（IC）。图 4 展示了不同参数下相关系数数值的分布规律。

图 4：日内分时段涨跌幅因子的 IC 分布

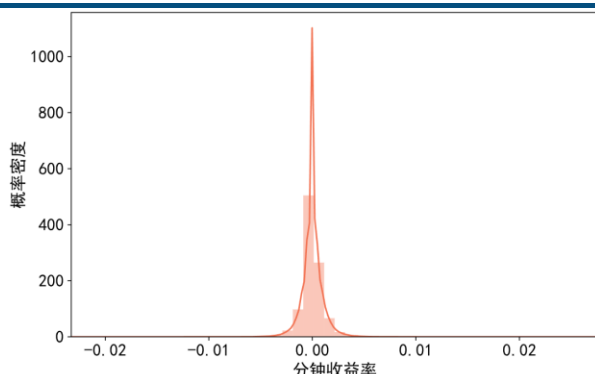


数据来源：Wind，中信建投证券；统计区间：20130102 至 20251105

以价格涨跌为例，我们发现在日内不同时间段在“反转”与“动量”特性上存在差异。由图 4 的数据分布特征可见，隔夜时段涨跌幅与开盘后首个半小时涨跌幅的 IC 为正值，其余日内交易时段尤其是临近收盘的 T8 时间段的 IC 显著为负值。早盘竞价是对隔夜信息的集中反映，涨跌幅对于后续价格的预示作用更强，尾盘竞价则容易出现价格过度反应从而出现短期反转。考虑到这一规律或与市场一字板连涨股票相关，我们在统计过程中剔除了该类股票样本，发现核心结论并未发生改变：隔夜时段与开盘后首个半小时的涨跌幅仍对未来涨跌具有正向预示作用，统计期间为 2013 年 1 月 4 日至 2025 年 11 月 5 日。

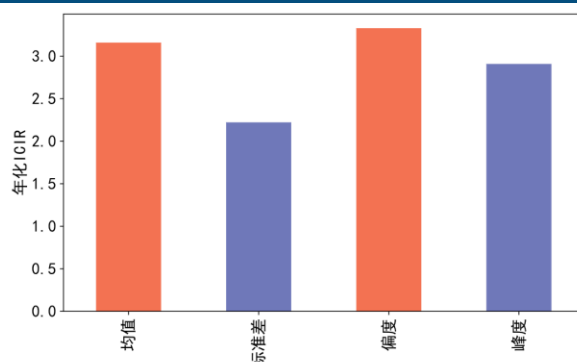
上一小节介绍了基于分钟频单笔成交金额数据，以分布特征为视角开发因子，该方法论可以迁移到其他常规变量。以反转为例，我们可以计算日内分钟涨跌幅的统计指标，以描述股票市场的反转特性（A 股存在显著的反转效应：价格过度涨跌，短期容易出现反向变动）。分钟涨跌幅呈现“尖峰厚尾”的分布特征，图 5 展示了分钟涨跌幅的均值、标准差、偏度以及峰度指标在全市场十分组下的测试表现。

图 5: 分钟涨跌幅呈现明显的“尖峰厚尾”特点



数据来源: Wind、中信建投证券

图 6: 分钟涨跌幅的统计指标因子 IC 测算结果



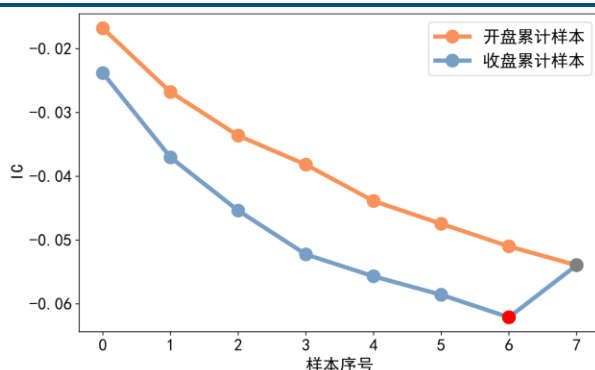
数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

作为描述数据分布特征的核心统计指标，一至四阶中心矩在截面选股因子的构造中表现出差异化的选股有效性：奇数阶中心矩（一阶、三阶）更擅长捕捉偏态样本的结构特征；偶数阶中心矩（二阶、四阶）用以刻画数据的分布离散程度，其选股效果在数据呈现显著对称分布特征的场景下更为突出。由于在分钟频率的观测窗口下，价格的变动通常不是特别明显，整体表现为“尖峰厚尾”的分布规律，在价格出现连续上涨或者连续下跌的时候才会表现为偏态分布。

收益率偏度因子是市场上使用较为广泛的中心矩类因子。该因子以股票日内分钟涨跌幅为基础数据，通过计算其偏度统计量构建截面选股指标，描述了股票“价格反转”特征：当偏度为正值时，说明日内收益率分布右偏，即少数极端上涨收益拉动了分布形态，这类股票往往因短期资金炒作而出现脉冲式上涨，后续价格出现回调概率较高；当偏度为负值时，分布左偏意味着股票近期可能因短期被恐慌性抛售，股价被过度压制，后续反弹潜力较大。

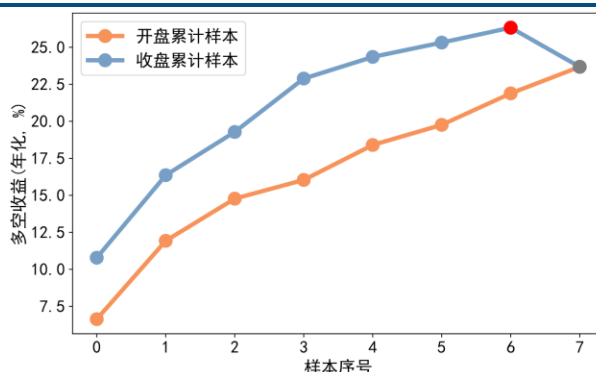
由于日内各交易时段的选股信息具备时段特异性，因子构建可基于其内在规律进行差异化调整，以匹配不同时段的市场微观结构特征。为了更加直观地展示日内分时段的差异，我们以半小时为单位累计样本计算分钟涨跌幅的偏度，测试每日指标的 20 日均值作为因子在月频调仓频率下的选股能力，如图 7 和图 8 所示。

图 7: 累计样本的偏度因子 IC 结果



数据来源: Wind、中信建投证券

图 8: 累计样本的偏度因子的多空对冲收益



数据来源: Wind、中信建投证券

综合上述参数测试结果，我们将收益率偏度因子改进，得到修正收益率偏度_CSKEW 因子。

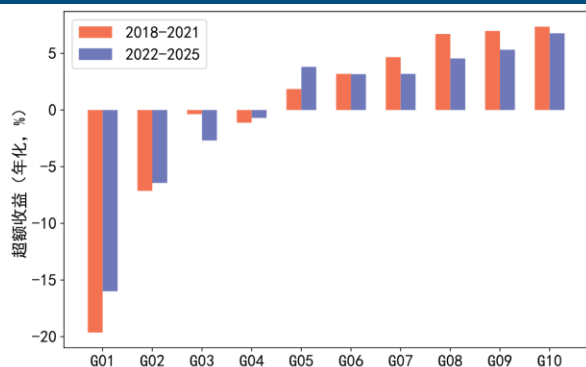
表 3: 修正收益率偏度_CSKEW 因子的计算步骤

步骤	说明
第一步	逐个交易日回顾分钟涨跌幅序列, 记为 R
第二步	剔除开盘后首个半小时内的分钟涨跌幅样本
第三步	构造指标 $S = (A_{10\%} - A_{min}) / (A_{max} - A_{min})$, 其中 A_{max} 和 A_{min} 分别为最大值和最小值, $A_{10\%}$ 为 10% 分位值
第四步	S 指标为 QUA 因子的日频信号, 例如, 滚 20 个交易日计算 S 指标的均值即为月度选股因子

资料来源: 中信建投证券

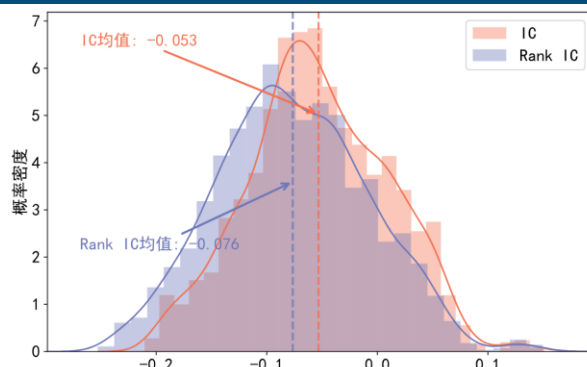
修正收益率偏度因子的 IC 为-0.053, 收益表现相比原始的收益率偏度因子有小幅提升, 多空对冲的年化收益超过 30.4%, 信息比率为 3.3。图 9 展示因子在全市场十分组的单调性, 在前后两个时段内均能够有不错的区分度。图 10 展示了中性化后因子的 IC 分布, 图 12 为中证 1000 指数成分股优选的超额净值曲线。

图 9: 修正收益率偏度_CSKEW 因子的分组单调性



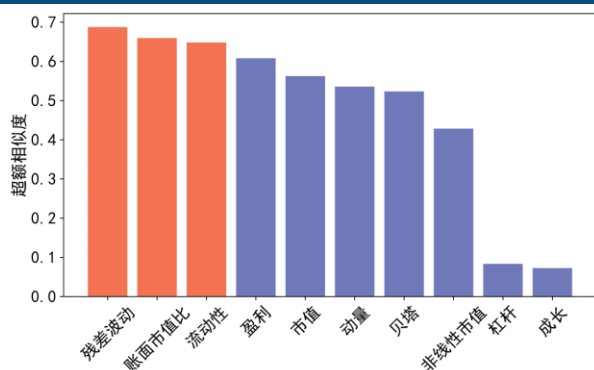
数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 10: 修正收益率偏度_CSKEW 因子 IC 均值为-0.053



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 11: 修正收益率偏度_CSKEW 的风格收益较高



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 12: 修正收益率偏度_CSKEW 因子优选组合超额净值



数据来源: Wind、中信建投证券

修正收益率偏度_CSKEW 因子与常规风格因子的暴露程度如图 11 所示, 暴露程度较高的因子分别是残差波动、账面市值比和流动性; 图 12 展示 CSKEW 因子在中证 1000 指数成分股内优选的 200 只股票, 相对指数的超额净值走势, 超额净值的夏普比率为 1.46。表 4 展示了更为详细的数据, 以月度窗口来统计的 CSKEW 因子相对中证 1000 指数的超额收益, 2025 年 6 月至 9 月, 因子的超额出现较为明显的回撤。

表 4: 中证 1000 指数成分股优选月度超额收益表现 (CSKEW 因子)

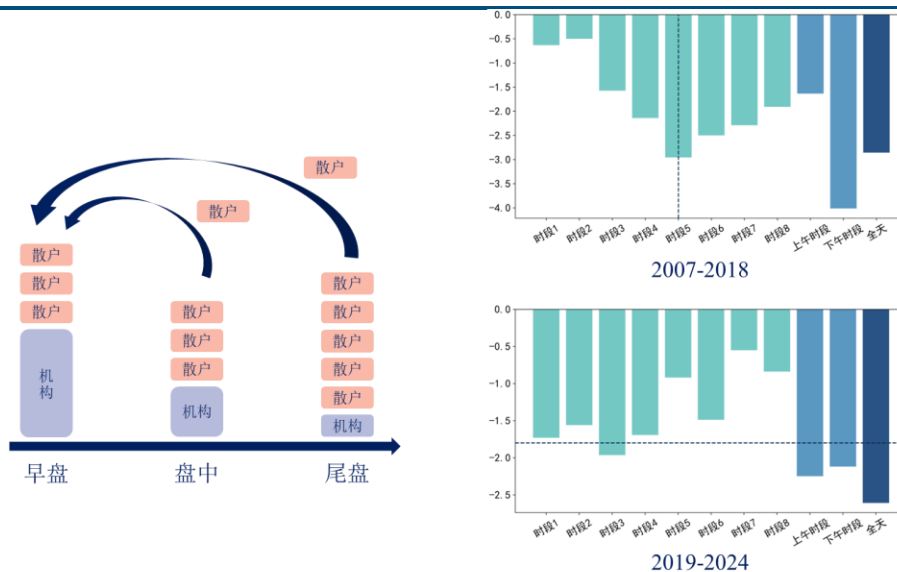
年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	0.37	-1.92	2.60	3.51	1.60	-1.43	-0.67	-2.26	-0.32	1.98	1.18		4.55
2024	1.34	-1.76	-0.02	-0.14	0.97	0.32	1.86	1.73	0.79	-1.94	2.80	2.66	8.83
2023	-0.03	0.63	0.80	0.21	0.05	0.20	3.04	1.05	-0.59	-0.21	0.44	-0.16	5.50
2022	2.35	-2.97	3.13	2.17	-0.53	-0.80	0.72	4.03	1.09	-0.68	0.55	0.77	10.07
20 这													
个 21	-0.92	2.87	1.48	0.99	-0.76	1.15	-2.93	-0.80	2.82	-1.36	0.90	4.06	7.53
2020	-1.70	-0.68	3.39	-0.30	-1.25	-1.25	0.70	3.08	0.95	0.29	2.59	-1.91	3.77
2019	0.29	-0.30	0.68	0.77	0.14	1.84	1.47	0.78	-0.03	-1.13	0.69	-0.87	4.37
2018	0.76	-0.34	0.41	0.47	1.23	1.73	2.69	1.06	0.95	0.51	1.25	0.65	11.96

资料来源: Wind、中信建投证券; 统计区间: 20180102-20251105; 单位: %

市场微观结构的稳定性是高频因子发挥作用的核心前提, 基于时序样本切割策略所构建的反转因子, 其有效性存在显著的时序波动特征, 并非始终保持稳健。A 股的反转之力微观来源应属参与市场交投最为活跃的散户投资者, 这部分交易者在全天均匀分布。根据我们的观测结果, 这一规律似乎在悄悄发生改变, 其中最为明显的变化就是早盘交易的拥挤, 大量原本应分散到全天任意时段交易的订单, 现在正在往早盘时段集中。

我们观察到一个值得关注的市场现象: 无论是流动性充裕的大市值股票, 还是流动性相对匮乏的小市值股票, 其交易日开盘后特定时段内的成交量占当日总成交量的比重均呈现持续攀升趋势。这一现象本质上反映了市场投资者形成的隐性交易共识: 若隔夜市场出现明确的利好或利空信息, 投资者次日需要执行相关买入或卖出决策时, 倾向于选择在开盘后尽早执行交易。其核心诉求在于通过抢占交易时点, 规避因其他投资者集中买入导致的买入成本抬升, 或因集中卖出引发的市场踩踏风险。

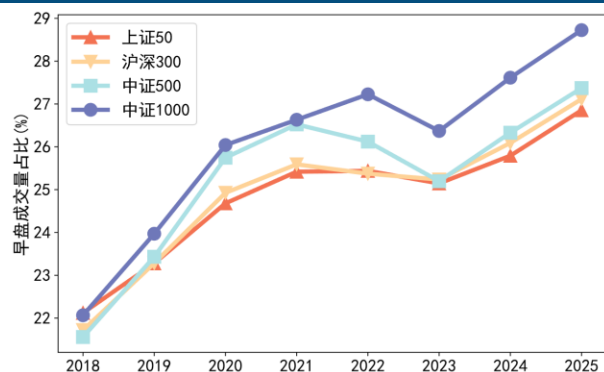
图 13: 分钟涨跌幅的反转效应变化 (讨论日内反转效应的变化)



数据来源: Wind、中信建投证券

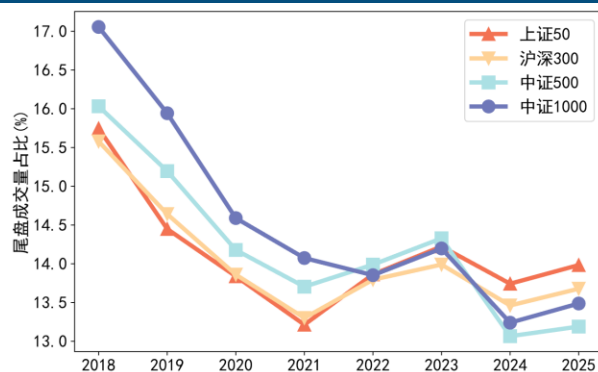
市场交易者“拥挤度”上升，早盘成交量的占比在逐渐增加。成交量上的转移可能是因为散户交易习惯发生了变化，导致在日内交易时段内其贡献了反转属性的因素集中在早盘阶段。这对于反转因子而言并非是一件好事，因为主导反转的因子在日内过于集中，并且集中在日内动量特征相对明显的早盘，二者相互抵消。

图 14: 早盘成交量占比近年来随着交易拥挤而抬升



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 15: 尾盘成交量占比 2021 年以来保持相对稳定



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

股票价量分析中另一个重要的维度是成交量。股票的成交活跃度在不同时刻是有差异的，成交量的日内分布大致呈现 U 型变化，关于这一现象，学术界和业界有诸多讨论。普遍的观点是，**A 股交易时间不连续，导致成交量较为容易地堆积在连续交易时间的两端**：开盘后的快速放量是对隔夜信息的集中消化，而后逐渐平复至正常水平；临近收盘时刻，投资者为准备次日的交易，提前将手上的现金兑换为筹码，造成交易量增加。

我们曾针对不同时段的交易主体差异展开系列讨论，核心结论明确：机构投资者更倾向于在早盘流动性充足、活跃筹码集中的时段交易，是因为大额委托不易产生明显滑点；尾盘交易的参与者则以散户为主，部分机构受交易条件限制，难以大量参与尾盘的竞价；盘中则是散户与游资的交易窗口，具备时间长、整体交易平均的特点。在缺乏市场炒作热点的情况下，早盘与尾盘的交易不会显著放量，全天的分时交易曲线变化平缓。

早盘成交占比：9:25 至 9:59 期间的成交量与 9:25 至 11:29 期间（上午）成交总量的比值；

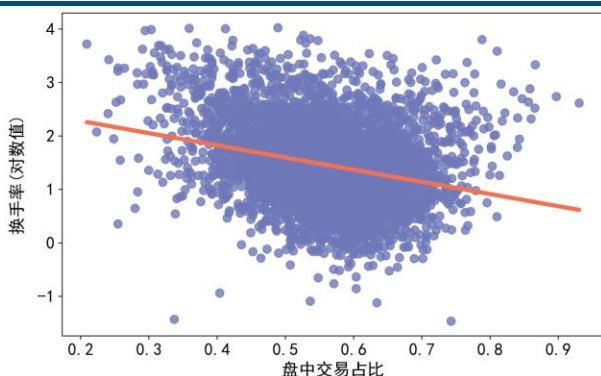
盘中成交占比：10:00 至 14:29 期间的成交量占全天成交总量的比值；

尾盘成交占比：14:30 至 14:59 期间的成交量与 13:00 至 14:59 期间（下午）成交总量的比值。

图 16 展示了盘中成交占比与当天换手率的关系，**盘中时段交易量占比越高，说明当日成交越不活跃，换手率相对偏低**；图 17 展示了上述三个指标在 20 日换仓的分组测试下，多空收益以及因子 IC 均值。三类成交占比因子整体相关性偏高，如盘中成交占比与早、尾盘的相关系数分别为 0.85 和 0.53。在剥离早盘和尾盘成交占比的收益之后，盘中成交量占比的超额收益显著变差，多空年化收益下降 30%，说明盘中成交占比仅是另外两个指标在交易总量维度的间接映射，并不是我们要寻找的逻辑主因。

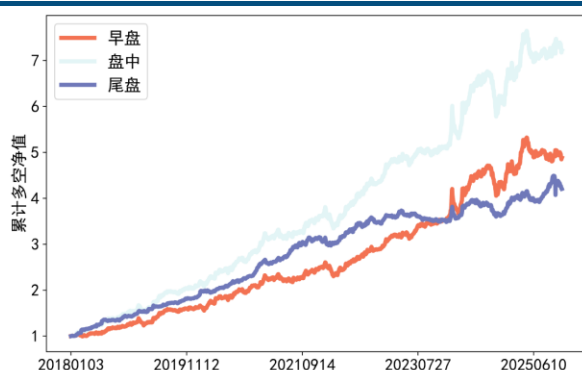
早盘和尾盘的成交占比指标，我们认为分别代表了成交量时序维度的不同信息。早盘属于全天交易量最为集中的一段区间，早盘成交占比的高低反映了全天交易的热情度，间接可用于判断是否为当天是否放量；尾盘成交量的交易不同于早盘，其主要贡献者为散户等以短线交易为主的个人投资者，在信息来源上也并非与是否放量相关的换手流动性完全重合。早盘成交占比因子与尾盘成交占比因子的 IC 序列相关系数为 0.38，若是将两者相互中性化，多空收益率分别下降至 19.4%（原值 23.6%）和 15%（原值 20.9%）。

