

# 金融工程专题研究

## 高频订单成交数据蕴含的 Alpha 信息

### 核心观点

**深入高频：从日频行情数据到逐笔成交数据。**

**逐笔成交数据介绍：**逐笔成交数据主要记录了每笔成交的数量、价格、时间信息以及本次成交涉及的买卖双方的订单数量、价格和订单号信息。

**考察高频订单的几个维度：**对高频订单的考察通常可从订单大小、成交时长、成交时间、出价高低等维度入手。

#### 基于订单大小的大单交易占比因子

**传统大单交易占比因子：**采用分位点的方式对大单进行划分，构建“传统大单交易占比”因子。该因子展现出一定的选股能力，但是总体来讲绩效较为平庸，且稳定性较差。

**传统大单交易占比因子拆解：**根据每笔成交买卖订单是否为大单，对传统大单交易占比因子进行拆解，发现子因子方向不同，导致因子效果变差。

**改进大单交易占比因子：**我们对构成大单交易占比因子的子因子方向进行调整构建“改进后大单交易占比”。该因子 RankIC 均值达到 7.6%，年化 RankICIR 为 3.73，月胜率 88.1%，月度自相关系数 0.81。

#### 基于订单成交时长的漫长订单交易占比因子

**漫长订单交易占比因子：**采用分位点的方式对订单成交时长进行划分，构建“漫长订单交易占比”因子，其 RankIC 均值为 7.1%，年化 RankICIR 为 3.88，月胜率 84.5%，月度自相关系数 0.82。

#### 衡量订单特征的其他维度

我们还可以从“早尾盘属性”及“高低价属性”出发，对订单包含的信息进行多特征维度的衡量。

#### 大单及漫长订单复合因子

**大单及漫长订单复合因子：**将“改进大单交易占比”因子和“漫长订单交易占比”进行等权合成，构建“大单及漫长订单”复合因子。复合因子的 RankIC 均值为 8.4%，年化 RankICIR 为 4.39，月胜率 88.1%。周频调仓复合因子的周度 RankIC 均值为 5.1%，年化 RankICIR 为 5.42，周胜率 78.4%。

**复合因子特征：**复合因子更倾向于低估值、大市值、低波动、低换手股票。在剥离掉传统选股因子后，纯净因子仍展现出稳健的选股能力。

#### 基于订单特征的成交量占比因子的统一框架

**基于订单特征的成交量占比因子的统一框架：**基于委买单和委卖单的“大单属性”和“漫长属性”，对不同类别的成交记录进行分类，分别检验不同类型成交记录交易占比因子的选股能力。

**基于订单特征的精选复合因子：**精选复合因子的 RankIC 均值为 8.5%，年化 RankICIR 为 5.06，月胜率 90.5%。周频调仓精选复合因子的周度 RankIC 均值为 5.2%，年化 RankICIR 为 6.57，周胜率 82.6%。

**风险提示：**市场环境变动风险，本报告基于客观数据，不构成投资建议。

### 金融工程专题报告

#### 金融工程·量化投资

证券分析师：张欣慰      证券分析师：张宇  
021-60933159              021-60875169  
zhangxinwei1@guosen.com.cn      zhangyu15@guosen.com.cn  
S0980520060001              S0980520080004

#### 相关研究报告

《超预期投资全攻略》——2020-09-30  
《基于优秀基金持仓的业绩增强策略》——2020-11-15  
《基于分析师认可度的成长股投资策略》——2021-05-12  
《基于分析师推荐视角的港股投资策略》——2021-05-17  
《券商金股全解析——数据、建模与实践》——2022-02-18  
《战胜机构投资者——再论主动股基业绩增强策略》——2022-12-07  
《超额图谱视角下的成长股投资策略》——2023-06-07  
《北向因子能否长期有效？——来自亚太地区的实证》——2021-05-17  
《基于风险预算的中证 500 指数增强策略》——2021-10-20  
《动量类因子全解析》——2021-12-13  
《寻找业绩与估值的错配：非理性估值溢价因子》2021-12-15  
《反转因子全解析》——2022-06-14  
《价值类风险因子挖掘初探》——2022-06-20  
《隐式框架下的特质类因子改进》——2022-08-17  
《风险溢价视角下的动量反转统一框架》——2022-12-07  
《寻找关键时刻的领头羊——时点动量全解析》——2023-05-17

## 内容目录

|                                |    |
|--------------------------------|----|
| 深入高频：从日频行情数据到逐笔成交数据 .....      | 5  |
| 深入高频：从日频行情数据到逐笔成交数据 .....      | 5  |
| 逐笔成交数据介绍 .....                 | 7  |
| 考察高频订单的几个维度 .....              | 9  |
| 基于订单大小的大单交易占比因子 .....          | 10 |
| 大单定义及识别 .....                  | 10 |
| 传统大单交易占比因子构建 .....             | 11 |
| 传统大单交易占比因子拆解 .....             | 13 |
| 改进大单交易占比因子表现 .....             | 15 |
| 基于订单成交时长的漫长订单交易占比因子 .....      | 16 |
| 订单成交大小与成交时长 .....              | 16 |
| 漫长订单定义及识别 .....                | 18 |
| 漫长订单交易占比因子构建 .....             | 19 |
| 漫长订单交易占比因子拆解 .....             | 21 |
| 衡量订单特征的其他维度 .....              | 22 |
| 基于早尾盘属性对订单进行划分 .....           | 22 |
| 基于高低价属性对订单进行划分 .....           | 23 |
| 大单及漫长订单复合因子 .....              | 24 |
| 改进大单交易占比因子 VS 漫长订单交易占比因子 ..... | 24 |
| 大单及漫长订单复合因子 .....              | 26 |
| 复合因子在不同样本空间中的表现 .....          | 27 |
| 周频调仓复合因子表现 .....               | 29 |
| 复合因子与常见选股因子相关性 .....           | 30 |
| 基于订单特征的成交量占比因子的统一框架 .....      | 31 |
| 基于订单特征的精选复合因子 .....            | 32 |
| 周频调仓精选复合因子表现 .....             | 35 |
| 参数敏感性检验 .....                  | 36 |
| 改进大单交易占比因子参数敏感性分析 .....        | 36 |
| 漫长订单交易占比因子参数敏感性分析 .....        | 37 |
| 总结 .....                       | 38 |
| 风险提示 .....                     | 39 |

## 图表目录

|                                                    |    |
|----------------------------------------------------|----|
| 图 1: 完全相同的日度 K 线对应着完全不同的日内走势 .....                 | 5  |
| 图 2: 数据的形成过程与因子挖掘的演进过程 .....                       | 6  |
| 图 3: 从订单数据到行情数据的演变 .....                           | 6  |
| 图 4: 股票每日委托订单及成交记录数量分年度统计 (单位: 笔) .....            | 8  |
| 图 5: 高频订单成交数据考察维度 .....                            | 9  |
| 图 6: 不同市值分档下股票订单成交金额分位数情况 (单位: 元) .....            | 10 |
| 图 7: 大单划分示意图 .....                                 | 11 |
| 图 8: 大单买入交易占比因子及大单卖出交易占比因子相关系数 .....               | 12 |
| 图 9: 传统大单交易占比因子 RankIC 走势 .....                    | 13 |
| 图 10: 传统大单交易占比因子分组月均超额收益 .....                     | 13 |
| 图 11: 传统大单交易占比因子拆解 .....                           | 13 |
| 图 12: 基于委买单及委卖单的“大单属性”构造的不同类型成交记录的成交笔数及成交量占比 ..... | 14 |
| 图 13: 改进大单交易占比因子 RankIC 走势 .....                   | 15 |
| 图 14: 改进大单交易占比因子分组月均超额收益 .....                     | 15 |
| 图 15: 改进大单交易占比因子多头相对基准净值走势 .....                   | 16 |
| 图 16: 改进大单交易占比因子多空分年度超额收益 .....                    | 16 |
| 图 17: 订单大小与订单成交时长关系示意图 .....                       | 17 |
| 图 18: 成交耗时最长的 10% 订单在不同成交量分位区间的订单数量占比 .....        | 17 |
| 图 19: 漫长订单划分示意图 .....                              | 18 |
| 图 20: 大单买入交易占比及大单卖出交易占比因子相关系数 .....                | 19 |
| 图 21: 漫长订单交易占比因子 RankIC 走势 .....                   | 20 |
| 图 22: 漫长订单交易占比因子分组月均超额收益 .....                     | 20 |
| 图 23: 漫长订单交易占比因子多头相对基准净值走势 .....                   | 20 |
| 图 24: 漫长订单交易占比因子多空分年度超额收益 .....                    | 20 |
| 图 25: 基于委买单及委卖单的“时长属性”构造的不同类型成交记录的成交笔数及成交量占比 ..... | 21 |
| 图 26: 早尾盘复合交易占比因子 RankIC 走势 .....                  | 23 |
| 图 27: 早尾盘复合交易占比因子分组月均超额收益 .....                    | 23 |
| 图 28: 改进大单交易占比因子与漫长订单交易占比因子相关系数 .....              | 24 |
| 图 29: 大单交易占比残差因子 RankIC 走势 .....                   | 25 |
| 图 30: 大单交易占比残差因子分组月均超额 .....                       | 25 |
| 图 31: 漫长订单交易占比残差因子 RankIC 走势 .....                 | 25 |
| 图 32: 漫长订单交易占比残差因子分组月均超额 .....                     | 25 |
| 图 33: 复合因子 RankIC 走势 .....                         | 26 |
| 图 34: 复合因子分组月均超额收益 .....                           | 26 |
| 图 35: 复合因子多头相对基准净值走势 .....                         | 27 |
| 图 36: 复合因子多空分年度超额收益 .....                          | 27 |
| 图 37: 不同风格型股票池样本数量变化 .....                         | 28 |

|                                                   |    |
|---------------------------------------------------|----|
| 图 38: 复合因子在不同样本空间中的多空净值走势 .....                   | 29 |
| 图 39: 周频调仓复合因子 RankIC 走势 .....                    | 29 |
| 图 40: 周频调仓复合因子分组周度平均超额收益 .....                    | 29 |
| 图 41: 周频调仓复合因子多头相对基准净值走势 .....                    | 30 |
| 图 42: 周频调仓复合因子多空分年度超额收益 .....                     | 30 |
| 图 43: 漫长订单与选股因子相关系数 .....                         | 30 |
| 图 44: 复合纯净因子 RankIC 走势 .....                      | 31 |
| 图 45: 复合纯净因子分组月均超额收益 .....                        | 31 |
| 图 46: 精选复合因子 RankIC 走势 .....                      | 34 |
| 图 47: 精选复合因子分组月均超额收益 .....                        | 34 |
| 图 48: 精选复合因子多头相对基准净值走势 .....                      | 35 |
| 图 49: 精选复合因子多空分年度超额收益 .....                       | 35 |
| 图 50: 周频调仓精选复合因子 RankIC 走势 .....                  | 35 |
| 图 51: 周频调仓精选复合因子分组周度平均超额收益 .....                  | 35 |
| 图 52: 周频调仓精选复合因子多头相对基准净值走势 .....                  | 36 |
| 图 53: 周频调仓精选复合因子多空分年度超额收益 .....                   | 36 |
| 图 54: 不同阈值下改进大单交易占比因子累计 RankIC 走势 .....           | 37 |
| 图 55: 不同阈值下改进大单交易占比因子月均超额 .....                   | 37 |
| 图 56: 1 秒及以上订单在全部订单数量占比分布情况 .....                 | 37 |
| 图 57: 不同阈值下漫长订单交易占比因子累计 RankIC 走势 .....           | 38 |
| 图 58: 不同阈值下漫长订单交易占比因子月均超额 .....                   | 38 |
| 表 1: 逐笔成交数据的数据条目及说明 .....                         | 7  |
| 表 2: 逐笔成交数据示例 .....                               | 7  |
| 表 3: 基于委买单及委卖单的“大单属性”构造的子因子回测结果 .....             | 15 |
| 表 4: 传统大单交易占比因子和改进大单交易占比因子绩效对比 .....              | 16 |
| 表 5: 改进大单交易占比因子和漫长订单交易占比因子绩效对比 .....              | 21 |
| 表 6: 基于委买单及委卖单的“时长属性”构造的子因子回测结果 .....             | 22 |
| 表 7: 基于委买单及委卖单的“早尾盘属性”构造的子因子回测结果 .....            | 23 |
| 表 8: 基于委买单及委卖单的“高低价属性”构造的子因子回测结果 .....            | 23 |
| 表 9: 复合因子在不同样本空间中的表现 .....                        | 28 |
| 表 10: 基于大单属性和漫长属性对成交类型进行划分 .....                  | 32 |
| 表 11: 不同订单类型成交量占比因子表现 .....                       | 34 |
| 表 12: 大单及漫长订单复合因子和基于订单特征的精选复合因子绩效对比 .....         | 35 |
| 表 13: 周频调仓大单及漫长订单复合因子和周频调仓基于订单特征的精选复合因子绩效对比 ..... | 36 |

## 深入高频：从日频行情数据到逐笔成交数据

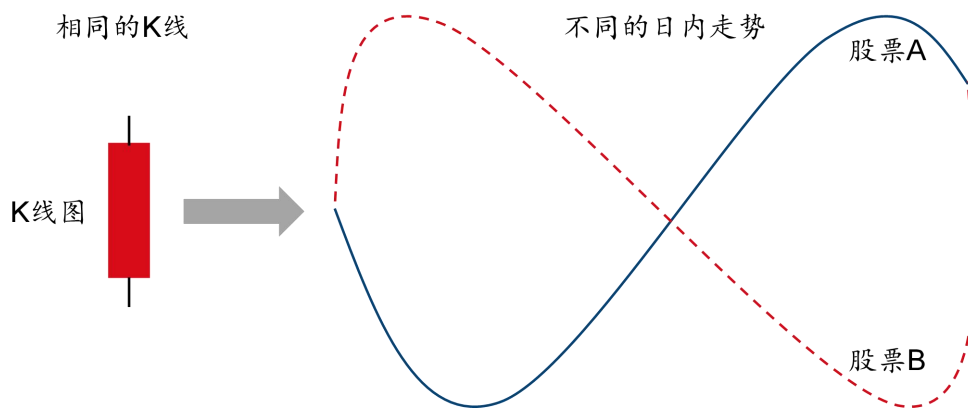
对于量化研究者而言，好的数据是一切策略的开始。在个股数据研究过程中，我们通常可以根据数据类别的不同将其划分为基本面数据、分析师预期数据、量价类行情数据和另类数据等。在这些数据中，个股量价类行情数据凭借其格式规整、覆盖度高、更新频率快、数据量大等特征，受到研究者的广泛关注。特别地，得益于机器学习、深度学习等前沿理论的发展和计算机算力的不断提升，机器学习、深度学习等算法被广泛地应用到量化策略的构建和实战过程中，量价类行情数据在这一应用领域发挥着重要的作用。

### 深入高频：从日频行情数据到逐笔成交数据

传统的量价类因子构建通常基于个股日频数据，通过高开低收及成交量等指标来描绘股价的波动幅度、成交的剧烈程度以及量价之间的关联程度等。国信金工团队在《动量类因子全解析-20211213》、《反转因子全解析-20220614》、《隐式框架下的特质类因子改进-20220817》和《风险溢价视角下的动量反转统一框架-20221207》等多篇专题报告中，对低频量价类因子的构建展开了讨论。然而随着时间的推移以及市场环境的变化，传统的低频量价因子已经不能满足研究者的需求，颗粒度更细、信息含量更为丰富的高频数据引起了市场的浓厚兴趣。

如图 1 所示，两个完全相同的日度 K 线可能对应着完全不同的日内走势。对于股票 A 和股票 B 而言，其当天开盘价、最高价、最低价和收盘价完全一致，因此两只股票的日 K 线完全相同。但是假如我们观察两只股票的日内走势，可以发现股票 A 在经历了开盘的短暂下跌后日内持续上扬，而股票 B 在经历了开盘的快速上涨后日内持续阴跌并于尾盘快速拉升。不同的日内价格走势意味着不同的资金博弈结果，这些结果对股价未来的走势具有一定的预测能力，而遗憾的是这些信息无法在日频数据中得以展现。

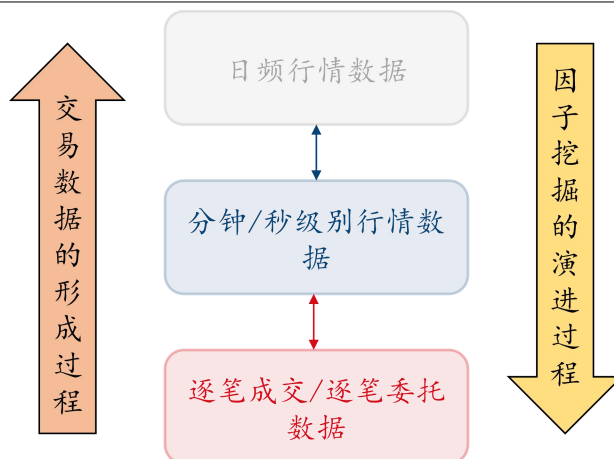
图1：完全相同的日度 K 线对应着完全不同的日内走势



资料来源：国信证券经济研究所绘制

为了进一步探索量价数据中的信息，近年来的量价因子构建开始向高频数据深入，例如使用个股的分钟频率数据、秒级别数据等。数据频率的提升使研究者们可以捕捉到股票更多的量与价的变动细节，考察到日频数据所不能刻画股票状态与中间过程。然而这些分钟级或秒级别的数据条目仍以高开低收及成交量等数据为主，均为集合竞价与连续竞价中订单经过撮合后形成的结果，与日频量价数据相比，没有发生根本性的变化。

