

数字的偏执：基于价格聚集效应的因子构建与实证分析

符昀烨，胡中山，李浩易，王一粟

【摘要】 在股票市场微观结构中，价格聚集效应（Price Clustering）反映了投资者在定价过程中对特定整数尾数的系统性偏好。本文基于中国 A 股市场的高频逐笔数据，对价格聚集现象的分布特征及其形成机制进行了系统分析。实证结果表明，中国市场不仅存在显著的整数聚集现象，还呈现出明显的“重 8 轻 4”文化偏好以及整数价位上的“虚假流动性”特征。基于上述发现，本文以具有显著偏好的整数尾数“058”作为非理性交易行为的代理指标，构建了一类高频微观结构因子。因子构建过程中，本文从分钟级聚集度均值出发，逐步引入波动率与分位筛选以降低微观噪音，最终提出基于日频历史均值偏离度的去趋势相对聚集度因子。回测结果显示，以 60 日历史均值作为基准的相对聚集因子（ $x = 60$ ）能够有效剥离个股固有的散户浓度差异，在截面上获得稳健的 Rank IC 与多空收益表现，为高频量化策略的设计提供了新的微观结构视角。

【关键词】 高频交易, 价格聚集, 市场微观结构, 微观结构因子

The Obsession with Numbers: Empirical Evidence on Price Clustering Effects

Fu Yunye, Hu Zhongshan, Li Haoyi, Wang Yisu

Abstract: Price clustering is a well-documented phenomenon in market microstructure, reflecting investors' systematic preference for specific integer price endings. Using high-frequency tick data from the Chinese A-share market, this report provides a comprehensive analysis of the distributional properties and underlying mechanisms of price clustering. The empirical evidence reveals pronounced integer clustering, a strong cultural preference for the digit “8” over “4,” and the presence of phantom liquidity at integer price levels. Motivated by these findings, we employ the biased integer endings “058” as proxies for irrational trading behavior and construct a set of high-frequency market microstructure factors. Starting from minute-level clustering averages, we progressively incorporate volatility measures and quantile filtering to mitigate microstructure noise. This process culminates in a de-trended relative clustering factor defined as the deviation from a stock's historical daily mean. Backtesting results indicate that the relative clustering factor benchmarked against a 60-day moving average ($x = 60$) effectively removes cross-sectional heterogeneity in retail investor concentration and delivers robust Rank ICs and long-short portfolio returns. Our findings suggest that price clustering contains exploitable information and offers a novel microstructure-based perspective for high-frequency quantitative strategies.

Key Words: High-Frequency Trading, Price Clustering, Market Microstructure, Quantitative Factors

1 引言

在股票市场中，价格不仅是交易达成的结果，更是交易者心理预期的直接投射。传统的技术分析与微观结构研究通常假设价格可以在最小变动单位上连续分布，然而大量的市场微观结构与行为金融学研究表明，投资者对特定的数字——尤

其是整数——表现出强烈的偏好。这种现象被称为“价格聚集效应”（Price Clustering），即交易价格和报价价格在整数及其倍数处出现的频率显著高于其他价格水平。不仅如此，价格聚集效应在不同文化背景下呈现出显著的异质性。

价格聚集效应的研究起源可以追溯到 20 世纪

60 年代。物理学家 Osborne (1962) 在应用布朗运动模型分析股价时，首次发现价格存在“偏好占据点”现象。Niederhoffer (1965, 1966) 随后通过系统分析纽交所专业交易商账簿，证实股票价格在整数处的聚集呈现明显的层级结构。

这一背景下，本文围绕价格整数聚集效应展开研究，系统梳理其理论机制与实证研究进展，并在此基础上提出一类基于高频微观结构信息的量化因子。通过构建规范的回测框架，本文进一步检验了该因子在量化投资中的预测能力与经济意义。全文结构安排如下：第二章回顾整数聚集效应的相关理论与行为解释；第三章基于中国 A 股 Level-2 数据，对该效应的存在性与市场特征进行实证分析；第四章至第六章分别介绍数据与回测方法、因子构建逻辑以及实证回测结果。

2 文献综述

价格聚集效应作为一种普遍且持久的市场微观结构特征，已在全球多个金融市场中得到广泛验证。早期研究主要集中在发达市场，而近年来的研究范围已扩展至新兴市场和加密货币市场。

在亚洲市场方面，Ke, Chen, Lin 和 Liu (2017) 基于台湾证券交易所的研究，证实末位数字分布高度非均匀：数字 0 和 5 出现频率最高，而数字 8 则显著高于数字 4，揭示了文化因素在价格形成中的独特作用。Ahn, Cai 和 Cheung (2005) 利用香港交易所数据，发现在所有 tick size 组中，偶数和整数价格的出现频率异常高，证实了该效应的跨市场稳健性。

欧美市场的研究提供了更为全面的证据。Bourghelle 和 Cellier (2007) 指出，投资者“自然被吸引”到整数点，在此积累潜在流动性并创造价格壁垒。Bhattacharya 等 (2012) 发现美国市场订单流在整数上下存在显著的不对称性 (Over-buying below/Over-selling above)。相关实证总结如表1所示。