图 16: 盘中成交占比与股票换手率负相关



数据来源: Wind、中信建投证券, 以 20251103 的数据为例

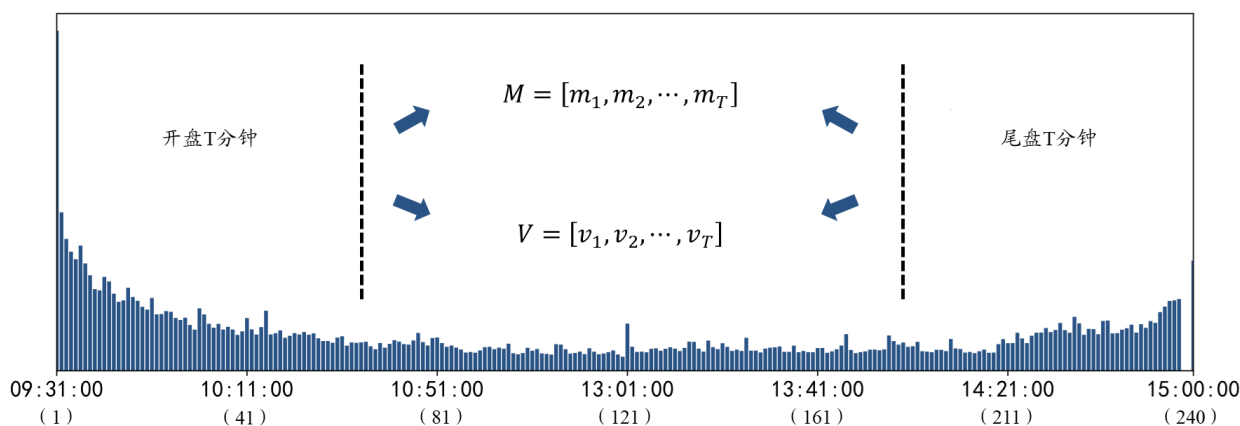
图 17: 早盘、盘中、尾盘成交占比因子横向对比



数据来源: Wind、中信建投证券

分时成交量占比指标的核心逻辑在于捕捉某种“低流动性”特征对应的风险溢价。早盘与盘中的成交量占比因子是换手率的近似代替, 尾盘交易反应的信息与早盘存在一定差异。当天放量的结果通常可以通过早盘交易情况预估, 而尾盘交易的情况则与投资者结构中偏好短线交易类的资金相关, 即流动性的来源不相同。本文创新性地构建一种刻画时序分布特征的时间重心算子: 首先, 将分钟频时序变量的时间戳转换为可用于运算的序号值, 再与原始变量执行内积运算, 最终得到时序变量在时间维度上的分布重心指标。

图 18: 日内分时成交量的特征提取步骤 (时间重心算子在分钟成交量上的应用)



第一步, 重新标注时间戳, 例如, 将09:31:00和15:00:00分别标注为1和240, 其余依次同理;

第二步, 选取开盘/尾盘T分钟的成交量时序数据, 统计时间序号M和成交量V两列向量;

第三步, 将成交量V标准化后与时间序号计算内积 (成交重心G)。 $\Rightarrow G = MV^T / \|V\|_1$

P.S. 开盘集合竞价成交量归于09:31 (第1根Bar), 尾盘集合竞价成交量归于15:00 (最后1根Bar)

数据来源: 中信建投证券

基于前述构建的时间分布重心指标, 可有效量化成交量在时间维度上的集中特征。但因股票市场成交量通常呈现典型U型分布, 数据层面表现为早盘与尾盘时段的双峰值形态。因而在实际分析中, 我们通常将早盘时段成交量与上午时段成交总量进行归一化对比, 而非直接对标全天成交总量, 以避免非同源交易群体的成交量相互干扰导致的指标失真。对于成交量的时间重心, 我们也是采取相同的处理方式: 上、下午分开处理。

请务必阅读正文之后的免责条款和声明。

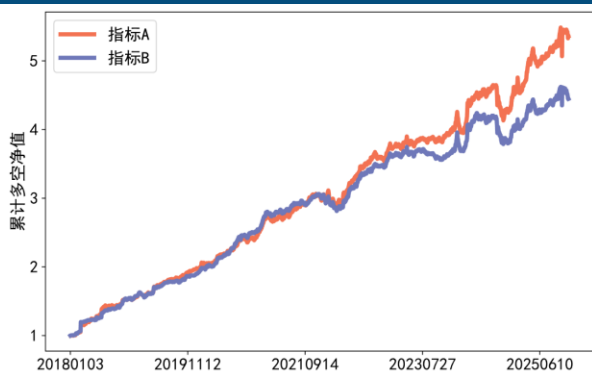
由于收盘竞价的特殊性，本文以尾盘的成交量聚集效应为例，构造如下两个指标作为选股因子：

指标 A：收盘前 30 分钟（14:31-15:00）为时间窗口，计算的时间重心指标；

指标 B：收市集合竞价阶段成交量与下午成交总量的比值。

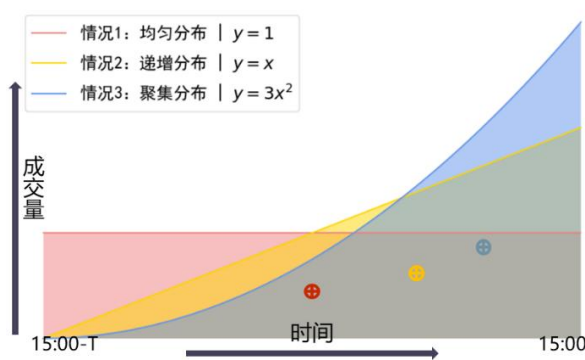
上述两个指标的计算参数均为寻优结果，出于篇幅原因不再展开介绍。图 19 分别展示了两个指标的多空收益净值曲线，二者的 IC 均值分别为-0.046 和-0.041。这两个指标的 IC 序列相关系数为 0.92，因其 Alpha 信息来源相同，高相似程度属于正常现象。笔者本意是对成交占比类因子进行算子优化，但这两个指标描述尾盘资金聚集现象的角度不同（其一从变化趋势角度，其二从占比角度），两类指标可互相补充，故而均予以保留。

图 19: 衡量尾盘聚集效应的指标 A 与指标 B 测试效果



数据来源：Wind、中信建投证券

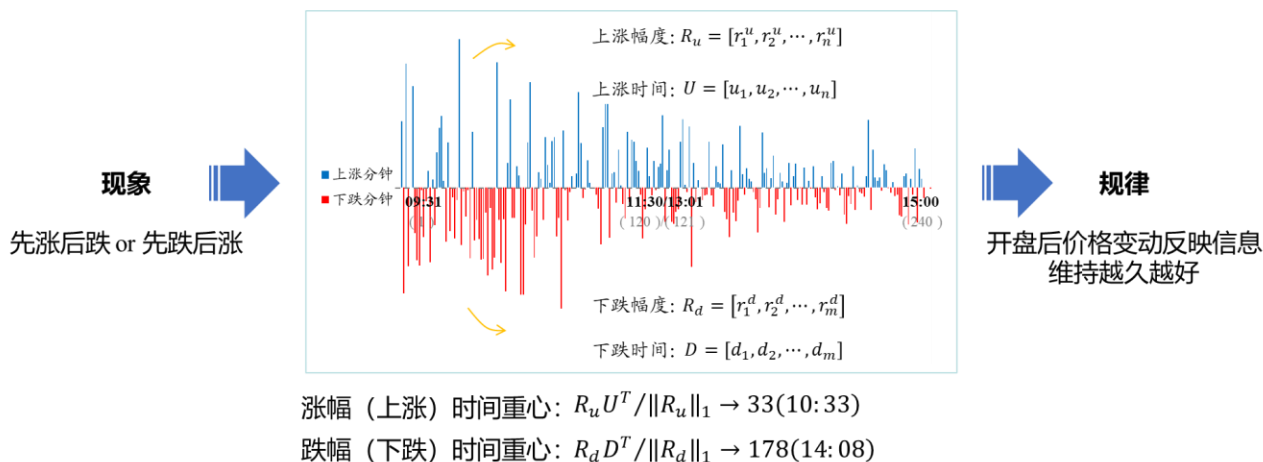
图 20: 成交重心与成交占比在信息上存在互补关系



数据来源：中信建投证券

最终，我们将指标 A 与指标 B 进行等权合成，构造短线交易拥挤度_STC 因子。因子的计算步骤及相关测试结果可见表 5、图 22 至图 25。除了成交量以外，本文研究过程中，笔者还尝试了该算子在其他分钟频特征上的迁移应用，但表现较好的因子乏善可陈，仅有基于分钟涨跌幅构造的时间重心偏离_TGD 因子表现尚可。篇幅有限，本文不对该因子的构造过程展开介绍，将其作为补充测试的内容进行展示。

图 21: 描述涨幅与跌幅在时间轴上分布位置的表示方法（时间重心算子在分钟涨跌幅上的应用）



数据来源：中信建投证券

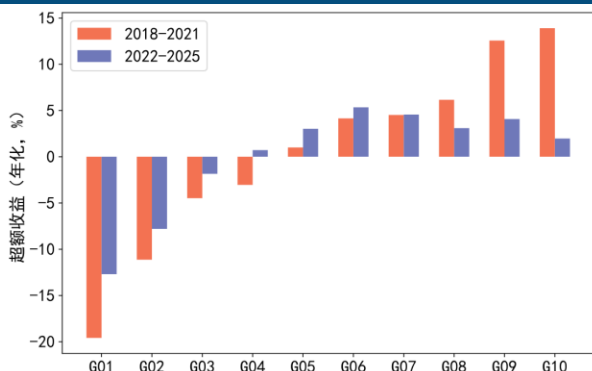
表 5: 短线交易拥挤度_STC 因子的计算步骤

步骤	说明
第一步	将日内的分钟序列时间戳转换为序号, 例如, 9:31 (1)、11:30 (120)、15:00 (240)
第二步	截取尾盘半小时计算成交重心 G : $G = MV^T / \ V\ _1$ 其中, M 和 V 分别表示分钟时间序号及对应的分钟成交量序列
第三步	计算个股尾盘集合竞价阶段 (14:57-15:00) 的成交量占全天总成交量的比值, 记为 P
第四步	将每日的尾盘半小时成交重心 G 和尾盘集合竞价成交占比 P 分别排序并相加, 得到 S 指标 $S = rank(G) + rank(P)$
第五步	S 指标为 STC 因子的日频信号, 例如, 滚 20 个交易日计算 S 指标的均值即为月度选股因子

资料来源: 中信建投证券

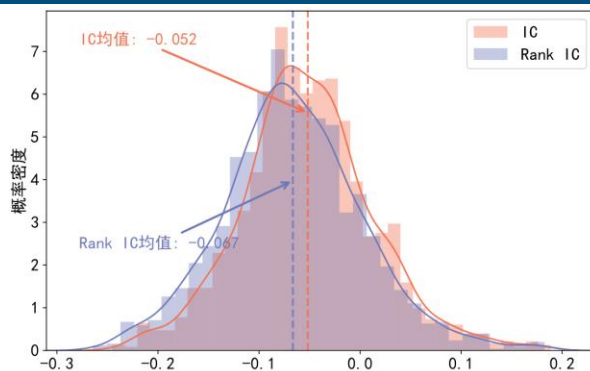
短线交易拥挤度衡量尾盘资金交易的拥挤情况。图 22 的测试结果显示该因子的分组效果在 2022 年至 2025 年期间有所降低, 多头部分收益减弱, 这与我们观测成交量日内分布漂移的现象相关: 由于尾盘整体的交易量减少, 基于成交重心构造的指标容易受到噪音的干扰, 从而降低了分组能力。

图 22: 短线交易拥挤度_STC 因子的分组单调性



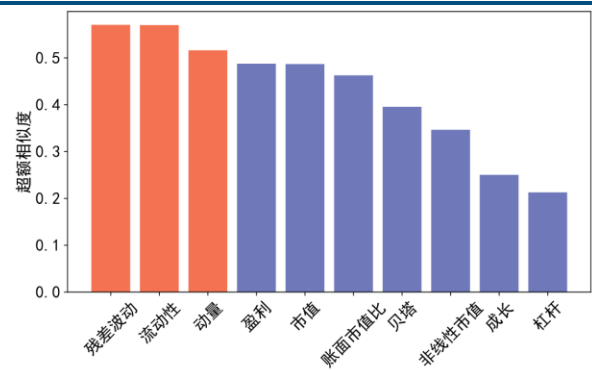
数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 23: 短线交易拥挤度_STC 因子的 IC 均值为 -0.052



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 24: 短线交易拥挤度_STC 因子与波动、流动性相关



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 25: 短线交易拥挤度_STC 因子优选组合超额净值



数据来源: Wind、中信建投证券

STC 因子在中证 1000 指数成分股范围内优选 200 只股票, 超额夏普比率为 2.19, 最大回撤为 6.23%。

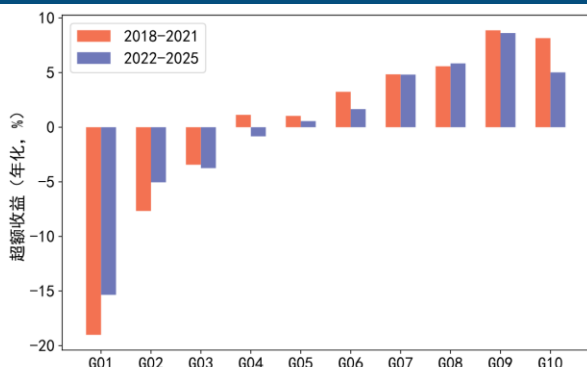
表 6: 时间重心偏离_TGD 因子的计算步骤

步骤	说明
第一步	将日内的分钟序列时间戳转换为序号，例如，9:31（1）、11:30（120）、15:00（240）
第二步	选取个股上涨分钟的涨幅序列 R_u 和下跌分钟的跌幅序列 R_d ，分别计算得到上涨时间重心 G_u 和 G_d ： $G_u = MR_u^T / \ R_u\ _1$ $G_d = MR_d^T / \ R_d\ _1$
第三步	“时间差”指标处理：将下跌时间重心回归上涨时间重心，取其残差项 ε $G_d = \alpha + \beta G_u + \varepsilon$
第五步	残差项 ε 为 TGD 因子的日频信号，例如，滚 20 个交易日计算残差项 ε 的均值即为月度选股因子

资料来源：中信建投证券

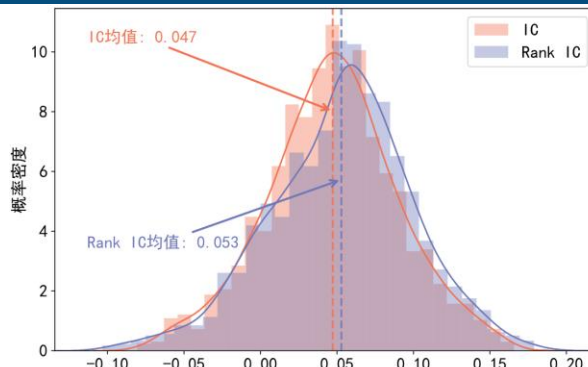
表 6 为 TGD 因子的计算步骤，图 26 至图 29 为 TGD 因子相关测试结果。我们测算了 TGD 因子在 2022 年前后的表现差异，结果发现整体多空收益衰减并不明显，分组的单调性依旧保持显著。这说明市场微观结构并未影响日内 A 股的定价逻辑，早盘阶段的价格涨跌信息依旧很重要：价格先涨后跌并且持续维持涨幅，具备这一特征的股票，未来的收益表现会更好。

图 26: TGD 因子的分组单调性



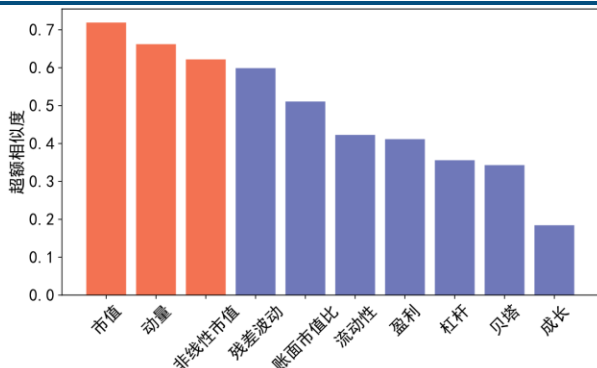
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 27: TGD 因子的 IC 分布图（测算 IC 和 Rank IC）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 28: TGD 与风格因子的超额相似度（暴露程度）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 29: TGD 单因子增强收益表现



数据来源：Wind、中信建投证券

TGD 因子在中证 1000 指数成分股范围内的有限表现相对一般，超额夏普为 1.52，最大回撤约为 7.9%。

表 7: 中证 1000 指数成分股优选月度超额收益表现 (STC 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	0.57	-0.39	1.72	3.42	-0.30	-0.02	-0.02	-1.59	-1.11	0.40	0.00		2.61
2024	2.67	-0.73	0.38	-0.27	2.06	0.36	0.38	0.48	0.72	-1.52	4.70	1.90	11.53
2023	1.27	1.35	0.46	-0.37	-0.24	1.95	1.62	0.58	0.53	-1.71	0.71	1.56	7.93
2022	2.48	-2.65	0.14	1.78	0.99	0.58	2.25	2.86	1.50	1.98	1.08	1.48	15.32
2021	1.99	2.47	1.74	-0.55	-2.22	0.58	-2.51	-0.53	3.03	1.49	4.20	4.51	14.85
2020	-0.49	-0.69	3.78	0.89	0.49	-1.08	1.67	3.25	1.06	-0.18	2.25	-1.49	9.69
2019	1.79	-1.69	0.94	0.95	0.22	1.53	2.83	-0.13	0.79	0.57	1.31	1.04	10.55
2018	1.00	0.97	0.70	0.60	1.77	2.00	2.08	0.84	0.61	0.44	0.26	1.35	13.35

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251105; 单位: %

表 7 和表 8 分别展示了短线交易拥挤度_STC 和时间重心偏离_TGD 两个因子在中证 1000 指数成分股中优选组合的月度超额收益率。其中, STC 因子在 2021 年和 2022 年表现均不错, 年化超额在 15% 左右; TGD 因子在经过中性化和标准化处后的表现整体不如 STC 因子, 但从过往较长时间观察结果来看, 在中证 1000 指数成分股中仍然能够优选出表现优于指数的组合, 且因子在 2022 年前后表现基本稳定, 仍然值得继续跟踪。

表 8: 中证 1000 指数成分股优选月度超额收益表现 (TGD 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	-0.99	0.97	0.52	-0.39	0.57	0.02	-0.74	-1.54	-0.39	1.76	1.12		0.87
2024	2.81	-0.93	-0.03	-0.46	1.79	0.20	1.70	-0.70	1.34	-0.20	1.29	0.33	7.30
2023	-0.37	1.66	0.29	1.13	0.42	0.90	3.43	0.32	1.69	-0.32	0.76	-0.31	9.96
2022	2.95	-0.79	3.00	-1.93	-1.38	1.01	1.79	3.66	-0.54	1.72	1.84	1.46	13.36
2021	-3.33	3.60	2.19	-0.71	0.70	3.25	-0.15	-0.50	3.18	-2.04	1.84	2.34	10.59
2020	-0.85	-0.85	3.27	-0.44	-1.99	-1.44	2.05	3.16	0.45	-0.48	0.99	-2.10	1.58
2019	-0.36	2.25	0.66	-0.29	1.77	-0.08	-0.27	-0.82	-0.01	-0.28	0.44	0.25	3.26
2018	-0.63	1.05	4.62	0.78	0.79	-0.12	1.95	0.62	0.87	0.71	2.14	1.10	14.70

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251105; 单位: %

本节聚焦高频时序变量在选股策略中的信息价值与作用机制。需明确的是, 日内不同交易时段的市场特征在截面选股上存在显著异质性, 这一现象的核心驱动因素在于机构与散户日内分布的不均衡。具体而言, 早盘市场流动性相对充裕, 而机构投资者往往偏好在流动性较好的时段完成交易, 因此形成了以机构为主导的交易结构; 尾盘时段受 A 股市场“T+1”交易机制约束, 散户与游资倾向于集中完成筹码兑换 (表现为尾盘成交量显著放大), 导致该时段的交易结构呈现散户化、短期化特征, 进而与早盘时段形成了明显的特征分化。

在因子体系构建中, 针对日内时段信息的差异化价值, 本文首先开发修正收益率偏度_CSKEW, 其在样本外测试中表现出优异的稳定性与超额收益捕获能力; 更具创新性的是, 本文提出的“时间重心”算子通过量化资金与价格变动在日内的时间分布集中度, 突破了传统高频因子的构造思路; 将其与分钟成交量、分钟涨跌幅数据深度融合, 分别构建了短线交易拥挤度_STC 与时间重心偏离_TGD 两大经典因子, 不仅实现了对信息的精准挖掘, 更在因子构造的创新性、表达式丰富度上形成突破, 完善了高频选股因子的覆盖维度。