图2：数据的形成过程与因子挖掘的演进过程

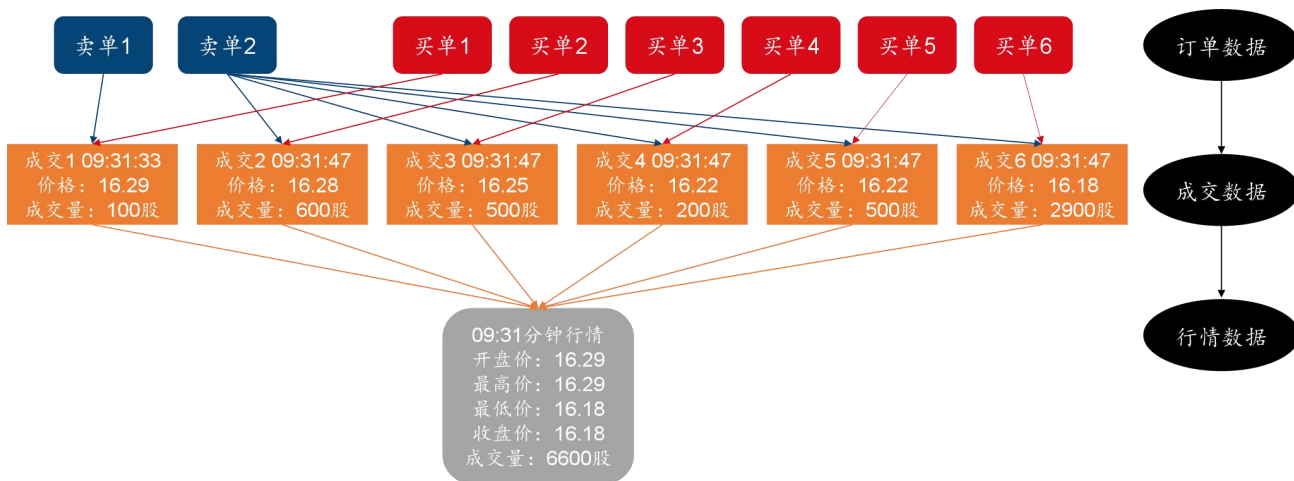


资料来源：国信证券经济研究所绘制

本文，我们将深入到行情数据形成的源头——逐笔成交数据，更细致地了解不同买卖订单的角力与碰撞、不同投资者之间的分歧与跟随，以及股票价格变化的全部过程，从而更深刻地理解过往量价数据的变化对未来股票收益的预测作用。

图3展示了从订单数据到行情数据的演变。市场如同深海，如果说订单数据是推动潮涨潮落的海水，那么行情数据则更像是随着海浪潮汐而高低涌动的帆船。股价行情的波动是由背后一个个投资者交易出来的结果，因此探寻每个投资者下单背后的逻辑及其信息优势，对于我们了解股价背后的资金博弈程度及未来可能的走势具有极强的指导意义。

图3：从订单数据到行情数据的演变



资料来源：国信证券经济研究所绘制

## 逐笔成交数据介绍

逐笔成交数据，作为颗粒度最细的数据，记录了集合竞价和连续竞价撮合成功的每一笔成交记录，是一切行情数据产生的源头。逐笔成交数据主要记录了每笔成交的数量、价格、时间信息以及本次成交涉及的买卖双方的订单数量、价格信息。如表 1 所示，本文使用的逐笔成交数据主要包含以下 12 个条目，其中对于同一只股票而言，其交易流水号是每一笔成交记录的唯一标识，因此所有交易流水号不重复。

表1: 逐笔成交数据的数据条目及说明

| 数据条目  | 说明                                        |
|-------|-------------------------------------------|
| 股票代码  | 股票代码                                      |
| 交易流水号 | 逐笔交易的流水号，所有号码不重复                          |
| 成交时间  | 该笔交易成交的时间                                 |
| 成交价格  | 该笔交易成交的价格                                 |
| 成交量   | 该笔交易成交的股票数量（单位：股）                         |
| 成交类型  | 该笔交易为主动买入（B）还是主动卖出（S），通常以买卖订单中委托时间靠后的订单为准 |
| 委卖量   | 委托卖出单的下单股票总数量（单位：股）                       |
| 委买量   | 委托买入单的下单股票总数量（单位：股）                       |
| 委卖 ID | 委托卖出单的 ID，相同 ID 表示同一笔委托订单                 |
| 委买 ID | 委托买入单的 ID，相同 ID 表示同一笔委托订单                 |
| 委卖价格  | 委托卖出单的挂单价                                 |
| 委买价格  | 委托买入单的挂单价                                 |

资料来源：国信证券经济研究所整理

表 2 以贵州茅台（600519.SH）在 2023 年 10 月 31 日的几笔交易数据为例，对逐笔数据的定义进行阐述。为了便于后文对相关概念的理解，我们进行如下定义：

- 一笔成交：是指由买卖双方共同撮合的一笔交易记录；
- 一笔订单：同一个委买 ID（或委卖 ID）表示相同的一笔买单（或卖单）。同一笔订单仅代表同一个投资者在同一时间下达的订单，若同一个投资者在不同时间下达多笔订单，则会采用不同的委托 ID 进行表示。

由于连续竞价阶段的订单成交遵循“价格优先、时间优先”规则，因此随着盘口价格的不断变化，同一笔订单可能会跟多个对手方进行交易、同一笔订单分成多笔成交才完成交易的情况。

表2: 逐笔成交数据示例

| 股票代码   | 交易流水号 | 成交时间     | 成交价格    | 成交量 | 成交类型 | 委卖量  | 委买量  | 委卖 ID    | 委买 ID    | 委卖价格    | 委买价格    |
|--------|-------|----------|---------|-----|------|------|------|----------|----------|---------|---------|
| 600519 | 15261 | 14:53:10 | 1686.34 | 200 | S    | 200  | 1000 | 11717480 | 11714165 | 1686.34 | 1686.34 |
| 600519 | 15267 | 14:53:20 | 1686.34 | 100 | S    | 100  | 1000 | 11724662 | 11714165 | 1686.34 | 1686.34 |
| 600519 | 15303 | 14:53:45 | 1686.6  | 100 | S    | 1100 | 100  | 11743118 | 11742486 | 1686.34 | 1686.6  |
| 600519 | 15304 | 14:53:45 | 1686.36 | 100 | S    | 1100 | 100  | 11743118 | 11726157 | 1686.34 | 1686.36 |
| 600519 | 15305 | 14:53:45 | 1686.36 | 100 | S    | 1100 | 100  | 11743118 | 11741046 | 1686.34 | 1686.36 |
| 600519 | 15306 | 14:53:45 | 1686.34 | 700 | S    | 1100 | 1000 | 11743118 | 11714165 | 1686.34 | 1686.34 |
| 600519 | 15307 | 14:53:45 | 1686.34 | 100 | B    | 1100 | 100  | 11743118 | 11743299 | 1686.34 | 1686.34 |

资料来源：WIND，国信证券经济研究所整理

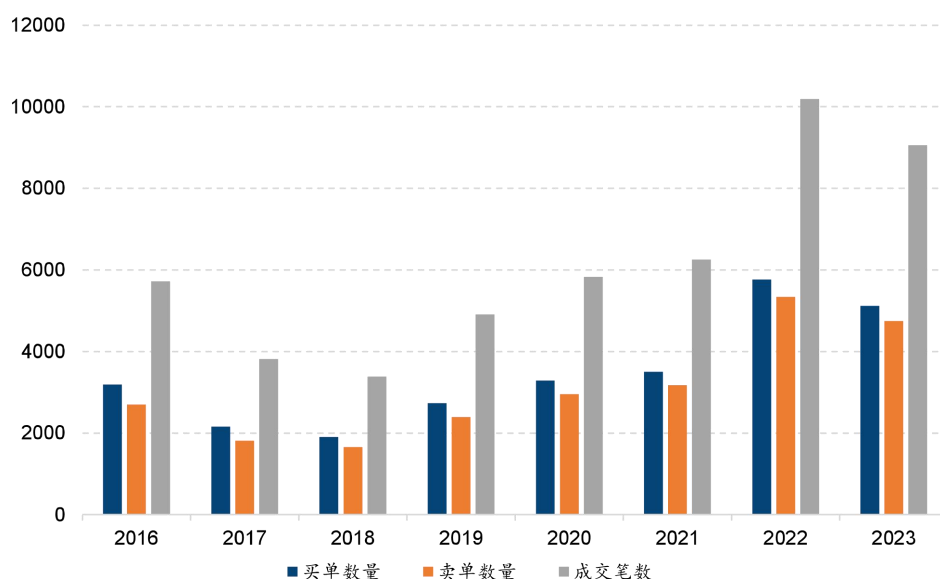
在表 2 中，红色方框标出的第一行、第二行和第六行记录了委买 ID 为 11714165（总委买量为 1000 股）的买入订单成交情况——在先后成交了 200 股和 100 股之后，股价上涨至高于该笔订单委买价的价格，此时其他委买价格更高的订单具有更高的成交优先级。直至 14:53:45 时刻股价回落至该笔订单委买价，才将剩余

的 700 股成交完成。在该案例中，委买 ID 为 11714165 的订单为“一笔订单”，交易流水号为 15261、15267 和 15306 的成交记录为三笔成交。

从委卖 ID 为 11743118（总委卖量为 1100 股）的卖出订单视角来看，由于其委卖量较大，因此最终分成 5 次完成交易，每次交易对手方都截然不同，表 2 中的后 5 行即对应其成交记录。

我们统计了每只股票每天的委买单数量、委卖单数量和总成交数量，随后对每只股票按照年度划分取其日度统计中位数，最后对所有股票取中位数，得到股票每日委托订单及成交记录数量的分年度统计，其结果如图 4 所示。可以看到，近年来委托数量和成交笔数均呈现出上升趋势。

图4：股票每日委托订单及成交记录数量分年度统计（单位：笔）



资料来源：WIND，国信证券经济研究所整理

## 考察高频订单的几个维度

不同类型的成交记录代表着不同类型投资者博弈之后的结果，因此不同类型的成交记录对个股未来预期收益的影响可能存在区别。本文中，我们将从多个维度对成交记录进行划分，并将相同类型的成交记录的实际交易量进行加总，除以全天的成交量，构建成交量占比因子，具体方式如下：

$$VolumeRatio_s = \frac{Volume_s}{VolumeAllDay}$$

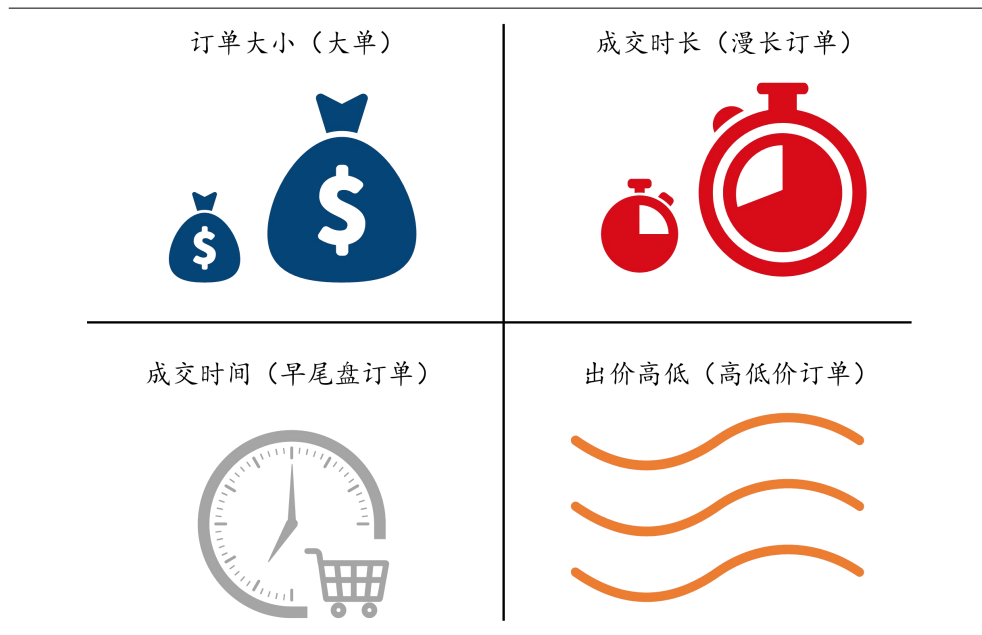
其中， $VolumeRatio_s$ 表示类型为  $s$  的成交记录的交易量之和占全天交易量的比例， $Volume_s$ 表示类型为  $s$  的成交记录的交易量之和， $VolumeAllDay$ 表示全天交易量之和。

对高频订单的考察通常可以分为如下几个维度：订单大小、成交时长、成交时间、成交价格。图 5 从不同维度对其进行了总结，具体来讲：

- **订单大小**：根据订单的成交大小可以将其划分为超大单、大单、小单等，通常来讲大单交易者可能拥有更多的信息优势及筹码优势；
- **成交时长**：根据订单的成交速度或成交时长可以将其划分为漫长订单、非漫长订单等，通常来讲漫长订单交易者更可能是长线资金；
- **成交时间**：根据订单的成交时间可以将其划分为早盘订单、尾盘订单等，通常来讲早尾盘订单交易者可能拥有更多的信息优势；
- **成交价格**：根据订单的委托价格可以将其划分为高价单、低价单等，通常来讲订单的委托价格反映了投资者对于股价“公允价值”的投票结果。

在后文中，我们将从**订单大小、成交时长、成交时间和成交价格等维度**对高频订单所蕴含的 Alpha 信息进行重点考察，其他维度的划分同样可以在本文提出的统一框架中进行分析。

图5：高频订单成交数据考察维度



资料来源：国信证券经济研究所绘制

## 基于订单大小的大单交易占比因子

对于高频订单的考察，最常用的角度即为订单大小。研究者通常依据订单成交额的大小划分为大单、小单，用以表示不同类型的投资者。例如在 Wind 数据库中，将单笔成交金额大于 100 万元的订单划分为机构投资者发起的超大单。

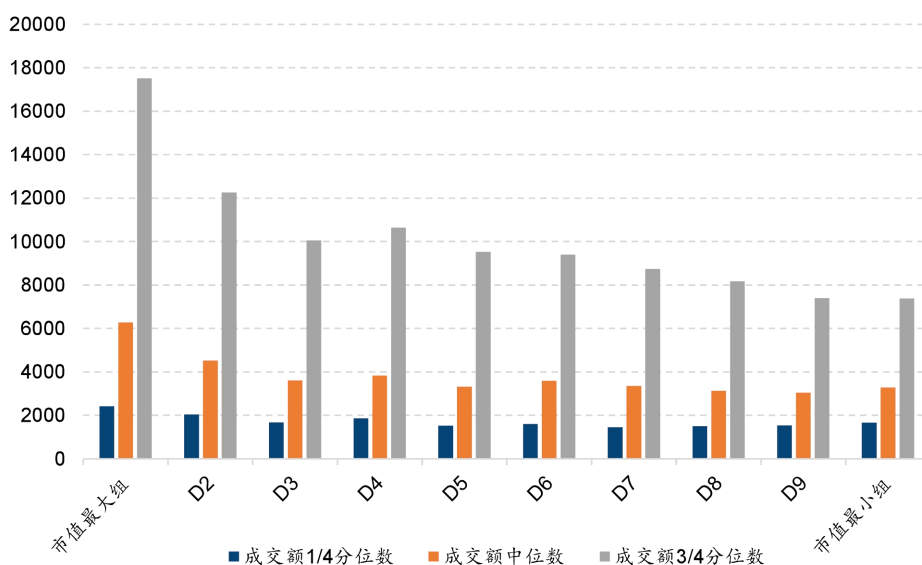
大单通常由机构投资者或资金体量大的个人投资者发起，这些大单的发起者由于其规模优势及专业能力，往往在获取信息的速度、处理信息的能力以及依据信息做出决策的能力上，胜过资金体量较小的个人投资者，因此大单行为的背后也隐藏着这类“优势投资者”的观点和对个股的判断。若能对这些“优势投资者”的行为动向进行跟踪，将有助于我们对个股未来的表现强弱进行判断。

### 大单定义及识别

近年来，研究人员通常根据订单的绝对金额作为标准来对大小单进行划分。例如在 Wind 数据库中，将单笔成交金额小于 4 万元的划分为小单，4 万元到 20 万元的订单划分为中单，20 万元至 100 万元的订单划分为大单，100 万元以上的订单划分为超大单。然而，对于不同市值、股价的股票，这种绝对金额的划分方法适用程度各不相同。

图 6 展示了不同市值分档下股票订单成交金额的 1/4 分位数、中位数和 3/4 分位数情况，可以看到大市值股票的订单金额普遍高于小市值股票。

图6：不同市值分档下股票订单成交金额分位数情况（单位：元）

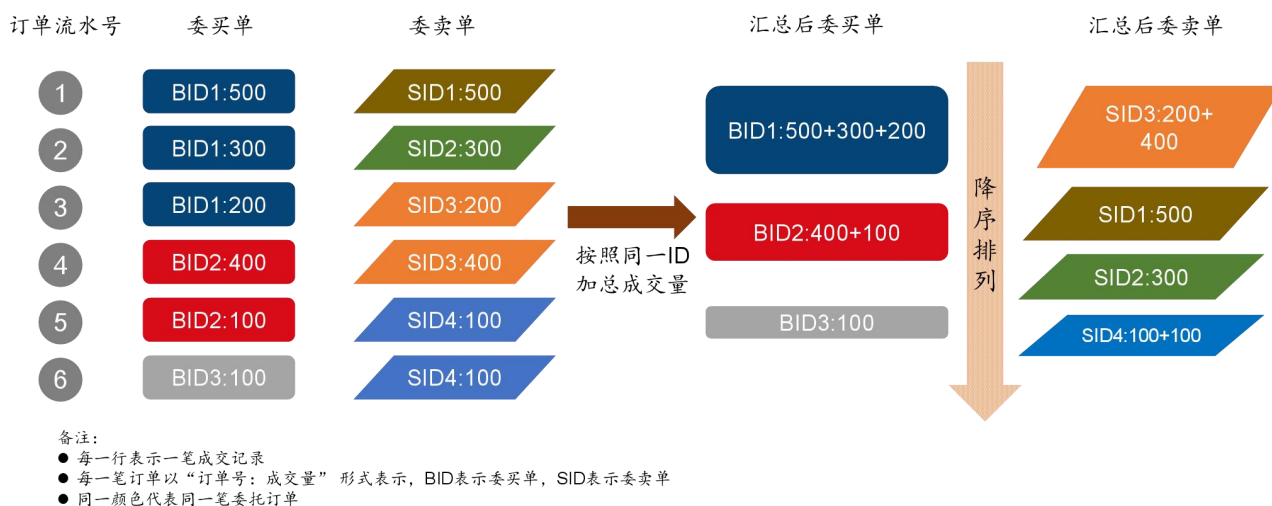


资料来源：WIND，国信证券经济研究所整理

为了增加大小单划分在不同股票中的适用性，本文采用分位点的方式对逐笔成交数据中的大单进行划分，示意图如图 7 所示。在该图中，每一行表示一笔成交记录；每一笔订单以“订单号：成交量”表示，BID 表示委买单，SID 表示委卖单，同一颜色表示同一笔委托订单。

