

证券研究报告 / 金融工程研究报告

日内成交量分布因子及 Logsig-Alpha 因子生成

---因子选股系列之六

报告摘要：

本报告分为两个部分，以人工构造与算法生成的角度全面地解析了日内成交量的分布，通过高频数据低频化方法构造选股因子。

第一部分从分布角度出发，分四个维度构建成成交量分布相关因子。由于成交量序列存在稀疏性和周期性特点，我们应用了日内 5min 的成交量序列，并且采用分段在日间对比的思想进行因子构造。

日内对数成交量分布因子主要描述对数成交量在日内的离散程度和偏移程度，其合成因子周度 Rank IC 为 6.31%，ICIR 为 0.97，五分组多头年化收益 15.35%，多空年化收益 30.72%。

日内成交量在时间上的分布因子主要描述日内成交量的 U 形形态，以及日内开盘收盘时段成交量的异动与稳定性，其合成因子周度 Rank IC 为 5.16%，ICIR 为 0.98，五分组多头年化收益 14.69%，多空年化收益 23.70%。

日内成交量在收益率上的分布因子主要关注在日内价格显著上涨时段中上行波动与推动价格上行的成交量稳定性或是成交量占比的稳定性，其合成因子周度 Rank IC 为 5.82%，ICIR 为 0.91，五分组多头年化收益 19.21%，多空年化收益 22.39%。

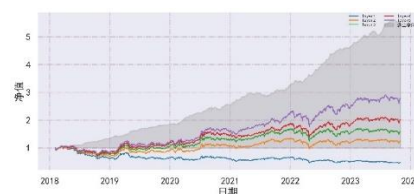
日内成交量在价格上的分布因子主要关注价格区间成交量分布的偏移，这反映日内公允价格区域偏低或是偏高，另外依据成交量在价格上的分布可以对日内反转因子做出改进，合成因子周度 Rank IC 为 2.81%，ICIR 为 0.44，五分组多头年化收益 13.57%，多空年化收益 16.12%。

第二部分构建了一个从序列到因子的高效模型 Logsig-Alpha。利用 log-signature 全面分析成交量序列，对成交量序列进行特征提取，直接提取到的因子一部分具有一定的选股能力，但另一部分线性性较弱。所构建的 Logsig-Alpha 因子生成器可以将序列简便高效地转化为选股因子，使用成交量序列构建的 Logsig-Alpha-v 周度因子和月度因子表现优异，其中周度因子 Rank IC 为 7.52%，ICIR 为 1.24，五分组多头年化收益 20.30%，多空年化收益 36.39%；月度因子 Rank IC 为 9.12%，ICIR 为 1.34，五分组多头年化收益 17.43%，多空年化收益 26.18%。Logsig-Alpha 首先计算序列的 log-signature 特征集，标准化后用因子正交转化模块将其转化为一批具有选股能力但两两相关性较低的基础因子，标准化后等权合成进行输出。高截断阶数的 log-signature 保留了序列中的大部分信息，使得模型具有优异的表现。

每类合成因子总体上表现良好且稳定，Logsig-Alpha 构造的成交量因子表现优异。整体上因子间相关性不高，另外各因子在 Barra 风险因子中略偏向流动性与特异性波动率。因子更契合于小市值股票池。

风险提示：以上分析基于模型结果和历史测算，存在模型失效风险。

Logsig-Alpha-v 周度因子表现



相关报告

《JASON' s alpha: 基本面+量价复合策略》

--20231121

《从 RNN 到 ChatGPT: 大模型的发展与应用》

--20231115

《精选 32 个技术指标在指数上的择时能力分析》

--20231114

《上月规模、成长、预期因子表现相对较优》

--20231103

《债基回报收窄，配置权益比重下降》

--20231030

证券分析师：王琦

执业证书编号：S0550521100001

021-61002390 wangqi_5636@nesc.cn

研究助理：贾英

执业证书编号：S0550122060006

13666061675 jiaying@nesc.cn

目 录

1.	引言：成交量分布中的信息	5
2.	日内成交量分布相关因子构造及测试	8
2.1.	日内对数成交量分布相关因子	8
2.2.	日内成交量在时间上的分布相关因子	10
2.3.	日内成交量在收益率上的分布相关因子	15
2.4.	日内成交量在价格上的分布相关因子	20
3.	Logsig-Alpha 因子生成	22
3.1.	成交量序列的 log-signature	22
3.2.	Logsig-Alpha：由序列到选股因子	23
4.	因子合成及稳健性测试	25
4.1.	相关性测试	25
4.2.	合成因子及其表现	26
4.3.	不同选股域的测试结果	28
5.	总结	29
6.	参考文献	29
7.	风险提示	30
8.	附录	30

图表目录

图 1：某股票日内成交量经验分布图	6
图 2：某股票日内对数成交量经验分布图	6
图 3：某股票日内 1min 成交量时间分布图	6
图 4：某股票日内 5min 成交量时间分布图	6
图 5：某股票日内价格走势与价格区间上的累计成交量	7
图 6：对数成交量日内波动因子分组回测结果	9
图 7：对数成交量日内波动因子 Rank IC	9
图 8：对数成交量日内偏度因子分组回测结果	9
图 9：对数成交量日内偏度因子 Rank IC	9
图 10：早盘成交量占比因子分组回测结果	12
图 11：早盘成交量占比因子 Rank IC	12
图 12：早盘成交量占比稳定性因子分组回测结果	12
图 13：早盘成交量占比稳定性因子 Rank IC	12
图 14：早盘成交量波动因子分组回测结果	12
图 15：早盘成交量波动因子 Rank IC	12
图 16：早盘成交量波动稳定性因子分组回测结果	12
图 17：早盘成交量波动稳定性因子 Rank IC	12
图 18：早午开盘成交量比例因子分组回测结果	13
图 19：早午开盘成交量比例因子 Rank IC	13
图 20：早午开盘成交量比例稳定性因子回测结果	13
图 21：早午开盘成交量比例稳定性因子 Rank IC	13
图 22：晚午收盘成交量比例因子分组回测结果	13
图 23：晚午收盘成交量比例因子 Rank IC	13
图 24：晚午收盘成交量比例稳定性因子回测结果	13
图 25：晚午收盘成交量比例稳定性因子 Rank IC	13

图 26: 价格显著上行价稳量比因子分组回测结果	17
图 27: 价格显著上行价稳量比因子 Rank IC.....	17
图 28: 价格显著上行价稳量比稳定性因子回测结果	17
图 29: 价格显著上行价稳量比稳定性因子 Rank IC.....	17
图 30: 价格显著上行价量波动因子分组回测结果	17
图 31: 价格显著上行价量波动因子 Rank IC.....	17
图 32: 价格显著上行价量波动稳定性因子回测结果	17
图 33: 价格显著上行价量波动稳定性因子 Rank IC.....	17
图 34: 价格显著上行价稳量比因子与其退化因子的绩效比较	19
图 35: 价格显著上行价稳量比稳定性因子与其退化因子的绩效比较	19
图 36: 价格显著上行价量波动因子与其退化因子的绩效比较	19
图 37: 价格显著上行价量波动稳定性因子与其退化因子的绩效比较	20
图 38: 成交量在价格上的分布偏移因子回测结果	21
图 39: 成交量在价格上的分布偏移因子 Rank IC.....	21
图 40: 成交量在价格上的分布反转因子回测结果	21
图 41: 成交量在价格上的分布反转因子 Rank IC.....	21
图 42: 成交量在价格上的分布反转因子与其日内因子的绩效比较	22
图 43: 成交量序列 log-signature 周度 Rank IC 分布	23
图 44: 成交量序列 log-signature 月度 Rank IC 分布	23
图 45: Logsig-Alpha 因子生成器的架构	24
图 46: Logsig-Alpha-v 周度因子分层回测结果	24
图 47: Logsig-Alpha-v 周度因子 Rank IC	24
图 48: Logsig-Alpha-v 月度因子分层回测结果	25
图 49: Logsig-Alpha-v 月度因子 Rank IC	25
图 50: Logsig-Alpha-v 大类因子分层回测结果	26
图 51: Logsig-Alpha-v 大类因子 Rank IC	26
图 52: 日内对数成交量分布因子分层回测结果	27
图 53: 日内对数成交量分布因子 Rank IC.....	27
图 54: 日内成交量在价格上的分布因子回测结果	27
图 55: 日内成交量在价格上的分布因子 Rank IC.....	27
图 56: 日内成交量在收益率上的分布因子回测结果	27
图 57: 日内成交量在收益率上的分布因子 Rank IC.....	27
图 58: 日内成交量在时间上的分布因子回测结果	27
图 59: 日内成交量在时间上的分布因子 Rank IC.....	27
表 1: 日内对数成交量分布相关因子测试结果	9
表 2: 对数成交量日内波动因子分年度测试结果	9
表 3: 对数成交量日内偏度因子分年度测试结果	9
表 4: 日内成交量在时间上的分布相关因子测试结果	11
表 5: 早盘成交量占比因子分年度测试结果	14
表 6: 早盘成交量占比稳定性因子分年度测试结果	14
表 7: 早盘成交量波动因子分年度测试结果	14
表 8: 早盘成交量波动稳定性因子分年度测试结果	14
表 9: 早午开盘成交量比例因子分年度测试结果	14
表 10: 早午开盘成交量比例稳定性因子分年度测试结果	14
表 11: 晚午收盘成交量比例因子分年度测试结果	15
表 12: 晚午收盘成交量比例稳定性因子分年度测试结果	15
表 13: 日内成交量在时间上的分布因子的相关性分析	15
表 14: 日内成交量在收益率上的分布相关因子测试结果	16
表 15: 价格显著上行价稳量比因子分年度测试结果	18
表 16: 价格显著上行价稳量比稳定性因子分年度测试结果	18