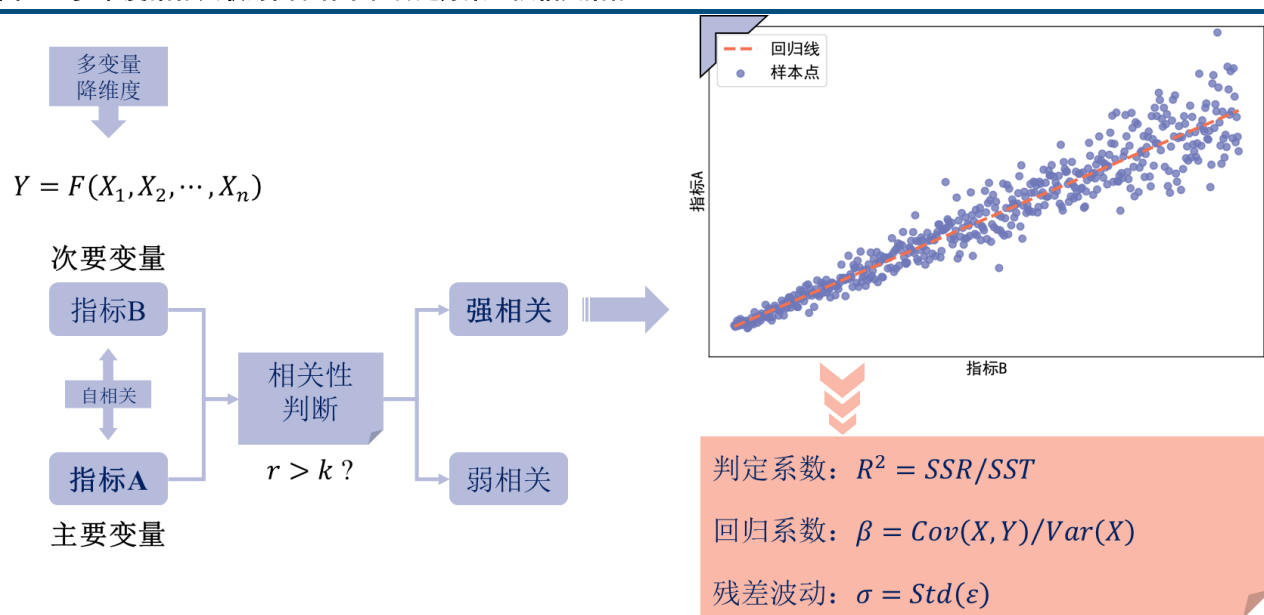
2.3 探究多指标关联的底层逻辑

规律并非总是通过单一指标呈现，多指标间的关联往往还能传递部分信息。单个指标只能反映事物的某个侧面，比如只看“销售额”无法判断盈利质量，只看“考试分数”不能体现学习能力的全貌。量价的 Alpha 具有隐蔽性，不会直接体现在某一个数据上，反而需要通过不同维度指标的搭配、对比才能显现较好的效果。

例如，价格与数量的协同变化可作为主力资金参与的参考信号，但这一信号并非绝对，仍存在一定不确定性。单一指标视角下，价格上涨或许只是散户跟风引发的短期波动，成交量放大也可能源于普通投资者的零散交易，两者单独出现时，均难以精准捕捉资金动向的核心逻辑；若价格呈现持续稳步攀升态势，且成交量同步温和放大，这种“价涨量增”的协同关联，能够传递出单一指标难以覆盖的关键信息：背后大概率存在主力资金持续入场吸筹、推动行情的行为，而非单纯由市场情绪主导的短期扰动。反之，若价格上涨但成交量持续萎缩，这一关联则可能暗示上涨缺乏足够资金支撑，后续行情回调的概率相对更高。

高频数据的特点在于数据丰富，我们利用好高频数据的优势，在以往的研究中，我们还覆盖对多个指标间关联性的分析，包括计算特征与特征之间的相关系数，基于两个以上的指标计算复合特征等等。我们融入统计学的方法构造了多维度指标的因子构建方法，如图 30 所示。

图 30: 多维度指标关联的因子体系化构建方案（强相关指标）



数据来源：中信建投证券

针对数据而言，笔者并没有进行刻意的筛选，而是选取自因子挖掘过程中较为常用的指标，其中既包括了基础的量价数据和资金流向数据，也包含其简单的扩展指标。该类数据无需额外特殊处理，部分数据服务商可直接提供，能够满足本研究尝试性分析与讨论的需求。

表 9 展示了上述指标的代码及简要计算公式。其中，资金流向数据的大单与小单分类依据我们以前测算的结论：单笔委托金额 2.5 万元，是划分大单与小单的较优阈值。此外，为了避免数据冗余（例如，大单净流入率等于小单净流出率），我们仅基于大单口径计算资金流入率和净流入额等指标。

表 9: 日内分钟指标梳理

类别	指标名称	指标代码	说明
基础量价	收盘价（价格）	PIC	时点指标，取开高低收任意一个即可
	成交量	VOL	该分钟所有交易的总成交量
	成交额	AMT	该分钟所有交易金额的加总值
	成交笔数	NUM	该分钟全部交易的笔数
衍生量价	涨跌幅	RET	分钟收盘价/上一分钟收盘价 - 1
	振幅	AMP	分钟最高价/分钟最低价 - 1
	VWAP	VWP	分钟成交额/分钟成交量
	非流动性	AHD	分钟涨跌幅绝对值/分钟成交额
	单笔成交金额	STA	分钟成交额/分钟成交笔数
资金流	大单/小单买入额	LBA/SBA	以 2.5 万元为划分标准，大于标准为大单，小于标准为小单 按分钟为频率统计成交量、成交额和成交笔数 衍生指标计算同理
	大单/小单主动买入额	LCB/SCB	
	大单/小单买入笔数	LBN/SBN	
	大单/小单卖出额	LSA/SSA	
	大单/小单主动卖出额	LCS/SCS	
	大单/小单卖出笔数	LSN/SSN	
资金流衍生	大单/小单单笔买入金额	LSB/SSB	买入额/买入笔数
	大单/小单单笔卖出金额	LSS/SSS	卖出额/卖出笔数
	大单/小单主动净流入额	LCA/SCA	大单/小单主动买入额-大单/小单主动卖出额
	大单/小单主动净流入率	LCR/SCR	大单/小单主动净流入额 / （大单/小单主动成交金额）
	净流入额	NIA	大单买入额-大单卖出额
	净流入率	NIR	大单净流入额/（大单买入额 + 大单卖出额）
	P.S. 大单与小单存在互补关系，为避免冗余此处仅考虑大单视角，净流入额与净流入率均基于大单口径计算		

资料来源：中信建投证券

价量相关指标呈现出显著的群聚特征，指标虽同属一个聚类集群，但各自表征的含义仍存在一定差异。如成交额与买入/卖出金额，图 31 的指标相关系数矩阵亦能验证。

相关性强弱是判断多维度指标关联因子有效性的标准之一。我们逐个交易日处理，选择任意两个指标的分钟时序数据，包含 9:25 集合竞价的时段，同时排除 14:58 与 14:59 的缺失数值（合并至 15:00），计算上述两组时间序列之间的 Pearson 相关系数，以量化其线性关联程度，将其作为因子的日频信号。依据全文的统一测试框架，我们滚动 20 个交易日计算日频信号的均值作为选股因子并在全市场内测算中性化后的 IC（取绝对值）。

图 31 展示了基于上述 31 个指标构造的 465 个“指标对”因子的 IC 与对应指标间相关系数分布情况，二者的相关系数达到 60% 以上，并存在明显的样本分层现象（虚线对应的横轴数值为 0.3），两个指标间相关性较低的“指标对”因子有效性整体偏低。从统计有效性的维度分析，这一现象的内在逻辑具有可解释性。相关系数本身即是判断两个指标线性关系的统计量（其二次方值为 OLS 回归中的 R 方），若两个指标并不存在线性相关性，也即回归模型不成立，基于相关假设的估计结果也将失去意义。此类数据挖掘方式，容易让我们忽视结果的逻辑合理性，进而陷入过度拟合历史样本的误区。

图 31：分钟指标相关性矩阵（资金流相关性较高）

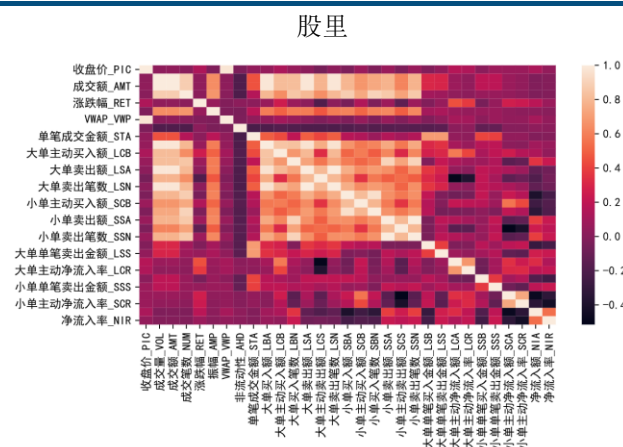
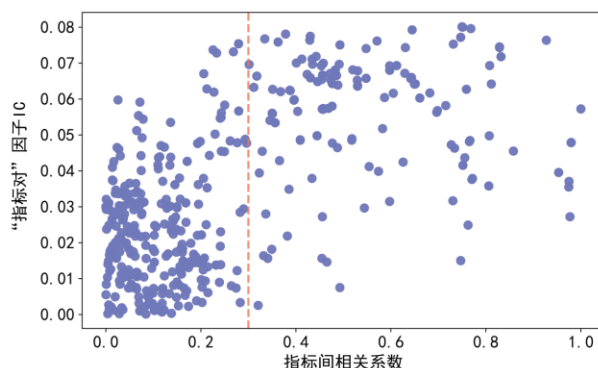


图 32：“指标对”因子 ICIR 与指标间相关系数正相关



数据来源：Wind、中信建投证券；以 20251103 数据为例

数据来源：Wind、中信建投证券；统计区间：20180102-20251103

基于上述分析，笔者选取部分有效性较强的因子（如 VOL_LBN 等）作为分析案例，并给出微观解释。

指标对：成交量（VOL）、大单买入笔数（LBN），IC 均值达到-0.08

微观解释：以机构为代表的大单资金是市场的核心力量，其交易行为直接带动成交笔数同步增加。以买入为例，若当前市场流动性充足，大单买入委托会快速与多笔小额卖出委托完成撮合。这一过程中会导致成交笔数与大额委托实际成交量的联动性显著增强，两者变化趋同的现象更为突出。成交额（AMT）与大单买入笔数（LBN）、大单买入额（LBA）与大单买入笔数（LBN）等等指标对同样适用此解释。

指标对：小单主动买入额（SCB）、小单主动卖出额（SCR），IC 均值为-0.073

微观解释：由于盘口存在价格博弈，主动买卖的订单数量在时序上通常是同增同减。例如，大单主动买卖金额的时序相关系数接近 0.21，而小单的主动买卖金额时序相关系数约为 0.27。这一现象在不同类型的股票中均有存在，本质上是买卖双方在争夺价格的主导权。然而，若是价格竞争过于激烈而短期无法决出强势方，价格则会小幅震荡或是维持不变（低波），侧面上看该正向相关的强度反映了股价的波动特征。

指标对：涨跌幅（RET）、振幅（AMP），IC 均值为-0.051，二者相关系数仅为 0.06（不显著）

微观解释：涨跌幅（Y轴）与振幅（X轴）本质上不具备显著线性相关性，二者散点的分布形态近似呈现侧倒 U 型，因为涨跌幅带方向而振幅则属于绝对值的范畴。影响二者线性相关系数测算结果的关键因素，在于日内上涨与下跌分钟的样本分布失衡。比如，日内上涨分钟（通常是上下影线比较短、实体较长的）样本偏多，涨跌幅与振幅相关系数会因样本“挤压”而增大；反之，若是下跌的分钟样本的数量占比更多，相关系数则会下降。这一相关系数间接统计了股票日内涨跌样本的分布结构，其核心反映的是市场上典型的反转效应。

尽管 RET_AMP 因子的实测表现尚可，但其属于典型的负面例子，因为相关系数并非刻画该类关系的直接且最优方法。对因子收益的理解需立足于更高维度，技术实现层面仅为次要因素。最终选股效果的优劣，则取决于具体的实现技巧，笔者并非反对将该指标直接作为特征变量，而是强调需先理解其背后所刻画的市场规律。

出于篇幅考虑，本文不再赘述其他“指标对”因子的表现，仅展示较为典型的主力交易强度_MTS 因子。

表 10: 主力交易强度_MTS 因子的计算步骤

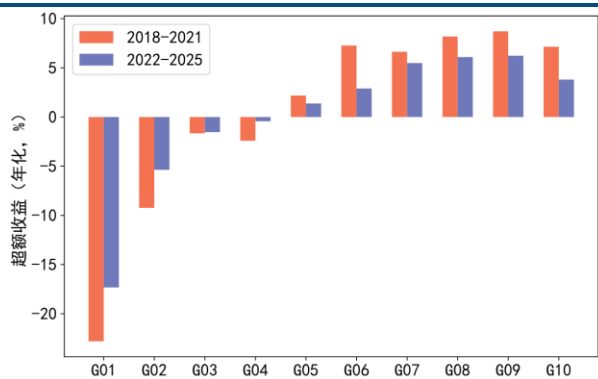
步骤	说明
第一步	选定股票，逐个交易日回顾分钟单笔成交金额序列，记为 A
第二步	计算单笔成交金额与成交额的相关系数，记为 TS 指标
第三步	TS 指标为 MTS 因子的日频信号，例如，滚 20 个交易日计算 TS 指标的均值即为月度选股因子

资料来源：中信建投证券

表 10 列示了主力交易强度_MTS 因子的具体计算步骤，图 33 至图 36 呈现其选股相关测试结果。该因子基于前文所述的单笔成交金额数据，结合分钟成交额指标构建而成，整体表现优良。

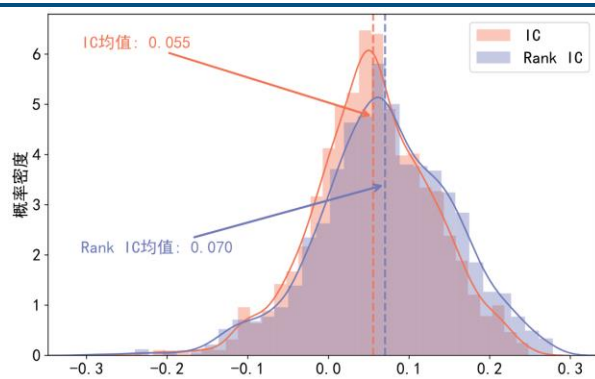
在全样本历史区间（20180102-20251105）内，MTS 因子的 IC 平均值为 0.06；全市场十分组测试中，其多空年化收益超 32%，信息比率达 2.6。在中证 1000 指数成分股优选测试中，其每年平均贡献约 9.8% 的超额收益率，超额夏普比率为 2.02。

图 33: 主力交易强度_MTS 因子的分组单调性



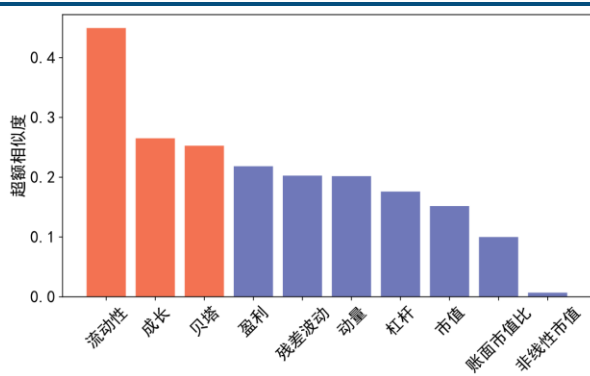
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 34: MTS 因子的 IC 均值为 0.06，Rank IC 均值为 0.07



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 35: MTS 因子与流动性的超额贡献重合度为 0.45



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 36: MTS 单因子增强收益表现



数据来源：Wind、中信建投证券

MTS 因子描述了主力资金交易影响成交量的变化的强度，与前文介绍的流动性规律基本一致，伴随着单笔成交金额增大的成交放量，说明是对手盘流动性不足；图 35 中因子超额贡献与流动性贡献重合度较高，从另一方面也能佐证我们对 MTS 因子 Alpha 收益来源的基本判断。

表 11: 中证 1000 指数成分股优选月度超额收益表现 (MTS 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	0.85	-1.15	2.00	2.52	0.43	-0.45	0.00	0.94	-0.93	1.28	-0.12		5.43
2024	-1.42	1.02	0.51	-0.30	2.73	-0.59	-1.10	-0.59	1.93	-3.24	1.32	2.69	2.81
2023	-0.05	1.14	-0.25	0.42	0.18	2.25	-0.10	-0.11	1.27	-1.31	0.63	0.80	4.94
2022	1.73	-1.76	-1.30	2.28	2.02	1.03	2.76	3.30	1.58	2.22	1.38	2.06	18.59
2021	2.42	4.13	-0.38	0.22	-1.70	0.58	-0.69	-0.71	1.14	1.15	5.10	1.68	13.51
2020	0.35	0.40	1.27	2.18	1.83	1.30	0.71	3.40	0.37	1.82	0.09	-1.09	13.30
2019	1.42	-2.77	0.84	0.41	0.33	2.08	2.31	0.23	1.21	0.28	1.65	0.68	8.92
2018	1.18	0.93	0.18	1.08	3.01	1.41	0.28	1.03	-0.10	1.10	-0.57	0.59	10.56

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251105; 单位: %

从表 11 的月度超额收益情况来看, MTS 因子在中证 1000 指数成分股内优选的表现不错, 每年都能累计正向收益, 尤其在 2022 年相对指数而言累计了近 20% 的收益安全垫, 值得继续关注。

我们强调了多指标的相关性, 但实际情况中依然会遇到两个或多个弱相关的指标, 其中隐含 Alpha 信息的情况, 例如, 前文提到的 RET 与 AMP 指标相关系数, 即便相关性偏弱, 但在截面上依旧有选股能力。对于这类弱相关的指标应该如何研究? 我们采取了与强相关情况不同的处理方式:

针对若干弱相关变量, 可采用聚类方法 (如 K-means 聚类), 分析不同簇间样本的**差异**特征。分析的核心思路是, 我们在 N 维空间标注每个样本的“位置”, 并且计算样本间的“距离”, 基于“簇内样本相似度高、簇间样本相似度低”的原则, 将样本划归为具有不同性质的若干簇, 并根据簇中心向量长度由大到小分类。

K-means 的优化过程: 给定 K 个初始簇中心, 通过计算样本与簇中心的距离, 将样本分配到最近的簇, 再更新簇中心为簇内样本的均值, 重复迭代直到簇中心稳定。簇中心的寻找过程与初始簇中心选择存在较强关联性, 因此我们固定每次的随机种子 (Seed=42), 以保障多次学习结果的稳定性。

K 值的数量可参考轮廓系数、CH 指数、DB 指数来确定, 计算表达式如下:

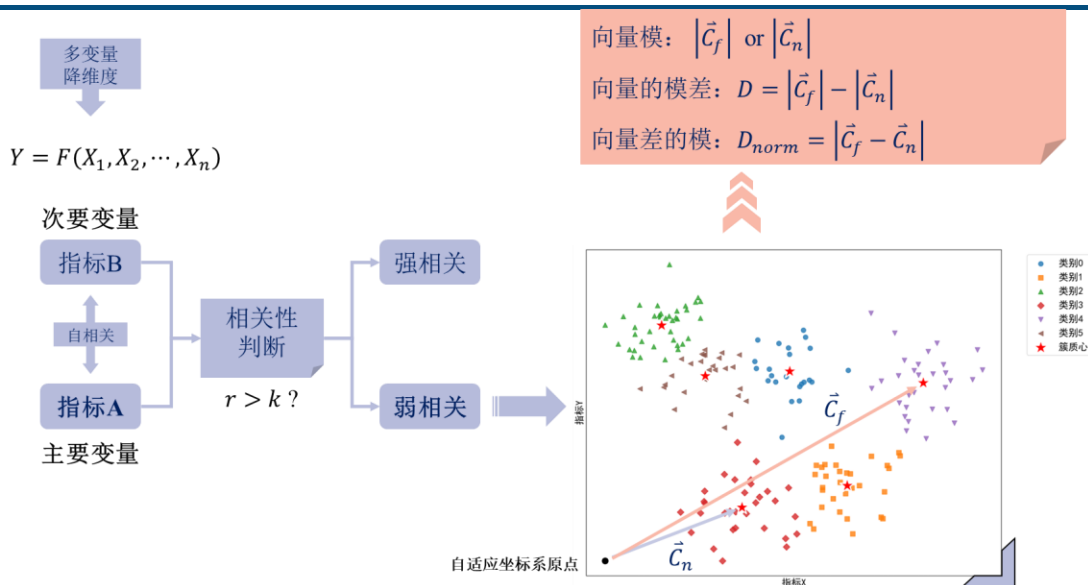
$$SC = \frac{1}{n} \sum_{i=1}^n \frac{b_i - a_i}{\max(a_i, b_i)}$$

$$CH(k) = \frac{B_k / (k-1)}{W_k / (n-k)}$$

$$DB(k) = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{\bar{a}_i + \bar{a}_j}{d_{ij}} \right)$$

其中, n 为样本数量, k 为簇的数量, $\bar{a}_i$ 表示簇内平均距离, d_{ij} 表示两个簇中心之间的距离。寻找上述几类指数的变化拐点可以确定针对样本最合适的 K 值。当然, K 值也可以人为指定, 如 K=10。