图7中共展示了6笔成交记录，以第1笔成交为例，该笔成交实际成交量为500股，由编号为BID1的委买单和编号为SID1的委卖单撮合完成。在该案例中，编号为BID1的委买单分成了3笔完成交易，总共成交1000股；编号为SID1的委卖单仅用1笔即完成交易，总共成交500股。

图7：大单划分示意图



资料来源：国信证券经济研究所绘制

在对大单进行判定时，我们对每只股票每天的委买单和委卖单分别进行统计。具体来讲，首先将同一委买ID（或委卖ID）对应的实际成交量进行加总，随后将全部实际成交的委买ID（或委卖ID）成交量进行降序排列，最后取前10%分位点作为大单划分阈值（剔除开盘集合竞价的成交记录），将成交量大于前10%分位点的委买单（或委卖单）记为“大买单”（或“大卖单”）。

大单的判定可用如下公式表示：

$$BigOrderFlag_i = \begin{cases} 1, & \text{if } Volume_i > quantile(Volume, 0.9) \\ 0, & \text{else} \end{cases}$$

其中， $BigOrderFlag_i$ 为订单*i*是否为大单的标识， $Volume_i$ 为订单*i*的实际成交量， $quantile(Volume, 0.9)$ 表示当日所有同类型订单实际成交量的前10%分位数。在判断委买单是否为大单时，我们仅在所有委买单中进行比较。在判断委卖单是否为大单时，我们仅在所有委卖单中进行比较。

### 传统大单交易占比因子构建

在对大单的定义进行界定后，我们即可以当日所有大买单成交量之和占个股当日成交量的比例，构建“大单买入交易占比”因子（ $VolumeBigBuy$ ），具体计算公式如下：

$$VolumeBigBuy = \frac{\sum VolumeBigBuy_i}{Volume}$$

其中， $VolumeBigBuy$ 表示股票的大单买入交易占比， $VolumeBigBuy_i$ 表示该股票当日大买单*i*的实际成交量， $Volume$ 表示股票当日全天总成交量。

同样地，我们可以当日所有大卖单的成交量之和占个股当日成交量的比例，构建“大单卖出交易占比”因子（ $VolumeBigSell$ ），具体计算公式如下：

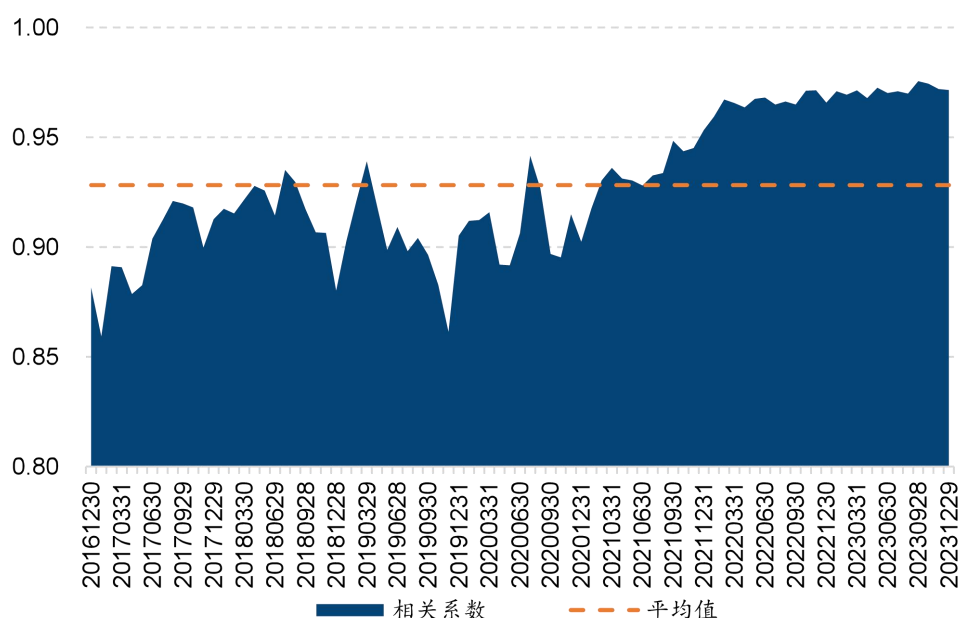
$$VolumeBigSell = \frac{\sum VolumeBigSell_i}{Volume}$$

其中， $VolumeBigSell$ 表示股票的大单卖出交易占比， $VolumeBigSell_i$ 表示该股票当日大卖单*i*的实际成交量， $Volume$ 表示股票当日全天总成交量。

在实际研究中我们发现，当一个股票日内成交中包含的“大买单”越多时，其对应发生的“大卖单”往往也会越多，二者呈现出显著的正相关关系。

图 8 展示了“大单买入交易占比”因子（ $VolumeBigBuy$ ）和“大单卖出交易占比”因子（ $VolumeBigSell$ ）之间在每个月月末的横截面相关系数，可以看到二者的平均相关系数高达 0.93。

图8: 大单买入交易占比因子及大单卖出交易占比因子相关系数



资料来源：WIND，国信证券经济研究所整理

基于此，我们将 $VolumeBigBuy$ 因子和 $VolumeBigSell$ 因子直接合成，构建“传统大单交易占比”因子（ $VolumeBigOrigin$ ），其计算方法如下：

$$VolumeBigOrigin = VolumeBigBuy + VolumeBigSell$$

$$= \frac{\sum VolumeBigBuy_i + \sum VolumeBigSell_i}{Volume}$$

其中， $VolumeBigOrigin$ 表示传统大单交易占比因子， $VolumeBigBuy_i$ 和 $VolumeBigSell_i$ 分别为大买单和大卖单的实际成交量， $Volume$ 表示股票在当日的全天总成交量。

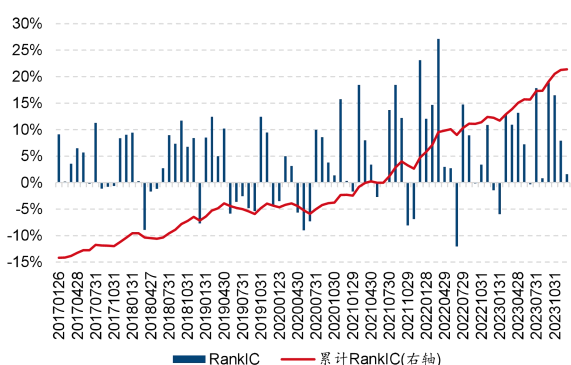
为了保证指标的稳定性，我们先根据每日数据计算个股的“传统大单交易占比”因子（ $VolumeBigOrigin$ ），随后以其过去 20 天的均值得到最终因子取值。

图9和图10展示了“传统大单交易占比”因子（*VolumeBigOrigin*）在全市场中的RankIC走势和分组月均超额收益情况（相对全市场等权组合）。在后续的因子测试中，如无特别说明，我们均已进行了行业及市值中性化处理，回测区间为2016年12月30日至2023年12月29日。

可以看到，自2017年以来，“传统大单交易占比”因子（*VolumeBigOrigin*）的RankIC均值为4.8%，年化RankICIR为2.05，月胜率66.7%。从分组月均超额收益分布来看，呈现出一定的单调性，多头月均超额收益0.65%，空头月均超额收益-0.66%。

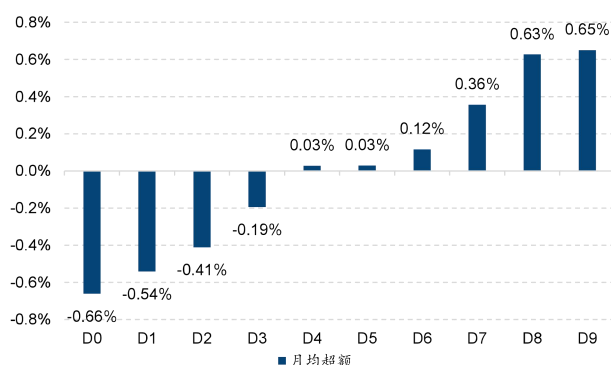
从因子的稳定性来看，“传统大单交易占比”因子月度自相关系数达到0.90，总体来看稳定性较强，从而使得因子表现不会出现过快的衰减。

图9：传统大单交易占比因子RankIC走势



资料来源：WIND，国信证券经济研究所整理

图10：传统大单交易占比因子分组月均超额收益

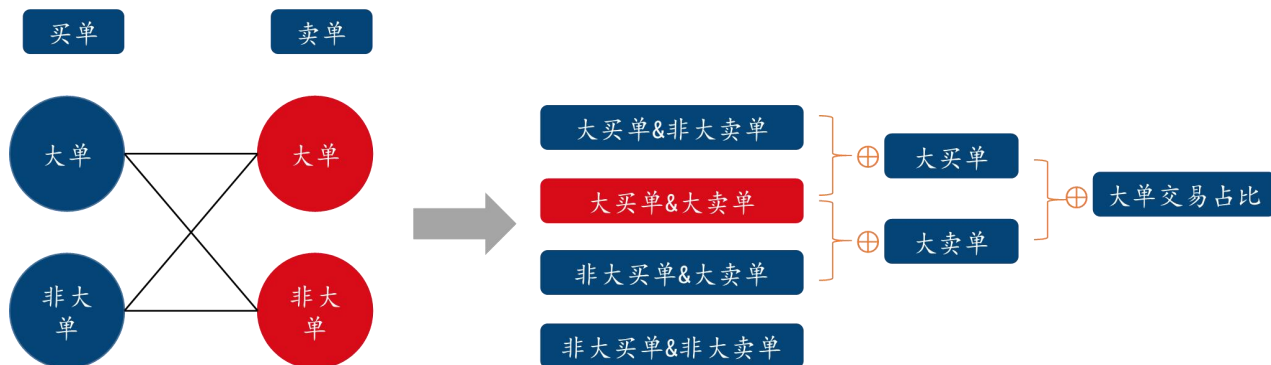


资料来源：WIND，国信证券经济研究所整理

## 传统大单交易占比因子拆解

由上一小节可知，“传统大单交易占比”因子（*VolumeBigOrigin*）展现出一定的选股能力，但是总体来讲绩效较为平庸，且月胜率相对较低。本小节，我们从促成成交的委买单和委卖单的“大单属性”出发，对传统大单交易占比因子进行进一步拆解。

图11：传统大单交易占比因子拆解

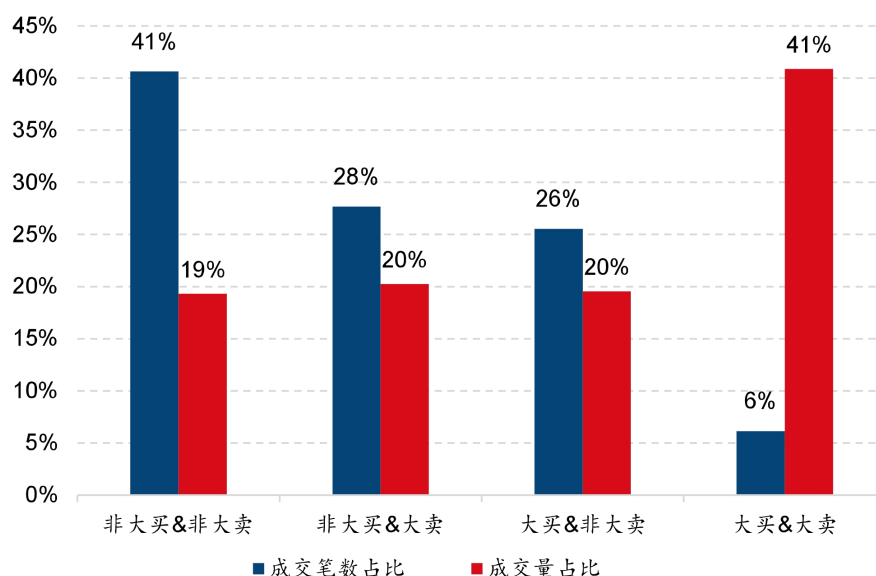


资料来源：国信证券经济研究所绘制

由于每一笔订单均是由委买单和委卖单双方撮合交易得到的结果，因此我们可以根据委买单和委卖单是否为“大单”，将每一笔成交记录划分为如图 11 所示的 4 种成交类型，对“传统大单交易占比”因子进行拆解。

图 12 展示了不同类型成交记录的成交笔数及成交量占比情况，以委买单和委卖单均为大单的成交记录为例，该类型订单的成交笔数占比仅有 6%，但其贡献的成交量占比却高达 41%；以委买单和委卖单均为非大单的成交记录为例，该类型订单的成交笔数占比高达 41%，但其贡献的成交量占比却仅有 19%。从各类型订单的成交量占比分布情况来看，不同类型成交记录的成交量占比相对均衡。

图12: 基于委买单及委卖单的“大单属性”构造的不同类型成交记录的成交笔数及成交量占比



资料来源：WIND，国信证券经济研究所整理

基于如上分析，我们可以根据每一笔成交的委买单和委卖单是否为大单，对“传统大单交易占比”因子（ $VolumeBigOrigin$ ）进行如下拆解：

$$\begin{aligned}
 \text{传统大单交易占比} &= \frac{\text{大买单} + \text{大卖单}}{\text{总成交量}} \\
 &= \frac{(\text{大买}\&\text{非大卖} + \text{大买}\&\text{大卖}) + (\text{非大买}\&\text{大卖} + \text{大买}\&\text{大卖})}{\text{总成交量}} \\
 &= \frac{\text{大买}\&\text{非大卖} + \text{非大买}\&\text{大卖} + 2 \times \text{大买}\&\text{大卖}}{\text{总成交量}}
 \end{aligned}$$

可以看到，“传统大单交易占比”因子（ $VolumeBigOrigin$ ）可以被拆解为三个子因子加总得到的结果：（1）委买单为大单、委卖单为非大单；（2）委买单为非大单、委卖单为大单；（3）委买单及委卖单均为大单。其中类型（3）订单的成交量在因子构建过程中被放大了 1 倍。

表 3 展示了拆解后的三个子因子在全市场中的回测表现，可以看到三个因子对股价未来的影响方向并不完全相同，其中大买&非大卖和非大买&大卖因子的 RankIC 均值为负，而大买&大卖因子的 RankIC 均值为正。由前述分析可知，“传统大单交易占比”因子为如上三个因子简单加总之后的结果。但是这三个因子之

间的方向并不相同，因此直接相加会导致复合因子的有效性出现削弱，我们需要对其方向进行调整。

表3: 基于委买单及委卖单的“大单属性”构造的子因子回测结果

| 因子名称   | RankIC 信息 |        |             | 月均超额收益 |        |
|--------|-----------|--------|-------------|--------|--------|
|        | 月胜率       | 均值     | 年化 RankICIR | 多头-基准  | 空头-基准  |
| 大买&非大卖 | 22.62%    | -5.26% | -2.39       | 0.50%  | -0.91% |
| 非大买&大卖 | 32.14%    | -3.29% | -1.47       | 0.40%  | -0.61% |
| 大买&大卖  | 79.76%    | 6.30%  | 3.03        | 0.71%  | -1.04% |

资料来源: WIND, 国信证券经济研究所整理

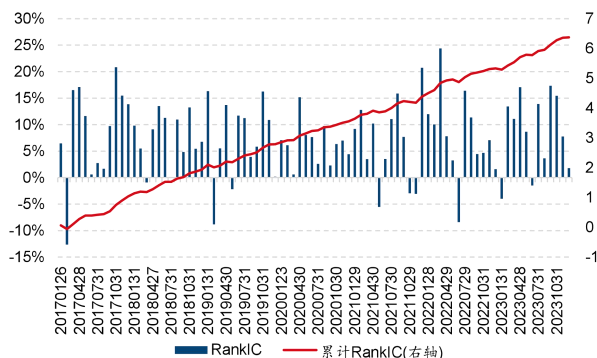
## 改进大单交易占比因子表现

我们对上一小节中提到的子因子方向进行调整，构建“改进大单交易占比”因子（*VolumeBig*），具体计算公式如下：

$$\text{改进大单交易占比} = \frac{(-1) \times \text{大买\&非大卖} + (-1) \times \text{非大买\&大卖} + \text{大买\&大卖}}{\text{总成交量}}$$

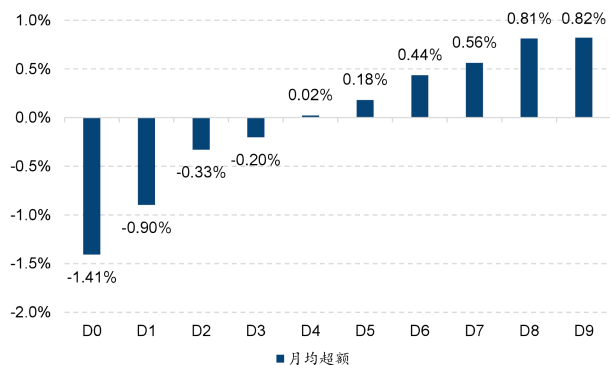
图 13 和图 14 展示了“改进大单交易占比”因子（*VolumeBig*）在全市场中的 RankIC 走势和分组月均超额收益情况。自 2017 年以来，该因子的 RankIC 均值达到 7.6%，年化 RankICIR 为 3.73，月胜率 88.1%。从分组月均超额收益分布来看，各组单调性特征十分明显，多头月均超额收益 0.82%，空头月均超额收益 -1.41%。总体来看，“改进大单交易占比”因子（*VolumeBig*）的绩效表现全面优于“传统大单交易占比”因子（*VolumeBigOrigin*）的表现。