价格聚集效应的形成机制是市场微观结构与行为金融学研究的核心理题。现有文献揭示了这一现象背后的多重机制，既包含理性的市场博弈，也反映了投资者的认知偏差与文化习惯。

表 1 价格聚集效应的全球实证证据

研究文献	市场/数据	主要发现
Osborne (1962)	美国股市	价格整数位置出现频率显著高于随机游走预测
Niederhoffer (1965)	纽交所	聚集呈层级结构：整数 > 半数 > 1/4 > 1/8
Ahn et al. (2005)	港交所	偶数和整数价格频率异常高
Bourghelle (2007)	Euronext	约 40% 交易价格末位为 0 或 5，形成价格壁垒
Ke et al. (2017)	台湾证交所	“8” 频率高于预期，“4” 最低（文化因素）
Blau et al. (2022)	美股多因子	高聚集股票面临负回报溢价

2.1 微观市场结构下的博弈行为

价格聚集研究的理论根源可追溯到 Ball、Torous 和 Tschoegl (1985) 提出的“价格解析假说” (Price Resolution Hypothesis)。该假说认为，当资产的基本面信息模糊不清时，价格就如同一张低像素的照片，只能在粗糙的网格（整数位）上成像。这种信息解析的不确定性导致交易者倾向于使用整数价格作为锚定点。Harris (1991) 在此基础上发展了“谈判假说” (Negotiation Hypothesis)，指出交易者使用离散价格集合是为了简化谈判过程——较小的价格集合限制了可能的报价数量，减少了达成交易所需的时间和认知成本，这与 Miller (1956) 和 Simon (1974) 关于人类短期记忆容量有限的认知研究相一致。Ahn 等 (2005) 在对香港股市的研究中进一步发现，如果强行将 tick size（最小变动价位）设置得过小，反而会阻碍价格的有效解析，导致更严重的聚集现象。这一发现表明，市场微观结构的制度安排与价格聚集之间存在复杂的非线性关系。

这种价格解析和非理性投资者的锚点意识，也因之带来了资产交易过程中的两个效应。其一是价格壁垒效应。Bourghelle 和 Cellier (2007) 的研究表明，在全电子限价市场中，整数和半数价格水平积累了大量的限价订单，形成了难以穿透的价格壁垒。股票价格在这些水平上停留的时间不成比例地长，这可能阻碍市场的高效信息处理。这种价格壁垒不仅影响价格发现的速度，也为高频交易者提供了可预测的流动性来源。

另一个效应是聚集削价效应 (Cluster Undercutting Effect)，其揭示了交易者的策略性行为。Bhattacharya 等 (2012) 以及 Bourghelle 和 Cellier (2009) 通过观察订单流上的聚集效应，发现理性的交易者为了获得执行优先权，会策略性地进行“抢跑”：买方倾向于在整数刚过一点的位置（如 10.01 元）挂单，以超越 10.00 元的大量堆积单；卖方倾向于在整数未到一点的位置（如 9.99 元）挂单，以确保成交。这种“聚集削价效应”导致了买卖订单流在整数位附近的不对称分布，形成了特定的微观价格形态，为量化交易策略的构建提供了重要启示。

尽管本文后续在因子构建过程中并未直接利用上述两类效应所反映的市场结构性偏差，但价格壁垒效应与聚集削价效应所揭示的订单博弈机制，对于理解高频环境下的价格发现过程具有重要理论与实证价值，亦为后续基于微观结构信息的因子设计与策略拓展提供了有意义的研究空间。

2.2 系统性认知偏差与文化因素

左位数效应 (Left-Digit Effect) 是价格聚集的重要心理机制。Thomas 和 Morwitz (2005) 首次正式命名并系统研究了这一效应，证明人们在判断多位数字大小时过度关注最左边数字。Bhattacharya 等 (2012) 将这一理论应用于金融市场，认为当股票价格从 7.00 美元跌至 6.99 美元时，尽管实际上只跌了 1 美分，但由于最左边的数字从“7”变成了“6”，投资者会直觉地感觉到这是一个接近 1 美元的巨大跌幅。这种认知偏差会导致投资者在整数价格附近做出非理性的交易决策，从而加剧了价格在整数处的聚集。

文化习俗对价格聚集也产生了显著影响。Ke 等 (2017) 在台湾股市的研究中发现了一个独特现象：由于谐音忌讳，投资者显著回避尾数“4”（谐音“死”），而偏好尾数“8”（谐音“发”）。更为重要的是，这种由迷信驱动的聚集主要存在于散户主导的交易中，机构投资者（聪明钱）则没有表现出这一特征。研究进一步发现，这种文化驱动的聚集在熊市和信息透明度低时尤为明显，这意味着特定数字的聚集程度可以作为衡量市场散户情绪化程度的代理变量。类似的文化效应也在中国 A 股市场得到证实 (Brown & Mitchell, 2008)，A 股价格以“8”结尾的可能性是“4”的两倍以上，而

B 股（外国投资者主导）则无此现象，进一步证明了文化因素的作用。

当前学术界的共识是，理性和行为因素共同作用于价格聚集现象。价格聚集既反映了降低谈判成本和应对不确定性的理性策略，也体现了人类认知能力的固有局限。值得注意的是，即使在十进制改革 (Bessembinder, 2003) 和电子交易普及后，价格聚集仍然存在，但程度有所下降，这支持了两种解释的互补性。对于高频交易策略的构建而言，理解这些机制有助于识别市场的有偏性质，进而或可用于高频因子的构建。