图 37: 多维度指标关联的因子体系化构建方案（弱相关指标）

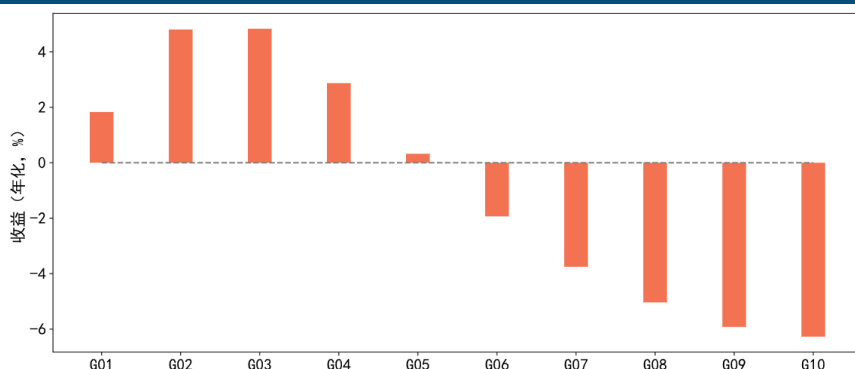


数据来源: 中信建投证券

首先，我们需要将指标 A 与指标 B 置于同一坐标系下，在对样本点进行聚类前，分别对两个指标做去量纲处理得到可比的样本点坐标值（X，Y）；然后，我们再根据 K-means 聚类方法将样本点分为若干个群类，从而得到 N 个簇中心，分别计算得到以自适应坐标系原点（ X_{min}, Y_{min} ）到上述簇中心的向量长度；最后，再按照向量长度由大到小的顺序对簇中心进行排序分类，其中距离原点最近的簇内样本为远样本，距离原点远的簇内样本为近样本。我们的目标在于分析这两类样本之间是否存在选股信息的差异。

不妨做个简单假设：X（主要变量）与 Y（次要变量）完全不相干。若如此，上述计算模型还可以进一步简化处理，样本维数将缩减至一维，聚类问题也简化为对主要指标分组并讨论有效性差异。图 38 展示了基于 AMP 进行样本聚类并分组计算 RET 求和的因子测试效果，在低波（AMP 近样本）呈现微弱的动量，而在高波（AMP 远样本）则表现出较强的反转特征。

图 38: 基于 AMP 样本分类结果：低波动量、高波反转



数据来源: 中信建投证券

这与开发过的一类因子方式类似，例如，我们曾以分钟单笔成交金额作为工具变量，对分钟涨跌幅样本进行“切割”，从中提取出更为有效的反转效应因子（强反转_SR）。

请务必阅读正文之后的免责条款和声明。

表 12: 强反转_SR 因子的计算步骤

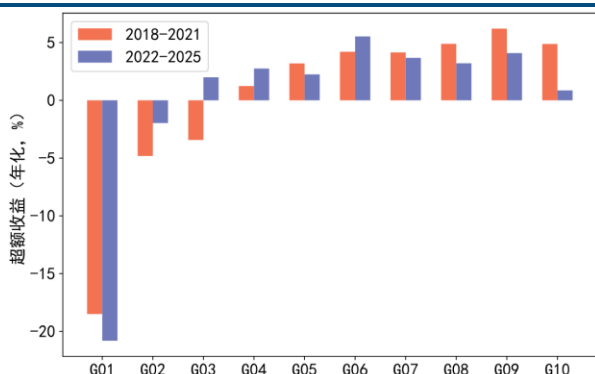
步骤	说明
第一步	选取日内每分钟的涨跌幅和单笔成交金额序列
第二步	将分钟涨跌幅序列按照对应单笔成交金额大小进行排序
第三步	选取单笔成交金额较大的 30% 样本, 求涨跌幅的加总值, 记为 R 指标
第四步	R 指标为 SR 因子的日频信号, 例如, 滚 20 个交易日计算 TS 指标的均值即为月度选股因子

资料来源: 中信建投证券

表 12 列示了强反转_SR 因子的具体计算步骤, 图 39 至图 42 呈现其选股相关测试结果。该因子基于前文所述的单笔成交金额数据, 结合分钟成交额指标构建而成, 整体表现优良。

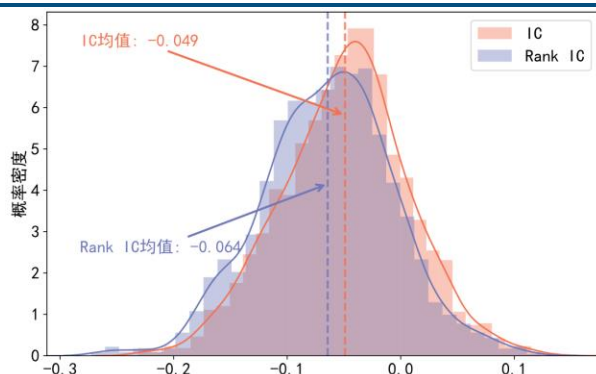
在全样本历史区间 (20180102-20251105) 内, SR 因子的 IC 平均值为-0.05; 全市场十分组测试中, 其多空年化收益超 32%, 信息比率达 2.6。

图 39: SR 因子的分组单调性



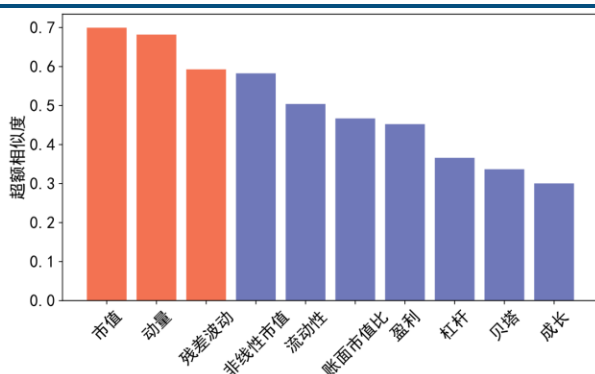
数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 40: SR 因子的 IC 分布图 (测算 IC 和 Rank IC)



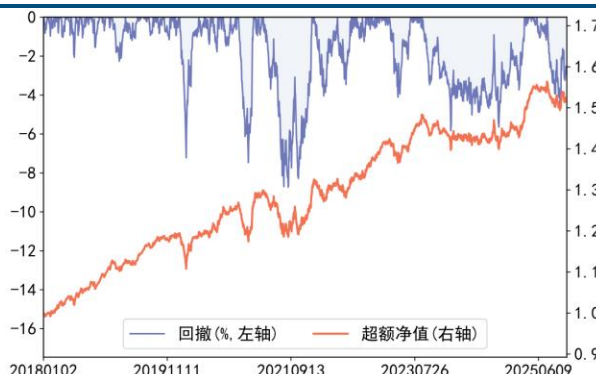
数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 41: SR 与风格因子的超额相似度 (暴露程度)



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 42: SR 单因子增强收益表现



数据来源: Wind、中信建投证券

表 13: 中证 1000 指数成分股优选月度超额收益表现 (SR 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	0.05	0.10	3.04	3.23	0.70	-0.62	-0.64	-2.10	-0.19	1.37	-0.01		4.91
2024	-1.61	0.43	0.00	0.24	0.09	-0.62	0.82	-1.15	2.43	-3.60	2.68	0.67	0.22
2023	1.22	0.87	-0.67	-1.96	1.91	0.55	3.09	1.90	-1.90	-0.74	1.15	-1.02	4.34
2022	4.87	-2.39	0.28	1.11	1.39	-1.10	-0.36	2.86	-0.30	0.01	2.43	1.25	10.30
2021	-1.02	6.07	0.98	0.19	-2.69	0.72	-5.26	-0.57	3.54	-3.06	2.30	3.10	3.81
2020	-1.61	-2.70	3.80	1.14	-0.41	-0.72	1.55	2.65	0.65	0.63	0.85	-3.99	1.58
2019	0.07	0.51	1.25	-0.77	0.85	1.95	1.69	-0.03	1.28	-0.37	0.49	0.52	7.67
2018	-0.18	0.47	1.80	0.04	1.59	0.44	1.60	0.05	1.06	0.25	1.63	1.48	10.70

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251105; 单位: %

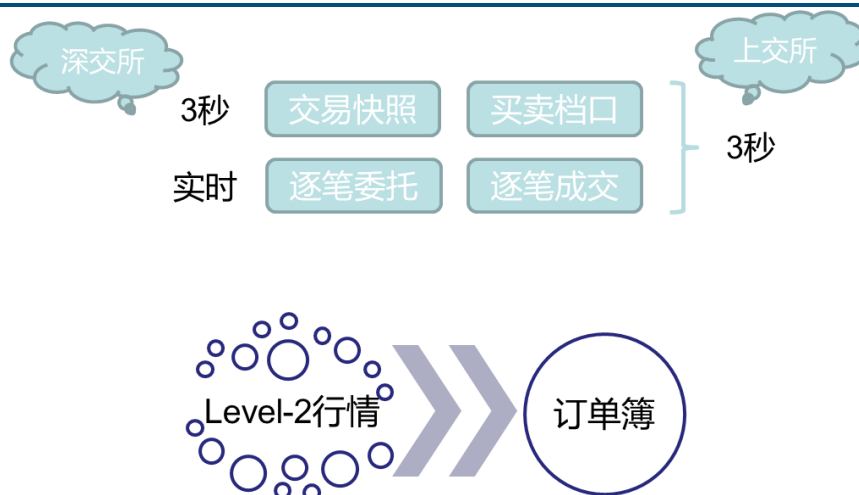
强反转因子在 2024 年表现不如人意 (全年超额仍然为正), 作为多指标研究方法的具体应用, 我们仍会跟踪该因子后续的表现; 作为高频指标构造的特征, 该因子在模型信息量上或能起到补充。

三、下篇：基于 Level2 数据的多元视角探究微观世界

市场微观结构是无法用肉眼（盯盘）观测的一方小世界。每笔订单就如同构成基本物质的分子与原子，不同类型的交易行为共同组成了 Alpha 模型中的各类量价规律和特征。从低频选股到高频研究，我们希望尽可能梳理出细致的规律，把握住因子生效的底层逻辑，这对于理解市场同样也会有帮助。

实际应用的 Level 2 数据包括逐笔成交、逐笔委托和 Tick 行情。与日频或分钟频的行情数据不同，逐笔成交和逐笔委托数据包含了每笔交易明细，通常能够更为直接地用以分析投资者的交易意图；Tick 行情则是对当前订单簿的“快照”，结构上也更加接近 Level 1 的行情数据，其中包含最新成交价和累计成交量等。因其描述每个时刻全部待成交的订单的量价信息，我们可以借此进一步观察交易的细节和筹码动态。对于日间选股策略开发而言，我们并不需要过度关注数据本身的“高频”属性，而应在逻辑层面深入理解并挖掘有效因子，利用高频数据丰富的结构特征，捕捉日频乃至分钟频无法捕捉到的交易细节与动态。

图 43: Level2 数据基本结构



数据来源：中信建投证券

Level2 数据的应用方向十分广阔。逐笔成交和逐笔委托数据既可用以还原得到实时订单簿，生成任意时刻的盘口数据，还可对任意笔交易委托在时序上汇总，识别出订单流的完整生命周期。这两类数据的无论是数据量还是信息量都更为丰富，也是我们订单流系列研究中的重点。在过往研究中，我们已尝试多种此类数据的开发方法，本节将结合三个典型的主题进行介绍，分别是：（1）探究订单背后的真实交易意图；（2）关于投资者性质的猜想；（3）交易算法演变与逆向策略面临的挑战。

3.1 探究真实的交易意图

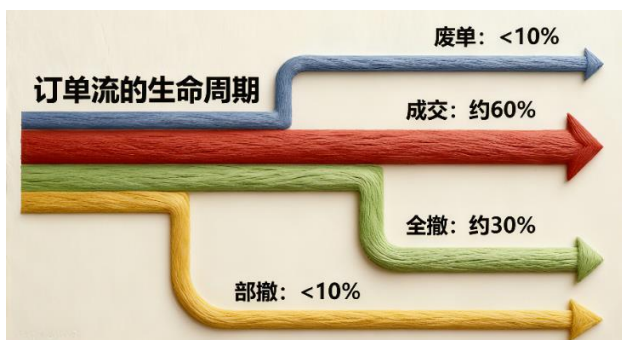
订单流是市场中所有买卖订单的实时动态集合，包括订单的价格、买卖方向、交易数量和执行状态，不仅是股票价格变动的根本驱动力，更是投资者交易意图的直接反映。不同于依赖价量的传统分析，订单流分析聚焦委托和交易动态，可直接透视机构投资者的隐蔽操作。提交给交易所的委托订单，仅有两种结果：要么全部撮合成交，要么被撤销。其中成交订单占绝大多数，撤单情况则更为复杂，主要分为三类：

- （1）全部撤回：投资者在交易过程中反悔，主动撤回全部未成交委托，这类订单约占全部委托的 30%；

(2) 部分撤回：订单部分撮合成交后，投资者撤回剩余未成交部分，这类情况占比接近 10%；

(3) 休市废弃：因交易时间结束（多为临近收盘时的委托），部分未成交订单会随休市自动失效。

图 44: 提交成功的委托订单有四种可能的退出方式



数据来源：中信建投证券

图 45: 从订单流中探寻投资者交易意图的蛛丝马迹

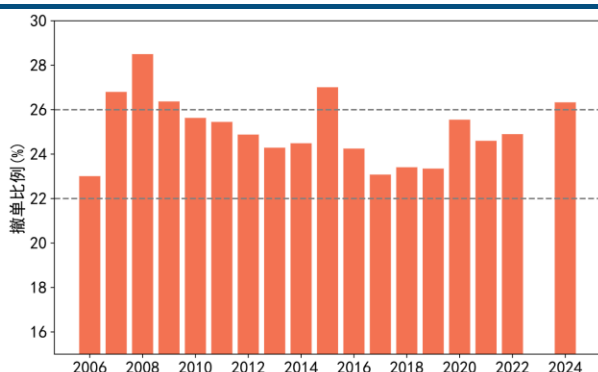
时间	交易操作	BS标志	价格	数量	卖方序号	买方序号
93014280	Order		11.8	12600		NaN
93014280	Deal		11.8	162		B1
93014280	Deal		11.8	1000		B2
93014280	Deal	S	11.8	300	896710	B3
93014280	Deal		11.8	100		B4
93014280	Deal		11.8	2000		B5
93014520	Cancel		NaN	9038		NaN

数据来源：Wind、中信建投证券

图 45 记录了某只股票在开盘后短时间内一笔特殊的委托挂单，这是一笔大额卖出委托，对应的订单 ID 是 896710，属于部分撤回的类型。该委托在 9 点 30 分 14.28 秒提交给交易所，因其价格在当前盘口附近，很快与其他订单撮合成交。但在 9 点 30 分 14.52 秒的时候，这笔委托剩余的部分就被迅速撤回，中间间隔时间甚至不超过 1 秒，实际成交的比例仅为 28.7%。从 A 股的订单流数据里，我们能够轻松筛选出此类委托，因此，我们自然会提出一个疑问：是否这些撤回的订单可直接或间接反映投资者的交易意图，并且能够作为 Alpha 的信息来源被构造成为有效的因子策略？围绕这一核心命题，本文展开了系统性的理论分析与探讨。

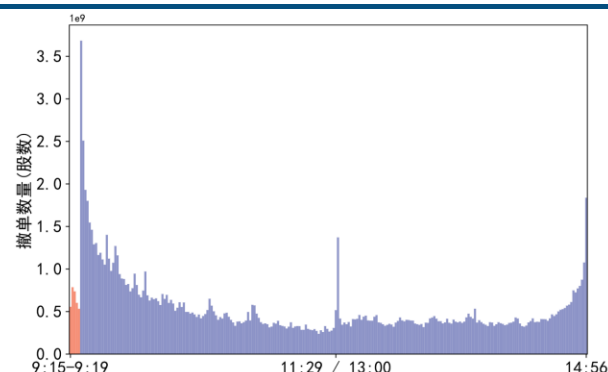
首先，需要明确的一点是，成交与撤回是孪生关系，股票交易过程中必定会伴随产生撤回订单。上交所每年披露的《市场运行质量报告》数据显示，撤单比例长期保持在大约 24% 的水平（见图 46）；日内交易不同时间段的撤单数量分布，与成交量的“U 型”分布基本一致。核心差异在于，9:15 至 9:19 时段允许撤回委托，而收盘集合竞价阶段仅存在休市后自动废弃的订单，不存在投资者主动撤回的情况。

图 46: 撤单比例长期保持稳定水平



数据来源：上交所、中信建投证券；按笔数计，2023 年数据暂缺

图 47: 撤单与成交量的时间分布规律基本一致

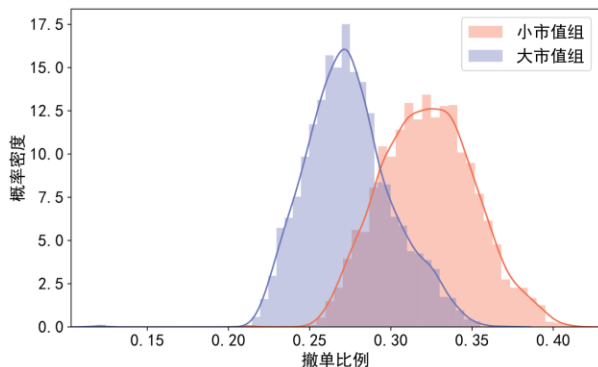


数据来源：Wind、中信建投证券，以 20251103 的数据为例

其次，撤单比例在不同风格维度的股票间呈现显著异质性，包括市值、价格涨跌等核心分类维度。在分析投资者交易意图的过程中，我们需重点留意不同属性股票的撤单规律是否存在一致性。例如，大市值的股票流

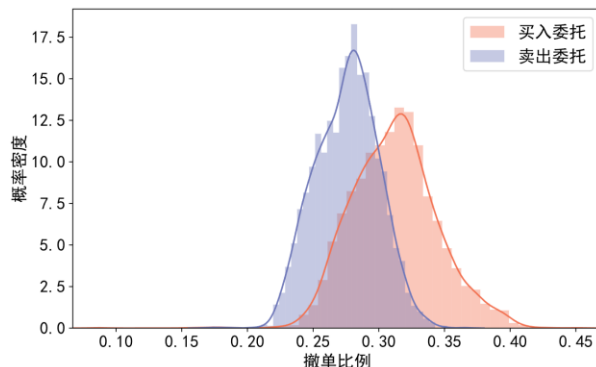
动性充足，出于正常交易目的的撤回委托相对较少，大部分委托都能够及时被撮合成交（见图 48）。

图 48: 小盘股相比大盘股流动性偏差，撤单比例相对较高



数据来源: Wind、中信建投证券; 统计区间: 20250103-20251103

图 49: 买卖委托的撤单比例随股价涨跌幅度变化



数据来源: Wind、中信建投证券; 此处统计剔除了涨跌停股票样本

买卖行为存在诸多不对称现象，例如买单数量多于卖单，且买单金额相对更小。造成这类现象的原因既包括交易规则限制，也涉及投资者决策的人为因素。从市场规则来看，A 股实行 T+1 交易制度，买卖的交易限制不对等，当天买入的股票无法立即卖出，进而导致可供卖出的股票数量在筹码流通过程中逐渐减少；从投资者的行为模式来看，散户在买入股票时倾向于降低成本，而在卖出股票时，则更多会考虑如何高效且快速地将变现筹码。一方面这会使得买入委托金额偏向小，另一方面价格敏感还会催生大量的买入撤单（见图 49）。

借助行情数据，我们仅能获取订单流中已成交部分的相关信息，而撤回订单则可以提供补充信息。成交的订单促成价格和交易量的变化，撤回的订单中往往隐含投资者的特殊交易意图，例如，以涨停价挂单的卖出委托、以跌停价格挂单的买入委托（类似于购买“彩票”等待兑现），抑或在盘口附近委托的高频撤回订单（实则提供有毒的流动性），或是包含在部分撤回中的大额委托（用于打压或者支撑目标价位）等等，我们对其进行数据切片，观察不同类型的撤回委托是否具备特殊的选股能力。

以部分撤回委托为例。这类委托的订单数量相对较少，占全部委托总笔数不足 10%，其中少数是以“最优五档成交剩余撤回”为规则的市价单，大部分仍然是以限价的形式，并基于主动策略撤回的委托。如表 14 所展示的数据，沪市委托的订单大部分是限价委托，按笔数或按股数统计均超过 99%。

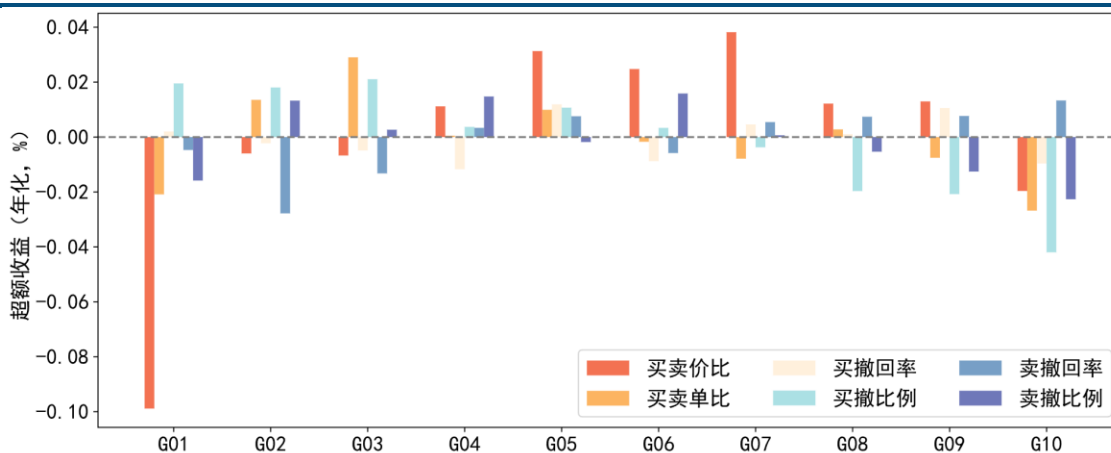
表 14: 订单类型在主要板块中应用的分布

板块	按笔数统计			按股数统计		
	限价订单	市价委托 剩余撤销	市价委托 剩余转限价	限价订单	市价委托 剩余撤销	市价委托 剩余转限价
上证 50	9.88	0.03	0.00	10.89	0.06	0.00
上证 180	25.28	0.08	0.00	26.33	0.15	0.00
上证 380	21.50	0.08	0.00	14.33	0.11	0.00
沪市主板	99.54	0.46	0.01	99.18	0.81	0.01