表 17: 价格显著上行价量波动因子分年度测试结果	18
表 18: 价格显著上行价量波动稳定性因子分年度测试结果	18
表 19: 日内成交量在收益率上的分布相关因子的相关性分析	18
表 20: 日内成交量在价格上的分布相关因子测试结果	21
表 21: 成交量在价格上的分布偏移因子分年度测试结果	21
表 22: 成交量在价格上的分布反转因子分年度测试结果	22
表 23: Logsig-Alpha-v 周度因子和月度因子测试结果	24
表 24: Logsig-Alpha-v 周度因子分年度测试结果	25
表 25: Logsig-Alpha-v 月度因子分年度测试结果	25
表 26: 大类合成因子的相关性分析	26
表 27: 大类合成因子与 Barra 风格因子的相关性分析	26
表 28: 大类合成因子测试结果	26
表 29: Logsig-Alpha-v 大类因子分年度测试结果	28
表 30: 日内对数成交量分布因子分年度测试结果	28
表 31: 日内成交量在价格上的分布因子分年度测试结果	28
表 32: 日内成交量在收益率上的分布因子分年度测试结果	28
表 33: 日内成交量在时间上的分布因子分年度测试结果	28
表 34: 大类合成因子在沪深 300 中的测试结果	29
表 35: 大类合成因子在中证 500 中的测试结果	29
表 36: 大类合成因子在中证 1000 中的测试结果	29
表 37: 成交量相关细分因子汇总	30

1. 引言：成交量分布中的信息

在因子选股系列前几篇中的高频数据低频化部分我们主要针对日内收益率分布来构造因子，如《高频数据下的已实现高阶矩因子及改进——因子选股系列之四》中构造了描述日内收益率分布偏移程度以及极端尾部风险的已实现超偏度与已实现超尾度因子，而后又对其进行优化改进；《基于高频数据的风险不确定性因子——因子选股系列之五》中构造了一系列描述日内风险日间不确定性的因子，比如日内特异性波动、特异性左尾风险及右尾风险在日间的稳定性因子，合成因子 UOIDR 表现优异且有较强的样本外表现。本篇报告则从不同角度聚焦日内成交量分布信息，介绍并构建成量分布相关因子，其中包括常见且表现不错的因子，也包括我们人工构建的特色因子以及 Logsig-Alpha 生成的成交量相关因子。本篇报告旨在提供一个日内成交量分布因子的集合与研究框架。

日内成交量序列指股票在交易日中每个可交易时段的成交数量构成的序列。成交量序列中蕴含着市场交易情绪的变动以及多空双方的博弈等重要信息，在交易方面，成交量分析是技术分析不可或缺的一部分；在选股方面，依据日内成交量构建的因子亦能捕捉到不同情景中的趋势或者反转等信息，从而实现对收益的预测。

成交量代表着股票的流动性（Liquidity），即在一定情况下订单被成交的难易程度，例如对于流动性较差的股票，大单可能难以被成交或是对成交价格产生较大的影响。通常来说，低流动性对于投资者是不利的，因而这类股票会为其流动性风险提供溢价，即流动性溢价（Liquidity premium），低流动性的股票预期收益可能会更高。

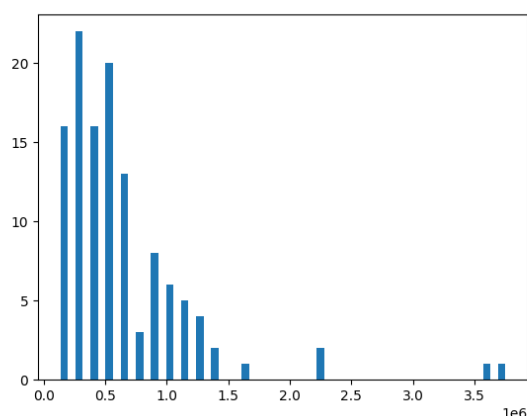
除了成交量本身，其变动与分布同样重要。前者反映流动性与交易情绪在日内或是日间的变化，后者考察交易的均匀或离散程度、偏移程度等统计指标。但通常在利用日内分钟级成交量计算分布特征时会面临两个问题：一是对于流动性较差的股票，其日内某些时段没有成交量；二是日内分钟成交量可能并非独立同分布，这两个问题将会导致分布指标的计算有所偏误。

我们按以下四个维度去考察日内成交量的分布。

1. 对数成交量的经验分布（Empirical distribution）：

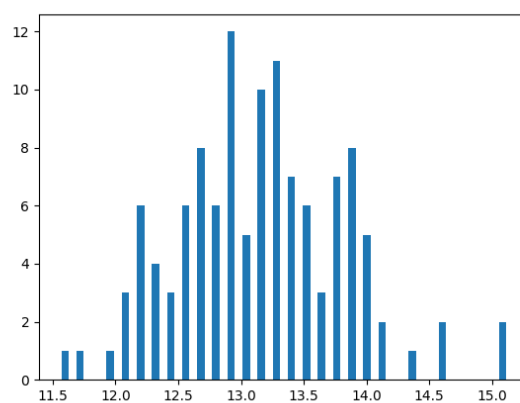
一方面，低流动性股票的分钟级成交量稀疏性在 5min 级别可能会有所改善，所以这里研究对象变为日内 5min 成交量序列。另一方面，由于对数成交量近似服从正态分布，所以我们着重考察对数成交量的一些分布特征。此部分我们介绍的有效因子为对数成交量日内波动因子、对数成交量日内偏度因子。

图 1：某股票日内成交量经验分布图



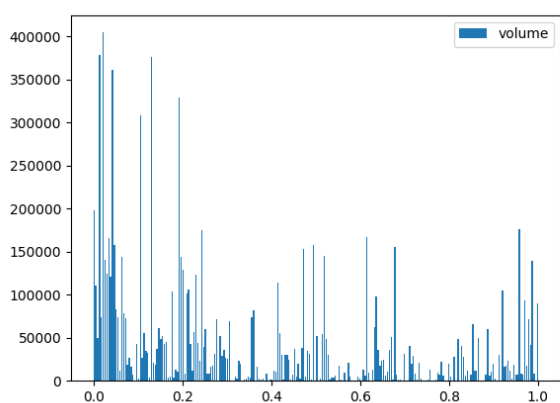
数据来源：东北证券，ricequant

图 2：某股票日内对数成交量经验分布图



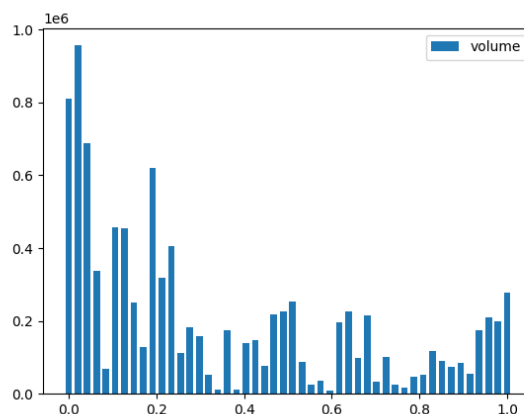
数据来源：东北证券，ricequant

图 3：某股票日内 1min 成交量时间分布图



数据来源：东北证券，ricequant

图 4：某股票日内 5min 成交量时间分布图



数据来源：东北证券，ricequant

2. 成交量在时间上的分布

通常来说，日内成交量在时间上呈现 U 形分布，即开盘和收盘时段的成交量显著更大，而日内其他时间段的成交量分布较为平坦。学术界使用统计方法建模成交量的日内模式，比较著名的是用 Beta 分布的概率密度函数（Probability density function; PDF）来拟合，例如[1]。这种日内 U 形分布给日间成交量序列带来了周期性，所以在实际的指标构建过程中需要剥离或降低这部分影响。另外也有研究从成交量序列模式中提取统计因子以应用于交易策略，例如[2]利用非负矩阵因子化方法（Non-negative matrix factorization; NMF）提取有效信息，其相比于主成分分析方法（Principal component analysis; PCA）得到的统计因子更具有直观的解释性。

在这部分我们对日内时间区间进行切分，再在日间进行对比，主要介绍早盘成交量占比因子、早盘成交量波动因子、早午开盘成交量比例因子、晚午收盘成交量比例因子及其对应的日间稳定性因子。

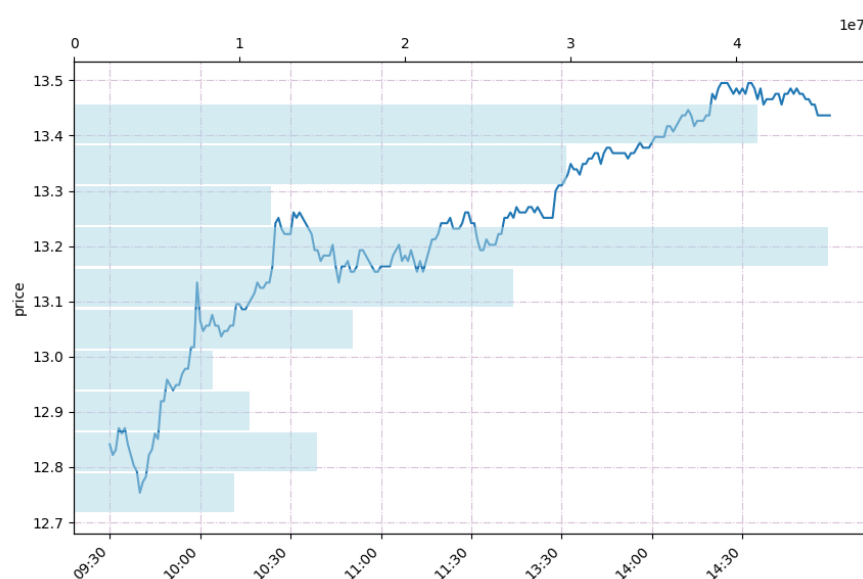
3. 成交量在收益率上的分布

在不同收益率对应的子区间中，成交量的分布也各不相同。另外价格异常波动时段的成交量可能含有更加丰富的信息。所以在因子构造时可以考虑在日内按收益率大小对成交量进行统计，例如统计在价格显著上行阶段中成交量的占比或波动，得到相关因子并对其进行改进，最终构造出价格显著上行价稳量比因子和价格显著上行价量波动因子，以及其对应的日间稳定性因子。