3 A 股市场概况：价格聚集效应的实证刻画

基于中国 A 股沪深 300 在 2025 年 7 月的高频逐笔数据，我们构建了针对价格聚集效应 (Price Clustering) 的完整实证图谱。本节旨在通过横截面统计检验，从全景分布、订单质量、策略博弈及成交效率四个维度，定量刻画中国市场的微观结构特征。实证结果不仅验证了经典的“整数效应”，更揭示了具有中国市场特色的文化偏好与策略性交易行为。

首先，我们考察了全市场价格尾数的分布规律。在理想的随机漫步假设下，价格尾数 0 至 9 出现的概率应当遵循均匀分布（即各为 10%）。然而，我们的统计推断结果显示，市场表现出极强的非均匀性。如下方热力图与全景聚类分布图所示，针对成交 (Trade)、限价买单 (Limit Buy) 及限价卖单 (Limit Sell) 的横截面 T 检验均在 1% 的显著性水平上拒绝了原假设。

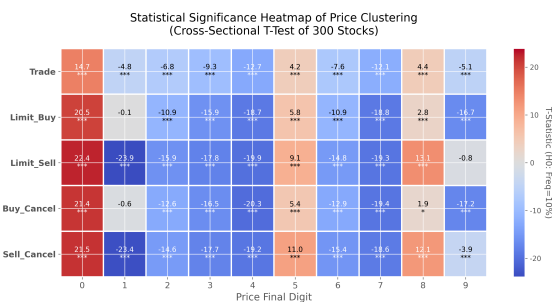


图 1 非均匀分布检验

表 2 价格聚集效应的理论机制与实证验证

理论类别	核心假说	代表文献	主要预测	实证支持
理性解释	价格解析假说	Ball et al. (1985)	不确定性高时聚集更严重	波动性与聚集正相关
理性解释	谈判假说	Harris (1991)	交易量小时聚集更明显	流动性高的股票聚集减少
理性解释	协调博弈理论	Schelling (1980)	整数作为谢林点协调交易	订单簿在整数处深度增加
理性解释	聚集削价效应	Bhattacharya et al. (2012)	整数附近订单流不对称	整数上下 1 美分买卖失衡
行为解释	左位数效应	Thomas & Morwitz (2005)	整数跨越引发过度反应	“9 结尾”价格后续回报偏低
行为解释	认知参考点理论	Rosch (1975)	整数作为判断锚点	价格在整数处停留时间长
行为解释	文化习俗影响	Ke et al. (2017)	特定数字偏好因文化而异	台湾市场“4”显著低于“8”

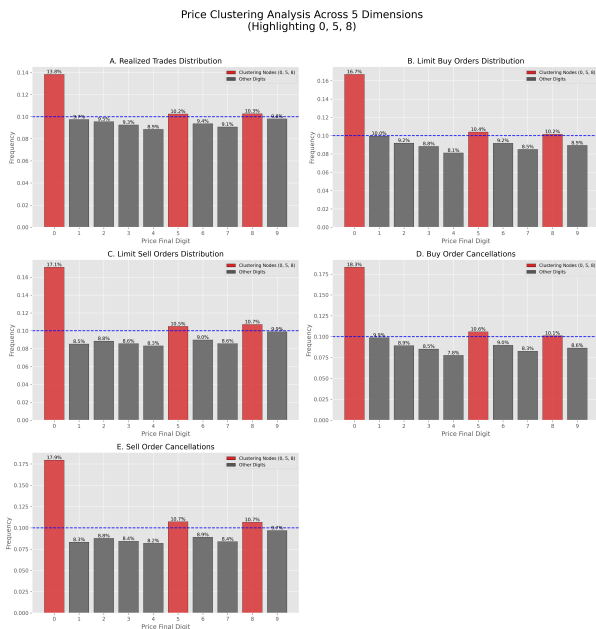


图 2 全市场价格尾数概况

具体而言，尾数“0”表现出最强的聚集特征，其在限价买单和卖单中的出现频率分别高达 15.75% 和 15.88%，T 统计量远超临界值，验证了 Harris (1991) 提出的假说，即交易者在处理信息时天然倾向于认知成本最低的整数位。此外，尾数“5”作为半整数位，其频率也显著高于 10% 的基准线，成为仅次于“0”的次级聚焦点。值得注意的是，中国市场表现出显著的文化效应：代表“发”的数字 8 在买卖订单中均呈现显著的正向偏离，而代表“死”的数字 4 则呈现显著的负向偏离（频率降至 8.4%-8.6% 区间）。这种“重 8 轻 4”的非对称分布证实了文化迷信（Superstition）对资产