资料来源: 上交所、中信建投证券; 单位: %

我们测算部分撤回委托的挂单价格及量比的相关指标，从结果上来看并无特别之处（见图 50）。

图 50: 大额部分撤回委托的因子构造效果一般



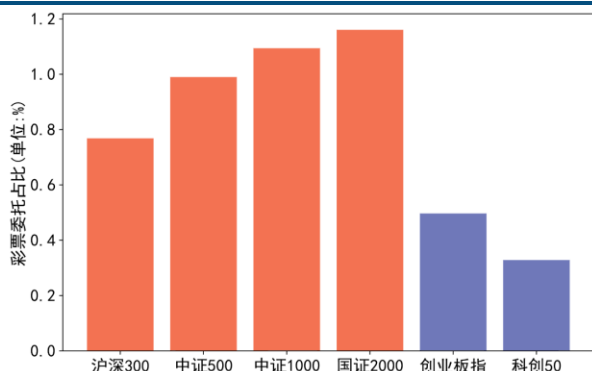
数据来源: Wind、中信建投证券

上述因子表现不佳的原因在于, 大额委托的有效性在于成交部分而非撤回, 真实交易的筹码具备可被定量的成本, 且后续的资金跟随效应可能持续验证前置交易; 大额部分撤回委托实则多为无逻辑、无固定规则的数据片段, 数据噪音较高。这类订单数量较少、来源杂乱, 且撤回后可重新委托, 其委托价格通常集中于盘口(即买一价或卖一价), 这表明此类委托多以实际交易为目的。因此其在因子层面的表现并不显著, 本文不再单独基于这类委托构造因子表达式, 也不进一步讨论其有效性。

极端价格委托等待成交, 本质上与买彩票无异, 都在赌运气而非理性投资。我们基于逐笔委托数据, 可以观察到一类非常有意思的挂单行为: 部分投资者选择以涨停价卖出股票或者以跌停价买入股票。这部分委托由于不会干扰当前价格, 并不会触发“价格笼子”(报价幅度不超过 3%)的限制条件, 依然能够正常申报; 此类投资者相当于以极低成本买入短时看跌期权(涨停价卖出)或是看涨期权(跌停价买入)。由于订单成交概率极低, 加之在价格变化至目标价之前均可撤销, 笔者将其称之为“彩票委托”。

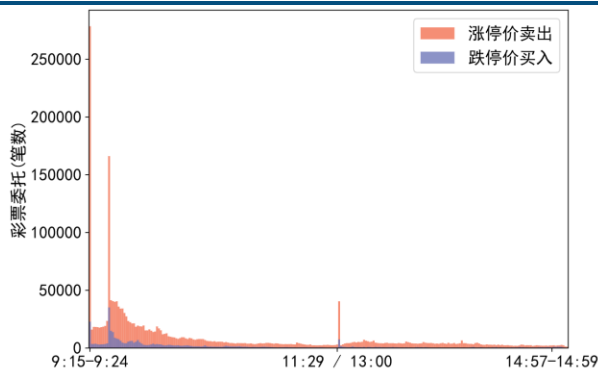
在同等涨跌幅限制下, 小市值股票的彩票委托数量更多; 而大市值股票及涨跌幅限制宽松的标的, 彩票委托占比明显更低。这是因为大市值股票流动性充足、市值规模大, 股价触板阻力更强, 而涨跌幅宽松的标的极端波动稀缺, 对散户的吸引力不足。

图 51: 核心指数成分股范围内彩票委托的占比



数据来源: Wind、中信建投证券; 按笔数统计

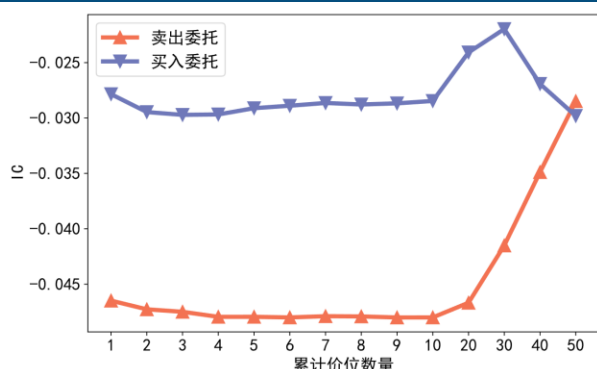
图 52: 集合竞价阶段“买彩票”行为较为活跃



数据来源: Wind、中信建投证券

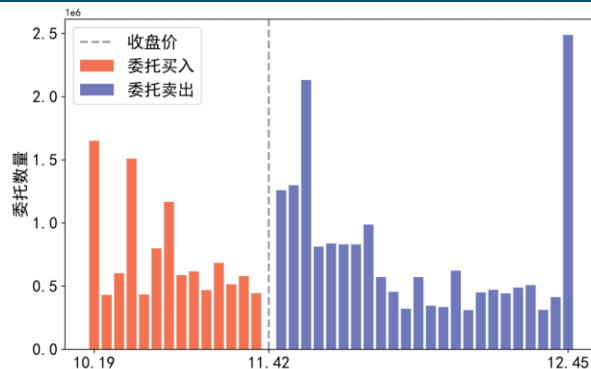
以简单的占比来统计跌停价买单与涨停价卖单的区别，构造两类指标：一是以当日涨停价卖单或是以跌停价买单的主动撤单比例；二是测算买卖的彩票委托各自占当天所有同向委托的比例。前者反映了市场上此类交易的主观意愿强弱，后者则是判断当前市场的流动性。从测试结果来看，图 53 的变化曲线与我们先前讨论的结论基本一致：相较于有“无限供应”属性的现金，当天仅能支撑一倍换手率的股票，才是实际可交易的真实筹码。如果我们进一步把彩票委托的定义放宽，选取低于涨停价或高于跌停价的卖出或买入委托，分别测试指标在全市场的选股效果，如图 53 所示。

图 53: 不同档位价格的差异（区分买单、卖单）



数据来源：Wind、中信建投证券

图 54: 废单的价格的分布情况（15:00 时刻订单簿快照）



数据来源：Wind、中信建投证券；以 20251103 的 000001.SZ 为例

结合图 53 的测试结果，我们最后将彩票委托_LOTTERY 因子的数据观察窗口限制在委以涨停价的卖出委托而非买入委托，主要基于逻辑与数据的两点考虑：一是相比买单，卖单作流动性的刻画更为准确；二是偏好彩票委托的投资者目的就是追寻涨停价，放宽极端价格的定义对于因子并不会提供增益，反而可能会增加了因子模型过拟合的风险，得不偿失。

废单为尾盘交易行为提供了独特的刻画维度。废单是一类蕴含投资者特殊交易意图的委托订单，其产生的原因是交易所收盘时间截止后，未完成撮合的委托订单被系统终止，因此 15:00 是观察该类订单的核心时间窗口（见图 54）。废单的规模（订单数量与金额）可以有效反映市场中追逐收盘价的流动性诉求，为刻画尾盘资金行为提供重要参考，为此，本文构造基于废单刻画了流动性指标，记为收盘参与度_PTO 因子。

根据一般理解，废单主要来源于两类典型场景：一是部分投资者为锁定成交概率或博弈极端行情，提前较长时间挂出极端价格委托（如以涨停价卖出或以跌停价买入），因其委托价格明显偏离合理区间而全天均未能匹配成交；二是临近收盘（如 14:57 后集合竞价阶段或连续竞价尾声），当日剩余交易时间极短，即便处于盘口档位的委托，也可能因未及时匹配而沦为废单。

表 15: 彩票委托_LOTTERY 因子的计算步骤

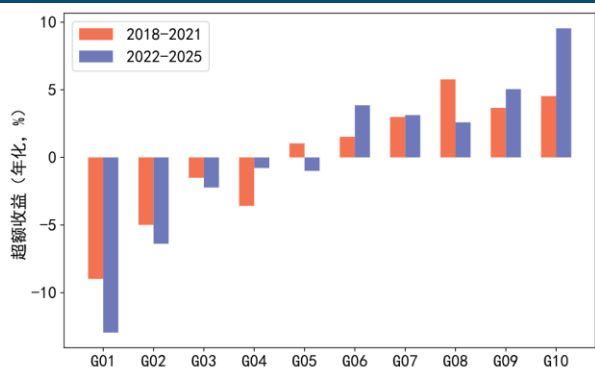
步骤	说明
第一步	逐个股票处理，基于逐笔委托统计当日所有价格的委托和撤回数量
第二步	选择以涨停价委托的卖出订单，统计其主动撤回的数量及总量，将二者的比值记为 L 指标
第四步	L 指标为 LOTTERY 因子的日频信号，例如，滚 20 个交易日计算 L 指标的均值即为月度选股因子

资料来源：中信建投证券

测试彩票委托_LOTTERY 因子的表现，如图 55 至图 58 所示。

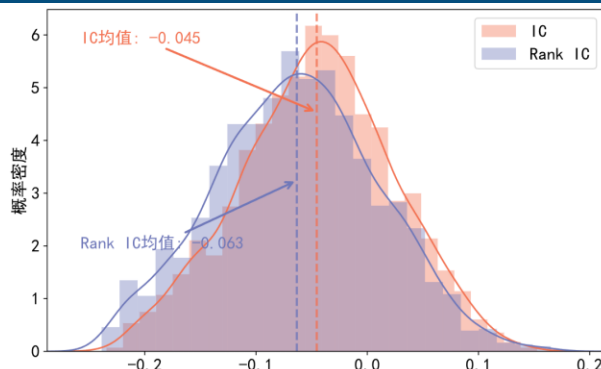
在全样本历史区间（20180102-20251105）内，LOTTERY 因子的 IC 平均值为 0.06；全市场十分组测试中，其多空年化收益超 20%，信息比率达 2.04。

图 55: LOTTERY 因子的分组单调性



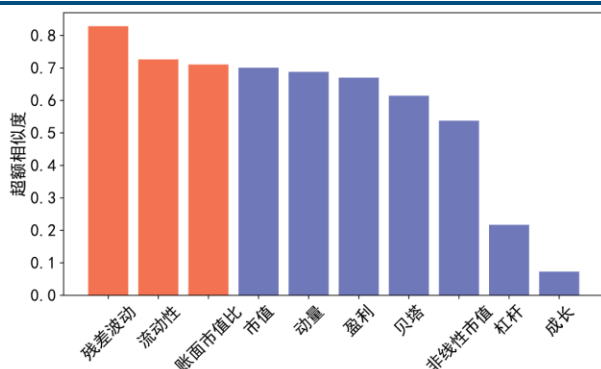
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 56: LOTTERY 因子的 IC 分布图



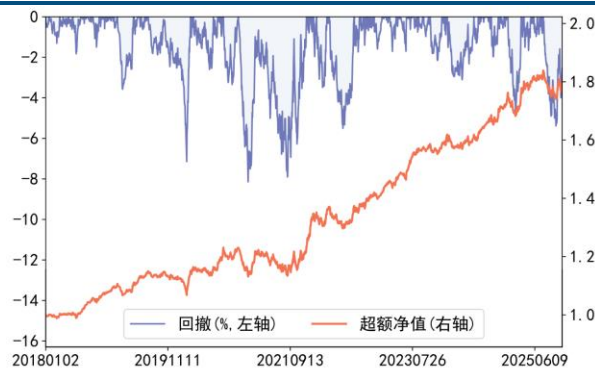
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 57: LOTTERY 与风格因子的超额相似度（暴露程度）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 58: LOTTERY 单因子增强收益表现



数据来源：Wind、中信建投证券

表 16: 收盘参与度_PTO 因子的计算步骤

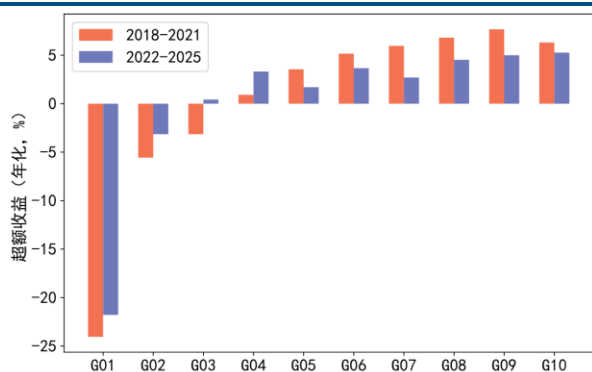
步骤	说明
第一步	基于逐笔委托和逐笔成交还原连续变动的订单簿
第二步	选择 15:00 收盘时刻的切片, 统计所有未成交的委托数量, 并除以当日自由流通股本得到比值, 记为 P 指标
第四步	P 指标为 PTO 因子的日频信号, 例如, 滚 20 个交易日计算 P 指标的均值即为月度选股因子

资料来源: 中信建投证券

测试收盘参与度_PTO 因子的表现, 如图 59 至图 62 所示。

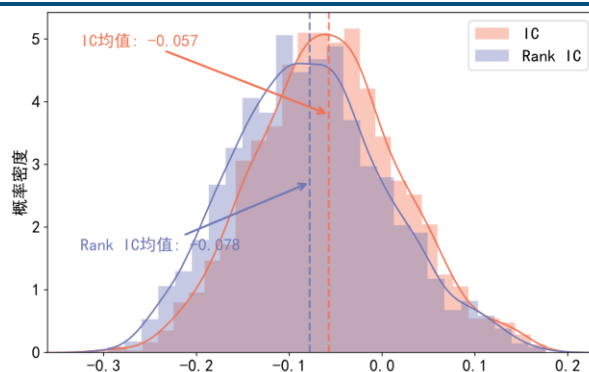
在全样本历史区间 (20180102-20251105) 内, PTO 因子的 IC 平均值为 0.06; 全市场十分组测试中, 其多空年化收益超 32%, 信息比率达 2.6。

图 59: PTO 因子的分组单调性



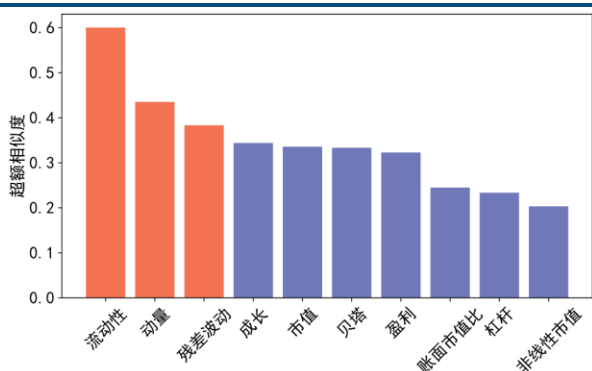
数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 60: PTO 因子的 IC 分布图



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 61: PTO 与风格因子的超额相似度 (暴露程度)



数据来源: Wind、中信建投证券; 统计区间: 20180102-20251105

图 62: PTO 单因子增强收益表现



数据来源: Wind、中信建投证券

表 17: 中证 1000 指数成分股优选分月度超额收益表现 (LOTTERY 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	-0.36	-1.56	3.14	1.05	2.09	-0.54	-0.11	-2.90	-0.57	1.82	1.01		2.96
2024	3.10	-1.82	-0.76	-0.28	2.54	-0.58	1.30	1.53	1.11	-0.34	3.10	1.57	10.82
2023	-0.46	2.48	0.56	1.02	1.17	0.82	3.98	0.70	0.09	0.79	-0.70	0.13	11.02
2022	4.77	-1.55	1.51	0.96	-0.09	-2.11	0.21	3.85	0.33	0.97	1.41	0.79	11.39
2021	-1.58	2.72	2.42	-0.55	-2.77	0.51	-2.00	-1.30	3.72	-1.84	3.83	5.67	8.74
2020	-1.33	-2.07	6.45	-0.99	0.26	-1.88	1.93	2.64	1.18	-0.43	2.82	-4.97	3.17
2019	2.14	-2.18	0.25	1.11	2.06	1.01	1.21	-0.66	0.33	-0.46	-0.24	-0.09	4.49
2018	-0.14	-0.55	0.76	0.16	-0.07	-0.05	2.26	1.83	0.53	1.54	1.05	0.77	8.35

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251103

表 18: 中证 1000 指数成分股优选分月度超额收益表现 (PTO 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	0.34	-1.52	1.97	2.14	0.75	-1.39	-0.26	-1.32	-0.65	1.57	0.58		2.13
2024	-0.98	-0.47	-1.04	0.36	2.38	-0.12	0.46	0.32	0.74	-3.32	3.36	2.52	4.10
2023	1.10	0.72	-1.46	-0.61	-0.77	1.43	1.91	0.53	-0.36	-1.07	1.11	0.40	2.92
2022	1.92	-3.44	-2.00	1.28	1.31	1.14	2.02	2.32	1.87	0.40	1.28	2.49	10.92
2021	2.44	3.32	2.10	1.36	-2.55	-0.03	-5.29	-1.50	1.90	1.03	3.45	3.38	9.58
2020	-0.20	-0.01	2.37	1.97	2.60	-0.54	1.27	3.38	0.91	0.36	1.80	-1.63	12.84
2019	1.15	-0.98	0.63	0.49	0.89	1.95	3.61	0.79	1.63	0.49	0.61	-0.61	11.09
2018	-0.08	0.09	1.96	1.41	0.98	0.61	2.37	1.28	0.15	0.85	1.48	0.79	12.52

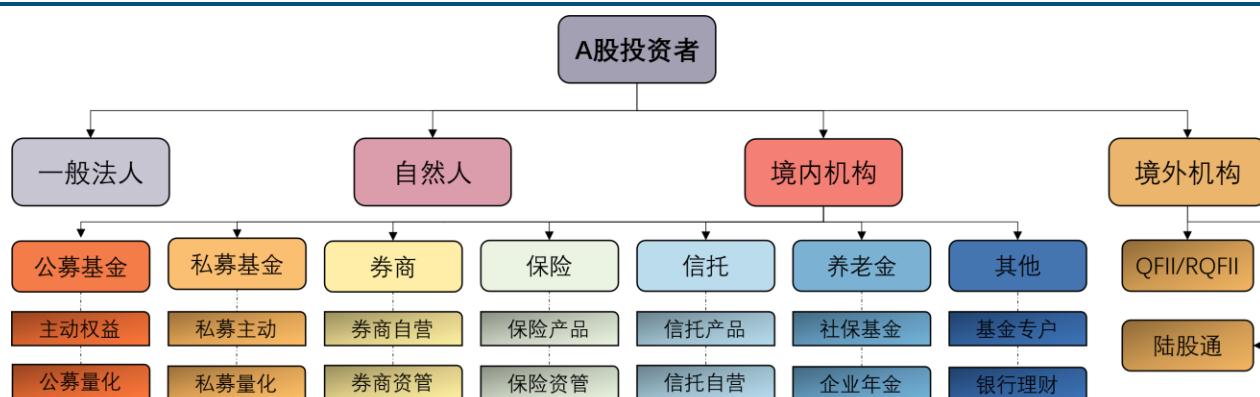
资料来源: Wind、中信建投证券; 测算区间: 20180102-20251103

3.2 关于投资者性质的猜想

在股票市场中，不同类型投资者的行为特征与决策逻辑存在显著差异。清晰认知这些投资者类型，不仅是把握市场交易动态的关键，更能为投资策略制定提供重要参考。

传统意义上，A股参与者包括一般法人、自然人和专业机构三类，其中，自然人和专业机构贡献了大部分交易量。自然人包含了公司个人持股部分和一般散户投资者，而专业机构按照属地不同可以分为境内机构和境外机构两种类型，前者参考Wind分类可分为以下六类：公募基金、私募基金、证券公司、保险、养老金和信托机构等，而境外投资者通常是指QFII、RQFII和陆股通等持牌投资者。

图 63: A 股的投资者结构示意图



数据来源：中信建投证券

从投资规模与专业维度划分，股市投资者又可划分为机构投资者、个人投资者（散户）两大类，此外还存在一类特殊的游资群体。

机构投资者凭借雄厚的资金实力、专业的研究团队及丰富的实战经验，是市场的重要影响力量。公募基金、私募基金、证券公司及保险等均属于此类，它们具备成规模的研究分析和交易能力，其投资决策往往基于系统性分析与长期布局，对市场走向具有较强的引导作用。

与之相对的是个人投资者（散户）。散户的资金规模普遍较小，多以个人积蓄参与投资，且普遍缺乏专业的投资知识与系统分析能力，信息获取渠道也相对狭窄。这导致散户的投资行为易受市场情绪左右：市场上涨时盲目跟风买入，市场下跌时又恐慌性抛售，决策过程往往带有主观性与随意性，缺乏明确的投资规划。

表 19: 投资者类型性质

投资者类型	资金规模	交易能力	信息获取	投资决策特点
机构投资者	雄厚	有专业化交易系统	渠道广泛，获取及时	相对理性
散户	整体偏小	交易页面内手动操作	信息来源有限	容易受市场涨跌影响

资料来源：中信建投证券

除上述两类之外，游资是股市中极具特殊性的投资群体。其核心目标是短期获利，擅长捕捉市场热点与题材进行快速炒作，具有资金流动性极强、操作手法灵活多变的特点。游资的介入往往能在短期内推动股价大幅