图13: 改进大单交易占比因子 RankIC 走势



资料来源: WIND, 国信证券经济研究所整理

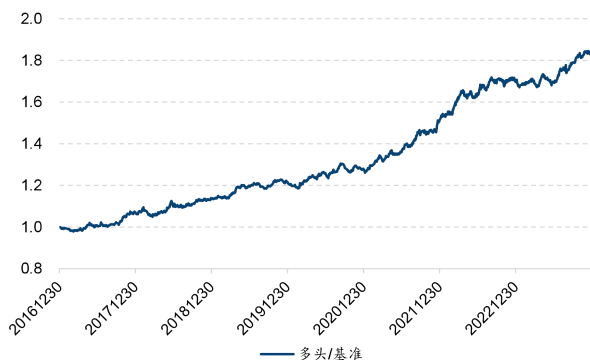
图14: 改进大单交易占比因子分组月均超额收益



资料来源: WIND, 国信证券经济研究所整理

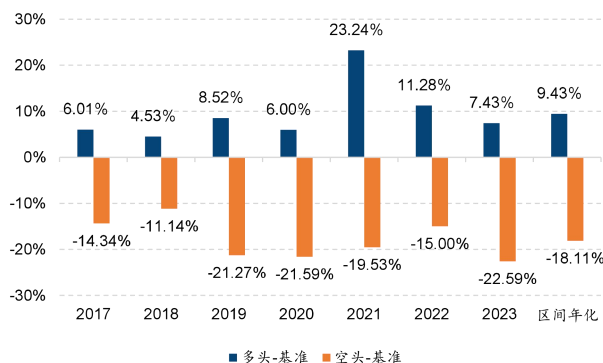
图 15 展示了“改进大单交易占比”（*VolumeBig*）因子多头相对基准的净值走势，图 16 展示了该因子多头和空头相对基准的分年度表现。其中，多头是指因子得分最高的前 1/10 组构建的等权组合，空头是指因子得分最低的 1/10 组构建的等权组合，基准指数采用全市场等权组合。在计算组合净值时，我们采用双边千分之三的费用进行扣除，且这一处理在后文中保持一致。可以看到该因子的多头表现依然十分稳健，自 2017 年以来每年都能够稳定战胜等权基准，多头组合年化超额收益 9.43%，空头组合年化超额收益-18.11%。

图15: 改进大单交易占比因子多头相对基准净值走势



资料来源: WIND, 国信证券经济研究所整理

图16: 改进大单交易占比因子多空分年度超额收益



资料来源: WIND, 国信证券经济研究所整理

表 4 展示了“传统大单交易占比”与“改进大单交易占比”因子在绩效上的表现, 可以看到改进后指标的绩效要全面优于传统指标的表现。

表4: 传统大单交易占比因子和改进大单交易占比因子绩效对比

|          | RankIC 均值 | RankICIR | RankIC 胜率 | 多头年化超额 | 空头年化超额  | 月度自相关系数 |
|----------|-----------|----------|-----------|--------|---------|---------|
| 传统大单交易占比 | 4.8%      | 2.05     | 66.7%     | 7.04%  | -9.03%  | 0.90    |
| 改进大单交易占比 | 7.6%      | 3.73     | 88.1%     | 9.43%  | -18.11% | 0.81    |

资料来源: WIND, 国信证券经济研究所整理

## 基于订单成交时长的漫长订单交易占比因子

### 订单成交大小与成交时长

衡量订单信息的另一个重要维度就是时长。我们定义订单成交时长为订单第一次成交到最后一次成交间隔的连续竞价时长, 具体如下:

$$\text{订单成交时长} = T_n - T_1$$

假设该笔订单总共分 $n$ 笔成交, 则上式中 $T_n$ 为最后一次成交的时间,  $T_1$ 为第一次成交的时间。在计算订单成交时长时, 我们剔除开盘集合竞价时成交的记录, 同时午休时间也进行剔除。

成交耗时不同的订单, 也蕴含着不同的信息。从直观的感受来看, 订单成交耗时的长短很大程度受到订单大小的影响, 但是否订单成交量越大, 订单成交耗时一定越长; 反之, 订单成交量越小, 订单成交耗时一定越短呢?

答案是否定的, 我们以 2023 年 10 月 31 日贵州茅台 (600519.SH) 的几笔订单为例, 初步感知一笔成交较快的大单和成交较慢的小单。

在图 17 中, 10:18:11 完成了一笔交易, 买卖双方出价一致, 下单股票数量都为 1000 股, 因此迅速成交, 耗时不足 1 秒。而 13:33:54 一笔下单量为 200 股的买入订单成交了 100 股后, 开始了漫长等待, 直到 5 分 12 秒后的 13:39:06 才和另一笔卖出订单成交了另外 100 股。在该案例中, 小单的成交时长比大单要长很多。

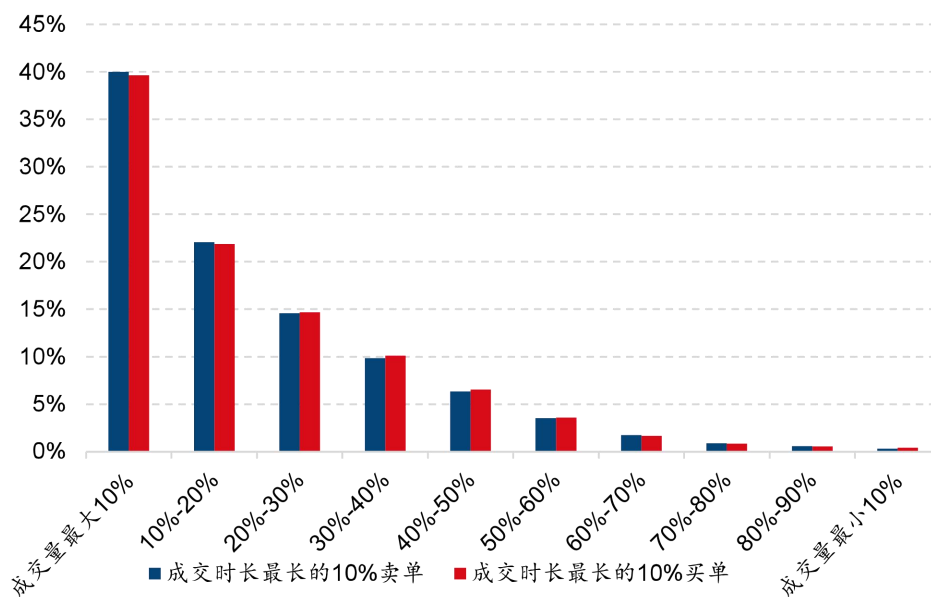
图17: 订单大小与订单成交时长关系示意图



资料来源: WIND, 国信证券经济研究所整理

那么订单成交时长与订单成交量大小之间的关系如何? 我们统计了每只股票成交耗时最长的10%订单在不同成交量分位区间的订单数量占比, 结果如图18所示。

图18: 成交耗时最长的10%订单在不同成交量分位区间的订单数量占比



资料来源: WIND, 国信证券经济研究所整理

由图 18 可以看到，成交耗时最长的 10% 订单中，仅有 40% 左右数量的订单处于成交量最大的前 10% 区间内，随着订单体量逐渐减小，其占比也逐渐降低。总体来看，订单的大小虽然与成交时长呈现出一定的正相关关系，但是二者之间仍然存在一定区别——成交耗时最长的订单并非全部都是成交量最大的订单，二者之间不能完全划等号。

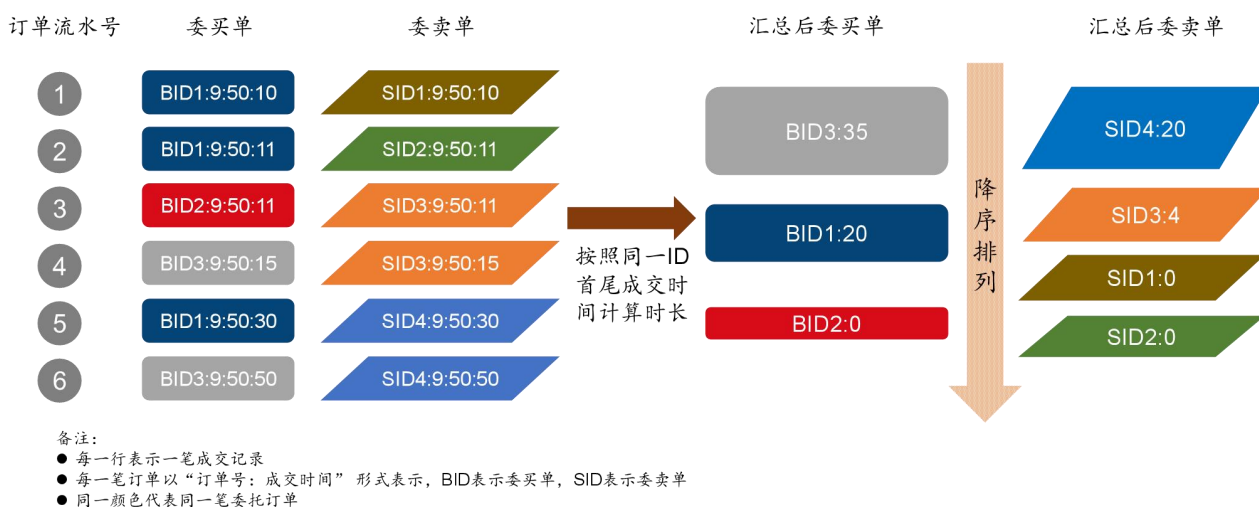
我们认为，订单的大小与成交时长之间最重要的区别在于“分歧”：分歧较大时，大单也可以迅速找到和自己交易方向相反的对手单，从而短时间内完成交易；分歧较小时，小单也很难找到和自己交易相反的对手单，所以需要花费更长的时间来完成交易。

## 漫长订单定义及识别

与大单划分方法类似，本文采用分位点的方式对逐笔成交数据中的漫长订单进行划分，示意图如图 19 所示。在该图中，每一行表示一笔成交记录；每一笔订单以“订单号：成交时间”表示，BID 表示委买单，SID 表示委卖单，同一颜色表示同一笔委托订单。

图 19 中共展示了 6 笔成交记录，以第 1 笔成交为例，该笔成交由编号为 BID1 的委买单和编号为 SID1 的委卖单撮合完成。在该案例中，编号为 BID1 的委买单第一次成交时间为 9:50:10，最后一次成交时间为 9:50:30，因此其成交时长为 20 秒；编号为 SID1 的委卖单仅用一次即完成交易，因此其成交时长为 0 秒。

图19：漫长订单划分示意图



资料来源：国信证券经济研究所绘制

在对漫长订单进行判定时，我们对委买订单和委卖订单分别进行统计，先计算不同委买 ID（或委卖 ID）订单的成交时长，随后将全部实际成交的委买 ID（或委卖 ID）成交时长进行降序排列，取前 10% 分位点作为漫长订单划分阈值（剔除开盘集合竞价的成交记录），最后将成交时长大于前 10% 分位点的委买单（或委卖单）记为“漫长买单”（或“漫长卖单”）。

漫长订单的判定可用如下公式表示：

$$LongOrderFlag_i = \begin{cases} 1, & \text{if } Duration_i > \text{quantile}(Duration, 0.9) \\ 0, & \text{else} \end{cases}$$

其中， $LongOrderFlag_i$ 为订单*i*是否为漫长订单标识， $Duration_i$ 为订单*i*的成交时长， $\text{quantile}(Duration, 0.9)$ 表示当日所有同类型订单成交时长的前 10%分位数。

### 漫长订单交易占比因子构建

在对漫长订单的定义进行界定后，我们即可以当日所有漫长买单成交量之和占个股当日成交量的比例，构建“漫长买单交易占比”因子（ $VolumeLongBuy$ ），具体计算公式如下：

$$VolumeLongBuy = \frac{\sum VolumeLongBuy_i}{Volume}$$

其中， $VolumeLongBuy$ 表示股票的漫长买单交易占比， $VolumeLongBuy_i$ 表示该股票当日漫长买单*i*的实际成交量， $Volume$ 表示股票当日全天总成交量。

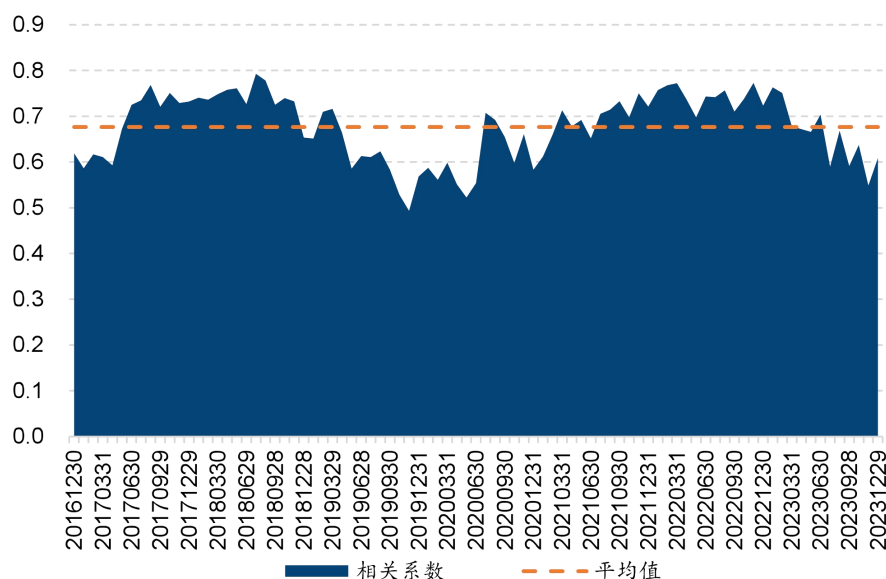
同样地，我们可以当日所有漫长卖单的成交量之和占个股当日成交量的比例，构建“漫长卖单交易占比”因子（ $VolumeLongSell$ ），具体计算公式如下：

$$VolumeLongSell = \frac{\sum VolumeLongSell_i}{Volume}$$

其中， $VolumeLongSell$ 表示股票的漫长卖单交易占比， $VolumeLongSell_i$ 表示该股票当日漫长卖单*i*的实际成交量， $Volume$ 表示股票当日全天总成交量。

同样的，在实际研究中我们发现，当一个股票日内成交中包含的“漫长买单”越多时，其对应发生的“漫长卖单”往往也会越多，二者呈现出显著的正相关关系。图 20 展示了“漫长买单交易占比”因子和“漫长卖单交易占比”因子之间在每个月月末的横截面相关系数，可以看到二者的平均相关系数达到 0.68。

图20: 大单买入交易占比及大单卖出交易占比因子相关系数



资料来源：WIND，国信证券经济研究所整理

基于此，我们将“漫长买单交易占比”因子（ $VolumeLongBuy$ ）和“漫长卖单交易占比”因子（ $VolumeLongSell$ ）直接合成，构建“漫长订单交易占比”因子（ $VolumeLong$ ），其计算方法如下：

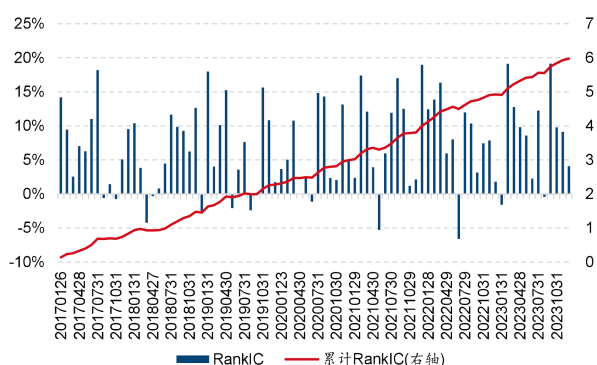
$$VolumeLong = VolumeLongBuy + VolumeLongSell$$

$$= \frac{\sum VolumeLongBuy_i + \sum VolumeLongSell_i}{Volume}$$

其中， $VolumeLong$  表示股票漫长订单交易占比因子， $VolumeLongBuy$  和  $VolumeLongSell_i$  分别为漫长买单和漫长卖单的实际成交量， $Volume$  表示股票在当日的全天总成交量。

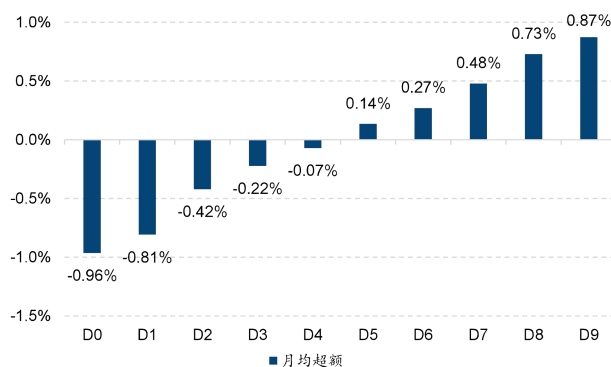
为了保证指标的稳定性，我们先根据每日数据计算个股的“漫长订单交易占比”因子，随后以其过去 20 天的均值得到最终因子取值。

图21：漫长订单交易占比因子 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

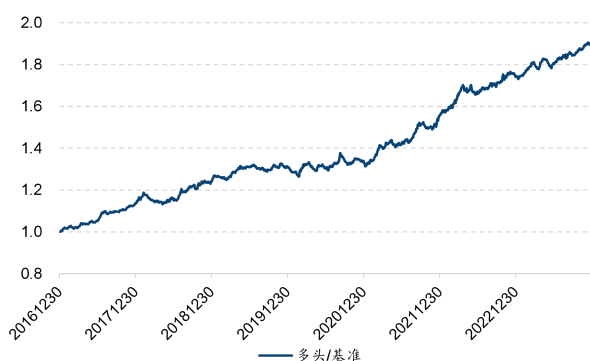
图22：漫长订单交易占比因子分组月均超额收益



资料来源：WIND，国信证券经济研究所整理

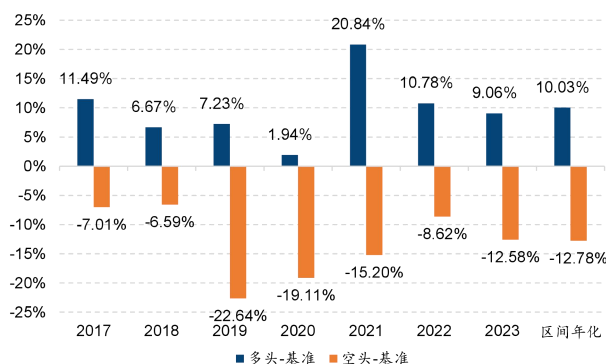
图 21 和图 22 展示了“漫长订单交易占比”因子（ $VolumeLong$ ）在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到，自 2017 年以来该因子的 RankIC 均值为 7.1%，年化 RankICIR 为 3.88，月胜率 84.5%。从分组月均超额收益分布来看，各组单调性特征良好，多头月均超额收益 0.87%，空头月均超额收益 -0.96%。