定价微观结构的实质性影响，表明中国市场的价格形成机制中包含了非理性的文化因子。

然而，挂单量的聚集并不等同于真实的流动性供给。为了辨析挂单的真实意图，我们进一步分析了不同尾数上的撤单行为。

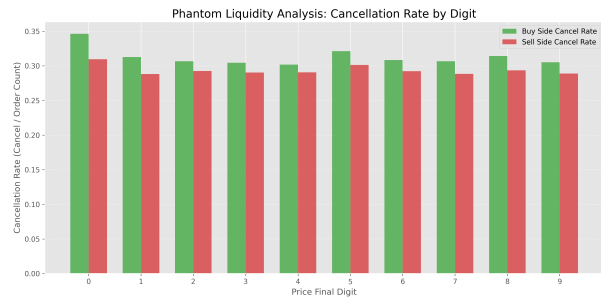


图 3 虚假流动性

撤单率分析揭示了整数价位流动性结构的复杂性。尽管尾数 0 聚集了全市场最为庞大的挂单规模，其撤单率亦相对高于其他尾数，表明该价位上确实存在一定比例的策略性挂撤行为。这一现象在一定程度上支持了“幻影流动性”（Phantom Liquidity）假说，即部分堆积于整数位的订单并非以实际成交为主要目的，而是用于算法交易的流动性探测或短期策略性博弈。然而，需要强调的是，较高的撤单率并不足以抵消尾数 0 在成交量与订单占比上的显著聚集特征。相较于其他尾数，整数位仍然承载了更为广泛的交易参与与持续的成交需求，反映出其作为非理性交易者定价锚点的核心地位。换言之，算法交易与策略性挂撤更可能是对既有聚集结构的“顺势利用”，而非价格聚集本身的根本驱动因素。因此，在整体层面上，尾

数 0 仍可被视为散户主导需求高度集中的价格区间，与尾数 5 和 8 共同构成对非理性交易行为具有代表性的价格聚集代理。

同时需要注意的是，这种流动性结构的差异又进一步诱发了知情交易者（Informed Traders）的策略性博弈。通过计算订单不平衡比率（Order Imbalance），我们捕捉到了显著的“抢跑”与“截胡”行为。



图 4 买卖压力

数据表明，在尾数 1 上存在极强的正向不平衡（绿色买压），而在尾数 9 上存在极强的负向不平衡（红色卖压）。这一现象生动地描绘了微观市场中的“跳价策略”（Penny Jumping）与“削价策略”（Undercutting）。知情交易者为了避免在拥挤的整数位（0）排队，倾向于支付额外的 1 分钱成本（在买单时挂在 1，或在卖单时挂在 9）以获得价格优先权（Priority Rule）。尾数 1 和 9 因此成为了知情资金与急切成交资金的主战场，而死板地挂在整数 0 上的往往是非知情交易者（Uninformed Traders）。

最后，我们将视角转向实际的成交效率，以评估在不同尾数挂单的经济后果。

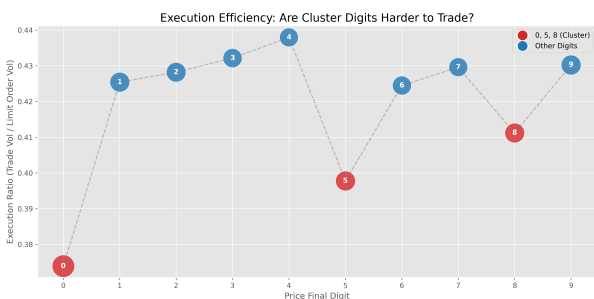


图 5 挂单转化率

气泡图清晰地展示了“拥挤效应”（Crowding Effect）带来的负面影响。尽管聚类数字（0, 5, 8）

的总成交规模（气泡大小）巨大，但其单位挂单的转化效率（Y 轴位置）却处于低位。特别是尾数 0，虽然吸引了全市场最大的关注度，但由于巨大的挂单存量导致时间优先队列极长，后来的订单极难成交。这一结果表明，整数位实际上构成了一个“流动性陷阱”：虽然它是最具吸引力的价格锚点，但在执行层面却是效率最低的交易位置。

综上所述，既有文献与上述实证结果共同刻画出中国市场价格聚集效应的多重面向。一方面，整数尾数，尤其是 0、5 和 8，体现了散户与非理性交易需求在定价过程中的集中表达（同时也包含策略性交易者往往是在既有聚集结构之上进行的“顺势利用”）；另一方面，尾数 1 与 9 附近所呈现的显著订单不平衡与抢跑行为，则更多反映了专业机构与高频交易者围绕执行优先权展开的策略性博弈。基于上述认识，本文后续的因子构建并不试图直接刻画整数位上的执行效率或策略性挂撤行为，而是将研究重心放在价格尾数 0、5、8 上所体现的相对聚集比例，将其作为非理性交易者（不知情交易者/散户）参与程度的代理变量。在此框架下，专业机构与量化交易者在聚类尾数上的占比相对较低，其主要交易活动更多集中于整数邻近的非聚类价位（如尾数 1 与 9）。因此，基于 058 聚集比例构建的高频因子，能够在不直接依赖微观执行效率的前提下，更为稳健地刻画市场中非理性交易力量的变化，为后续的因子设计与实证分析奠定逻辑基础。

4 回测框架设计与数据来源

本研究采用贴近 A 股实盘环境的回测引擎，确保因子测试结果的真实性与可执行性。

4.1 核心回测机制与交易逻辑

该引擎模拟了真实的交易环境，具体机制如下：

- **交易执行**：因子信号在 T 日生成，以 T 日收盘价进行买卖。引擎通过计算 `close_to_close_ret`（即 T+1 收盘价相对于 T 收盘价的收益）来实现这一点。
- **A 股涨跌停特殊处理**：自动识别涨跌停状态（默认阈值为 9.8%）。实现了持仓锁定逻辑：如果在调仓日股票发生涨跌停（无法买入或卖