波动，投资风格激进，在追求高收益的同时，也承担着极高的市场风险。

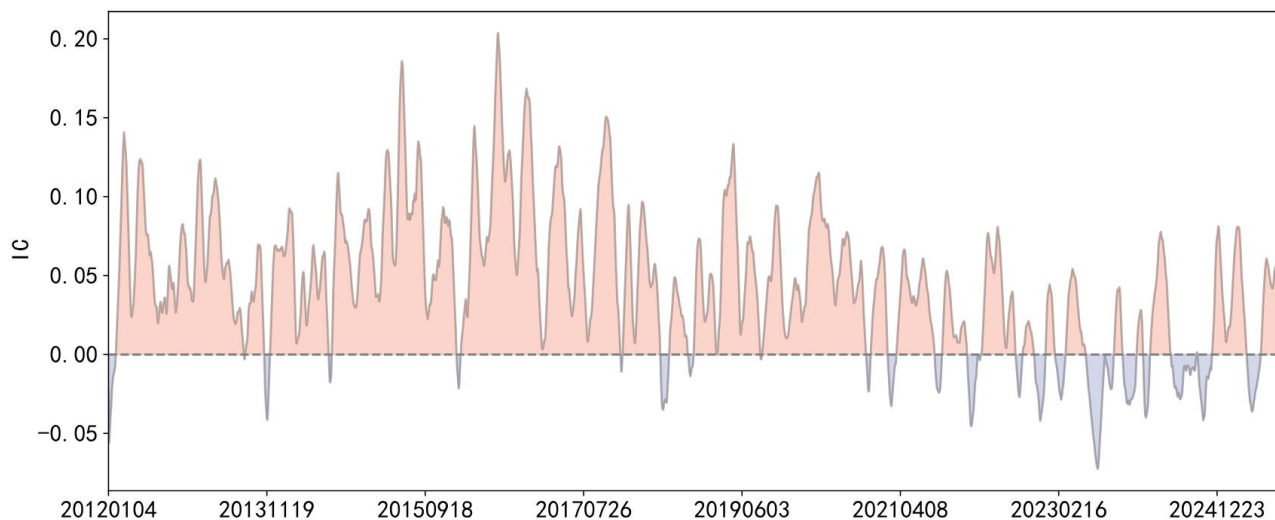
理解不同类型投资者的特点和行为模式，有助于投资者更好地认识市场。在 A 股投资者类型的界定与分类研究领域，我们一直在深入探索与挖掘。目前基于 Level2 数据，我们可以识别出部分投资者的类型，判断依据主要源自两个方面的抓手：资金规模与交易能力。

依据 1：投资者的资金实力越强，属于机构的可能性越大。

资金流从历史长期来看，属于非常有效的多头 Alpha 来源：以机构为代表的主力资金在股票市场中能够影响价格，甚至能够决定股价的中长期走势。若能够跟踪市场上主力投资动向，我们则可能获取超越市场平均水平的投资回报率。笔者曾基于 Wind 的资金流向开发了系列因子，核心逻辑在于精确构造主力（大单）资金流入或流出股票程度的衡量指标，其中，样本内最优的是广义主力净流入率_CNIR 因子（其计算方式见附录）。

资金流在长期被证实是有效的 Alpha 来源，但在近几年的收益表现明显不如以往，其背后原因来源于机构投资者交易算法的演变。图 64 展示了广义主力净流入率_CNIR 因子自 2012 年以来的 IC 变化趋势，比较突出的区间有 2014 年至 2017 年的市场活跃期，以及 2019 至 2021 年主动管理业绩较好的区间；在 2021 年以来，因子的 IC 水平出现明显下降，尽管在 2024 年底以来的上涨行情催动下，胜率依旧不高。

图 64: 广义主力资金净流入率_CNIR 因子 2012 年以来的 IC 变化情况



,数据来源: Wind、中信建投证券

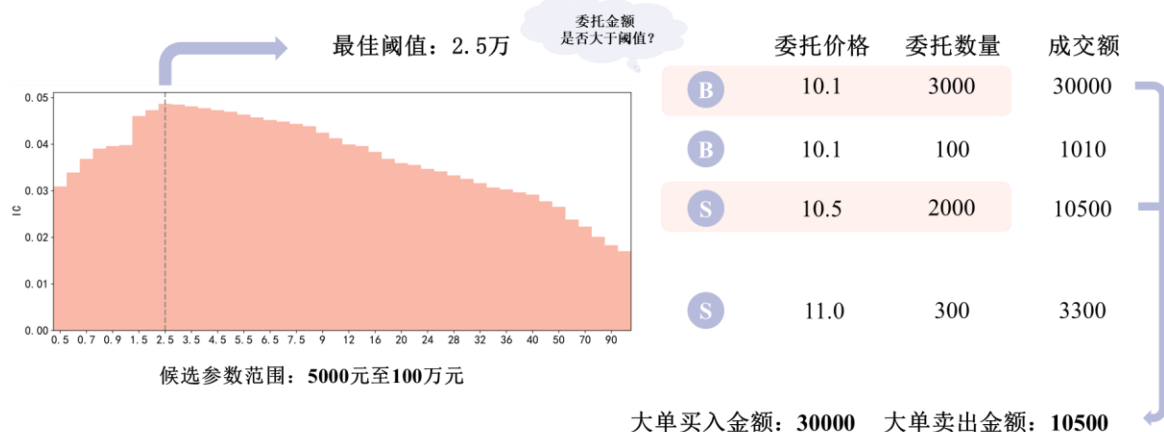
资金流 Alpha 衰退的原因，主要是伴随交易的“拆单”算法不断普及和演化，机构类大单的识别准确度正在逐渐降低。为此，我们曾尝试过对资金流的划分标准进行修正，分别使用绝对阈值法和相对比例法，确定资金流中的大单交易动向。以绝对阈值法为例，我们遵循“由低到高、先密后疏”的原则，人为设定了一系列划分阈值，具体涵盖 0.5 万元、0.6 万元至 100 万元，以此作为区分机构大单与散户小单的核心界限。为探究区分大小单的阈值标准，我们选取形式较为简单的净流入率 NIR 作为分析载体：

$$NIR = (B - S) / (B + S)$$

基于不同参数划分的资金流数据，我们分别计算大单买入金额 B 和大单卖出金额 S，进一步计算 20 日的净

流入率作为选股因子，测算各参数下因子的 IC 指标作为评判阈值有效性的依据。

图 65: 依据不同参数下因子测试效果来确定划分大单的最佳阈值

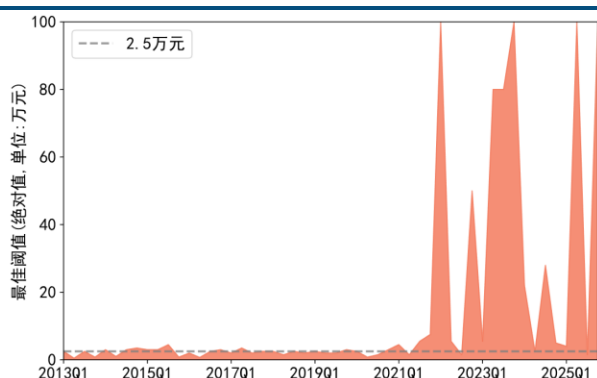


数据来源: Wind、中信建投证券; 统计区间: 20120102-20251107

我们在一段时间内统计 NIR 因子的 IC 均值，取 IC 最高的参数作为区间内的最佳阈值，跟踪“最优值”在历年以来不同区间内的变化情况。图 66 显示了最佳阈值的变化情况，该指标在 2021 年以前基本围绕在 2.5 万元的标准上下调整，但在 2021 年以来则出现明显变化，最佳阈值的稳定性大幅降低。

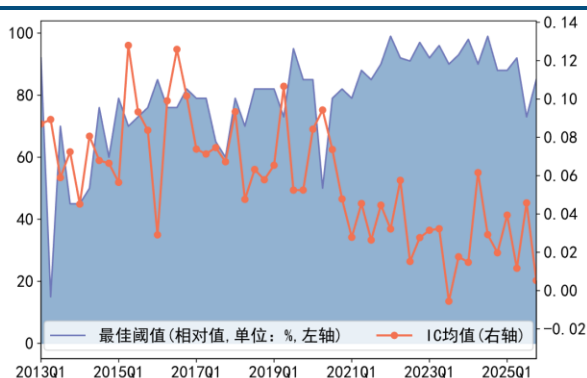
类似地，在相对阈值的方法下，我们进行了类似的讨论。两者结论基本一致：并非是绝对值下移，但有几年的方法出现了明显的分组混乱的情况。前后结果的差异为何如此显著，其核心原因究竟是什么？

图 66: 2021 年以来，绝对阈值方法下标准值大幅度漂移



数据来源: Wind、中信建投证券

图 67: 相对阈值方法的结论与绝对阈值方法基本一致



数据来源: Wind、中信建投证券

结合表 20 与表 21 的统计数据，我们在历史观测区间内发现一个显著现象：2022 年、2023 年及 2024 年的订单结构中，小额委托的数量占比呈现持续偏高的特征。这一结构性变化直接导致大单与小单的理论划分边界趋于模糊，使得基于传统金额阈值的订单分类效果显著弱化。需要明确的是，该现象的成因并非源于拆单颗粒度进一步细化导致的识别失效，而是由于市场中机构拆单行为的发生比例显著提升，且拆分后的委托金额大多集中于 2 万至 10 万元这一传统分类标准下的模糊区间。此类集中的拆单操作，使得原本用于区分机构与散户交易行为的订单金额边界失去了明确的辨识度，最终造成机构与散户订单分类的有效性下降。

上述测试结果，都指向一个关键点：市场上订单金额的分布在 2021 年以后发生了明显漂移。

表 20: 买入订单委托金额分位数统计结果

分位数	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
50%	1.05	1.19	1.37	1.18	1.12	0.86	0.77	0.83	0.76	0.62	0.65	0.63	0.77
80%	3.4	3.92	4.52	3.76	3.56	2.83	2.63	2.79	2.45	2	2.07	2.08	2.52
90%	6.78	7.93	9.6	7.86	7.31	5.66	5.29	5.61	4.86	3.99	3.95	3.95	4.87
95%	11.91	14.2	18.14	14.81	13.55	10.23	9.62	10.23	8.78	7.28	7	7.03	8.78
99%	35.19	42.74	64.16	48.85	42.78	31.07	30.07	32.14	27.76	23.26	22.6	22.95	28.81

资料来源：中信建投证券；单位：万元；2025 年数据统计截至 11 月 3 日

表 21: 卖出订单委托金额分位数统计结果

分位数	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
50%	1.18	1.34	1.58	1.36	1.29	1.01	0.86	0.91	0.83	0.64	0.67	0.64	0.8
80%	3.94	4.56	5.34	4.41	4.14	3.36	2.99	3.15	2.77	2.15	2.23	2.2	2.72
90%	7.8	9.13	11.22	8.98	8.24	6.66	5.92	6.25	5.5	4.28	4.31	4.2	5.24
95%	13.65	16.16	20.79	16.31	14.75	11.84	10.59	11.24	9.84	7.7	7.59	7.33	9.28
99%	38.87	47.4	68.67	50.9	45.09	34.66	32.17	34.59	30.57	24.23	23.65	22.68	29.61

资料来源：中信建投证券；单位：万元；2025 年数据统计截至 11 月 3 日

金额分布出现漂移，我们判断核心驱动因素有二：其一，机构采用算法交易模式处理委托，将原本的大额订单拆分为多笔小额订单，在单笔委托绝对金额下降的同时，小单总量实现大幅增长；其二，低延时、高频化的日内交易策略对价差波动较为敏感，这类策略通常不会发起大额委托，这一特征也使得 2022 年至 2024 年期间，小额委托数量呈现异常增多的态势。上述两大因素共同导致统计结果出现失真，最终影响了我们所构造资金流因子的选股效果，使其未能达到预期水平。

表 22: 不同订单规模组下投资者群体订单占比

区间	投资者类型	账户规模	按笔数统计			按金额统计		
			小单	中单	大单	小单	中单	大单
全样本	小散	小于 300 万	73%	66%	38%	77%	63%	31%
	大散	大于 300 万	5%	23%	49%	8%	26%	55%
	机构	-	22%	11%	13%	15%	11%	14%
2018 年前	小散	小于 300 万	81%	66%	35%	82%	63%	28%
	大散	大于 300 万	6%	25%	53%	8%	28%	58%
	机构	-	13%	9%	12%	10%	9%	14%
2018 年后	小散	小于 300 万	58%	65%	42%	69%	63%	37%
	大散	大于 300 万	4%	19%	44%	6%	22%	49%
	机构	-	38%	16%	14%	25%	15%	14%

资料来源：《Tracking Retail and Institutional Investors Activity in China》、上交所、中信建投证券；统计区间：201401-202005

然而，另外一种观点认为，资金流本质上是主力资金的交易行为，相关指标表现较弱，原因在其市场主体

请务必阅读正文之后的免责条款和声明。

本身寻找超额收益的能力在不断减弱，也即存在“牛熊切换”。本文认为这一观点不失为一种理解方式，但从数据角度观察，这一论点似乎又站不住脚。结合前面图 66 展示的数据，资金流的 IC 在 2013 年至 2021 年期间始终都能保持较高的水平，这说明资金流背后的逻辑（跟随效应）具备穿越牛熊的能力。市场上的机构投资者选股能力（如，超额收益的中位数）出现波动或许能在某些时段与资金流因子的 IC 变化趋势吻合，但也只是策略相互博弈得到的结果，而非是内在原因。市场微观结构改变使得资金流 Alpha 减弱，反其道而行会遇到不少阻力。可见，隐藏与识别是未来市场微观结构讨论的一个重要课题。

依据 2：散户可接触的交易条件较为有限，高频交易指令大概率源自机构。

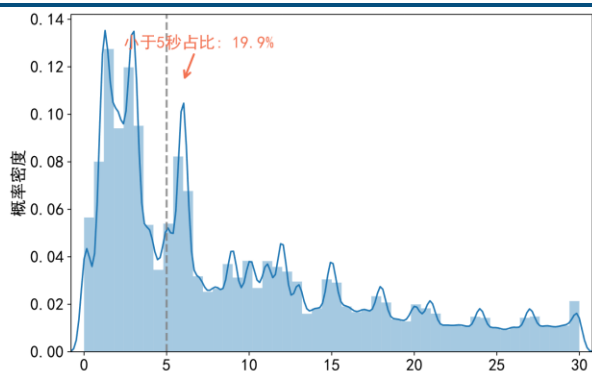
我们在上一章节中讨论了订单中的撤单行为，分析了其中部分撤回委托、废单等。然而，撤单类型中主动撤回委托的情况比较复杂，其中亦不乏属于投资者有目的地撤回订单。这类主动撤单并非随机行为，而是大量投资者基于特定交易目标的策略性操作：例如，T0 策略中，投资者为捕捉日内价格波动，常会在买入后快速卖出委托（持有底仓交易），若短期未达成预期成交价格便会主动撤销委托以避免持仓风险或释放资金用于下一笔交易；又如，冰山委托等算法交易中，程序化交易算法会根据市场冲击成本等指标动态调整委托单，当预设条件触发时，自动执行撤单操作以优化交易效率。

无论是 T0 套利，还是算法交易，这类撤单行为都具备同一个特征：高速。这种高速性不仅体现在订单提交与撤销的毫秒级响应效率，更依赖于低延迟交易通道、近距离托管服务等专业基础设施的支撑。对于一般的散户投资者而言，由于缺乏接入这类专业交易通道的资源与资质，其交易指令发出及执行的速度往往处于显著劣势。因此，在订单流中表现出的高频、高速撤单行为，通常源自具备技术与资源优势的机构投资者，这也成为区分散户与机构交易行为的重要微观特征之一。我们对**毒流动性**进行定义：

真实流动性与毒流动性，本质上是基于投资者视角界定的一组相对概念。真实流动性指的是订单簿中价格优于投资者目标价且能实际成交的对手单，而毒流动性则是指流动性中较难匹配委托的部分订单。

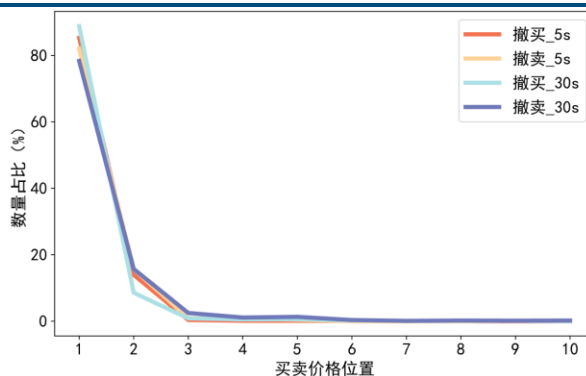
结合实际经验我们知道，根据对手盘口委托的订单并非最终都能成交，委托与成交的差额部分可能是真实流动性，也有可能是毒流动性，区别在于对手单“拒绝”当前交易的事实是否为主观策略所致，核心要素是交易对手方寻租意愿的强弱：寻租的过程会频繁出现挂单后快速撤单的现象，导致盘口的结构出现快速变化，基于盘口结构委托的优先限价单成交概率随之降低。在股票市场中，部分“活跃”订单实则反向制造流动性不足的市场矛盾，此类订单由此被定义为毒流动性。

图 68：时长在较短（5 秒以内）的撤回订单较多



数据来源：Wind、中信建投证券

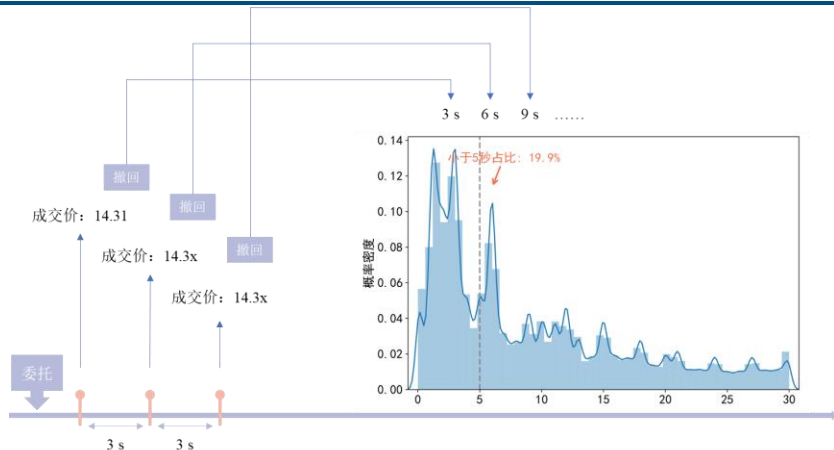
图 69：高频订单通常集中在盘口附近，快速流转



数据来源：Wind、中信建投证券

撤单的时间长度有些特殊的数据，来源于我们交易所的数据报送方式，每 3 秒报送一次数据，于是在观测之前可以报送一次数据，然后再在内部观测。

图 70: 挂单时间分布规律来源于交易所的 3 秒一次的报价规则



数据来源：中信建投证券

关于高频撤单的来源，是否是因为某种交易规则导致快速的撤回？表 23 中展示了上交所和深交所可接收的几种委托类型，其中，市价剩余转撤回的委托（FAK）是指，投资者以市场当前对手方最优价格提交委托，主动与对手方挂单进行成交匹配；若当前对手盘挂单量不足以全额满足该委托订单的申报数量，剩余未成交部分将被系统自动撤回，不再参与后续市场撮合，此时，发生部分交易的同时会产生剩余部分的撤回委托。

表 23: 沪深交易所接受的市价委托类型

交易所	委托类型	交易规则
上交所	最优五档即时成交剩余撤销申报	以对手方价格为成交价，与申报时对手方 最优五个价位 的申报队列依次成交，未成交部分自动撤销
	最优五档即时成交剩余转限价申报	与最优五档成交规则相同，但未成交部分自动 转为限价委托 ，限价为最后成交价
	对手方最优价格申报	以申报进入交易主机时， 对手方队列 的最优价格（即卖一价或买一价）为成交价
	本方最优价格申报	以申报进入交易主机时， 本方队列 的最优价格（即卖一价或买一价）为成交价
深交所	即时成交剩余撤销申报 (FAK)	以对手方价格为成交价，与申报时对手方 所有 申报队列依次成交，未成交部分自动撤销
	全额成交或撤销申报 (FOK)	要求全部成交，否则立即自动撤销
	最优五档即时成交剩余撤销申报	与上交所一致
	对手方最优价格申报	与上交所一致
	本方最优价格申报	与上交所一致

资料来源：上交所、深交所、中信建投证券

尽管特定类型的市价委托会产生瞬时撤单，但高频撤单的核心驱动因素并非此类订单。以上交所 2024 年公布的市场统计数据为例，限价订单在全市场委托结构中占据绝对主导地位：从申报笔数维度统计，限价订单占比仅为 0.5%；从申报股数维度观测，其占比为 0.8% 上。上述两组统计数据与此前提出的猜想并不相符，可见，高频交易更多采用的是“限价+主动撤回”的方式，间接产生了数量庞大的高频撤单。