图23：漫长订单交易占比因子多头相对基准净值走势



资料来源：WIND，国信证券经济研究所整理

图24：漫长订单交易占比因子多空分年度超额收益



资料来源：WIND，国信证券经济研究所整理

图 23 展示了“漫长订单交易占比”因子（VolumeLong）多头相对基准的净值走势，图 24 展示了该因子多头和空头相对基准的分年度表现。可以看到该因子的多头表现依然十分稳健，自 2017 年以来每年都能够稳定战胜等权基准，多头组合年化超额收益 10.03%，空头组合年化超额收益-12.78%。

表 5 展示了“改进大单交易占比”因子与“漫长订单交易占比”因子在绩效上的表现，可以看到二者都展现出较强的选股能力，且月度自相关系数较高，稳定性较强。

表5：改进大单交易占比因子和漫长订单交易占比因子绩效对比

|          | RankIC 均值 | 年化 RankICIR | RankIC 月胜率 | 多头年化超额 | 空头年化超额  | 月度自相关系数 |
|----------|-----------|-------------|------------|--------|---------|---------|
| 改进大单交易占比 | 7.6%      | 3.73        | 88.1%      | 9.43%  | -18.11% | 0.81    |
| 漫长订单交易占比 | 7.1%      | 3.88        | 84.5%      | 10.03% | -12.78% | 0.82    |

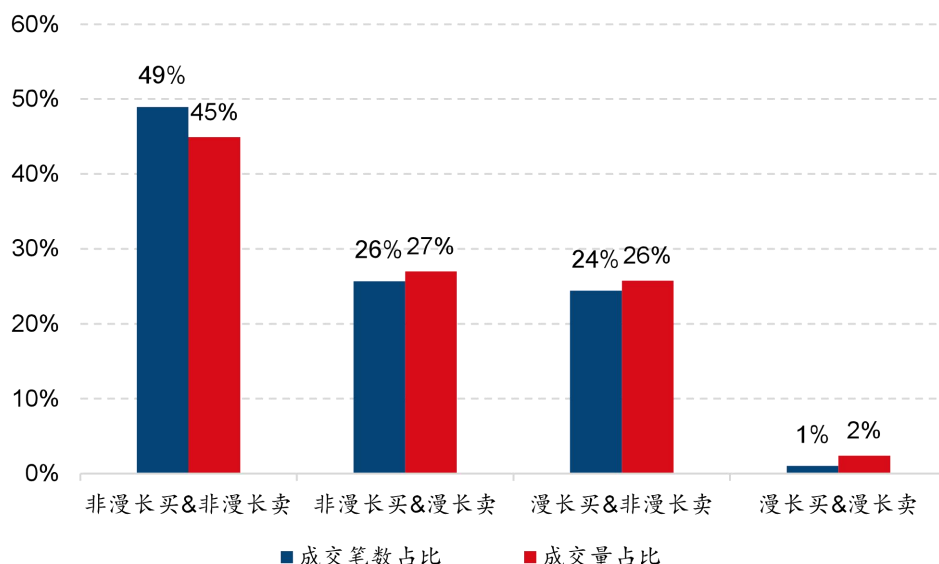
资料来源：WIND，国信证券经济研究所整理

## 漫长订单交易占比因子拆解

与“传统大单交易占比”因子的拆解类似，我们也可基于委买单和委卖单的“时长属性”对不同成交记录的类型进行划分。

图 25 展示了不同类型成交记录的成交笔数及成交量占比情况，以委买单和委卖单均为非漫长订单的成交记录为例，该类型订单的成交笔数占比达到 49%，贡献的成交量占比同样达到 45%；对于单边订单为漫长订单的成交记录而言，其成交笔数占比和成交量占比基本保持在 1/4 的水平。

图25：基于委买单及委卖单的“时长属性”构造的不同类型成交记录的成交笔数及成交量占比



资料来源：WIND，国信证券经济研究所整理

同样的，我们将“漫长大单交易占比”因子拆解为如下三个子因子加总得到的结果：

- 委买单为漫长订单、委卖单为非漫长订单；
- 委买单为非漫长订单、委卖单为漫长订单；
- 委买单及委卖单均为漫长订单。

表6：基于委买单及委卖单的“时长属性”构造的子因子回测结果

| 因子名称     | RankIC 信息 |       |             | 月均超额收益 |        |
|----------|-----------|-------|-------------|--------|--------|
|          | 月胜率       | 均值    | 年化 RankICIR | 多头-基准  | 空头-基准  |
| 漫长买&非漫长卖 | 70.24%    | 5.07% | 2.20        | 0.60%  | -0.69% |
| 非漫长买&漫长卖 | 78.57%    | 6.08% | 3.10        | 0.74%  | -0.73% |
| 漫长买&漫长卖  | 76.19%    | 3.82% | 2.33        | 0.52%  | -0.55% |

资料来源：WIND，国信证券经济研究所整理

表 6 展示了拆解后的三个子因子在全市场中的回测表现，可以看到三个因子对股价未来的影响方向均为正向影响。因此，我们无需对“漫长订单交易占比”因子进行调整。

## 衡量订单特征的其他维度

在前面的两个部分中，我们分别从订单的“大单属性”和“时长属性”两个维度出发，构建了“改进大单交易占比”和“漫长订单交易占比”因子，这两个指标对个股未来收益的预测均有着较为明显的区分作用。事实上，我们还可以从其他特征维度对不同订单特征包含的信息进行衡量，在本小节中，我们从订单的“早尾盘属性”和“高低价属性”两个维度展开讨论。

### 基于早尾盘属性对订单进行划分

施红俊（2004）采用具有操纵嫌疑的股票分笔数据作为研究样本发现，具有操纵嫌疑的股票收益率、换手率、收益波动率在早盘、尾盘表现出明显的异常，这种异常现象起源于早盘操纵和尾盘操纵。因此，“早尾盘属性”也是分析订单的一个维度。

我们基于如下方式对“早尾盘订单”进行划分：

- 早盘订单：第一笔成交时间在 10:00 之前的订单；
- 尾盘订单：第一笔成交时间在 14:30 之后的订单。

通常来讲，一笔成交记录的买卖双方通常同为早盘订单或同为尾盘订单，因此我们根据委买单和委卖单的“早尾盘属性”将其划分为“早盘买&早盘卖”、“尾盘买&尾盘卖”和“非早盘买&非早盘卖”三种类型。

表 7 展示了基于委托订单的“早尾盘属性”构建的三个子因子在全市场中的回测结果。总体来看，“早盘买+早盘卖”成交量占比越高，其未来反转效应越明显，但是该因子的多头表现相对较弱；“尾盘买+尾盘卖”成交量占比越高，其未来也呈现出一定的反转效应，但是该因子的 RankIC 均值相对较低。“非早盘买+非早盘卖”成交量占比越高，其未来的动量效应越强。

表7：基于委买单及委卖单的“早尾盘属性”构造的子因子回测结果

| 因子名称      | RankIC 信息 |        |             | 月均超额收益 |        |
|-----------|-----------|--------|-------------|--------|--------|
|           | 月胜率       | 均值     | 年化 RankICIR | 多头-基准  | 空头-基准  |
| 早盘买&早盘卖   | 16.67%    | -6.33% | -2.72       | 0.03%  | -1.01% |
| 尾盘买&尾盘卖   | 42.86%    | -1.24% | -0.71       | 0.25%  | -0.71% |
| 非早盘买&非早盘卖 | 90.48%    | 7.53%  | 3.80        | 0.43%  | -1.41% |

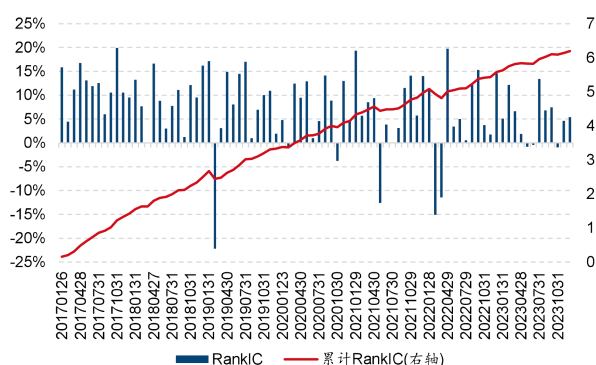
资料来源：WIND，国信证券经济研究所整理

我们将表7中有效性较高的“非早盘买&非早盘卖”成交量占比减去“早盘买&早盘卖”成交量占比，构建“早尾盘复合交易占比因子”：

$$\text{早尾盘复合交易占比} = \text{非早盘买\&非早盘卖} - \text{早盘买\&早盘卖}$$

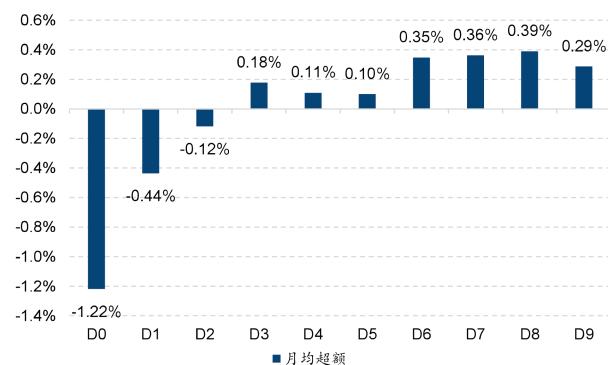
在样本区间内，该指标 RankIC 均值 7.4%，年化 RankICIR 为 3.36。

图26：早尾盘复合交易占比因子 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

图27：早尾盘复合交易占比因子分组月均超额收益



资料来源：WIND，国信证券经济研究所整理

## 基于高低价属性对订单进行划分

订单的委托价格代表着买卖双方对于当前股价公允价值的看法，因此订单的“高低价属性”也是分析订单的一个维度。我们基于如下方式对“高低价订单”进行划分：

- 高价单：委托价格处于同类型订单委托价格前 10%分位点的订单；
- 低价单：委托价格处于同类型订单委托价格后 10%分位点的订单。

通常来讲，一笔成交记录的买卖双方通常同为高价单或同为低价单，因此我们根据委买单和委卖单的“高低价属性”将其划分为“高价买&高价卖”、“低价买&低价卖”和“非高低价买&非高低价卖”三种类型。

表8：基于委买单及委卖单的“高低价属性”构造的子因子回测结果

| 因子名称        | RankIC 信息 |        |             | 月均超额收益 |        |
|-------------|-----------|--------|-------------|--------|--------|
|             | 月胜率       | 均值     | 年化 RankICIR | 多头-基准  | 空头-基准  |
| 高价买&高价卖     | 36.90%    | -4.16% | -1.38       | 0.34%  | -0.15% |
| 低价买&低价卖     | 41.67%    | -2.34% | -0.73       | 0.07%  | 0.14%  |
| 非高低价买&非高低价卖 | 64.29%    | 3.89%  | 1.33        | 0.25%  | -0.24% |

资料来源：WIND，国信证券经济研究所整理

表 8 展示了基于委托订单的“高低价属性”构建的三个子因子在全市场中的回测结果。总体来看，“高价买&高价卖”成交量占比越高，其未来反转效应越明显；“非高低价买+非高低价卖”成交量占比越高，其未来的动量效应越强。

到目前为止，我们从订单大小、成交时长、成交早晚和出价高低等四个维度对订单进行分类并构建了对应的成交量占比因子。从回测结果来看，基于订单大小的“改进大单交易占比”和基于成交时长的“漫长订单交易占比”因子对个股未来收益的区分作用更为明显，因此在后文中，我们将重点围绕以上两个维度展开分析和讨论。读者若能够找到其他更具区分能力的特征维度，可以采用本文提出的框架进行延伸和拓展。

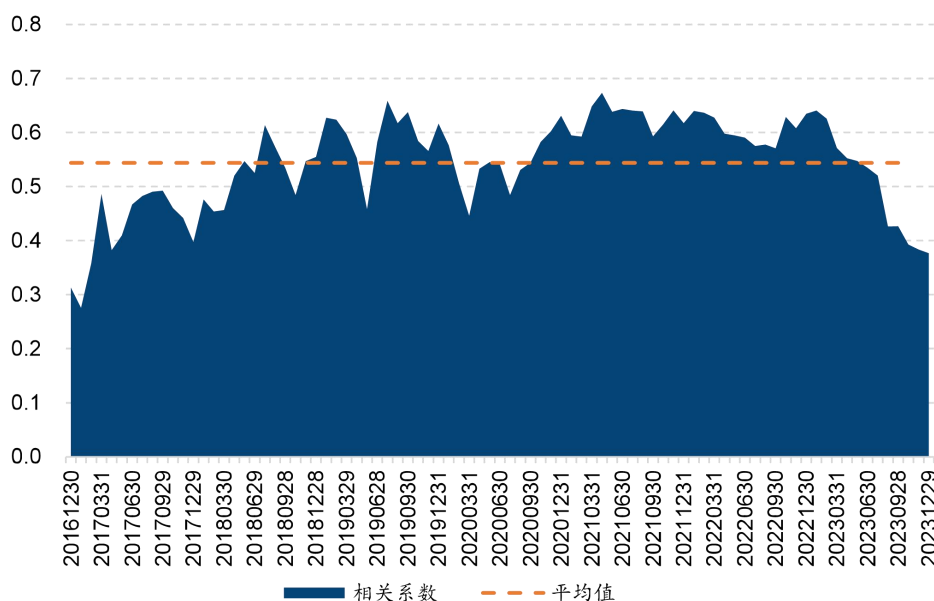
## 大单及漫长订单复合因子

### 改进大单交易占比因子 VS 漫长订单交易占比因子

由前述分析可知，漫长订单与大单之间具有一定的正相关关系，且“改进大单交易占比”因子（*VolumeBig*）和“漫长订单交易占比”因子（*VolumeLong*）的构造过程中都采用的是成交量占比的方式，因此二者之间包含的信息可能具有一定的相似性。

为了对二者之间的相关性进行验证，图 28 展示了二者在横截面上的相关系数。可以看到，在样本区间段内二者平均相关系数达到 0.54。

图28: 改进大单交易占比因子与漫长订单交易占比因子相关系数



资料来源：WIND，国信证券经济研究所整理

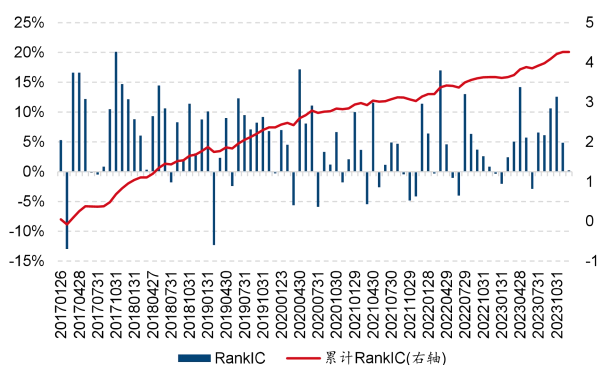
为了探究“改进大单交易占比”因子（*VolumeBig*）和“漫长订单交易占比”因子（*VolumeLong*）互相之间是否能提供增量信息，我们采用互相正交的方式观察在剥离掉另一个因子的影响之后，目标因子的有效性是否仍然存在。

首先将“改进大单交易占比”（*VolumeBig*）对“漫长订单交易占比”因子（*VolumeLong*）进行剥离，采用如下方式进行回归：

$$VolumeBig = \alpha + \beta \cdot VolumeLong + \varepsilon$$

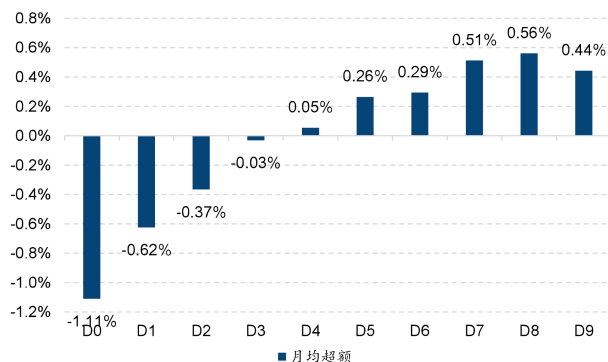
图 29 和图 30 展示了回归后残差因子在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到，自 2017 年以来，残差因子的 RankIC 均值为 5.1%，年化 RankICIR 为 2.63，月胜率 75.0%。从分组月均超额收益分布来看，各组单调性特征良好，多头月均超额收益 0.44%，空头月均超额收益-1.11%。尽管有一定程度的下降，但仍然保持了较为稳定的绩效表现。

图29：大单交易占比残差因子 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

图30：大单交易占比残差因子分组月均超额



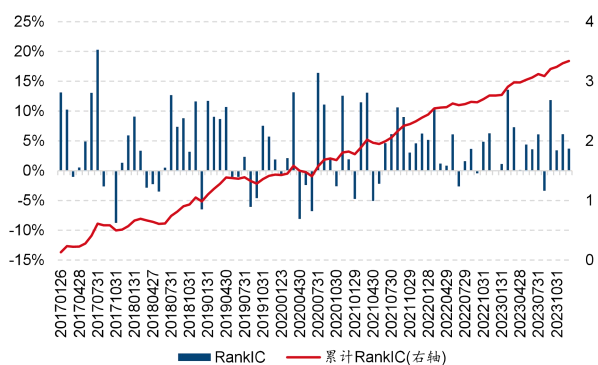
资料来源：WIND，国信证券经济研究所整理

接下来我们将“漫长订单交易占比”因子（*VolumeLong*）对“改进大单交易占比”（*VolumeBig*）进行剥离，采用如下方式进行回归：

$$VolumeLong = \alpha + \beta \cdot VolumeBig + \varepsilon$$

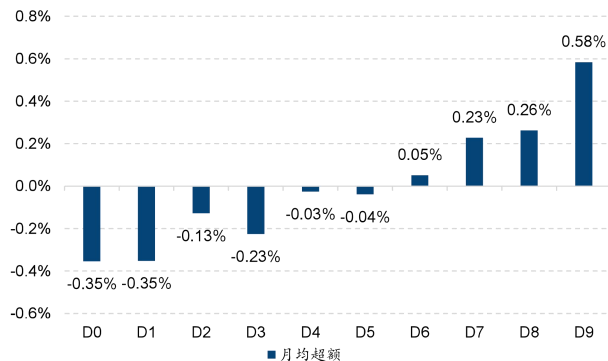
图 31 和图 32 展示了残差因子在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到，自 2017 年以来，残差因子的 RankIC 均值为 4.0%，年化 RankICIR 为 2.23，月胜率 72.6%。从分组月均超额收益分布来看，各组单调性特征良好，多头月均超额收益 0.58%，空头月均超额收益-0.35%。