4. 成交量在价格上的分布

价格区间的成交量是指对日内价格分段划分后，处于该价格段内的成交量之和。价格区间成交量的高低表示该价格附近多空博弈的激烈程度以及该价格范围是否被广泛认可。考虑日内各个区间内的成交量，即成交量在价格上的分布，不同的偏移程度可能表示不同的价格偏好，影响未来的收益。本部分我们介绍并构造成成交量在价格上的分布偏移因子以及成交量在价格上的分布反转因子。

图 5：某股票日内价格走势与价格区间上的累计成交量



数据来源：东北证券，ricequant

另一种全面提取成交量分布信息的方法是 log-signature。不同于 Fourier 变换(Fourier transform)从频率中提取信息，signature 以及 log-signature 方法从阶数与面积中提取路径或序列信息，将原始序列转化为特征集，并且特征集含有原始序列中的所有信息。Log-signature 可以看作是 signature 更紧凑的表示，相同阶数下两者含有的信息一致但 log-signature 维数更低。与 signature 不同，log-signature 并没有 universality 性质，所以通常需要搭配非线性模型使用。另外经验上来说，log-signature 对于稀疏样本数据的处理更为稳健，并且更擅于处理可变长度或者不均匀的长序列数据。关于 signature 与 log-signature 的介绍与性质可以参考之前报告《基于 Logsig-RNN 的高频数据低频化选股因子——机器学习系列之三》，相关文章可以参考[3][4][5]。

近年来，signature 与 log-signature 越来越受到学术界以及业界的关注，尤其是在机器学习与深度学习的应用上。[3]将循环神经网络 RNN 与 log-signature 相结合，构建出一类新的模型 Logsig-RNN，它通过时间维度的降维以及数据序列局部结构的提取提升了原始 RNN 的表现，更适用于处理高频率序列。[4]中还介绍了 Logsig-RNN 在生成任务中的应用，在序列生成方面其相对于传统模型有更好的表现。[4]的另一个贡献是建立了一个新的 Sig-WGAN 框架来解决 GAN 在序列数据生成方面的效率问题和训练不稳定性问题。[6]将 signature 与线性模型应用在美国 GDP 增长的即时预测方面，结果显示其相对基于 New York Fed 的动态因子模型有更低的误差。另外

在原油价格数据的预测上也有不错的表现。[7]使用 signature 和 log-signature 去度量分布之间的相似性，并应用到 GAN 生成数据的评估上。在具体实践时，Signatory[8]模块可以帮助我们计算路径的 signature 或 log-signature。

本报告的内容安排如下：

第 2 章分四个维度介绍日内成交量分布相关因子的构造，并对因子进行测试分析，对部分因子进行改进。第 3 章首先介绍成交量序列的 log-signature，其有效实现了对成交量序列信息的提取与转化，直接提取到的部分特征也具有一定的选股效果。接下来利用 Logsig-Alpha 因子生成器融合成交量 log-signature 中的信息生成选股因子，即仅含有成交量信息的因子，该因子全面地总结了成交量序列的分布，表现优异。第 4 章对因子进行合成，并测试大类因子之间的相关性以及其与风险因子的相关性，另外做稳健性测试。第 5 章总结。

2. 日内成交量分布相关因子构造及测试

2.1. 日内对数成交量分布相关因子

如引言中所述，日内的对数成交量近似服从正态分布，另外采样频率为 5min 时可以较好的缓解成交量稀疏性问题，所以这里我们使用日内 5min 对数成交量来构建一些经验分布因子。成交量反映投资者的行为特征，对数成交量日内波动因子描述日内交易情绪的持续性和稳定性，而对数成交量日内偏度因子描述交易情绪的偏移程度。下面是因子的计算方式。

(1) 对数成交量日内波动因子 (logvolume_intraday_std):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \sigma(\log V_s).$$

(2) 对数成交量日内偏度因子 (logvolume_intraday_skew):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t skew(\log V_s).$$

其中， $V_s = \{V_s^i\}_{i=1}^n$ 表示第 s 天的日内成交量序列， V_s^i 表示第 s 天第 i 个分段对应的成交量， $\sigma(\cdot)$ 与 $skew(\cdot)$ 表示计算样本标准差和计算样本偏度， τ 表示用来平滑信号的窗口期长度。

对数成交量日内波动较低时，当日订单成交更加均匀，交易情绪也相对稳定，价格受到的信息冲击较小，通常这类资产具有更高的预期收益。对数成交量日内偏度较低时，通常表明日内对数成交量的分布形态左偏，出现短时间极高成交量的概率较小，分段成交量一般集中在中间位置，对外部信息的敏感度较弱，预期收益也可能更高。所以这两个因子在逻辑上均是负向的。

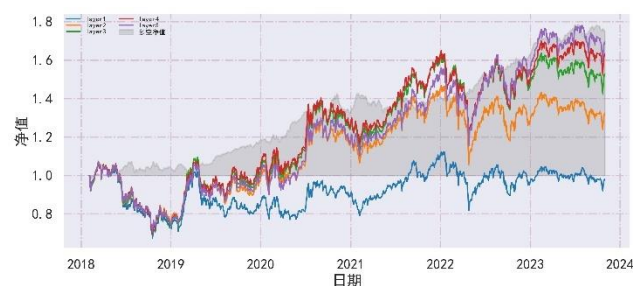
我们使用 2018 年 1 月至 2023 年 11 月的数据进行回测，回测范围是全市场剔除 ST/ST* 以及当日停牌的股票，分五组进行回测，周频调仓，因子平滑的窗口期长度为 20，因子回测前进行去极值、标准化以及行业市值中性化操作，最后乘以因子方向。

表 1：日内对数成交量分布相关因子测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
对数成交量日内波动因子	1.55%	0.22	9.93%	9.35%	0.44	10.66%	6.35%	1.68	7.40%
对数成交量日内偏度因子	5.94%	0.72	16.43%	15.82%	0.55	26.29%	8.33%	3.16	7.90%

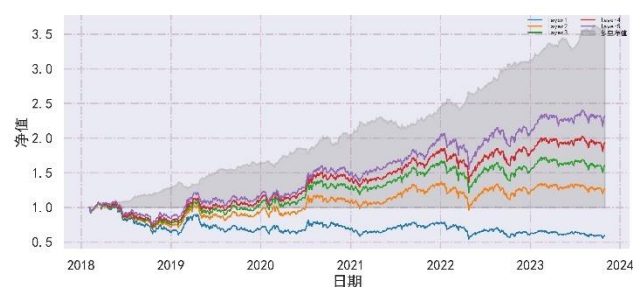
数据来源：东北证券，ricequant

图 6：对数成交量日内波动因子分组回测结果



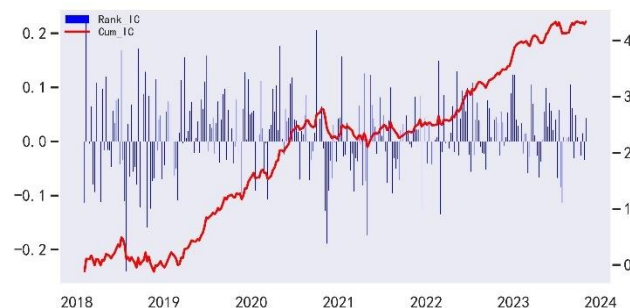
数据来源：东北证券，ricequant

图 8：对数成交量日内偏度因子分组回测结果



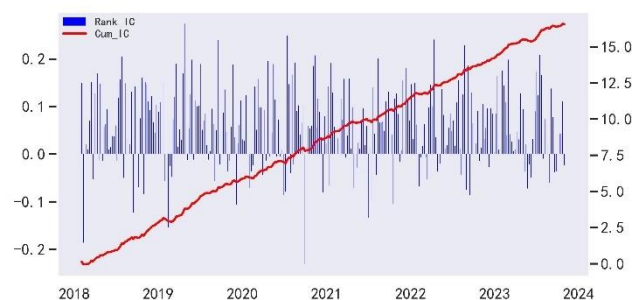
数据来源：东北证券，ricequant

图 7：对数成交量日内波动因子 Rank IC



数据来源：东北证券，ricequant

图 9：对数成交量日内偏度因子 Rank IC



数据来源：东北证券，ricequant

表 2：对数成交量日内波动因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	0.00%	0.00	-27.68%	5.49%	4.70%	6.93%	0.68
2019	3.13%	0.52	34.45%	-0.85%	14.17%	5.56%	2.55
2020	1.55%	0.21	17.69%	-5.57%	12.98%	7.35%	1.76
2021	0.52%	0.08	28.47%	19.22%	5.59%	7.14%	0.78
2022	2.48%	0.44	-1.48%	21.89%	16.34%	5.93%	2.76
2023	1.45%	0.27	9.26%	16.20%	9.14%	4.63%	1.97
汇总	1.55%	0.22	9.93%	9.35%	10.66%	6.36%	1.68