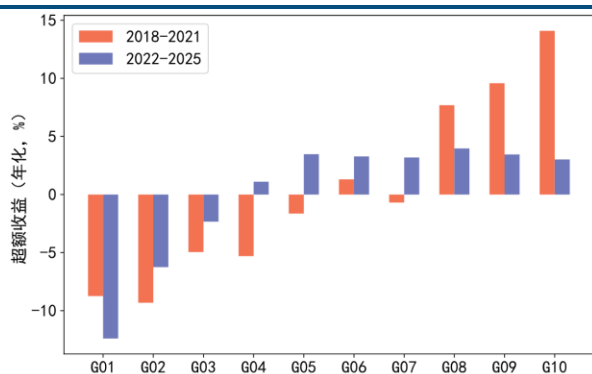
表 24: 毒流动性_TOX 因子的计算步骤

步骤	说明
第一步	逐笔订单统计自下单时起至主动撤回的间隔时长
第二步	分别筛选时长短于 5 秒以及时长短于 30 秒的订单，分别汇总全天的委托数量，记为 O_5 和 O_{30}
第三步	计算两个标准下委托数量 O_5 和 O_{30} 的比值，记为 T 指标
第四步	T 指标为 TOX 因子的日频信号，例如，滚 20 个交易日计算 T 指标的均值即为月度选股因子

资料来源：中信建投证券

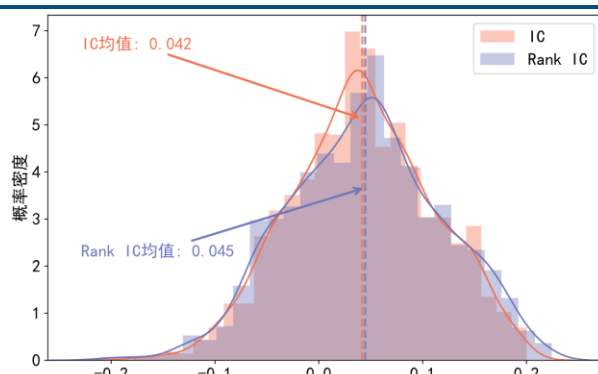
本文考虑将高频撤单的范围定义在 5 秒以内，但这并不一定是最有效的方式。高频交易的挂单主要集中在盘口（一档或者二档的价格），我们可以采取更为准确的定义方式，如统计盘口总撤单量作为比值的分母，从而减少样本中可能包含的非盘口撤单的干扰，表 24 的因子构造步骤仅作为抛砖引玉，供读者参考。

图 71: TOX 因子的分组单调性



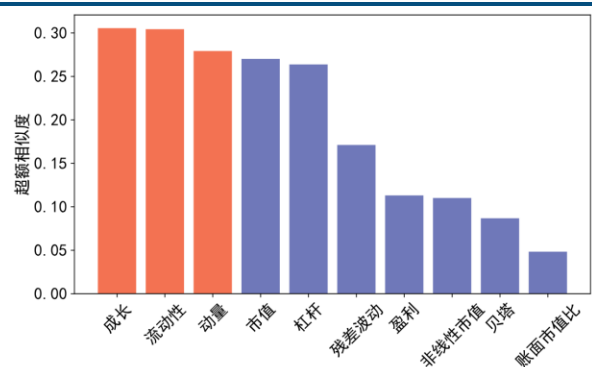
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 72: TOX 因子的 IC 分布图（测算 IC 和 Rank IC）



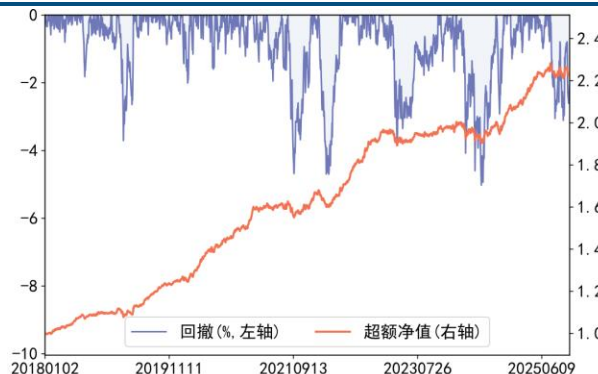
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 73: TOX 与风格因子的超额相似度（暴露程度）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 74: TOX 单因子增强收益表现



数据来源：Wind、中信建投证券

毒流动性在策略中的风格暴露相对偏弱。整体上是因为近两年，市场微观结构出现了明显的干扰。测试结果分析。T0 交易的市场容量逐渐达到饱和，市场交易拥挤度较高，过往 T0 策略的高收益逐渐降低，此外，行业规范性也在逐渐增强，监管约束使得部分低效扰动的高频交易逐渐出清，整体的微观结构受到多维度的复杂变化的影响，整体稳定性不如早些时候。

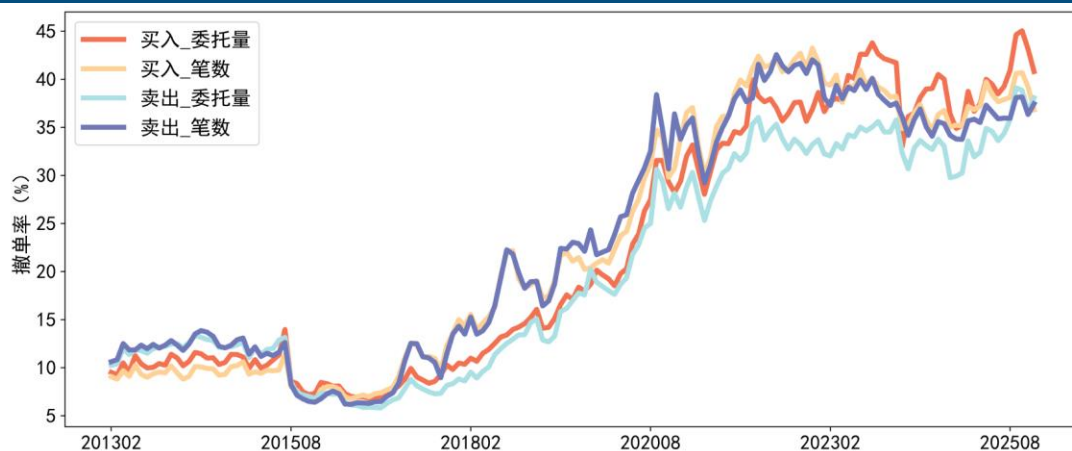
表 25: 中证 1000 指数成分股优选月度超额收益表现 (TOX 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	1.21	1.16	1.49	2.77	1.09	0.41	-0.08	0.19	-1.31	0.59	0.33		8.07
2024	0.38	0.23	-0.56	-1.15	0.49	-1.65	1.11	1.09	1.13	-1.55	3.85	1.04	4.38
2023	1.42	0.21	-2.13	-0.33	-0.50	0.61	1.16	0.68	-0.47	-0.03	1.64	0.24	2.47
2022	2.63	-2.38	-1.91	2.00	2.25	1.82	3.66	2.98	0.87	2.66	1.43	1.00	18.18
2021	3.38	2.00	-0.04	1.47	-1.66	1.34	-0.96	-0.93	-0.10	-0.19	1.97	2.10	8.57
2020	0.18	0.83	2.99	3.39	2.86	1.85	0.37	1.84	1.03	1.67	0.44	1.46	20.56
2019	1.25	-1.23	0.31	1.03	0.92	2.38	2.44	1.19	2.61	0.20	0.50	0.99	13.29
2018	-0.24	2.18	0.95	0.18	1.61	2.41	0.85	0.62	-0.09	0.73	0.09	0.67	10.38

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251105; 单位: %

高频交易受到市场的监管, 短期观测到的特征易发生变化。2024 年初, 监管部门要求量化交易者提前报告账户信息、资金来源及杠杆率等核心信息, 且需“先报告、后交易”。该举措在一定程度上遏制了市场中高频交易的无序扩张与过度交易现象, 引导其回归理性发展轨道。我们观测高频撤单率的时序变化, 自 2017 年开始快速增加, 并在 2023 年逐渐达到历史峰值, 监管约束使得撤单率出现小幅下降, 目前整体仍然维持较高水平。

图 75: 高频撤单占比的时序变化



数据来源: Wind、中信建投证券; 按笔数统计

3.3 与时俱进的算法交易

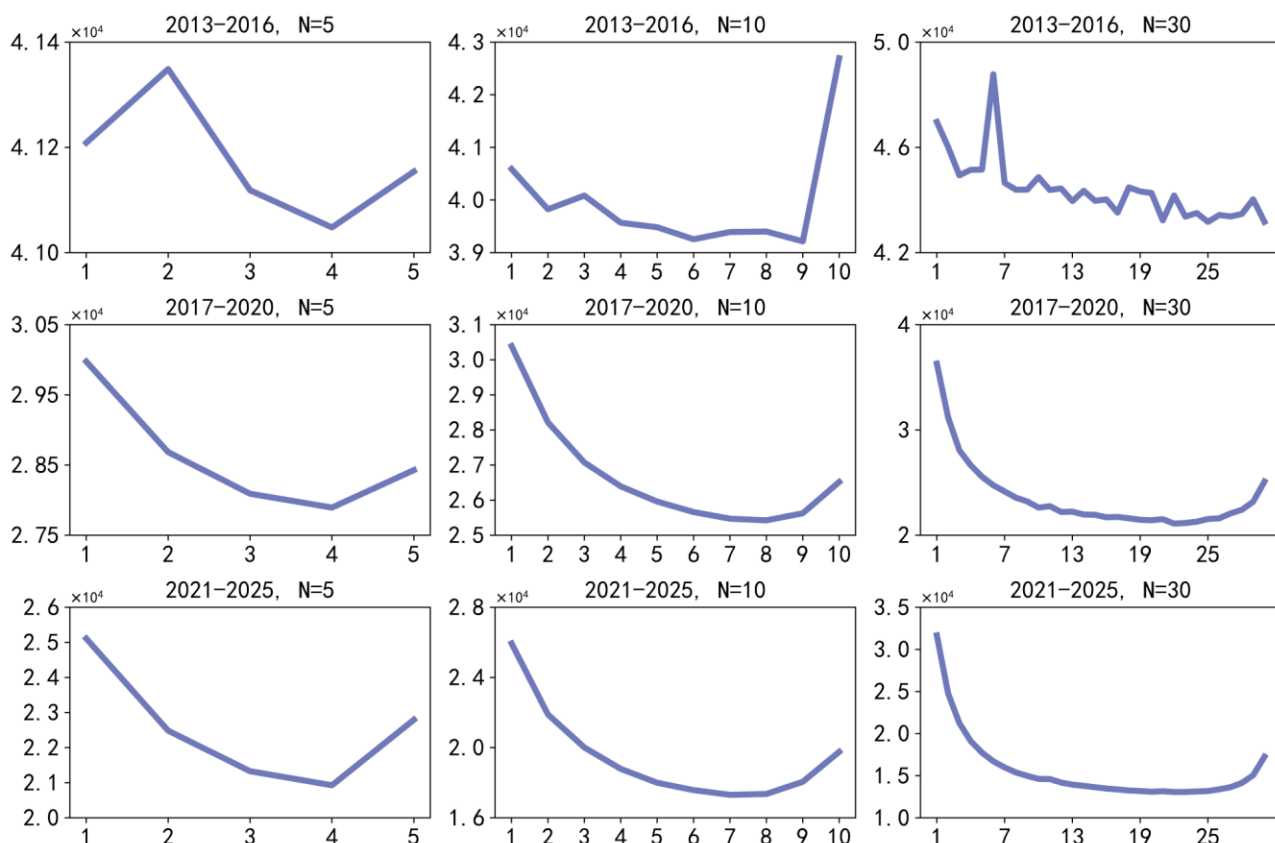
算法交易作为金融科技与资本市场深度融合的核心产物，依托程序化、智能化的交易逻辑，实现了效率与收益的双重优化，但在 A 股独特的市场生态下，也面临不小的挑战。

2024 年 10 月《证券市场程序化交易管理规定（试行）》实施后，2025 年 7 月沪深北交易所同步出台《程序化交易管理实施细则》，明确了高频交易的定义范畴：每秒 300 笔以上或日交易超 2 万笔，建立了“申报速率限制+撤单频率管控”的监管体系，明确高频交易定义与差异化监管要求，为算法交易划定合规边界。

对于策略研究而言，我们能够直观感受到，高频交易在市场微观结构变迁的过程中，扮演了重要角色。在上一案例中本文讨论了大小单识别遇到的难题，由于机构在逐渐采用算法交易执行买卖订单，我们在识别机构订单的难度在不断上升；甚至最简单的“大单化小”的策略，亦能够使得 A 股市场中的大单资金流的定义标准出现偏差。尽管如此，我们仍然可以利用高频数据工具对市场的真实情况进行逆向还原。

订单流中存在特殊规律：卖单后大概率仍跟随卖单，类似地，买单同样具备时间维度的连续性。为直观展示这一现象，可截取任意连续 N 笔同方向订单，观察其委托金额从首笔到末笔的变化趋势。

图 76: 相比早期，2021 年以来的连续订单金额衰减的现象更明显



数据来源: Wind、中信建投证券

从图 76 展示的结果中，我们可以发现，在 2013 年至 2016 年（算法交易还未兴起）期间，连续订单的委托

金额并没有特别明确的变化规律。此时市场上的挂单行为相对并不透明，跟踪订单行为并以此为依据指定的交易策略较少，机构投资者大多是基于过去或预估未来一段时间内的委托成交情况制定交易清单。然而，在 2016 年后，算法交易的应用逐渐铺开，更加高级的盘口交易策略不断迭代出新，其中一类较为典型的策略是盘口动态调整委托价格。这类策略利用低延时、高效率的交易通道，在有大单改变当前盘口交易状态时候，迅速在己方盘口挂单以争取更优先的序列。这些机构类订单的委托金额通常较小，申报的时间通常在同方向的大额交易委托挂单之后，表现为图中委托金额呈现逐渐减小的变化趋势。这部分委托逐渐减少以后，后续跟随的非机构类的订单会使得委托金额重新恢复到一般水平。而对方发现己方流动性充足，会主动委托吃掉当前己方的委托订单，从而形成相反方向的连续委托，如此循环往复。

Lillo 等学者（2004）曾指出，由连续委托的挂单方向组成的序列基本符合长期记忆过程的特征，即序列的自相关强度随着距离变远而减小的速度较慢，即便在间隔很长的样本之间仍保留一定的关联性。后续学者则在不同市场内检验此结论的合理性，例如，Doojin Ryu（2012）在韩国期货市场中确认该规律显著存在，Yuki Sato 等（2023）则在日本股市中找寻了能够定量刻画指标。

我们沿用前辈们的做法，将每笔买入委托的标识记作“1”，而卖出委托的标识记作“-1”，从而得到了一组连续的数值序列。图 77 展示了某只股票订单流信息，以及挂单方向经数值处理后的结果。从图中不难发现，时序相邻的两笔委托之间的相似度通常比较高，包括但不限于委托方向、价格以及数量等。

图 77: 机构算法拆单（大单化小）的具体示例

数值处理: Embedding						
交易所委托号	时间	委托代码	委托价格	委托数量	挂单方向	
14071094	103027150	B	10.71	500	1	
14071219	103027210	S	10.7	100	-1	
14071247	103027230	S	10.68	39100	-1	
14071277	103027240	S	10.68	2800	-1	
14071280	103027240	S	10.68	2800	-1	
14071285	103027240	S	10.68	3000	-1	
14071286	103027240	S	10.68	2700	-1	
14071287	103027240	S	10.68	1500	-1	
14071289	103027240	S	10.68	3000	-1	
14071290	103027240	B	10.68	100	1	
14071376	103027260	S	10.75	900	-1	
14071749	103027410	S	11.18	5000	-1	
14071790	103027420	S	10.68	200	-1	
符号序列					数值序列	

数据来源: Wind、中信建投证券

针对上述现象进行定量刻画，笔者利用自相关系数描述变量的长期记忆性。为了方便讨论，我们先处理所需的数学符号，若股票单日订单数量为 N ，每笔订单挂单方向组成的数值序列为 $\{X_n\}$ ，其中， $n = 1, 2, \dots, N$ 。对于上述变量在间隔 k 期后的自相关程度计算如下。

自协方差:

$$\gamma_k = \text{cov}(X_n, X_{n-k}) = \frac{1}{N-k} \sum_{n=k+1}^N (x_n - \bar{x})(x_{n-k} - \bar{x})$$

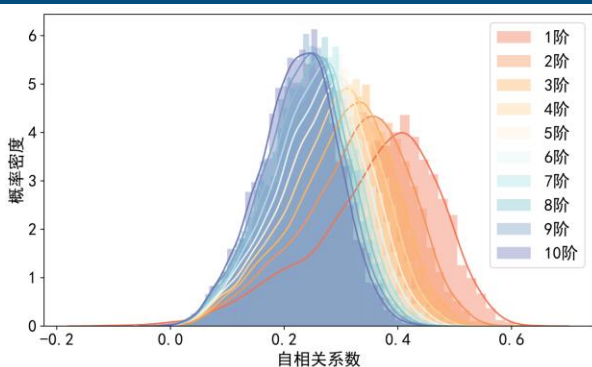
自相关系数:

$$\rho_k = \gamma_k / \gamma_0$$

其中, $\bar{x}$ 表示数值序列 $\{x_n\}$ 的均值。

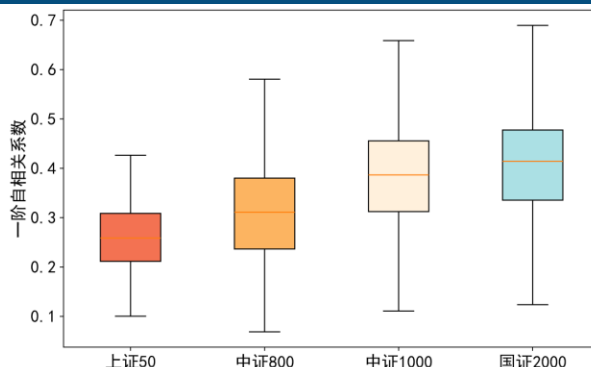
在时间序列分析中, 自相关函数 (ACF) 和偏自相关函数 (PACF) 可以用来衡量时间序列的趋势性和周期性的规律, 我们采用的自相关系数在统计上具有显著性。并且我们以较长的滞后窗口取任意两笔订单, 其委托的方向大概率仍然是一样的 (见图 78), 该结论与国外学者们研究基本一致。

图 78: 滞后多阶的两笔委托, 在交易方向上同样存在关联



数据来源: Wind、中信建投证券; 以 20251103 的数据为例

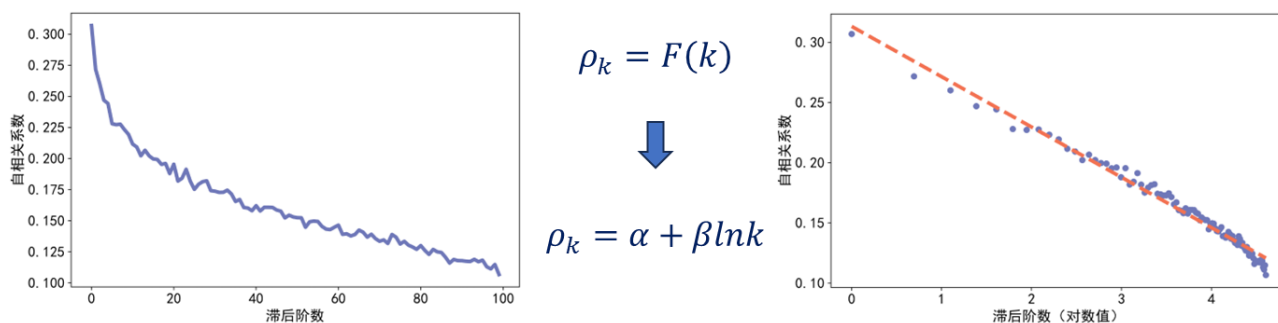
图 79: 交易方向的关联性在小市值上显著性更高



数据来源: Wind、中信建投证券; 以 20251103 的数据为例

订单流的连续性并非是由散户在时间上的拥挤行为, 而是某种或者多种算法共同作用的条件下所实现的表象特征。尽管市场环境的变化让因子策略开发面临挑战, 但仍需客观来看, 算法交易的存在并不会使得 Alpha 完全消失, 我们希望在现有的市场环境中寻找这些特征的蛛丝马迹。(引出因子构造)

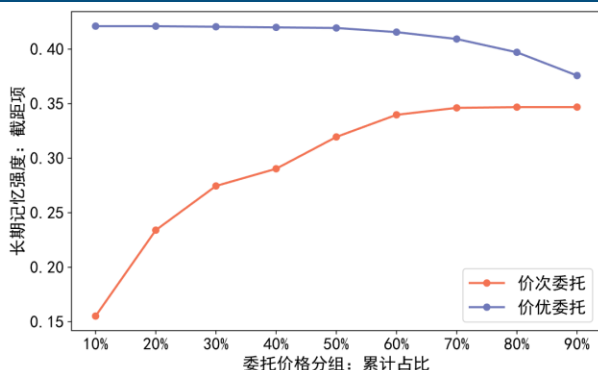
图 80: 自相关系数随滞后阶的变化符合幂律分布的规律, 取滞后阶的对数值方便线性化处理



数据来源: Wind、中信建投证券; 以 20251103 的 000001.SZ 为例

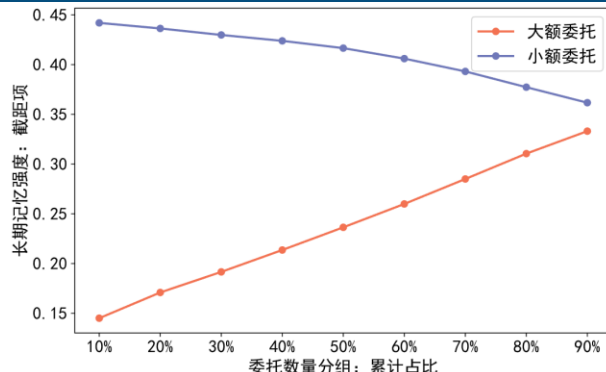
价格、小额是连续订单的重要特点。

图 81: 靠近盘口的委托更容易同向连续出现



数据没用来源: Wind、中信建投证券

图 82: 小额委托的交易方向连续性更强



数据来源: Wind、中信建投证券

委托价格较为接近盘口的订单（价优委托）以及委托数量较小的订单（小额委托）在时序上的相似性较高，说明这一交易的目的是为了快速成交，而非从干扰订单簿的角度，并且小额的订单受到流动性的局限更小。