出)，代码会强制延续上一期的持仓分组，模拟无法交易的情况。

- **收益计算**：使用每日“收盘到收盘”收益进行复利计算，包含了隔夜跳空和日内波动。

4.2 因子评价指标

框架计算了两大类核心指标来评估因子质量：

- **Rank IC**：计算每日因子值与未来收益率的斯皮尔曼等级相关系数。
- **IC 统计**：计算 IC 均值、IC 标准差、ICIR (信息比率) 以及 IC 胜率 (IC > 0 的占比)。
- **分层回测**：将股票池根据因子值每日划分为 N 个组（默认为 10 组）。
- **多空收益 (Long-Short)**：计算 Top 组（多头）减去 Bottom 组（空头）的收益。

4.3 策略绩效分析

对于每个分组（以及多空对冲组合），引擎计算了标准的量化风险收益指标：

- **年化收益率 (Annual Return)**
- **年化波动率 (Volatility)**
- **夏普比率 (Sharpe Ratio)**
- **最大回撤 (Max Drawdown)**
- **日胜率 (Win Rate)**

4.4 数据来源

本文后续研究采用中国 A 股市场的 Level-2 高频逐笔数据作为主要研究样本，数据时间跨度为 2024 年 10 月至 2025 年 12 月，覆盖沪深 300 指数成分股。数据类型包括逐笔成交数据 (Trade)、逐笔委托数据 (Insert)、逐笔撤单数据 (Cancel) 以及 3 秒频率的盘口快照数据 (Snapshot)。上述多源高频数据能够从成交、报价与订单行为等多个维度刻画市场微观结构特征，为价格聚集效应的系统分析提供了精细粒度的信息基础。

5 因子构建：从微观噪音到有效信号

基于前文的实证发现，我们尝试将 A 股市场的价格聚集这一微观现象转化为可交易的量化因

子。我们的核心逻辑假设是：价格尾数的聚集程度反映了非理性散户（倾向于聚集在“058”）与理性机构（基本保持对于尾数的无差异，或由于跳价/削价策略有一定规避整数的倾向）之间的博弈力量对比。

5.1 基础定义与初步探索

首先，我们将高频 Tick 数据降采样为分钟级统计量，定义“聚集度比率” (Clustering Ratio, CR) 为特定时间窗口内，落在聚类数字 (0, 5, 8) 上的成交量占比：

$$CR_{t,type} = \frac{\sum_{i \in t} \mathbb{I}(d(P_i) \in \{0, 5, 8\}) \cdot V_i}{\sum_{i \in t} V_i} \quad (1)$$

其中 $d(P_i)$ 代表价格尾数， V_i 代表成交量。

在早期的探索阶段，我们直接使用了分钟级的聚集度均值作为因子（特定时间窗口 t 为 1 分钟）。然而，无论是买单或者卖单的挂/撤单数，其实证结果都不理想，Rank IC 接近于零。深入分析后我们发现，短周期内的价格聚集往往受制于该分钟“起始报价”的随机惯性，更多体现的是微观结构的机械噪音，而非有效的情绪信号。

$$\text{Volatility} = \sqrt{\frac{1}{N} \sum_{i=1}^N (CR_i - \bar{CR})^2} \quad (2)$$

$$\text{Quantile Filter} = \mathbb{I}(CR_i > Q_{75} \text{ or } CR_i < Q_{25})$$

为了剔除这些噪音，我们随后引入了波动率算子和分位筛选算子：使用历史 N 个聚集度比率，计算其波动率以衡量是否其在过去一段时间存在大幅度的变化。在此，若存在大幅度的变化，说明存在非理性投资者的交易高峰；而 Quantile 算子可以帮助提取出数据差异最大的部分，以突出该因子想要测度的内容。基于我们的回测框架显示，虽然信噪比有所提升，但其 IC 和多空收益依然缺乏稳定性。

5.2 逻辑重构：去趋势的相对聚集度

为了排除短期噪音的干扰，我们在后续研究中将时间粒度提升至日频，重新定义了因子（时间窗口 t 为当日开盘截止到对应时间点，对于目前的回测引擎，时间窗口 t 即为 1day）。日频的聚集度均值 (ts_mean_day) 在统计上表现出更强的 Rank IC，具体结果如表3所示。此处 Cancel 和 Insert 分别表示逐笔撤单数据和逐笔委托数据，Buy 和 Sell 区分了买方和卖方数据。

表 3 日频因子 (ts_mean_day) 回测统计

Factor	Mean IC	ICIR	Ann. Ret	Sharpe	MaxDD	Win Rate
Cancel Buy Ratio	-0.0207	-0.121	18.34%	0.78	20.04%	52.63%
Cancel Sell Ratio	-0.0328	-0.194	7.37%	0.31	23.81%	52.63%
Insert Buy Ratio	-0.0209	-0.125	27.15%	1.20	16.71%	53.68%
Insert Sell Ratio	-0.0341	-0.191	12.75%	0.54	24.43%	54.39%

可以看出, 这些因子的 Rank IC, 都能达到 0.02 以上, 但如此定义的因子, Rank IC 较高的 Insert Sell 和 Cancel Sell 其多空收益却低于 Insert Buy 和 Cancel Buy, 而且回撤明显, IC 同 Return 的不保序性让我们对因子的稳健性产生了质疑。于是我们以 Cancel Sell 和 Cancel Buy 为例, 日频调仓, 每日以收盘价做多因子较低的 30 支股票, 做空因子较高的 30 支股票, 绘制出其多空净值曲线。不难看出, 曲线波动剧烈, 缺乏稳健性, 而且在 7 月前出现了明显的风格切换。