数据来源：东北证券，ricequant

表 3：对数成交量日内偏度因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	6.31%	0.73	-17.70%	20.05%	34.27%	7.02%	4.88
2019	6.28%	0.72	41.00%	3.98%	29.50%	8.11%	3.64
2020	5.78%	0.62	26.45%	1.46%	25.87%	10.42%	2.48
2021	5.68%	0.72	36.02%	26.23%	21.18%	9.49%	2.23
2022	5.71%	0.76	-0.21%	23.46%	28.62%	7.84%	3.65
2023	5.88%	0.77	13.37%	20.57%	21.26%	5.85%	3.63
汇总	5.94%	0.72	16.43%	15.82%	26.29%	8.34%	3.15

数据来源：东北证券，ricequant

对数成交量日内波动因子有一定的选股表现，但周度表现欠佳，可能还需进一步优化。对数成交量日内偏度因子表现较优，因子五分组多头收益稳定，且持续贡献超

额收益，周度 Rank IC 达到 5.94%，多头超额达到了 15.82%，周度因子多头年化换手约单边 14 倍。

2.2. 日内成交量在时间上的分布相关因子

通常来说，每个交易日中的日内成交量序列在时间上的分布为 U 形，这使得成交量序列存在周期性。由于日内开盘与收盘时段的成交量较为显著，所以我们对时间进行切分，着重考察上午开盘及下午收盘阶段的成交量，另外也关注早午开盘与晚午收盘成交量的比例。将日内特征在时间窗口期中做日间对比和信息融合，对比每个交易日对应时段的特征，在一定程度上规避了周期性影响。

我们介绍并测试了如下因子。

(1) 早盘成交量占比因子 (volume_ratio_1):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \frac{\sum_{i \in J_1} V_s^i}{\sum_{i=1}^n V_s^i}$$

其中， J_1 为早盘对应的分段，我们这里取上午开盘后的 30min。令 $R_s^{(1)} = \frac{\sum_{i \in J_1} V_s^i}{\sum_{i=1}^n V_s^i}$ 表示第 s 天的早盘成交量占比。

(2) 早盘成交量占比稳定性因子 (volume_ratio_1_stability):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \left(R_s^{(1)} - \frac{1}{\tau} \sum_{u=t-\tau+1}^t R_u^{(1)} \right)^2$$

(3) 早盘成交量波动因子 (volume_intraday_std_1):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \frac{\sigma(V_s^{J_1})}{\sum_{i=1}^n V_s^i}$$

其中， $V_s^{J_1} = \{V_s^i\}_{i \in J_1}$ 表示第 s 天的日内早盘成交量序列，这里取上午开盘后的 30min。令 $STD_s^{(1)} = \frac{\sigma(V_s^{J_1})}{\sum_{i=1}^n V_s^i}$ 表示第 s 天的早盘成交量波动。

(4) 早盘成交量波动稳定性因子 (volume_intraday_std_1_stability):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \left(STD_s^{(1)} - \frac{1}{\tau} \sum_{u=t-\tau+1}^t STD_u^{(1)} \right)^2$$

(5) 早午开盘成交量比例因子 (volume_ratio_1_5):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \frac{\sum_{i \in J_1} V_s^i}{\sum_{i \in J_5} V_s^i}$$

其中， J_5 为下午开盘对应的分段，这里都取开盘后的 30min。令 $R_s^{(1,5)} = \frac{\sum_{i \in J_1} V_s^i}{\sum_{i \in J_5} V_s^i}$ 表示第 s 天的早午开盘成交量比例。

(6) 早午开盘成交量比例稳定性因子 (volume_ratio_1_5_stability):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \left(R_s^{(1,5)} - \frac{1}{\tau} \sum_{u=t-\tau+1}^t R_u^{(1,5)} \right)^2$$

(7) 晚午收盘成交量比例因子 (volume_ratio_8_4):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \frac{\sum_{i \in \mathcal{I}_8} V_s^i}{\sum_{i \in \mathcal{I}_4} V_s^i}$$

其中, $\mathcal{I}_4$ 和 $\mathcal{I}_8$ 分别代表中午收盘以及下午收盘对应的分段, 这里都取收盘前的 30min。

令 $R_s^{(8,4)} = \frac{\sum_{i \in \mathcal{I}_8} V_s^i}{\sum_{i \in \mathcal{I}_4} V_s^i}$ 表示第 s 天的晚午收盘成交量比例。

(8) 晚午收盘成交量比例稳定性因子 (volume_ratio_8_4_stability):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \left(R_s^{(8,4)} - \frac{1}{\tau} \sum_{u=t-\tau+1}^t R_u^{(8,4)} \right)^2$$

早盘成交量占比较高的股票, 可能在开盘前或短期内公布过重要的信息, 交易情绪在开盘时段急剧释放, 而后成交量迅速下降。另外噪声交易者和非理性交易者在开盘时段的参与较为活跃, 可能会助推开盘期间成交量的不合理上升, 使得对信息的反应过度。过去一段时间内早盘成交量占比更加稳定的股票, 相对来说不容易出现交易情绪冷热不定或是对信息过度反应的情况, 另外非理性投资者参与的比例也较低, 这类股票往往有更好的预期表现。

相比于早盘成交量占比因子整体考察早盘时段的成交量总和, 早盘成交量波动因子更关注于在早盘时段内部成交量是否出现异常变动的情况。早盘是前日收盘后累积信息的集中释放阶段, 早盘成交量波动越低, 反映价格对信息的反应过程持续稳定, 且早盘中没有短时内突然激增再下降的异常情况。而早盘成交量波动稳定性因子则是想要筛选出早盘成交异常波动持续较低的股票, 即早盘中信息的释放与反应需要更加均匀, 且具有一定的可持续性。

早午开盘成交量比例和晚午收盘成交量比例则是从另一个角度, 即日内成交量 U 形分布的陡峭程度来考察日内信息分布的均匀程度。通常日内开盘时段和收盘时段的成交量会相对高一些, 对比早午开盘成交量比例以及晚午收盘成交量比例可以观察信息释放反应的均衡程度。其比例越低, 表明日内整体交易相对均衡。两者的稳定性因子均衡量了各自在日间的稳定性与可持续性。

以上分析表明, 本节介绍因子的方向均为负。因子测试设定与前一节保持一致, 下面是测试结果。

表 4: 日内成交量在时间上的分布相关因子测试结果

因子名称	Rank	IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
早盘成交量占比因子	4.61%	0.54	9.25%	8.68%	0.54	13.40%	8.91%	1.5	12.53%	
早盘成交量占比稳定性因子	4.89%	1.06	13.95%	13.35%	0.73	22.62%	5.31%	4.26	5.14%	
早盘成交量波动因子	4.33%	0.86	12.87%	12.27%	0.62	19.21%	5.93%	3.24	6.32%	
早盘成交量波动稳定性因子	4.03%	0.85	15.61%	15.01%	0.67	22.74%	4.89%	4.65	5.22%	
早午开盘成交量比例因子	5.56%	0.92	15.93%	15.32%	0.60	27.38%	6.69%	4.09	6.93%	
早午开盘成交量比例稳定性因子	4.35%	0.78	17.09%	16.47%	0.59	27.45%	5.39%	5.09	4.34%	
晚午收盘成交量比例因子	2.58%	0.53	14.78%	14.18%	0.63	19.58%	5.02%	3.9	4.47%	
晚午收盘成交量比例稳定性因子	1.70%	0.36	13.99%	13.39%	0.62	13.74%	4.50%	3.05	2.95%	