基于对长期记忆性的规律分析，本文构造并测试了多个有效因子，但因为对每个因子分别详细地介绍并不现实，此处仅展示部分因子的测试效果与关键结论分析。文中主要提供两种计算方法，分别是：

（1）自相关系数回归法

在时序上相邻的两笔订单的关联性可以通过相关系数来刻画，我们观察到，随着滞后阶的对数值增大，相关系数呈现线性衰减的变化趋势。

结合我们对时间序列数据处理的经验，该方法核心思路是计算方向数值序列 1 至 100 阶自相关系数，并通过对滞后阶的对数值进行回归，得到 OLS 模型估计参数，将回归参数作为选股因子的日频信号。通过最小二乘回归的方法，长期记忆性可被描述得更加立体，如用截距项表示长期记忆性的强度以及用斜率反映其持续性等等。

（2）基于“傅里叶变换”的频谱分析

任何事物在两个不同时刻都不可能保持完全相同的状态，但很多变化往往存在着一定的规律，例如 24 小时日出日落，潮起潮落，这些现象通常称为“周期”。针对时间序列的周期性检测问题，“傅里叶变换”和“自相关系数”是两种在解决实际问题的常用方法，前者是将原本时域（常见时序特征）信息转变为频域（由数据周期性决定的）信息，方便我们计算得到更深层的因子信号。

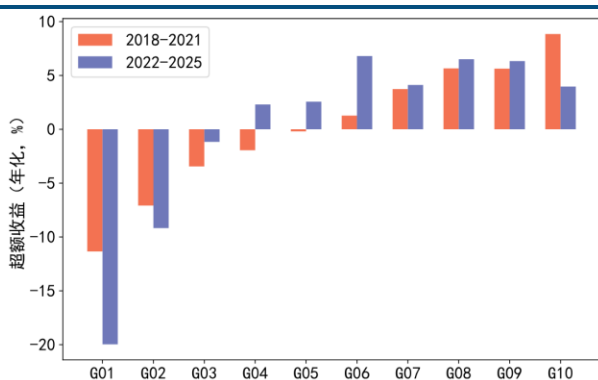
表 26: 长期记忆强度_LMS 因子的计算步骤

步骤	说明
第一步	选定股票某个交易日的逐笔数据，将连续订单流挂单方向编码为数值序列（买入记为 1，卖出记为-1）
第二步	计算数值序列滞后 1 阶至 100 阶的自相关系数，记为 ρ_k ，滞后阶 $k \in [1,100]$
第三步	针对每只股票的一组数据，利用 OLS 估计 ρ_k 对 $\ln(k)$ 的回归系数，记为 B 指标
第四步	B 指标为 LMS 因子的日频信号，例如，滚 20 个交易日计算 B 指标的均值即为月度选股因子

资料来源：中信建投证券

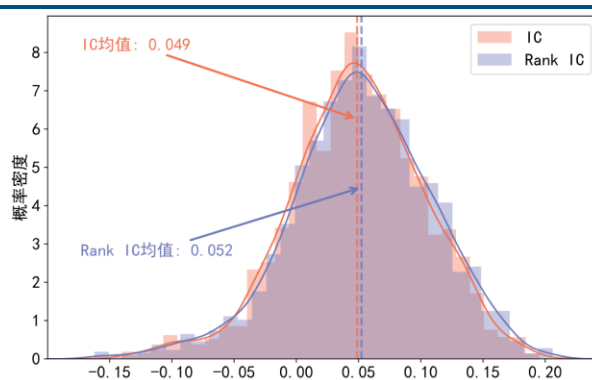
表 26 展示了长期记忆强度_LMS 因子的计算步骤；图 83 至图 86 为 LMS 因子的相关测试结果。

图 83: LMS 因子的分组单调性



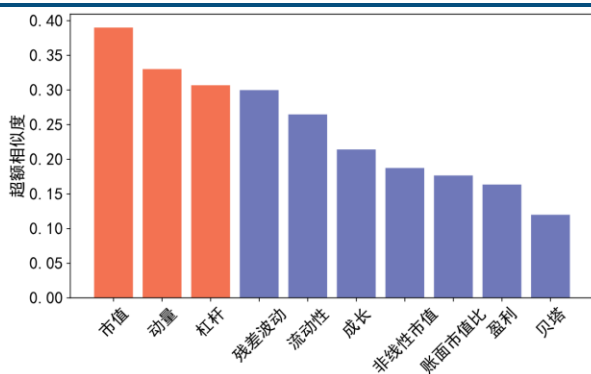
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 84: LMS 因子的 IC 分布图（测算 IC 和 Rank IC）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 85: LMS 与风格因子的超额相似度（暴露程度）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 86: LMS 单因子增强收益表现



数据来源：Wind、中信建投证券

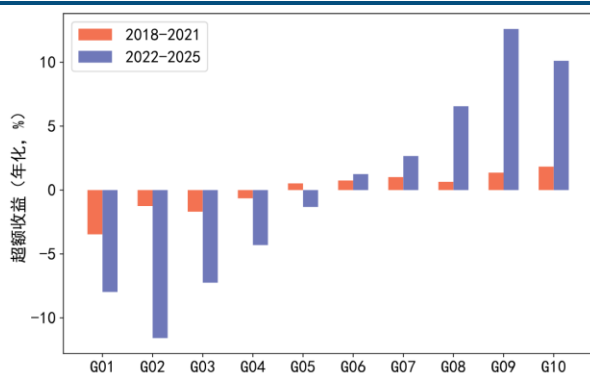
表 27: 分拆痕迹_OST 因子的计算步骤

步骤	说明
第一步	选定股票某个交易日的逐笔数据，将连续订单流挂单方向编码为数值序列（买入记为 1，卖出记为-1）
第二步	利用傅立叶变换将时序上的数值序列转化为频域的振幅序列
第三步	针对每只股票的一组数据，将所有频率的振幅当作样本计算其分布峰度，记为 K 指标
第四步	K 指标为 OST 因子的日频信号，例如，滚 20 个交易日计算 K 指标的均值即为月度选股因子

资料来源：中信建投证券

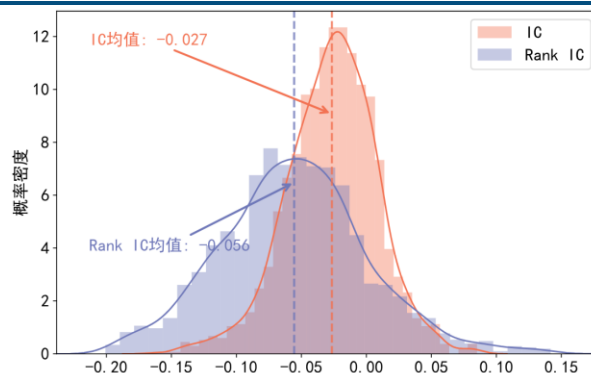
表 27 展示了分拆痕迹_OST 因子的计算步骤；图 87 至图 90 为 OST 因子的相关测试结果。

图 87: OST 因子的分组单调性



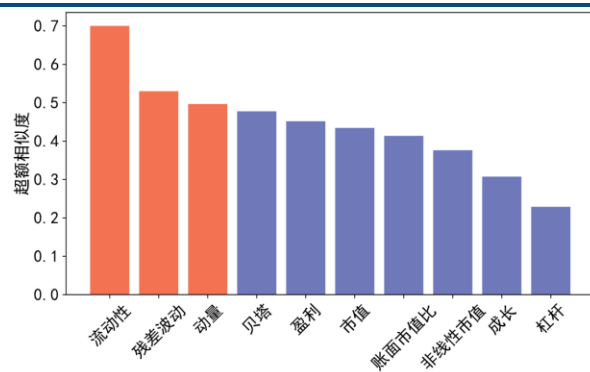
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 88: OST 因子的 IC 分布图（测算 IC 和 Rank IC）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 89: OST 与风格因子的超额相似度（暴露程度）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 90: OST 单因子增强收益表现



数据来源：Wind、中信建投证券

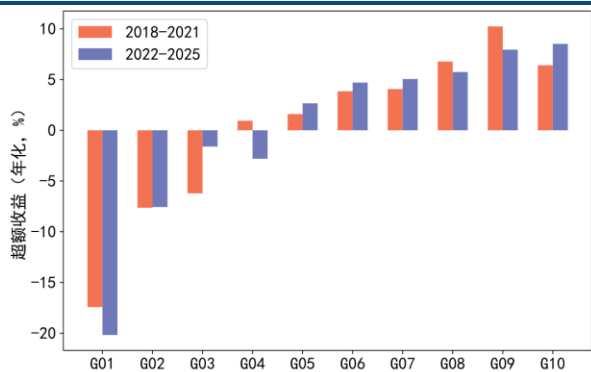
表 28: 高维记忆_MEMO 因子的计算步骤

步骤	说明
第一步	选定股票某个交易日的逐笔数据，将连续订单流挂单方向编码为数值序列（买入记为 1，卖出记为-1）
第二步	计算数值序列滞后 1 阶至 100 阶的自相关系数，记为 ρ_k ，滞后阶 $k \in [1,100]$
第三步	针对每只股票的一组数据，将所有自相关系数当作样本计算其分布偏度和峰度，并将其加总得到 SK 指标
第四步	SK 指标为 MEMO 因子的日频信号，例如，滚 20 个交易日计算 SK 指标的均值即为月度选股因子

资料来源：中信建投证券

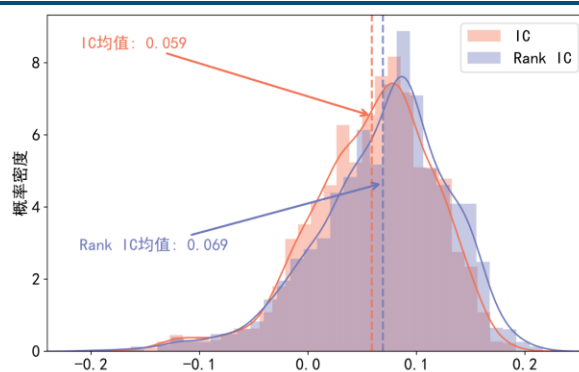
表 28 展示了高位记忆_MEMO 因子的计算步骤；图 91 至图 94 为 OST 因子的相关测试结果。为了避免单个超参数对结论的影响，我们考虑将不同参数的因子进行“平均”，类似于模型训练的多维度处理方式。

图 91: MEMO 因子的分组单调性



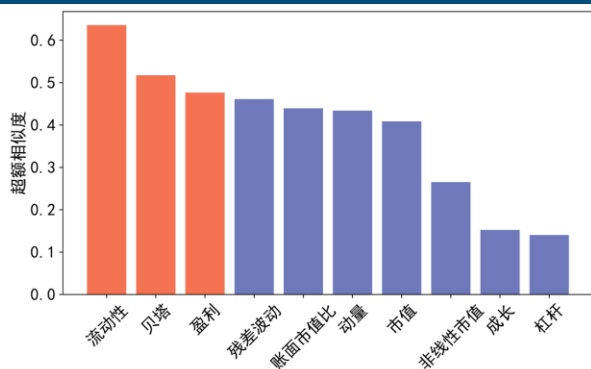
数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 92: MEMO 因子的 IC 分布图（测算 IC 和 Rank IC）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 93: MEMO 与风格因子的超额相似度（暴露程度）



数据来源：Wind、中信建投证券；统计区间：20180102-20251105

图 94: MEMO 单因子增强收益表现



数据来源：Wind、中信建投证券

表 29: 中证 1000 指数成分股优选分月度超额收益表现 (LMS 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	-0.78	-0.43	1.49	1.76	0.57	0.08	0.27	-0.42	-1.01	0.88	-0.12		2.26
2024	-0.11	0.10	1.07	-0.65	1.54	0.23	1.44	2.80	1.07	-2.67	2.37	1.39	8.79
2023	-0.06	0.42	2.16	3.20	2.14	0.58	0.62	0.39	0.24	0.78	0.46	0.23	11.70
2022	1.46	-0.97	-0.10	-0.70	1.04	-0.91	2.64	0.99	0.14	2.45	0.95	1.69	8.92
2021	1.64	2.03	0.56	0.81	-0.47	0.51	-0.77	0.57	1.38	-0.72	4.57	1.08	11.65
2020	1.11	0.21	3.09	2.42	1.97	2.70	1.19	1.77	0.30	0.90	-0.15	-0.15	16.40
2019	0.95	0.08	-1.26	-0.08	2.22	1.12	2.32	0.96	3.24	0.15	0.25	1.34	11.81
2018	-1.66	1.23	3.81	-0.41	0.92	0.66	0.98	0.64	-0.34	0.20	0.50	1.38	8.10

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251103

表 30: 中证 1000 指数成分股优选分月度超额收益表现 (OST 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	1.45	-1.45	2.87	3.16	0.03	-0.98	-0.75	-1.13	-0.56	1.03	0.16		3.76
2024	0.75	-0.03	-0.62	1.45	2.21	-1.00	-0.11	1.36	0.67	-4.80	1.69	2.31	3.75
2023	-0.24	0.75	-1.43	0.21	-0.39	0.00	1.58	0.85	-0.12	-0.83	1.54	0.21	2.09
2022	3.89	-2.88	-1.66	0.92	1.57	-0.28	0.81	1.80	1.35	1.92	0.40	1.68	9.74
2021	0.10	2.71	1.92	-0.80	-1.44	0.15	-2.68	-0.70	2.66	-0.24	3.13	3.29	8.16
2020	-1.75	-1.39	2.13	0.80	1.02	-1.18	1.75	2.29	0.39	-0.38	1.36	-2.02	2.93
2019	-0.12	-0.26	0.71	0.99	0.22	1.14	-0.15	-0.37	1.00	-1.09	-0.51	-0.44	1.10
2018	0.53	-0.74	2.34	1.23	0.65	-0.38	1.95	0.37	-0.31	1.30	0.65	0.71	8.58

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251105

表 31: 中证 1000 指数成分股优选分月度超额收益表现 (MEMO 因子)

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	全年
2025	0.72	-2.81	3.37	2.50	0.94	-0.96	0.00	-3.35	-0.47	1.97	0.54		2.25
2024	1.84	-0.58	-0.60	0.20	2.34	-0.63	-0.50	1.96	2.19	-2.32	1.80	1.47	7.29
2023	0.82	1.58	0.87	1.70	1.33	0.15	-0.05	0.89	0.59	-0.24	1.12	-0.01	9.09
2022	4.33	-2.11	0.13	1.99	1.01	1.39	2.46	1.64	1.12	2.69	-0.03	1.14	16.81
2021	1.00	3.05	1.15	0.44	-1.67	0.44	-2.33	0.84	3.56	-1.62	2.93	4.73	12.97
2020	0.04	0.21	1.89	1.55	1.30	1.65	1.21	2.99	1.61	0.53	1.05	-0.77	14.04
2019	0.59	-0.61	-0.04	1.38	0.65	1.30	1.39	0.93	1.64	0.69	0.84	1.10	10.30
2018	0.14	0.67	1.48	0.53	1.00	1.06	0.63	0.90	0.13	0.10	0.66	-0.68	6.82

资料来源: Wind、中信建投证券; 测算区间: 20180102-20251103

四、附录

4.1 MOD 修正法

主力资金买入股票或卖出股票会驱使价格上涨或下跌，由于交易金额较大，交易的冲击成本会造成资金流跟随策略的收益降低。因此，MOD 修正法的核心步骤是通过截面回归的方式调整因价格变动带来的影响。

表 32: MOD 修正法处理资金流的冲击的计算步骤

步骤	说明
第一步	统计每个交易日的大单买入和大单卖出的成交金额，记为 B 和 S
第二步	逐个交易日计算下列回归方程，并取残差项 ε 作为修正系数 $\ln(B/S) = \alpha + \beta \times R + \varepsilon$ 其中，R 为当日涨跌幅。
第三步	基于修正系数反算大单买入和大单卖出的比例关系，并重新分配大单买入和卖出的成交金额。 $\hat{B} = \frac{e^{\varepsilon}}{1 + e^{\varepsilon}} \times (B + S)$ $\hat{S} = \frac{1}{1 + e^{\varepsilon}} \times (B + S)$

资料来源：中信建投证券

若是仅需对资金流的相关指标进行反转修正，则 MOD 修正法的第三步并非必须。

4.2 CNIR 因子计算公式

我们基于 Level2 数据确认了大小单的划分阈值接近 4 万元的数值，基于 Wind 的资金流向数据，资金流入率指标的构造方式可以改进为如下：

表 33: 广义主力净流入率_CNIR 因子的计算步骤

步骤	说明
第一步	分别计算主力资金的买入金额 B 和卖出金额 S： $B = B_{Extra} + B_{Large} + B_{Med}$ $S = S_{Extra} + S_{Large} + S_{Med}$
第二步	逐个交易日处理，利用当日涨跌幅来修正主力资金的买入金额 B 和卖出金额 S，取残差项作为修正后的资金流指标： $\ln(B/S) = \alpha + \beta \times R + \varepsilon$ $\varepsilon = \ln(\hat{B}/\hat{S})$
第三步	通过 MOD 修正法第三步得到 $\hat{B}$ 和 $\hat{S}$，进而计算主力资金的净流入率，即为 CNIR 因子： $CNIR = \frac{\sum_i^N (\hat{B}_i - \hat{S}_i)}{\sum_i^N (\hat{B}_i + \hat{S}_i)}$ 其中，N 取 20 时，CNIR 为对应的月度选股因子，无需再次时间窗口平滑。

资料来源：中信建投证券

风险分析

本报告中所有数据结果是基于历史统计结果的展示，未来有可能发生风格切换导致因子失效的风险。模型运行存在一定的随机性，初始化随机数种子会对结果产生影响，单次运行结果可能会有一定偏差。历史数据的区间选择会对结果产生一定的影响。模型参数的不同会影响最终结果。模型对计算资源要求较高，运算量不足会导致结果存在一定的欠拟合风险。本文所有模型结果均来自历史数据，模型存在统计误差，不保证模型未来的有效性，对投资不构成任何建议。

分析师介绍

姚紫薇

金融工程及基金研究团队首席分析师，上海财经大学硕士，厦门大学统计学与数据科学系行业导师、中证报金牛奖专家评委。历任兴业证券金融工程分析师、招商证券基金评价业务负责人，2024 年加入中信建投研究发展部，深耕基金研究、资产配置、财富管理、数字化建设、大模型应用等领域，主导开发招商证券研基实验室、中信建投智研多资产配置+平台（网址：<https://fund.csc.com.cn>）。多次获得团队荣誉，获得 2020 年新财富最佳分析师金融工程方向第 3 名，2021 年新财富最佳分析师金融工程方向第 2 名。

苏良

金融工程及基金研究组分析师，复旦大学金融学硕士，曾在私募、券商等从事量化策略开发工作，2025 年加入中信建投证券研究发展部，主要负责因子研究、机器学习与深度学习模型、市场微观结构、量化基金研究等领域，工作地在上海。

研究助理

评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后6个月内的相对市场表现，也即报告发布日后的6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作为基准。A股市场以沪深300指数作为基准；新三板市场以三板成指为基准；香港市场以恒生指数作为基准；美国市场以标普500指数为基准。	股票评级	买入	相对涨幅 15%以上
		增持	相对涨幅 5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅 5%—15%
		卖出	相对跌幅 15%以上
	行业评级	强于大市	相对涨幅 10%以上
		中性	相对涨幅-10-10%之间
		弱于大市	相对跌幅 10%以上

分析师声明

本报告署名分析师在此声明：（i）以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，结论不受任何第三方的授意或影响。（ii）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。

法律主体说明

本报告由中信建投证券股份有限公司及/或其附属机构（以下合称“中信建投”）制作，由中信建投证券股份有限公司在中华人民共和国（仅为本报告目的，不包括香港、澳门、台湾）提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格，本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

在遵守适用的法律法规情况下，本报告亦可能由中信建投（国际）证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础，不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料，但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断，该等观点、评估和预测可能在不发出通知的情况下有所变更，亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件，而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策，中信建投都不对该等投资决策提供任何形式的担保，亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内，中信建投可能持有并交易本报告中所提公司的股份或其他财产权益，也可能在过去12个月、目前或者将来为本报中所提公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点，分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系，分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可，任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分内容，亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分内容。版权所有，违者必究。

中信建投证券研究发展部

北京
 朝阳区景辉街16号院1号楼18层
 电话：（8610）56135088
 联系人：李祉瑶
 邮箱：lizhiyao@csc.com.cn

上海
 上海浦东新区浦东南路528号南塔2103室
 电话：（8621）6882-1600
 联系人：翁起帆
 邮箱：wengqifan@csc.com.cn

深圳
 福田区福中三路与鹏程一路交汇处广电金融中心35楼
 电话：（86755）8252-1369
 联系人：曹莹
 邮箱：caoying@csc.com.cn

中信建投（国际）

香港
 中环交易广场2期18楼
 电话：（852）3465-5600
 联系人：刘泓麟
 邮箱：charleneliu@csci.hk