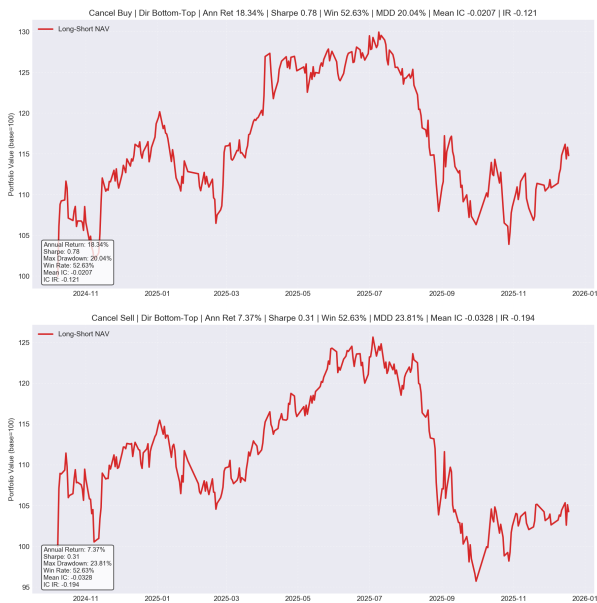


图 6 买卖撤单日频因子多空净值曲线

经过对“聚集”本质的重新思考。我们发现, 不同股票天生具有不同的“散户浓度”——小盘股的自然聚集度往往高于大盘股。如果直接比较不同股票的绝对聚集度 (CR), 本质上是在做多或做空“市值风格”或“流动性风格” (Beta), 而非捕捉个股层面的情绪异动 (Alpha)。

因此, 我们确立了“相对聚集度” (Relative Clustering Ratio, RCR) 的构建思路: 因子的核心不

在于当前的聚集度有多高, 而在于它相对于自身历史水平偏离了多少。我们构建了去趋势因子:

$$RCR_x = \frac{CR_{t, \text{today}}}{\text{RollingMean}(CR_{t-1 \dots t-x})} \quad (3)$$

其中绝对聚集度的时间粒度保持为日频, 以避免短期噪音的干扰。这一逻辑有效地剥离了个股固有的微观结构特征, 使得因子能够纯粹地反映当日市场情绪的异常波动。

5.3 参数确定

在确定历史基准窗口 x 时, 我们选取了 CR 中, Mean IC 表现更好的 Sell Insert 的 RCR, 进行了广泛的参数敏感性测试 ($x \in \{5, 10, 30, 60\}$)。

表 4 不同窗口 (x) 下的去趋势因子绩效对比

Window (x)	Mean IC	ICIR	L-S Ret	Sharpe	MaxDD	Win Rate
5 Days	-0.0212	-0.296	-15.26%	-1.75	14.38%	45.27%
10 Days	-0.0257	-0.350	-20.80%	-2.43	18.78%	44.03%
30 Days	-0.0305	-0.389	-24.46%	-2.57	24.19%	42.39%
60 Days	-0.0317	-0.399	-21.45%	-2.17	22.84%	43.62%

结果如表4 所显示, 随着窗口长度的增加, 因子的稳定性显著提升。由于数据跨度有限, 为避免时间跨度过长造成的数据浪费, 最终我们选定 $x = 60$ 作为最佳参数。60 日均值能够有效表征个股“常态化”的投资者结构, 以此为基准计算出的偏离度, 最能精准捕捉短期内的非理性情绪积聚。

6 实证回测: 策略表现与优化

利用前述的回测框架, 我们对基于 60 日均值偏离度的卖单挂单 (Sell Insert) 聚集因子 RCR_{60} 进行了多空策略的实证回测。

6.1 多空收益与 IC 表现

策略构建逻辑为: 假设起始资金为 1, 一半资金做多因子值最低的前 15 支股票 (即聚集度显著低于历史均值, 暗示机构资金活跃), 一半资金做空因子值最高的 15 支股票 (即聚集度显著高于历史均值, 暗示散户情绪拥挤)。

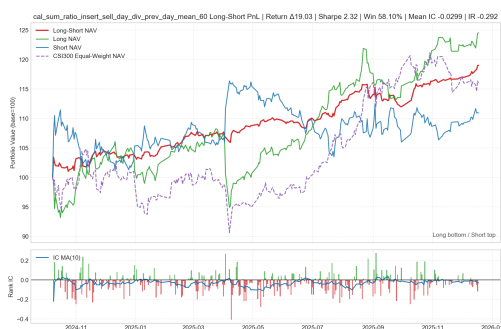


图 7 Sell Insert 60 日相对偏离度策略多空净值曲线

回测结果（如图10所示）显示，因子展现出显著的负向预测能力（Mean IC ≈ -0.03 ），有力地验证了“高整数聚集预示未来负收益”的理论假说。其次，随着大盘 CSI300 指数（图中紫色虚线）的变动，多空净值曲线（红线）呈现出稳健的上升趋势，显著减少了早期版本中因市场风格切换导致的剧烈回撤。这说明，通过剥离个股固有的散户浓度风险，我们成功提取出了更为纯粹的 Alpha 收益。