数据来源: 东北证券, ricequant

图 10：早盘成交量占比因子分组回测结果



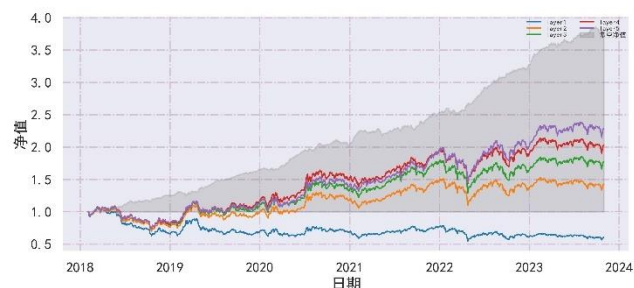
数据来源：东北证券，ricequant

图 12：早盘成交量占比稳定性因子分组回测结果



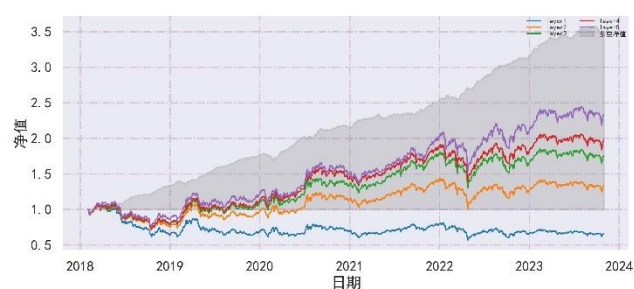
数据来源：东北证券，ricequant

图 14：早盘成交量波动因子分组回测结果



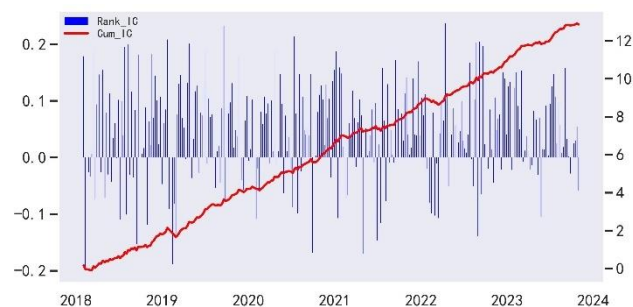
数据来源：东北证券，ricequant

图 16：早盘成交量波动稳定性因子分组回测结果



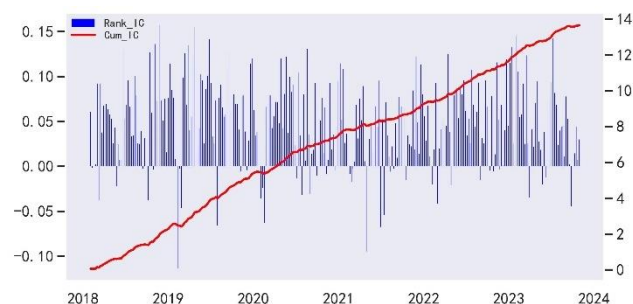
数据来源：东北证券，ricequant

图 11：早盘成交量占比因子 Rank IC



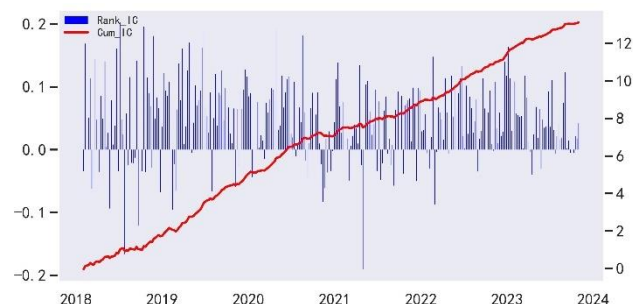
数据来源：东北证券，ricequant

图 13：早盘成交量占比稳定性因子 Rank IC



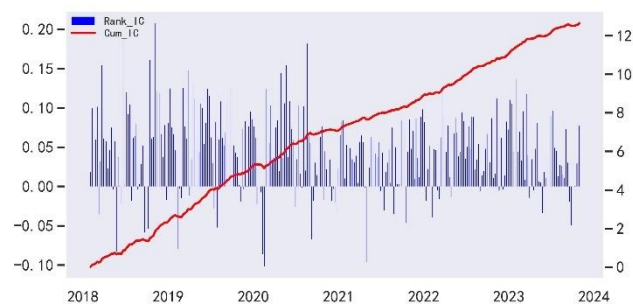
数据来源：东北证券，ricequant

图 15：早盘成交量波动因子 Rank IC



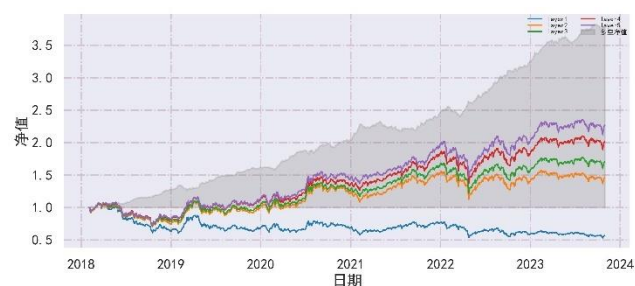
数据来源：东北证券，ricequant

图 17：早盘成交量波动稳定性因子 Rank IC



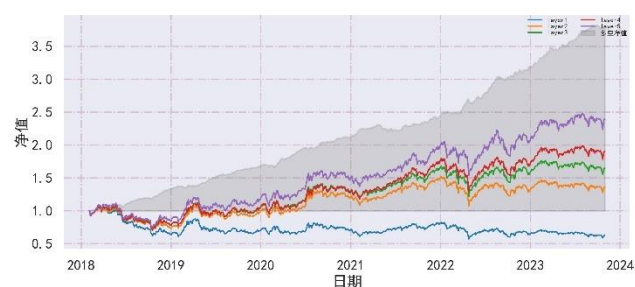
数据来源：东北证券，ricequant

图 18：早午开盘成交量比例因子分组回测结果



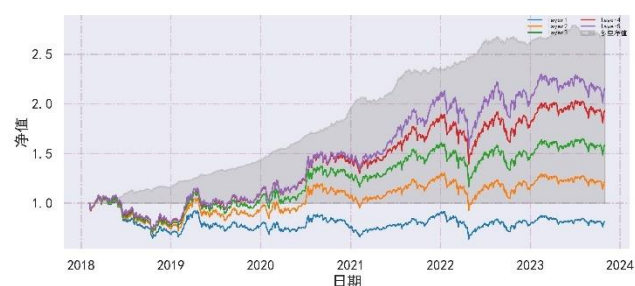
数据来源：东北证券，ricequant

图 20：早午开盘成交量比例稳定性因子回测结果



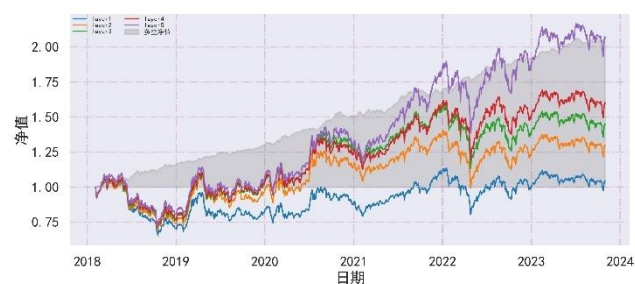
数据来源：东北证券，ricequant

图 22：晚午收盘成交量比例因子分组回测结果



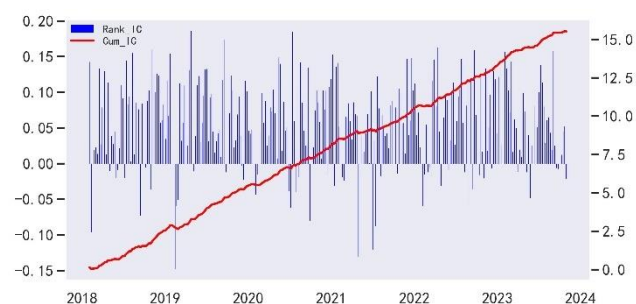
数据来源：东北证券，ricequant

图 24：晚午收盘成交量比例稳定性因子回测结果



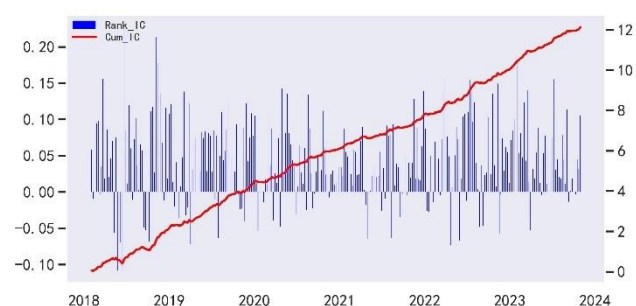
数据来源：东北证券，ricequant

图 19：早午开盘成交量比例因子 Rank IC



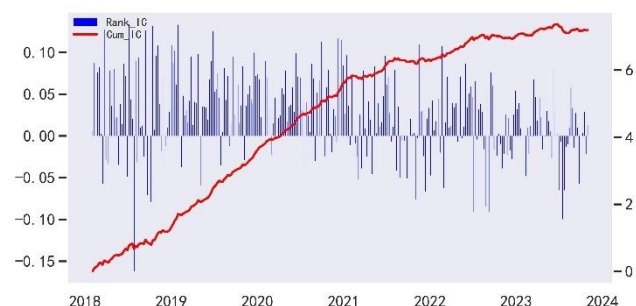
数据来源：东北证券，ricequant

图 21：早午开盘成交量比例稳定性因子 Rank IC



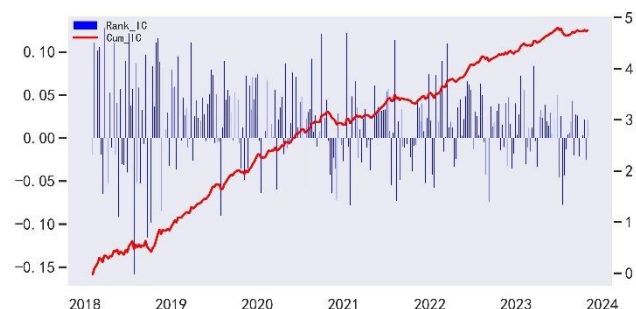
数据来源：东北证券，ricequant

图 23：晚午收盘成交量比例因子 Rank IC



数据来源：东北证券，ricequant

图 25：晚午收盘成交量比例稳定性因子 Rank IC



数据来源：东北证券，ricequant

表 5：早盘成交量占比因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.21%	0.42	-27.61%	5.59%	10.14%	9.02%	1.12
2019	4.81%	0.52	33.50%	-1.55%	14.85%	8.77%	1.69
2020	4.92%	0.61	20.47%	-3.34%	12.48%	9.47%	1.32
2021	4.30%	0.51	31.10%	21.65%	12.78%	10.37%	1.23
2022	4.59%	0.53	-6.97%	15.10%	19.98%	8.69%	2.30
2023	4.83%	0.76	9.67%	16.64%	14.23%	6.34%	2.25
汇总	4.61%	0.54	9.25%	8.68%	13.40%	8.92%	1.50

数据来源：东北证券，ricequant

表 6：早盘成交量占比稳定性因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	5.04%	1.14	-21.83%	14.03%	31.16%	4.67%	6.67
2019	6.38%	1.20	42.86%	5.35%	31.23%	5.77%	5.41
2020	4.39%	0.97	21.15%	-2.79%	18.20%	5.86%	3.11
2021	3.61%	0.79	30.94%	21.51%	13.04%	5.90%	2.21
2022	4.90%	1.24	-0.92%	22.58%	27.08%	4.96%	5.46
2023	5.07%	1.09	12.17%	19.30%	17.36%	4.23%	4.11
汇总	4.89%	1.06	13.95%	13.35%	22.62%	5.32%	4.26