图31：漫长订单交易占比残差因子 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

图32：漫长订单交易占比残差因子分组月均超额



资料来源：WIND，国信证券经济研究所整理

由以上分析可知，尽管“改进后大单交易占比”因子和“漫长订单交易占比”因子展现出一定的正相关关系，但是将二者互相进行正交化剥离后，残差因子仍然保持了较为稳健的选股能力，其 RankIC 的走势持续向上，分组单调性特征也较为明显。

因此，我们认为无论是“改进后大单交易占比”因子（*VolumeBig*）还是“漫长订单交易占比”因子（*VolumeLong*），都能够为对方提供较多的信息增量。基于此，我们在下一小节中对二者进行合成，构建“大单及漫长订单”复合因子。

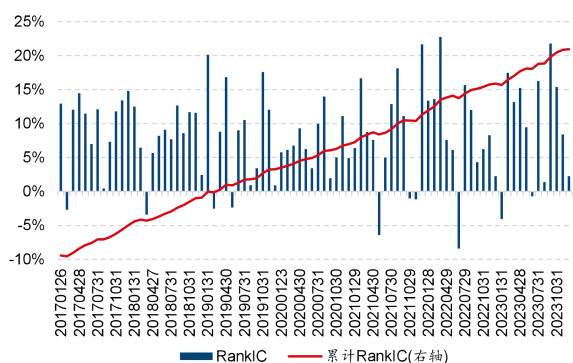
### 大单及漫长订单复合因子

本部分，我们将“改进大单交易占比”因子（*VolumeBig*）和“漫长订单交易占比”因子（*VolumeLong*）进行等权合成，构建“大单及漫长订单”复合因子（*VolumeLongBig*）：

$$VolumeLongBig = VolumeBig + VolumeLong$$

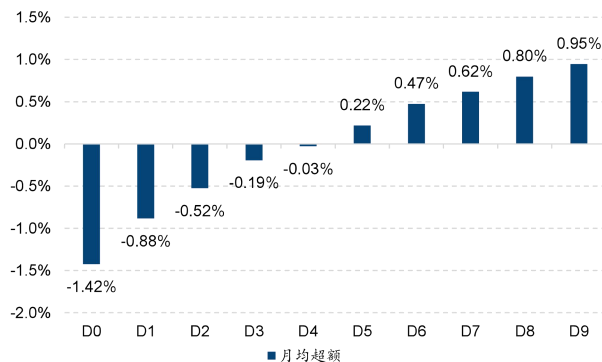
图 33 和图 34 展示了该复合因子在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到，自 2017 年以来，复合因子的 RankIC 均值为 8.4%，年化 RankICIR 为 4.39，月胜率 88.1%。从分组月均超额收益分布来看，各组展现出严格的单调性，多头月均超额收益 0.95%，空头月均超额收益-1.42%。总体来看，复合因子的绩效表现均优于单个指标的表现。

图33: 复合因子 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

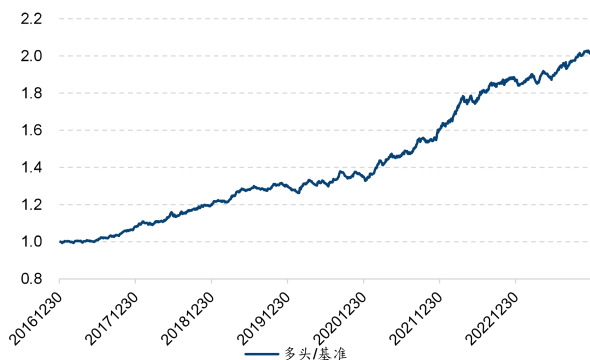
图34: 复合因子分组月均超额收益



资料来源：WIND，国信证券经济研究所整理

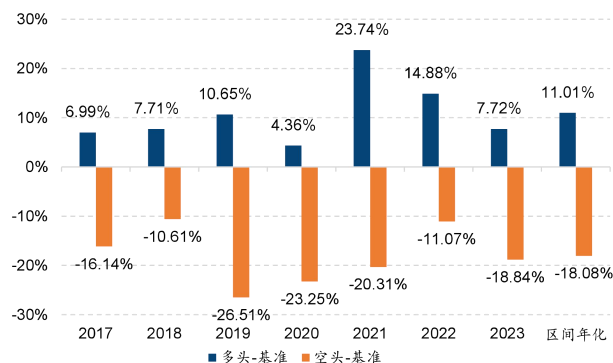
图 35 展示了复合因子多头相对基准的净值走势，图 36 展示了该因子多头和空头相对基准的分年度表现。可以看到该因子的多头表现依然十分稳健，自 2017 年以来每年都能够稳定战胜等权基准，多头组合年化超额收益 11.01%，空头组合年化超额收益-18.08%。

图35: 复合因子多头相对基准净值走势



资料来源: WIND, 国信证券经济研究所整理

图36: 复合因子多空分年度超额收益



资料来源: WIND, 国信证券经济研究所整理

## 复合因子在不同样本空间中的表现

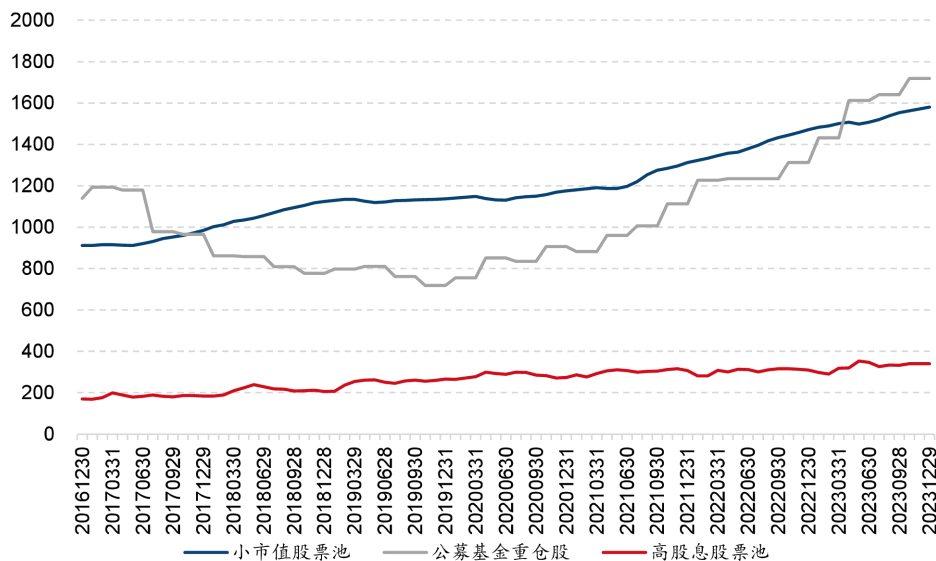
本部分我们考察“大单及漫长订单”复合因子 (*VolumeLongBig*) 在不同样本空间中的表现, 包括常见的宽基指数如沪深 300 指数、中证 500 指数、中证 1000 指数和国证 2000 指数, 以及近几年市场关注度较高的不同风格型股票池, 如小市值股票池、公募基金重仓股和高股息股票池。

首先我们对近几年市场关注度较高的不同风格型股票池进行如下定义:

- 小市值股票池: 沪深股票池中剔除上市时间不满 1 年的新股、ST 及\*ST 等风险警示股、三个月内存在 ST 或\*ST 等风险警示的股票后, 按照其总市值取最小的 1/3 的股票池;
- 公募基金重仓股: 沪深股票池中最近一个基金季报中, 被任意一只主动股基持有为前十大重仓股的股票取并集构建的股票池;
- 高股息股票池: 沪深股票池中连续三年均进行现金分红且连续三年的股息支付率均大于 10% 的股票池中, 股息率处于该股票池前 1/3 或处于该股票池中同行业公司前 1/3 的股票。

图 37 展示了不同风格型股票池每期样本数量变化情况, 总体来看小市值股票池和公募基金重仓股股票在 1000-1500 只股票左右, 高股息股票池在 300 只股票左右。

图37: 不同风格型股票池样本数量变化



资料来源: WIND, 国信证券经济研究所整理

表 9 展示了复合因子在不同的样本空间中的绩效:

- 在宽基指数中, 复合因子在中证 1000 指数和国证 2000 指数中的年化 RankICIR 均达到 4 以上, 在中证 500 指数中的多头超额收益表现较为稳健。与大部分量价类指标类似, 其在指数成分股越多、市值区间越小的样本中展现的选股能力越强。
- 在不同风格型股票池中, 无论是分析师覆盖度少、公募基金持仓较低的小市值股票池, 还是分析师覆盖度高、公募基金持仓较高的公募重仓股股票池, 还是偏低估值、低波动的高股息股票池中, 该复合因子都展现出较为稳定的选股能力。

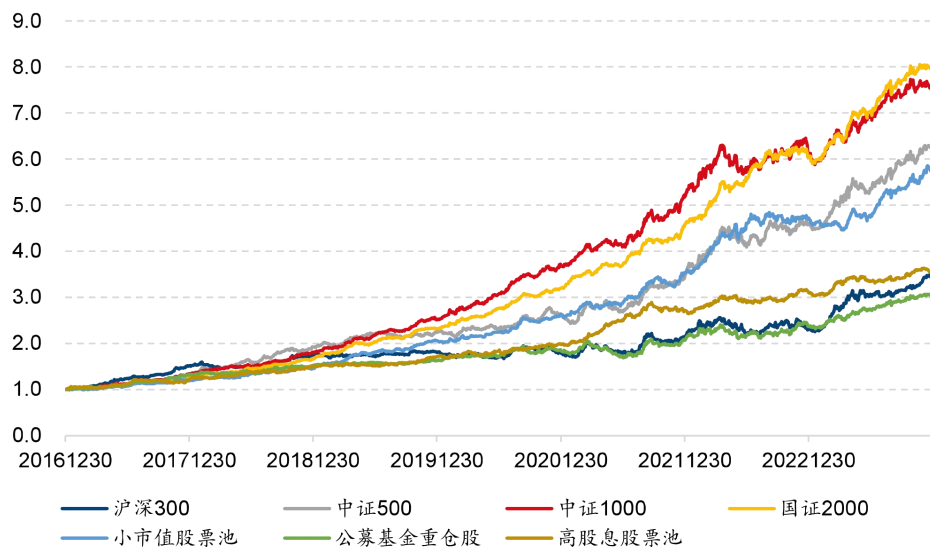
表9: 复合因子在不同样本空间中的表现

| 因子名称          | RankIC 信息 |       |             | 月均超额收益 |        |
|---------------|-----------|-------|-------------|--------|--------|
|               | 月胜率       | 均值    | 年化 RankICIR | 多头-基准  | 空头-基准  |
| <b>宽基指数</b>   |           |       |             |        |        |
| 沪深 300 指数     | 66.67%    | 6.58% | 1.58        | 0.56%  | -0.85% |
| 中证 500 指数     | 82.14%    | 7.89% | 3.00        | 0.91%  | -1.20% |
| 中证 1000 指数    | 85.71%    | 8.54% | 4.26        | 0.93%  | -1.42% |
| 国证 2000 指数    | 89.29%    | 8.86% | 4.65        | 0.79%  | -1.62% |
| <b>风格型股票池</b> |           |       |             |        |        |
| 小市值股票池        | 83.30%    | 7.50% | 3.30        | 0.75%  | -1.29% |
| 公募重仓股股票池      | 73.80%    | 4.40% | 2.06        | 0.67%  | -0.64% |
| 高股息股票池        | 76.20%    | 4.90% | 2.32        | 0.69%  | -0.80% |

资料来源: WIND, 国信证券经济研究所整理

图 38 展示了复合因子在不同样本空间中的多空净值走势, 整体来看在各个指数成分股中的多空表现都呈现出稳健向上的趋势。

图38: 复合因子在不同样本空间中的多空净值走势



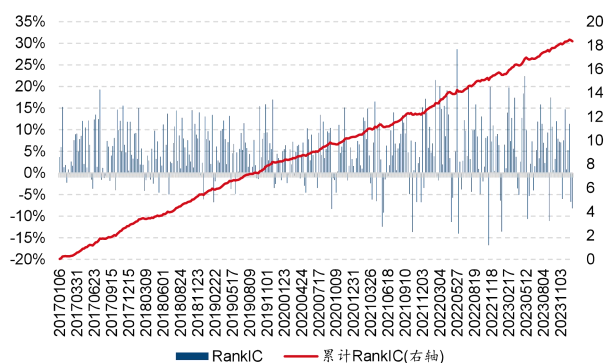
资料来源: WIND, 国信证券经济研究所整理

## 周频调仓复合因子表现

对于量化研究者而言, 对基于高频数据构建的因子, 我们不仅关注其在月频调仓下的因子表现, 还会关注在更高频率的调仓周期下的因子表现。本部分, 我们考察周频“大单及漫长订单”复合因子 (*VolumeLongBigWeek*) 的选股能力。

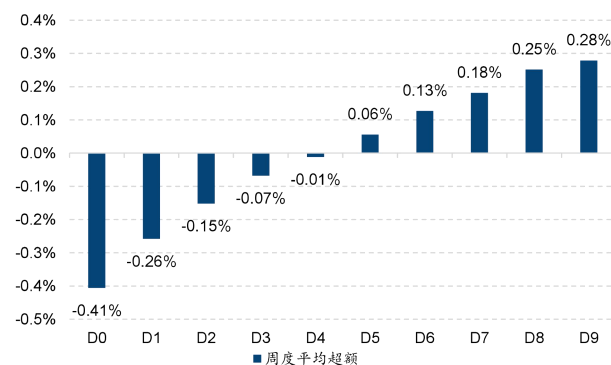
图 39 和图 40 展示了该周频调仓复合因子在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到, 自 2017 年以来, 周频调仓复合因子的周度 RankIC 均值为 5.1%, 年化 RankICIR 为 5.42, 周胜率 78.4%。从分组周度平均超额收益分布来看, 各组展现出严格的单调性, 多头周度平均超额收益 0.28%, 空头周度平均超额收益-0.41%。

图39: 周频调仓复合因子 RankIC 走势



资料来源: WIND, 国信证券经济研究所整理

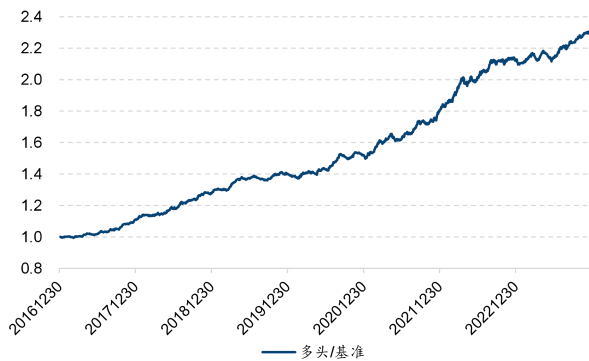
图40: 周频调仓复合因子分组周度平均超额收益



资料来源: WIND, 国信证券经济研究所整理

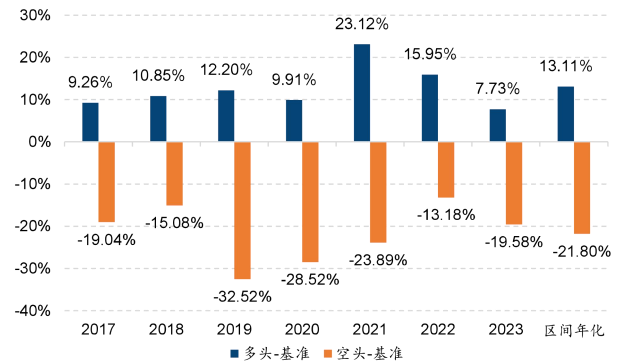
图 41 展示了周频调仓复合因子多头相对基准的净值走势, 图 42 展示了该周频调仓因子多头和空头相对基准的分年度表现。可以看到周频调仓下多头组合年化超额收益 13.11%, 空头组合年化超额收益-21.80%。总体来看, 周频调仓下的组合绩效表现比月频调仓下的组合绩效表现更为稳健。