6.2 因子优化

进一步的分析表明，Buy Insert 也表现出类似的负向预测能力，验证了因子构建逻辑的稳健性。我们推测，一致的负向预测能力主要源于 A 股市场的做空限制：高质量的资产在多头端的交易不受限制，而空头端尽管识别出下跌趋势，亦很难参与交易。因而，知情交易者参与较少（空头受限）、对应非知情交易者比例更高的资产有更强的下跌趋势；知情交易者参与更多、对应非知情交易者比例更低的资产有更强的上升动力。



图 8 Buy Insert 60 日相对偏离度策略多空净值曲线

基于此，我们尝试将 Buy Insert 和 Sell Insert 数据求和，计算 Total Insert 的 60 日相对偏离度因子，结果显示该合成因子在多空收益和 IC 表现上均优于单一方向的因子，进一步提升了策略的稳定性与收益水平。

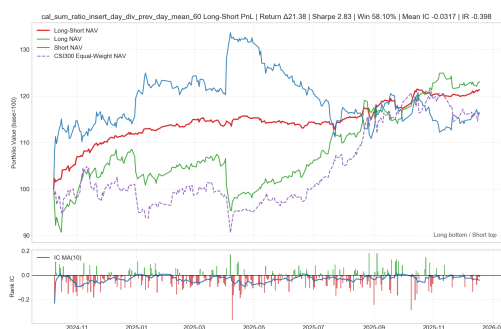


图 9 Total Insert 60 日相对偏离度的多空净值曲线

6.3 风格检验与分组异象

本文选择以 1 个交易日（T+1）作为因子的主要预测周期，核心原因在于价格聚集所刻画的交易者行为更接近于散户与非理性投资者的短期情绪状态。相关文献与前文实证结果均表明，此类情绪冲击在日内具有一定的延续性，但其对价格的影响往往需要通过隔夜价格调整才能充分体现。因此，将预测周期设定为 1 日，有助于在不过度拉长时间跨度的前提下，于中频层面捕捉价格聚集所蕴含的行为性信息。然而，当预测周期进一步延长至 1 日以上时，资产收益中往往会逐渐叠加显著的风格暴露，例如动量（Momentum）、订单不平衡（Imbalance）以及波动率（Volatility）等因子在时间维度上的累积效应。这类风格因子本身具有较强的预测能力，但在实际应用中通常伴随较高的拥挤度，其收益来源也具有较强的可替代性。

基于上述考虑，我们对其进行 de-risk 的独立性检验。在方法上，针对风格暴露的检验通常有两种思路：一是将未来收益（label）对常见风格因子进行回归，并基于残差计算因子的预测能力；二是直接考察所构建因子与典型风格因子之间的相关性。本文采用第二种方法，对因子与动量及订单不平衡等常见风格指标的相关性进行检验。具体而言，我们选取前一交易日的收盘到收盘收益率作为 Momentum 因子（Mom），并以当日最优买卖挂单数的订单不平衡率作为 Imbalance 因子

(OFI237m), 计算其与 60 日相对聚集度因子的相关系数。

表 5 Insert Day 与 Mom、OFI237m 的相关性

Window (x)	Mom	OFI237m
5 Days	0.0148	0.0903
10 Days	0.0091	0.0910
30 Days	0.0070	0.0936
60 Days	0.0066	0.0863

实证结果显示, 该因子与 Momentum 和 Imbalance 因子的相关系数均低于 0.1, 表明其与常见风格因子之间的线性相关性较弱。在此意义下, 可以认为该因子所反映的信息并非主要来源于风格性 Beta 暴露, 而是在去风险检验后仍保留了独立的预测能力, 从而具备一定的 Alpha 属性与实际应用价值。

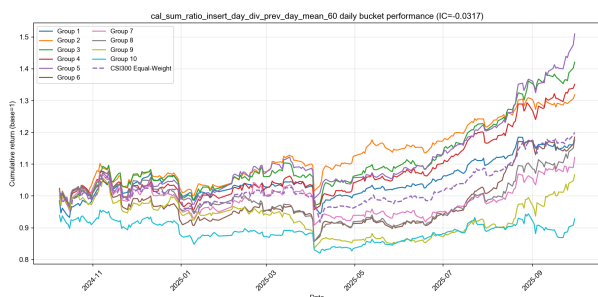


图 10 Buy Insert 60 日相对偏离度策略多空净值曲线

除此之外, 我们还根据既有因子从低到高, 将 300 只股票分为 10 组, 绘制了分组收益率曲线, 结果显示该因子选出的高低组在收益率上存在显著差异, 而且因子高组收益率 (空头端) 明显低于其他组, 进一步验证了其负向预测能力; 不过因子较低组的收益率 (多头端) 并未显著高于其他组。从制度层面看, 多头端收益表现相对受限的最主要原因在于 A 股市场长期存在的卖空约束。由于空头受限, 市场中具备正向预期的股票往往面临更为激烈的多头竞争, 价格中所包含的正向信息更容易被迅速反映, 从而压缩了低聚集度股票在简单分组框架下所能获得的超额收益空间。该制度特征使得多头收益的形成机制本身较为空头端更为复杂, 也更容易受到市场整体风险偏好与资金拥挤程度的影响。