数据来源：东北证券，ricequant

表 7：早盘成交量波动因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.03%	0.89	-21.28%	14.83%	23.06%	4.92%	4.69
2019	5.17%	0.95	36.38%	0.58%	24.37%	6.11%	3.99
2020	4.23%	0.84	21.07%	-2.86%	16.50%	6.23%	2.65
2021	4.21%	0.76	31.75%	22.26%	15.06%	7.24%	2.08
2022	4.34%	0.92	-2.34%	20.82%	24.28%	5.54%	4.38
2023	3.88%	0.78	11.25%	18.32%	12.78%	5.09%	2.51
汇总	4.33%	0.86	12.87%	12.27%	19.21%	5.94%	3.24

数据来源：东北证券，ricequant

表 8：早盘成交量波动稳定性因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.38%	0.75	-16.74%	21.45%	33.22%	4.85%	6.85
2019	5.07%	1.06	40.11%	3.33%	27.45%	5.15%	5.33
2020	3.21%	0.61	20.79%	-3.08%	17.96%	5.37%	3.35
2021	3.64%	0.91	35.07%	25.34%	20.11%	5.09%	3.95
2022	4.12%	1.09	0.67%	24.55%	25.77%	4.51%	5.72
2023	3.74%	0.83	11.22%	18.29%	12.27%	4.19%	2.93
汇总	4.03%	0.85	15.61%	15.01%	22.74%	4.89%	4.65

数据来源：东北证券，ricequant

表 9：早午开盘成交量比例因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	5.73%	0.92	-19.68%	17.16%	32.99%	5.90%	5.60
2019	5.95%	0.93	39.16%	2.63%	28.42%	6.71%	4.23
2020	5.40%	0.93	27.67%	2.44%	27.32%	7.26%	3.76
2021	4.95%	0.79	35.40%	25.64%	22.51%	7.93%	2.84
2022	5.73%	0.94	-0.05%	23.66%	33.50%	6.62%	5.06
2023	5.63%	1.05	13.63%	20.85%	21.92%	5.07%	4.32
汇总	5.56%	0.92	15.93%	15.32%	27.38%	6.70%	4.09

数据来源：东北证券，ricequant

表 10：早午开盘成交量比例稳定性因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.70%	0.61	-17.10%	20.92%	38.16%	5.66%	6.74
2019	4.73%	0.89	39.98%	3.23%	27.86%	5.28%	5.28
2020	3.51%	0.74	25.61%	0.78%	26.16%	5.30%	4.94
2021	3.45%	0.82	32.11%	22.59%	16.25%	4.95%	3.29
2022	4.62%	0.76	1.51%	25.59%	31.42%	6.48%	4.85
2023	5.28%	1.09	17.77%	25.25%	25.71%	4.28%	6.01
汇总	4.35%	0.78	17.09%	16.47%	27.45%	5.39%	5.09

数据来源：东北证券，ricequant

表 11：晚午收盘成交量比例因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	2.77%	0.43	-22.53%	13.00%	19.34%	5.98%	3.24
2019	4.78%	1.15	39.73%	3.05%	24.29%	3.92%	6.19
2020	4.19%	1.09	32.14%	6.02%	36.91%	4.99%	7.40
2021	1.47%	0.33	42.78%	32.49%	21.64%	5.71%	3.79
2022	1.40%	0.31	-5.61%	16.78%	11.34%	5.33%	2.13
2023	0.47%	0.13	4.97%	11.64%	2.78%	3.34%	0.83
汇总	2.58%	0.53	14.78%	14.18%	19.58%	5.02%	3.90

数据来源：东北证券，ricequant

表 12：晚午收盘成交量比例稳定性因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	1.90%	0.26	-21.73%	14.17%	18.02%	6.47%	2.79
2019	2.89%	0.70	32.70%	-2.14%	12.77%	3.65%	3.50
2020	1.33%	0.30	22.61%	-1.62%	15.89%	4.68%	3.40
2021	1.26%	0.30	40.75%	30.61%	12.71%	4.53%	2.81
2022	1.69%	0.44	-1.73%	21.58%	13.19%	3.96%	3.33
2023	1.04%	0.32	10.87%	17.92%	7.57%	2.95%	2.57
汇总	1.70%	0.36	13.99%	13.39%	13.74%	4.50%	3.05

数据来源：东北证券，ricequant

从测试结果来看，早盘成交量占比因子空头表现更为显著，但多头区分度并不高，这跟因子逻辑一致，即对于早盘成交量占比高的股票可能非理性投资者参与度高，对信息的反应较为剧烈甚至过度，但早盘最低占比不一定说明交易均匀稳定。另外晚午收盘比例的两个因子表现一般，可能是日内两个收盘时段含有的信息不如开盘时段丰富。

从 Rank IC 和 ICIR 来看，早午开盘成交量比例因子（Rank IC: 5.56%，ICIR: 0.92）与早盘成交量占比稳定性因子（Rank IC: 4.89%，ICIR: 1.06）均占优，从五分组因子多头收益来看，早午开盘成交量比例稳定性因子与早午开盘成交量比例因子表现优异，分别实现 17.09% 和 15.93% 的年化收益。

表 13：日内成交量在时间上的分布因子的相关性分析

	早午开盘成交量比例	早午开盘成交量比例稳定性	早盘成交量占比	早盘成交量占比稳定性	早盘成交量波动	早盘成交量波动稳定性	晚午收盘成交量比例	晚午收盘成交量比例稳定性
早午开盘成交量比例	1.00	0.54	0.30	0.35	0.56	0.52	0.16	0.11
早午开盘成交量比例稳定性	0.54	1.00	0.01	0.03	0.03	0.02	0.09	0.16
早盘成交量占比	0.30	0.01	1.00	0.68	0.78	0.61	0.11	0.01
早盘成交量占比稳定性	0.35	0.03	0.68	1.00	0.77	0.84	0.14	0.05
早盘成交量波动	0.56	0.03	0.78	0.77	1.00	0.87	0.15	0.05
早盘成交量波动稳定性	0.52	0.02	0.61	0.84	0.87	1.00	0.13	0.04
晚午收盘成交量比例	0.16	0.09	0.11	0.14	0.15	0.13	1.00	0.77
晚午收盘成交量比例稳定性	0.11	0.16	0.01	0.05	0.05	0.04	0.77	1.00

数据来源：东北证券，ricequant

从相关性表来看，因子与其对应的稳定性因子相关性较高，稳定性因子是保证因子在日间稳定可持续，所以可以考虑将因子与其稳定性因子合成，效果可能更优。另外，早盘信息类的因子间相关性较高，早午开盘比例类因子与其他相关性相对较低，晚午收盘比例类因子由于数据源的差异使得与其余因子相关性低。

2.3. 日内成交量在收益率上的分布相关因子

一般来说，高成交量对应的收益率更为重要，那么反过来，在收益率显著高的区间对应的成交量是否蕴含着重要的信息？换句话说，不同收益率时段的成交量分布是否具有差异性？本节我们考虑在日内不同区间收益率情况下的成交量分布。

考虑日内价格显著上升时段的成交量，结合收益率波动信息，构造如下因子：

(1) 价格显著上行价稳量比因子（sig_up_p_v_ratio）：

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \sigma(r_s^{j_0^s}) \frac{\sum_{i \in j_0^s} V_s^i / |j_0^s|}{\sum_{i=1}^n V_s^i / n}.$$

其中, j_0^s 为第 s 天收益率大于当日均值加 1 倍标准差且大于 0 的区间指标集, 即 $j_0^s = \{i \in J: r_s^i > 0, r_s^i > \sum_{j=1}^n r_s^j / n + \sigma(r_s^j)\}$, 意为第 s 天中价格显著上行的时段, 这里 $J = \{1, \dots, n\}$ 。 $r_s^{j_0^s} = \{r_s^i\}_{i \in j_0^s}$ 为第 s 天价格显著上行的区间收益率集合。令 $R_s^{(0)} = \sigma(r_s^{j_0^s}) \frac{\sum_{i \in j_0^s} V_s^i / |j_0^s|}{\sum_{i=1}^n V_s^i / n}$ 表示第 s 天价格显著上行时收益率波动乘以成交量占比。

(2) 价格显著上行价稳量比稳定性因子 (sig_up_p_v_ratio_stability):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \left(R_s^{(0)} - \frac{1}{\tau} \sum_{u=t-\tau+1}^t R_u^{(0)} \right)^2.$$

(3) 价格显著上行价量波动因子 (sig_up_p_v_intraday_std):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \frac{\sigma(r_s^{j_0^s}) \sigma(V_s^{j_0^s})}{\sum_{i=1}^n V_s^i / n}.$$

其中, $V_s^{j_0^s}$ 表示第 s 天中价格显著上行的时段对应的成交量集合。令 $STD_s^{(0)} = \sigma(r_s^{j_0^s}) \sigma(V_s^{j_0^s})$ 表示第 s 天价格显著上行时收益率波动与成交量波动之积除以成交量均值。

(4) 价格显著上行价量波动稳定性因子 (sig_up_p_v_intraday_std_stability):

$$\frac{1}{\tau} \sum_{s=t-\tau+1}^t \left(STD_s^{(0)} - \frac{1}{\tau} \sum_{u=t-\tau+1}^t STD_u^{(0)} \right)^2.$$

当价格显著上行时, 通常希望这些时段的收益率波动较低, 即期望在日内实现稳定上涨。同时在这些时段中, 对过高成交量占比是厌恶的, 因为异常高的成交量推动的显著涨幅可能为过度反应, 未来出现反转的概率相对较大。另外, 价格显著上行对应成交量占比在日间的波动越小, 则说明每个交易日价格上行过程中都有一定比例的成交量推动, 且比例持续稳定。