图41：周频调仓复合因子多头相对基准净值走势



资料来源：WIND，国信证券经济研究所整理

图42：周频调仓复合因子多空分年度超额收益



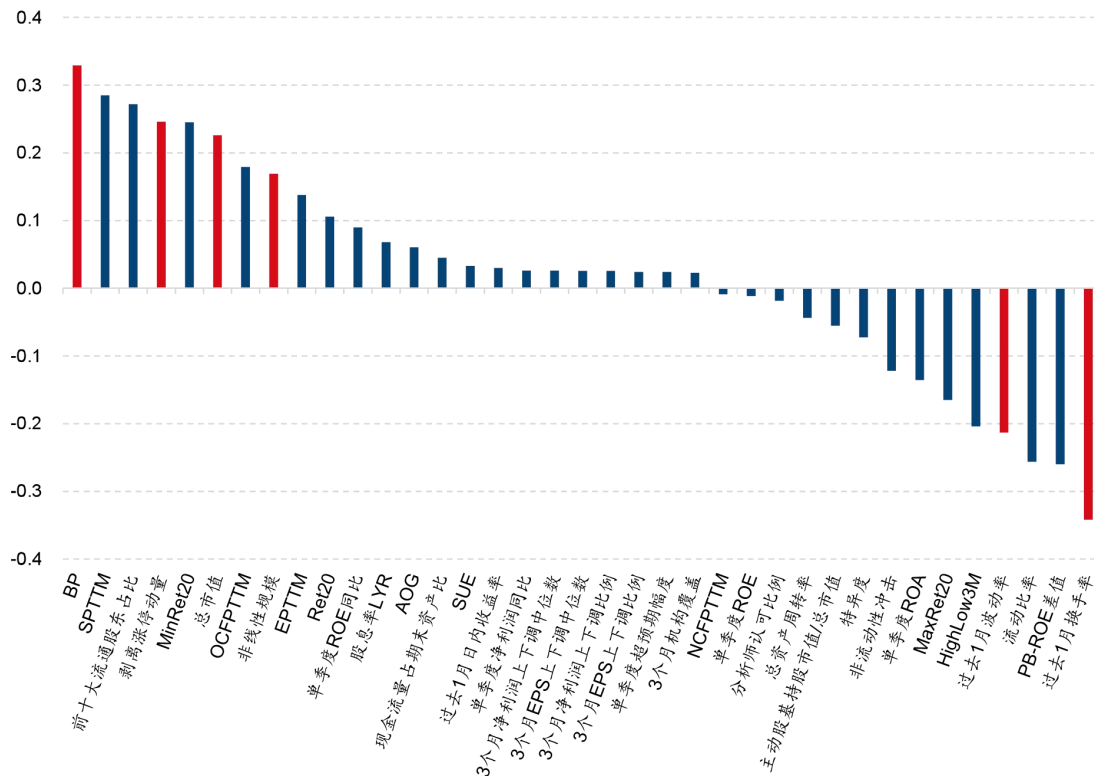
资料来源：WIND，国信证券经济研究所整理

## 复合因子与常见选股因子相关性

为了进一步探究“大单及漫长订单”复合因子对已有因子库的增量信息，本部分我们考察其与常见选股因子的相关性。

图 43 展示了复合因子与常见选股因子的横截面平均相关系数，总体来看“大单及漫长订单”复合因子多头组合更倾向于低估值、大市值、低波动、低换手股票。该因子与成长类及分析师类因子的相关性较弱。

图43：漫长订单与选股因子相关系数



资料来源：WIND，国信证券经济研究所整理

进一步地，我们将该复合因子对上述所有选股因子进行正交剥离，采用如下方式进行回归：

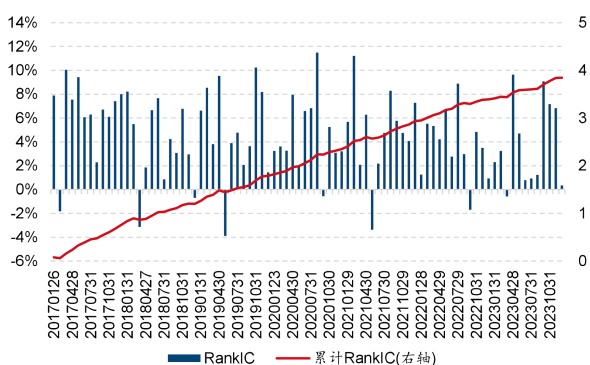
$$VolumeLongBig = \alpha + \beta_1 \cdot Alpha_1 + \dots + \beta_n \cdot Alpha_n + \varepsilon$$

其中， $Alpha_i$ 表示图 43 中列出的 Alpha 因子， $\varepsilon$ 即为通过正交回归得到的纯净因子，我们对该纯净因子进行行业市值中性化，观察其选股能力。

图 44 和图 45 展示了该纯净因子在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到，自 2017 年以来，纯净因子的 RankIC 均值为 4.6%，年化 RankICIR 为 4.63，月胜率 90.5%。

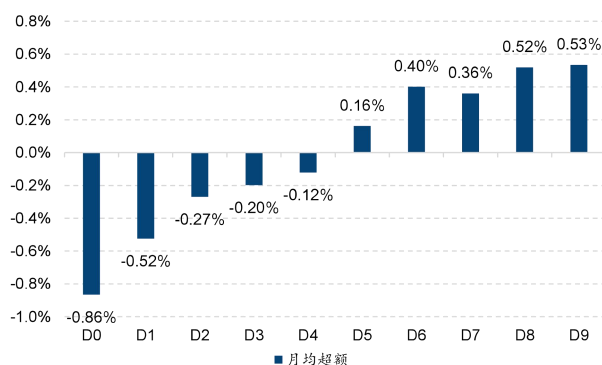
从分组月均超额收益分布来看，各组展现出严格的单调性，多头月均超额收益 0.53%，空头月均超额收益-0.86%。总体来看，纯净因子依然具备较为稳健的选股能力，说明复合因子能够为现有因子库贡献一定的增量信息。

图44: 复合纯净因子 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

图45: 复合纯净因子分组月均超额收益



资料来源：WIND，国信证券经济研究所整理

## 基于订单特征的成交量占比因子的统一框架

在前文中，我们从订单大小、成交时长、成交时间和出价高低等多个维度出发，基于委买单和委卖单的不同特征，将每一笔成交划分为不同的类型。随后，将相同类型的成交记录所产生的成交量进行加总，除以全天的成交量，构建了特定类型订单的成交量占比因子，公式表达如下：

$$VolumeRatio_s = \frac{Volume_s}{VolumeAllDay}$$

其中， $VolumeRatio_s$ 表示类型为  $s$  的成交记录的交易量之和占全天交易量的比例， $Volume_s$ 表示类型为  $s$  的成交记录的交易量之和， $VolumeAllDay$ 表示全天交易量之和。

很自然地我们会想到，同一笔订单的不同维度存在交叉，例如同一笔订单可能属于大单的同时也属于漫长订单，另一笔订单可能属于大单的同时也属于非漫长订单，那么如何对同一笔订单的不同维度同时进行分析呢？本部分，我们提出一种基于订单特征构建成交量占比因子的统一研究框架。

从前文的回测结果来看，基于订单大小的“改进大单交易占比”和基于成交时长的“漫长订单交易占比”因子对个股未来收益的区分作用更为明显，因此本部分

我们以订单大小和成交时长两个维度为例对该统一研究框架进行分析，其他类似的特征维度可以简单地运用到该框架的拓展中。

### 基于订单特征的精选复合因子

以订单大小和成交时长两个维度为例，我们根据委托单的“大单属性”将订单分为大单、非大单两类，根据委托单的“漫长属性”将订单分为漫长订单和非漫长订单两类，因此同一笔订单在“大单”和“漫长”属性上即拥有 4 种取值。由于每一笔成交由买卖双方撮合决定，因此一笔成交记录从类型上可以被划分为  $4 \times 4 = 16$  种类别，如表 10 所示。

表10: 基于大单属性和漫长属性对成交类型进行划分

| 成交类型 | 大单属性 (Big) |          | 漫长属性 (Long) |          | 综合标识            |
|------|------------|----------|-------------|----------|-----------------|
|      | 委买单 Buy    | 委卖单 Sell | 委买单 Buy     | 委卖单 Sell |                 |
| 1    | 1          | 0        | 0           | 0        | BB1_BS0_LB0_LS0 |
| 2    | 1          | 0        | 0           | 1        | BB1_BS0_LB0_LS1 |
| 3    | 1          | 0        | 1           | 0        | BB1_BS0_LB1_LS0 |
| 4    | 1          | 0        | 1           | 1        | BB1_BS0_LB1_LS1 |
| 5    | 1          | 1        | 0           | 0        | BB1_BS1_LB0_LS0 |
| 6    | 1          | 1        | 0           | 1        | BB1_BS1_LB0_LS1 |
| 7    | 1          | 1        | 1           | 0        | BB1_BS1_LB1_LS0 |
| 8    | 1          | 1        | 1           | 1        | BB1_BS1_LB1_LS1 |
| 9    | 0          | 0        | 0           | 0        | BB0_BS0_LB0_LS0 |
| 10   | 0          | 0        | 0           | 1        | BB0_BS0_LB0_LS1 |
| 11   | 0          | 0        | 1           | 0        | BB0_BS0_LB1_LS0 |
| 12   | 0          | 0        | 1           | 1        | BB0_BS0_LB1_LS1 |
| 13   | 0          | 1        | 0           | 0        | BB0_BS1_LB0_LS0 |
| 14   | 0          | 1        | 0           | 1        | BB0_BS1_LB0_LS1 |
| 15   | 0          | 1        | 1           | 0        | BB0_BS1_LB1_LS0 |
| 16   | 0          | 1        | 1           | 1        | BB0_BS1_LB1_LS1 |

资料来源：国信证券经济研究所绘制

为了便于理解，我们采用如下方式对不同类型的成交记录进行标记：

- “BB” 表示 “Big Buy”，代表该笔成交的委买单是否为大单；
- “BS” 表示 “Big Sell”，代表该笔成交的委卖单是否为大单；
- “LB” 表示 “Long Buy”，代表该笔成交的委买单是否为漫长订单；
- “LS” 表示 “Long Sell”，代表该笔成交的委卖单是否为漫长订单；

为了表征订单类型是否属于该类型，我们以 0-1 后缀进行表示。后缀为 1 代表属于该类别，为 0 代表不属于该类别。由此，“BB1\_BS0\_LB1\_LS0”即代表该笔成交记录中委买单为大单+漫长订单，委卖单为非大单+非漫长订单。我们将每只股票每日具有相同标签的成交记录的实际成交量进行加总，除以当日股票总成交量，即可构建该标签下的成交量占比因子。

事实上，前述部分提到的“改进大单交易占比”和“漫长订单交易占比”，均可由如上 16 个子因子中的某些部分通过简单的线性加总得到。例如，构成“改进大单交易占比”因子的三个子因子中，“大买&非大卖”即为表 10 中所有包含 BB1\_BS0 类型加总的结果，“非大买&大卖”即为所有包含 BB0\_BS1 类型加总的结果，“大买&大卖”即为所有包含 BB1\_BS1 类型加总的结果。

$$\text{改进大单交易占比} = \frac{(-1) \times \text{大买\&非大卖} + (-1) \times \text{非大买\&大卖} + \text{大买\&大卖}}{\text{总成交量}}$$

$$\text{大买\&非大卖} = \text{BB1\_BS0\_LB0\_LS0} + \text{BB1\_BS0\_LB0\_LS1} + \text{BB1\_BS0\_LB1\_LS0} + \text{BB1\_BS0\_LB1\_LS1}$$

$$\text{非大买\&大卖} = \text{BB0\_BS1\_LB0\_LS0} + \text{BB0\_BS1\_LB0\_LS1} + \text{BB0\_BS1\_LB1\_LS0} + \text{BB0\_BS1\_LB1\_LS1}$$

$$\text{大买\&大卖} = \text{BB1\_BS1\_LB0\_LS0} + \text{BB1\_BS1\_LB0\_LS1} + \text{BB1\_BS1\_LB1\_LS0} + \text{BB1\_BS1\_LB1\_LS1}$$

类似的，构成“漫长订单交易占比”因子的三个子因子中，“漫长买&非漫长卖”即为表 10 中所有包含 LB1\_LS0 类型加总的结果，“非漫长买&漫长卖”即为所有包含 LB0\_LS1 类型加总的结果，“漫长买&漫长卖”即为所有包含 LB1\_LS1 类型加总的结果。

漫长订单交易占比

$$= \frac{\text{漫长买\&非漫长卖} + \text{非漫长买\&漫长卖} + 2 \times \text{漫长买\&漫长卖}}{\text{总成交量}}$$

漫长买&非漫长卖

$$= \text{BB0\_BS0\_LB1\_LS0} + \text{BB0\_BS1\_LB1\_LS0} + \text{BB1\_BS0\_LB1\_LS0} + \text{BB1\_BS1\_LB1\_LS0}$$

非漫长买&漫长卖

$$= \text{BB0\_BS0\_LB0\_LS1} + \text{BB0\_BS1\_LB0\_LS1} + \text{BB1\_BS0\_LB0\_LS1} + \text{BB1\_BS1\_LB0\_LS1}$$

$$\text{漫长买\&漫长卖} = \text{BB0\_BS0\_LB1\_LS1} + \text{BB0\_BS1\_LB1\_LS1} + \text{BB1\_BS0\_LB1\_LS1} + \text{BB1\_BS1\_LB1\_LS1}$$

表 11 展示了不同订单类型成交量占比因子在全市场中的 RankIC 信息及月均超额收益，总体来看各因子的表现存在一定的差异，颜色标黄的 5 个因子是我们基于测试结果得到的较为有效的指标。

表11: 不同订单类型成交量占比因子表现

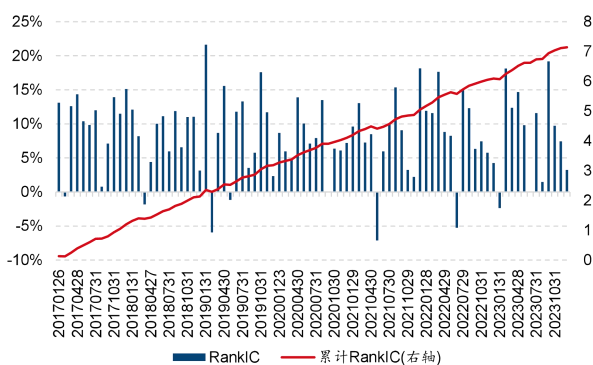
| 不同订单类型          | RankIC 信息 |        |             | 月均超额收益  |         |
|-----------------|-----------|--------|-------------|---------|---------|
|                 | 月胜率       | 均值     | 年化 RankICIR | D9 组-基准 | D0 组-基准 |
| BB1_BS1_LB1_LS1 | 83.33%    | 4.82%  | 3.40        | 0.73%   | -0.68%  |
| BB1_BS1_LB0_LS1 | 80.95%    | 6.11%  | 3.09        | 0.77%   | -0.92%  |
| BB1_BS1_LB1_LS0 | 72.62%    | 5.83%  | 2.77        | 0.70%   | -0.96%  |
| BB0_BS1_LB0_LS1 | 69.05%    | 2.41%  | 1.38        | 0.32%   | -0.21%  |
| BB1_BS1_LB0_LS0 | 58.33%    | 1.80%  | 1.09        | 0.08%   | -0.41%  |
| BB0_BS1_LB1_LS0 | 52.38%    | 1.16%  | 0.62        | 0.07%   | 0.06%   |
| BB0_BS1_LB1_LS1 | 50.00%    | 0.57%  | 0.25        | -0.04%  | -0.10%  |
| BB1_BS0_LB1_LS0 | 54.76%    | 0.12%  | 0.05        | 0.05%   | -0.14%  |
| BB0_BS0_LB1_LS0 | 57.14%    | -0.07% | -0.03       | -0.25%  | 0.03%   |
| BB0_BS0_LB0_LS1 | 54.76%    | -0.09% | -0.04       | -0.19%  | 0.05%   |
| BB1_BS0_LB1_LS1 | 45.24%    | -0.51% | -0.25       | -0.06%  | 0.10%   |
| BB1_BS0_LB0_LS1 | 48.81%    | -0.58% | -0.27       | -0.01%  | 0.43%   |
| BB0_BS0_LB1_LS1 | 44.05%    | -1.96% | -0.71       | -0.48%  | 0.21%   |
| BB0_BS0_LB0_LS0 | 42.86%    | -3.14% | -1.04       | -0.35%  | 0.55%   |
| BB0_BS1_LB0_LS0 | 14.29%    | -7.14% | -3.35       | -1.12%  | 0.79%   |
| BB1_BS0_LB0_LS0 | 5.95%     | -8.21% | -5.05       | -1.36%  | 0.70%   |

资料来源: WIND, 国信证券经济研究所整理

我们根据表 11 中测试得到的 5 个有效因子进行等权合成(将因子方向进行调整), 构建基于订单特征的**精选复合因子**(VolumeLongBigSelect)。

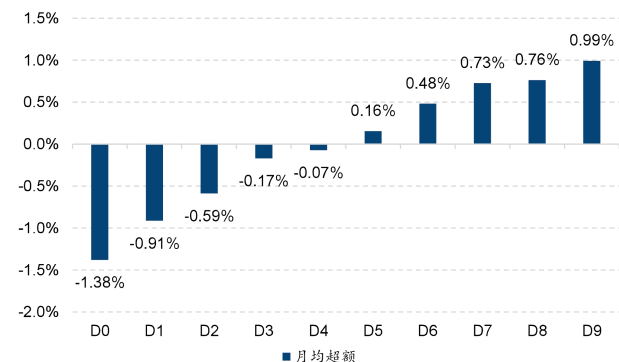
图 46 和图 47 展示了该精选复合因子在全市场中的 RankIC 走势和分组月均超额收益情况。可以看到, 自 2017 年以来, **精选复合因子的 RankIC 均值为 8.5%, 年化 RankICIR 为 5.06, 月胜率 90.5%**。从分组月均超额收益分布来看, 各组展现出严格的单调性, 多头月均超额收益 0.99%, 空头月均超额收益-1.38%。总体来看, 精选复合因子的绩效表现优于前一小节中基于“大单交易占比”因子和“漫长订单交易占比”因子等权复合的因子效果。

图46: 精选复合因子 RankIC 走势



资料来源: WIND, 国信证券经济研究所整理

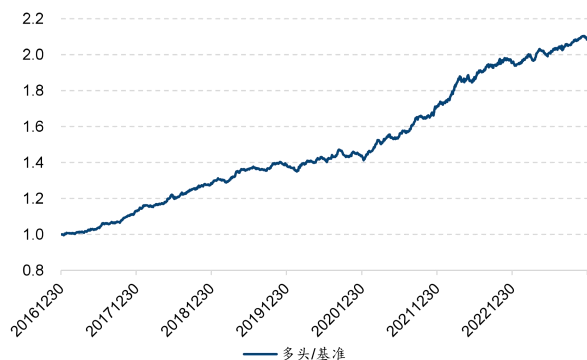
图47: 精选复合因子分组月均超额收益



资料来源: WIND, 国信证券经济研究所整理

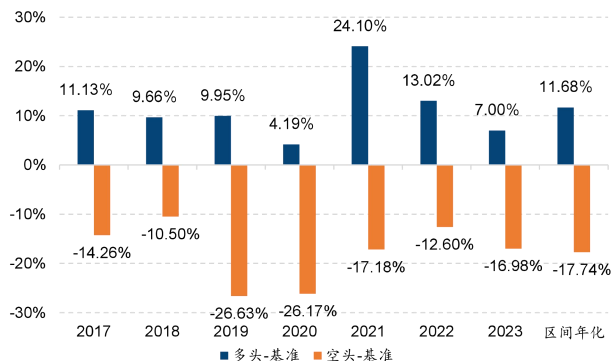
图 48 展示了精选复合因子多头相对基准的净值走势, 图 49 展示了该因子多头和空头相对基准的分年度表现。可以看到精选复合因子的多头表现十分稳健, 自 2017 年以来每年都能够稳定战胜等权基准, 多头组合年化超额收益 11.68%, 空头组合年化超额收益-17.74%。