7 结语

本文从市场微观结构视角出发, 围绕价格聚集效应在中国 A 股市场中的表现与可交易含义展开研究。基于沪深 300 成分股的高频逐笔 (Level-2) 数据, 我们首先在横截面层面刻画了价格尾数分布的显著非均匀性, 验证了整数聚集的普遍存在, 并进一步揭示了具有本土特征的“重 8 轻 4”文化偏好以及整数价位上更为复杂的流动性结构。相关证据表明, 价格聚集并非单一机制的结果, 而是认知锚定、文化偏好与策略性交易行为共同作用的产物。

在此基础上, 本文将具有显著偏好的尾数“058”视为非理性与散户交易参与程度的代理变量, 尝试将这一微观结构现象转化为可用于量化投资的截面信号。因子构建经历了从分钟级聚集度均值出发、引入波动率与分位筛选以缓解微观噪音干扰, 再到将时间粒度提升至日频并进行去趋势处理的迭代过程。最终提出的去趋势相对聚集度因子 RCR_{60} (以 60 日历史均值为基准) 能够有效剥离个股固有的“散户浓度”异质性, 获得较为稳健的 Rank IC 与多空组合收益表现; 并且与动量、订单不平衡等常见微观因子的相关性较低, 显示出一定的独立信息含量。上述结果表明, 价格聚集所蕴含的交易者结构信息在统计意义上具有可利用性, 为基于微观结构的因子设计提供了新的视角。

当然, 本文亦存在若干局限。其一, 在数据处理方面, 样本期相对有限, 因子的构建和最终的评测很难剥离该时段市场风格的固有特征; 同时, 对于上交所逐笔委托的 Insert 数据, 由于上交所提供的是首次撮合后的委托数据, 因而更精细化的测度需要基于逐笔成交的 Trade 数据对上交所的 Insert 数据进行还原。其二, 本文主要聚焦于“058”聚集比例这一参与结构信号, 尚未对整数邻近价位 (如尾数 1 与 9) 以及其他的有偏效应进行系统建模, 未来研究可在更长时间跨度与更丰富市场环境的数据基础上进行更多元化的测试与研究。而在因子构建维度, 目前的因子生成基于较为粗糙的主观经济学直觉, 未来可考虑采用 DL 模型、遍历搜索等方式进一步探究该有偏现象在因子工程领域的应用。其三, A 股卖空约束使多头端收益更易被竞争与拥挤效应压缩, 导致因子在空头端的预测

能力更为突出。对此,我们认为,简单的线性分组或极端分位策略可能不足以充分刻画低聚集区间内的有效信号。在更长时间跨度的数据支持下,未来研究可考虑通过非线性拟合等模型强化因子的预测能力,弱化极端样本的干扰,突出具有稳定预测能力的多头收益成分。

参考文献

- [1] Ahn, H.-J., Cai, J., and Cheung, Y.-L. (2005). Price clustering on the limit-order book: Evidence from the Stock Exchange of Hong Kong. *Journal of Financial Markets*, 8(4):421–451.
- [2] Ball, C. A., Torous, W. N., and Tschoegl, A. E. (1985). The degree of price resolution: The case of the gold market. *Journal of Futures Markets*, 5(1):29–43.
- [3] Bessembinder, H. (2003). Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis*, 38(4):747–777.
- [4] Bhattacharya, U., Holden, C. W., and Jacobsen, S. (2012). Penny wise, dollar foolish: Buy-sell imbalances on and around round numbers. *Management Science*, 58(2):413–431.
- [5] Blau, B. M. and Griffith, T. G. (2016). Price clustering and the stability of stock prices. *Journal of Business Research*, 69(10):3933–3942.
- [6] Blau, B. M., Griffith, T. G., and Whitby, R. J. (2022). Price clustering, preferences for round prices, and expected returns. *Journal of Behavioral Finance*, 23(3):301–315.
- [7] Brown, P. and Mitchell, J. (2008). Culture and stock price clustering: Evidence from the Peoples' Republic of China. *Pacific-Basin Finance Journal*, 16(1-2):95–120.
- [8] Cellier, A. and Bourghelle, D. (2007). Limit order clustering and price barriers on financial markets: empirical evidence from Euronext. *SSRN Electronic Journal*. Available at SSRN 966454.
- [9] Harris, L. (1991). Stock price clustering and discreteness. *Review of Financial Studies*, 4(3):389–415.
- [10] Johnson, E., Johnson, N., and Shanthikumar, D. (2007). Round numbers and security returns. *SSRN Electronic Journal*. Available at SSRN 972802.
- [11] Ke, W.-C., Chen, H., Lin, H.-W. W., and Liu, Y.-C. (2017). The impact of numerical superstition on the final digit of stock price. *The North American Journal of Economics and Finance*, 39:145–157.
- [12] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97.
- [13] Niederhoffer, V. (1965). Clustering of stock prices. *Operations Research*, 13(2):258–265.
- [14] Niederhoffer, V. (1966). A new look at clustering of stock prices. *Journal of Business*, 39(2):309–313.
- [15] Osborne, M. F. (1962). Periodic structure in the Brownian motion of stock prices. *Operations Research*, 10(3):345–379.
- [16] Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7(4):532–547.
- [17] Schelling, T. C. (1980). *The Strategy of Conflict: with a new Preface by the Author*. Harvard University Press.
- [18] Simon, H. A. (1974). How big is a chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science*, 183(4124):482–488.
- [19] Thomas, M. and Morwitz, V. (2005). Penny wise and pound foolish: The left-digit effect in price cognition. *Journal of Consumer Research*, 32(1):54–64.