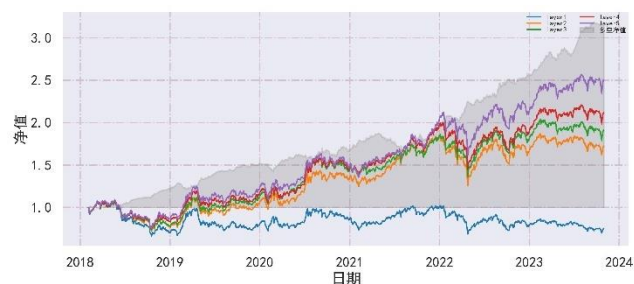
另一方面, 在日内这些时段中同样需要保证成交量是稳定的, 即稳定的成交量推动平稳的上涨, 事实上可以规避掉区间内成交量异动带来的风险。同样, 稳定性因子则是希望这种日内性质在窗口期内保持不变。

表 14: 日内成交量在收益率上的分布相关因子测试结果

因子名称	Rank	IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
价格显著上行价稳量比因子	6.30%	0.83		18.01%	17.39%	0.59	23.14%	7.85%	2.95	10.68%
价格显著上行价稳量比稳定性因子	5.20%	0.93		20.59%	19.95%	0.63	21.89%	5.36%	4.08	5.40%
价格显著上行价量波动因子	5.93%	0.90		18.46%	17.84%	0.60	22.50%	6.73%	3.34	8.75%
价格显著上行价量波动稳定性因子	4.84%	0.93		20.39%	19.75%	0.63	19.62%	4.86%	4.04	5.34%

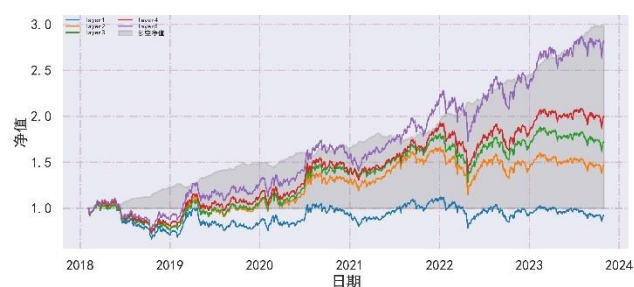
数据来源: 东北证券, ricequant

图 26：价格显著上行价稳量比因子分组回测结果



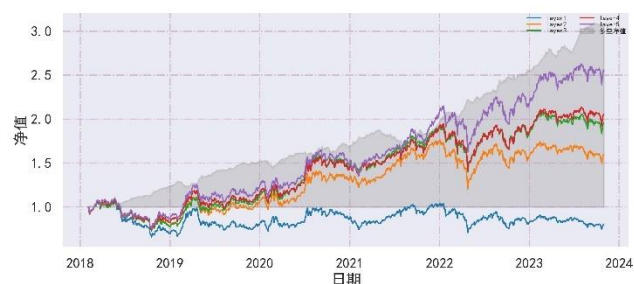
数据来源：东北证券，ricequant

图 28：价格显著上行价稳量比稳定性因子回测结果



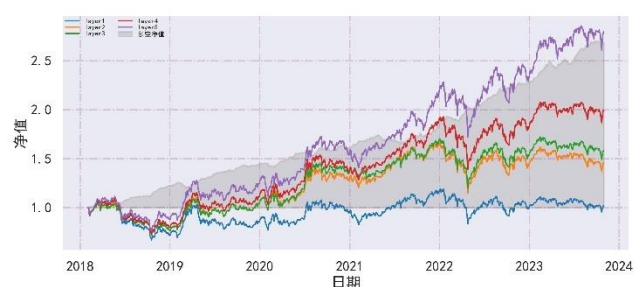
数据来源：东北证券，ricequant

图 30：价格显著上行价量波动因子分组回测结果



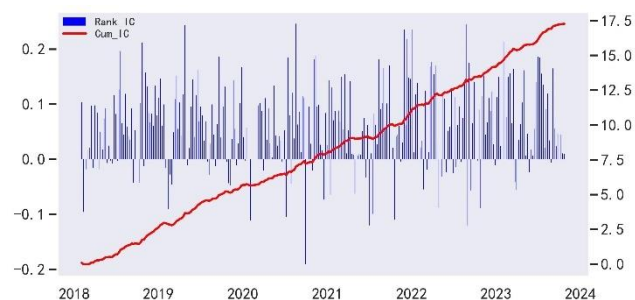
数据来源：东北证券，ricequant

图 32：价格显著上行价量波动稳定性因子回测结果



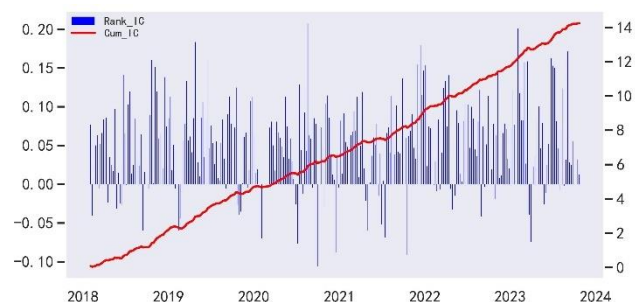
数据来源：东北证券，ricequant

图 27：价格显著上行价稳量比因子 Rank IC



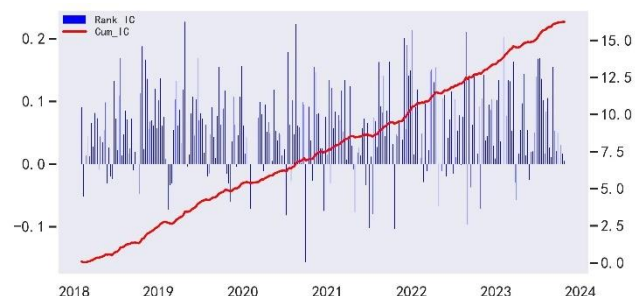
数据来源：东北证券，ricequant

图 29：价格显著上行价稳量比稳定性因子 Rank IC



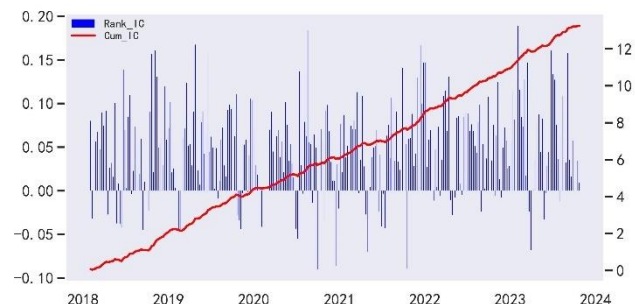
数据来源：东北证券，ricequant

图 31：价格显著上行价量波动因子 Rank IC



数据来源：东北证券，ricequant

图 33：价格显著上行价量波动稳定性因子 Rank IC



数据来源：东北证券，ricequant

表 15：价格显著上行价稳量比因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	5.82%	0.90	-15.53%	23.22%	26.04%	4.60%	5.67
2019	6.33%	0.90	42.98%	5.44%	24.95%	6.15%	4.06
2020	5.13%	0.63	21.37%	-2.62%	11.93%	9.24%	1.29
2021	6.15%	0.76	35.74%	25.96%	16.03%	10.39%	1.54
2022	6.63%	0.78	1.72%	25.85%	33.19%	8.86%	3.74
2023	7.94%	1.09	19.99%	27.61%	30.19%	5.46%	5.53
汇总	6.30%	0.83	18.01%	17.39%	23.14%	7.85%	2.95

数据来源：东北证券，ricequant

表 16：价格显著上行价稳量比稳定性因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.82%	0.90	-13.68%	25.91%	26.63%	3.83%	6.95
2019	5.16%	0.96	44.84%	6.82%	22.72%	4.49%	5.06
2020	4.06%	0.69	22.94%	-1.36%	11.47%	6.20%	1.85
2021	5.43%	0.96	38.87%	28.86%	17.34%	6.58%	2.64
2022	5.39%	1.07	4.01%	28.68%	26.80%	5.72%	4.68
2023	6.48%	1.01	23.81%	31.68%	28.53%	4.44%	6.42
汇总	5.20%	0.93	20.59%	19.95%	21.89%	5.36%	4.08

数据来源：东北证券，ricequant

表 17：价格显著上行价量波动因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	5.43%	0.97	-15.45%	23.33%	27.28%	4.22%	6.46
2019	6.01%	0.96	43.53%	5.85%	24.40%	5.49%	4.44
2020	4.88%	0.71	21.89%	-2.20%	12.34%	7.71%	1.60
2021	5.82%	0.82	36.95%	27.09%	16.16%	8.80%	1.84
2022	6.25%	0.88	1.79%	25.93%	30.89%	7.52%	4.10
2023	7.33%	1.12	19.98%	27.61%	26.42%	4.93%	5.36
汇总	5.93%	0.90	18.46%	17.84%	22.50%	6.74%	3.34

数据来源：东北证券，ricequant

表 18：价格显著上行价量波动稳定性因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.60%	0.85	-13.44%	26.27%	26.02%	3.68%	7.06
2019	4.78%	0.96	43.27%	5.66%	19.25%	4.20%	4.58
2020	3.76%	0.70	24.15%	-0.38%	12.25%	5.46%	2.24
2021	5.06%	0.94	39.70%	29.64%	14.67%	5.91%	2.48
2022	5.01%	1.14	3.38%	27.91%	24.17%	5.06%	4.77
2023	5.94%	1.01	22.54%	30.33%	22.90%	4.23%	5.41
汇总	4.84%	0.93	20.39%	19.75%	19.62%	4.86%	4.03