图48: 精选复合因子多头相对基准净值走势



资料来源: WIND, 国信证券经济研究所整理

图49: 精选复合因子多空分年度超额收益



资料来源: WIND, 国信证券经济研究所整理

表 12 展示了“大单及漫长订单复合因子”与“基于订单特征的精选复合因子”在绩效上的表现, 可以看到精选复合因子的 RankIC 均值略微提升, 但是其 RankICIR 及月度胜率提升较为明显。

表12: 大单及漫长订单复合因子和基于订单特征的精选复合因子绩效对比

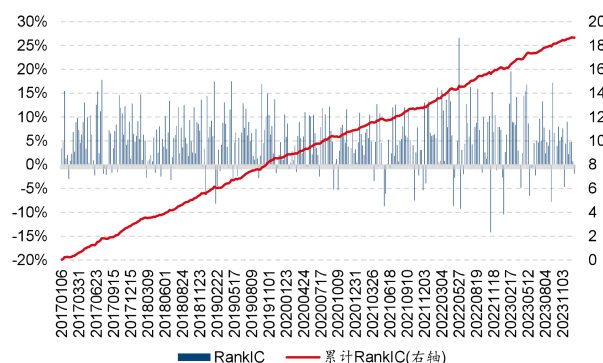
|        | RankIC 均值 | 年化 RankICIR | RankIC 月胜率 | 多头年化超额 | 空头年化超额  | 月度自相关系数 |
|--------|-----------|-------------|------------|--------|---------|---------|
| 复合因子   | 8.4%      | 4.39        | 88.1%      | 11.01% | -18.08% | 0.83    |
| 精选复合因子 | 8.5%      | 5.06        | 90.5%      | 11.68% | -17.74% | 0.80    |

资料来源: WIND, 国信证券经济研究所整理

## 周频调仓精选复合因子表现

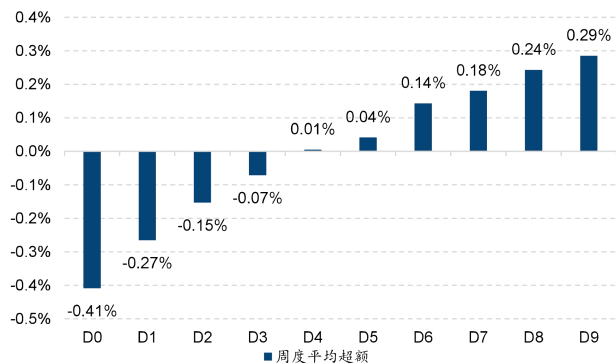
本部分, 我们考察精选复合因子在周度调仓频率下的选股能力。图 50 和图 51 展示了该周频调仓精选复合因子 (*VolumeLongBigSelectWeek*) 在全市场中的 RankIC 走势和分组周度平均超额收益情况。可以看到, 自 2017 年以来, 周频调仓精选复合因子的周度 RankIC 均值为 5.2%, 年化 RankICIR 为 6.57, 周胜率 82.6%。从分组周度平均超额收益分布来看, 各组展现出严格的单调性, 多头周度平均超额收益 0.29%, 空头周度平均超额收益-0.41%。

图50: 周频调仓精选复合因子 RankIC 走势



资料来源: WIND, 国信证券经济研究所整理

图51: 周频调仓精选复合因子分组周度平均超额收益

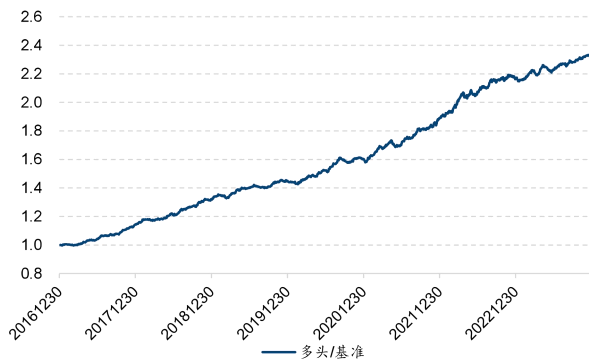


资料来源: WIND, 国信证券经济研究所整理

图 52 展示了周频调仓精选复合因子多头相对基准的净值走势, 图 53 展示了该周频调仓精选复合因子多头和空头相对基准的分年度表现。可以看到周频调仓下多

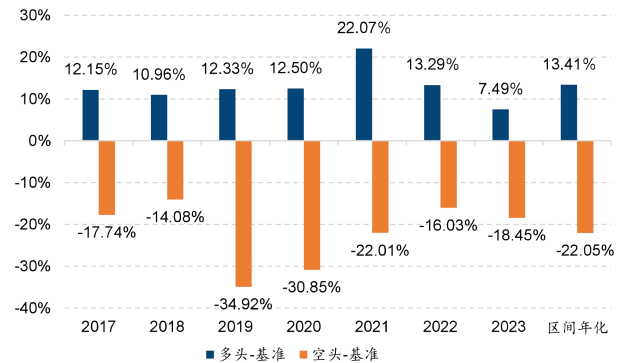
头组合年化超额收益 13.41%，空头组合年化超额收益-22.05%。总体来看，周频调仓下的组合绩效表现比月频调仓下的组合绩效表现更为稳健。

图52：周频调仓精选复合因子多头相对基准净值走势



资料来源：WIND，国信证券经济研究所整理

图53：周频调仓精选复合因子多空分年度超额收益



资料来源：WIND，国信证券经济研究所整理

表 13 展示了“周频大单及漫长订单复合因子”与“周频基于订单特征的精选复合因子”在绩效上的表现。同样的，我们观察到精选复合因子的 RankIC 均值略微提升，但是其 RankICIR 及周度胜率提升较为明显。

表13：周频调仓大单及漫长订单复合因子和周频调仓基于订单特征的精选复合因子绩效对比

|           | RankIC 均值 | 年化 RankICIR | RankIC 周胜率 | 多头年化超额 | 空头年化超额  | 周度自相关系数 |
|-----------|-----------|-------------|------------|--------|---------|---------|
| 复合因子-周频   | 5.1%      | 5.41        | 78.4%      | 13.11% | -21.80% | 0.98    |
| 精选复合因子-周频 | 5.2%      | 6.57        | 82.6%      | 13.41% | -22.05% | 0.97    |

资料来源：WIND，国信证券经济研究所整理

## 参数敏感性检验

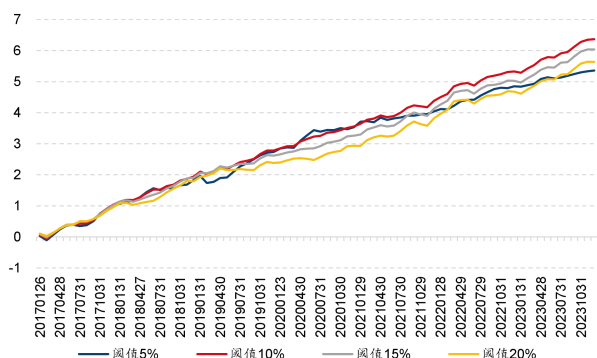
在前述分析中，我们对“大单”及“漫长订单”的判断，均是基于前 10%分位点的方式进行划分。本小节，我们对划分阈值参数的敏感性进行检验。

### 改进大单交易占比因子参数敏感性分析

为了检验“改进大单交易占比”因子对不同阈值划分的敏感性，我们将划分阈值设置为 5%、10%、15%和 20%，观察不同阈值参数下该因子的累计 RankIC 走势和分组月均超额收益，如图 54 和图 55 所示。

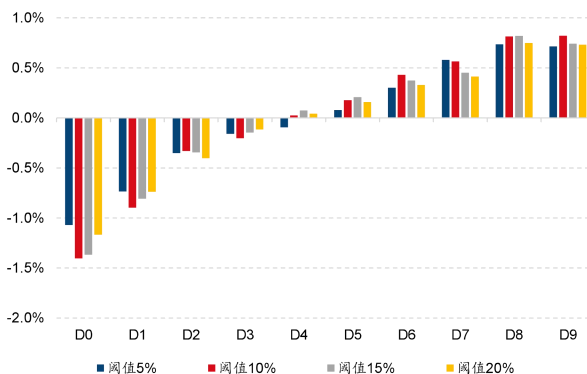
可以看到，不同参数阈值设定下，“改进大单交易占比”因子的累计 RankIC 走势较为稳健，分组月均超额也展现出较好的区分程度，表明该因子的稳定性并不随着参数的变化而发生明显的变化。

图54: 不同阈值下改进大单交易占比因子累计 RankIC 走势



资料来源: WIND, 国信证券经济研究所整理

图55: 不同阈值下改进大单交易占比因子月均超额



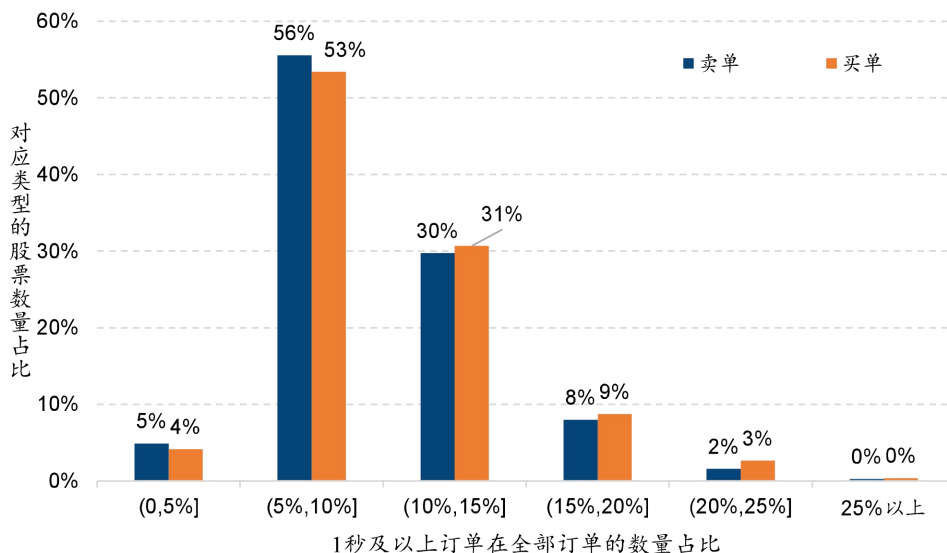
资料来源: WIND, 国信证券经济研究所整理

## 漫长订单交易占比因子参数敏感性分析

在对“漫长订单交易占比”因子的不同阈值划分敏感性进行检验时,我们通过检验每只股票 1 秒及以上订单数量在全部订单数量中的占比分布情况发现,对于“漫长订单”的划分采用 10%阈值是一个相对合理的划分方式。

具体来讲,我们计算每只股票的全部委买单和委卖单中,成交时长大于 0 秒的订单数量占各自委托订单的比例;随后,我们将该比例分为不同的区间,观察其落在不同区间的股票数量占比情况,其结果如图 56 所示。

图56: 1 秒及以上订单在全部订单数量占比分布情况

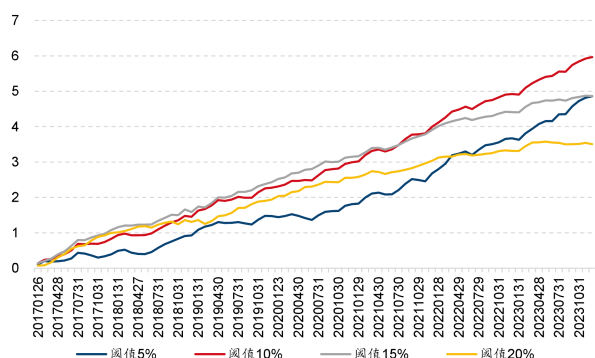


资料来源: WIND, 国信证券经济研究所整理

可以看到,约 60%的股票 1 秒及以上订单占全部订单的比例不足 10%,约 90%的股票 1 秒及以上订单占全部订单的比例在 15%以内。因此,对于“漫长订单”的划分而言,其阈值不建议大于 15% (因为对于绝大多数股票而言,前 15%分位数已经是 0 秒了)。

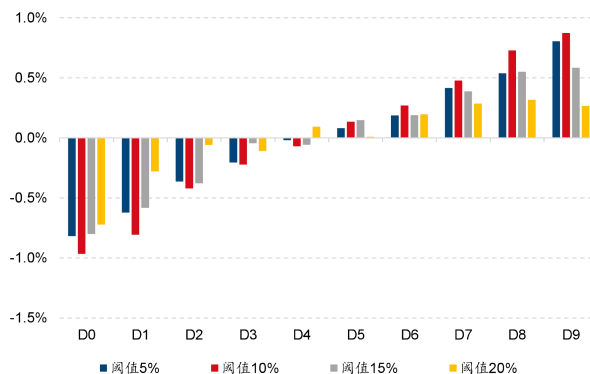
图 57 和图 58 展示了不同参数下，月频调仓的“大单及漫长订单”复合因子的累计 RankIC 走势及分组月均超额收益情况，可以看到不同参数下的复合因子均展现出较强的选股能力。当划分阈值从 10% 增长至 15% 及 20% 时，因子的表现呈现出一定的衰减，这一点与前面提到的个股 1 秒及以上订单数量占比分布情况相吻合。

图57：不同阈值下漫长订单交易占比因子累计 RankIC 走势



资料来源：WIND，国信证券经济研究所整理

图58：不同阈值下漫长订单交易占比因子月均超额



资料来源：WIND，国信证券经济研究所整理

## 总结

**深入高频：从日频行情数据到逐笔成交数据。**

**逐笔成交数据介绍：**逐笔成交数据主要记录了每笔成交的数量、价格、时间信息以及本次成交涉及的买卖双方的订单数量、价格信息。

**考察高频订单的几个维度：**对高频订单的考察通常可从订单大小、成交时长、成交时间、出价高低等维度入手。

**基于订单大小的长单交易占比因子**

**传统大单交易占比因子：**采用分位点的方式对大单进行划分，构建“传统大单交易占比”因子。该因子展现出一定的选股能力，但是总体来讲绩效较为平庸，且稳定性较差。

**传统大单交易占比因子拆解：**根据每笔成交买卖订单是否为大单，对传统大单交易占比因子进行拆解，发现子因子方向不同，导致因子效果变差。

**改进大单交易占比因子：**我们对构成大单交易占比因子的子因子方向进行调整构建“改进后大单交易占比”。该因子 RankIC 均值达到 7.6%，年化 RankICIR 为 3.73，月胜率 88.1%，月度自相关系数 0.81。

**基于订单成交时长的漫长订单交易占比因子**

**漫长订单交易占比因子：**采用分位点的方式对订单成交时长进行划分，构建“漫长订单交易占比”因子，其 RankIC 均值为 7.1%，年化 RankICIR 为 3.88，月胜率 84.5%，月度自相关系数 0.82。

### 衡量订单特征的其他维度

我们还可以从“早尾盘属性”及“高低价属性”出发，对订单包含的信息进行多特征维度的衡量。

### 大单及漫长订单复合因子

**大单及漫长订单复合因子：**将“改进大单交易占比”因子和“漫长订单交易占比”进行等权合成，构建“大单及漫长订单”复合因子。复合因子的 RankIC 均值为 8.4%，年化 RankICIR 为 4.39，月胜率 88.1%。周频调仓复合因子的周度 RankIC 均值为 5.1%，年化 RankICIR 为 5.42，周胜率 78.4%。

**复合因子特征：**复合因子更倾向于低估值、大市值、低波动、低换手股票。在剥离掉传统选股因子后，纯净因子仍展现出稳健的选股能力。

### 基于订单特征的成交量占比因子的统一框架

**基于订单特征的成交量占比因子的统一框架：**基于委买单和委卖单的“大单属性”和“漫长属性”，对不同类别的成交记录进行分类，分别检验不同类型成交记录交易占比因子的选股能力。

**基于订单特征的精选复合因子：**精选复合因子的 RankIC 均值为 8.5%，年化 RankICIR 为 5.06，月胜率 90.5%。周频调仓精选复合因子的周度 RankIC 均值为 5.2%，年化 RankICIR 为 6.57，周胜率 82.6%。

## 风险提示

### 市场环境变动风险

市场环境变动可能会导致模型失效。

### 因子失效风险

因子失效、因子拥挤等情况的发生可能导致模型失效。

## 免责声明

### 分析师声明

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

### 国信证券投资评级

| 投资评级标准                                                                                                                                                                                                                                               | 类别         | 级别 | 说明                      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|----|-------------------------|
| 报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即报告发布日后的 6 到 12 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A 股市场以沪深 300 指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普 500 指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。 | 股票<br>投资评级 | 买入 | 股价表现优于市场代表性指数 20%以上     |
|                                                                                                                                                                                                                                                      |            | 增持 | 股价表现优于市场代表性指数 10%-20%之间 |
|                                                                                                                                                                                                                                                      |            | 中性 | 股价表现介于市场代表性指数±10%之间     |
|                                                                                                                                                                                                                                                      |            | 卖出 | 股价表现弱于市场代表性指数 10%以上     |
|                                                                                                                                                                                                                                                      | 行业<br>投资评级 | 超配 | 行业指数表现优于市场代表性指数 10%以上   |
|                                                                                                                                                                                                                                                      |            | 中性 | 行业指数表现介于市场代表性指数±10%之间   |
|                                                                                                                                                                                                                                                      |            | 低配 | 行业指数表现弱于市场代表性指数 10%以上   |

### 重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中所意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

### 证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

## 国信证券经济研究所

### 深圳

深圳市福田区福华一路 125 号国信金融大厦 36 层

邮编：518046 总机：0755-82130833

### 上海

上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层

邮编：200135

### 北京

北京西城区金融大街兴盛街 6 号国信证券 9 层

邮编：100032