数据来源：东北证券，ricequant

从 Rank IC 来看，价格显著上行价稳量比因子与价量波动因子更高，分别达到 6.30% 和 5.93%。从五分组因子多头收益的角度来看，价格显著上行价稳量比与价量波动两个稳定性因子表现更好，分别实现 20.59% 与 20.39%。

表 19：日内成交量在收益率上的分布相关因子的相关性分析

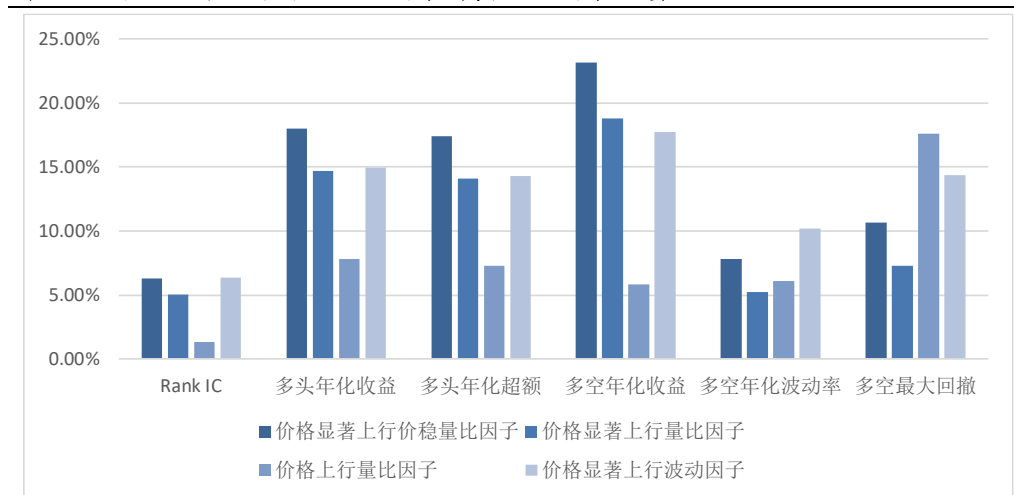
	价格显著上行价稳量比	价格显著上行价稳量比稳定性	价格显著上行价量波动	价格显著上行价量波动稳定性
价格显著上行价稳量比	1.00	0.90	0.93	0.81
价格显著上行价稳量比稳定性	0.90	1.00	0.85	0.89
价格显著上行价量波动	0.93	0.85	1.00	0.92
价格显著上行价量波动稳定性	0.81	0.89	0.92	1.00

数据来源：东北证券，ricequant

因子相关性矩阵说明 4 个因子有着较强的相关性，原因在于 4 个因子均反映成交量稳定持续配合价格平稳上涨现象，后续将进行因子合成。

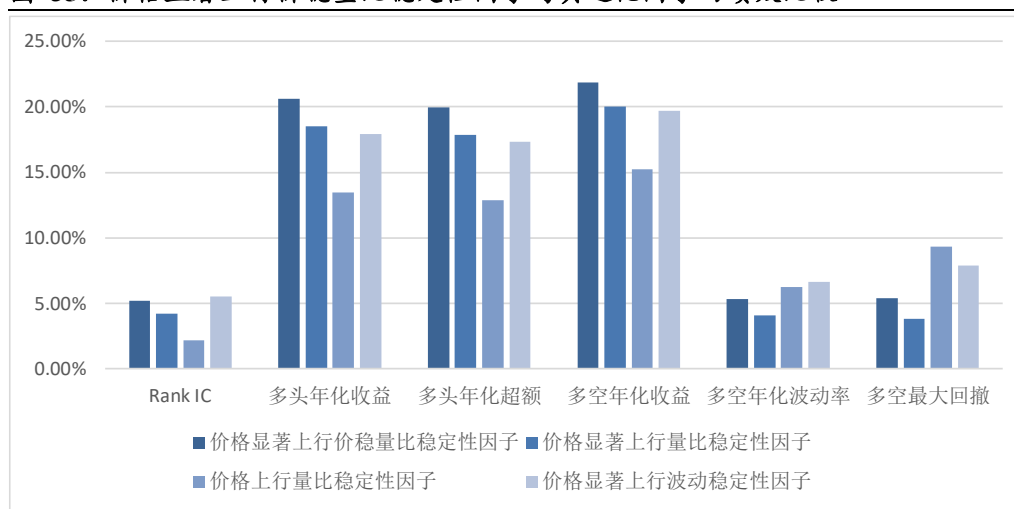
本节最后将因子与部分退化因子进行对比。

图 34：价格显著上行价稳量比因子与其退化因子的绩效比较



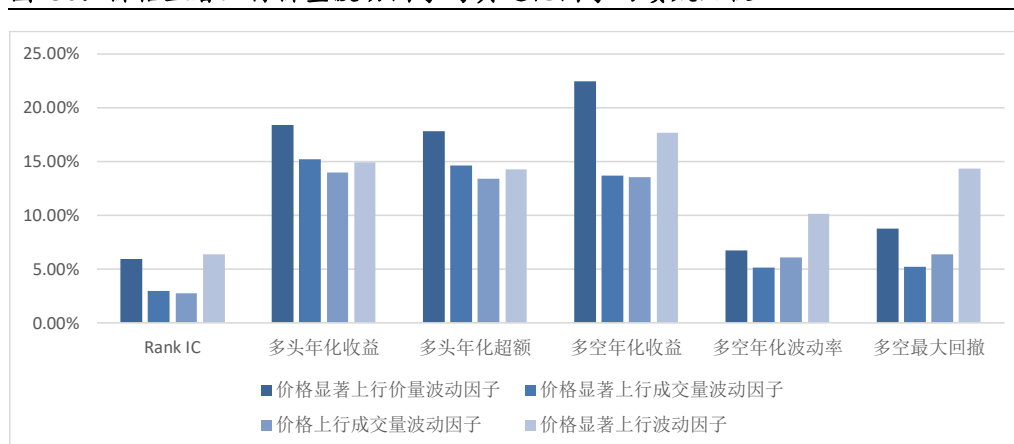
数据来源：东北证券，ricequant

图 35：价格显著上行价稳量比稳定性因子与其退化因子的绩效比较



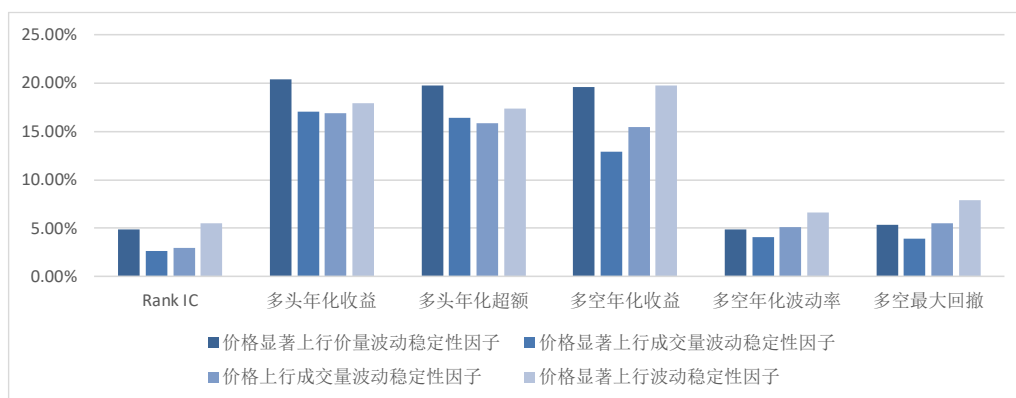
数据来源：东北证券，ricequant

图 36：价格显著上行价量波动因子与其退化因子的绩效比较



数据来源：东北证券，ricequant

图 37：价格显著上行价量波动稳定性因子与其退化因子的绩效比较



数据来源：东北证券，ricequant

退化因子考虑 3 个类型，第一类是仍考虑价格显著上行情况，但不计算收益率波动部分，这类因子整体的表现相对原始因子较弱。第二类是在第一类的基础上将显著上行改为上行，在成交量比例类型因子中更改后表现大幅削弱，但在成交量波动类因子中第二类与第一类差别不大。第三类是仅考虑价格显著上行情况的收益率波动部分，相比于原始因子多头收益有显著下降。

2.4. 日内成交量在价格上的分布相关因子

我们将日内的价格区间进行分段，然后计算每个区间的成交量和，称为价格区间成交量。价格区间成交量的高低表示该价格附近多空博弈的激烈程度以及该价格范围是否被广泛认可。较高的价格区间成交量表明当前价格区间为价格接受区或公允区域，反之为价格拒绝区或非公允区域。日内价格的运动便是从一个公允区域转到下一个公允区域，对于非公允区域仅会短暂停留。另一个重要的概念是 POC (Point of control)，它指最高价格区间成交量对应的价格，通常机构在 POC 附近积累了最多数量的成交量，所以是市场参与者一个重要的参考点。

考虑日内成交量在价格上的分布，不同的偏移程度可能表示不同的价格偏好，影响未来的收益。另外考虑 POC 与收盘价的关系，可以对传统反转因子进行改进。以下是因子构造方法。

(1) 成交量在价格上的分布偏移因子 (v_p_skewness):

将日内价格按照 0, 0.1, ..., 0.9, 1 分位数分成 10 组，每组用位置 0, 0.1, ..., 0.9 进行标号，计算每组价格区间对应的成交量和后除以当日总成交量标准化。然后将标准化后的成交量作为标号的频率，计算其 Pearson 中值偏度 (Pearson's median skewness) 再在窗口期内取均值。

其中，对于随机变量 X ，其均值为 μ ，中位数为 m ，标准差为 σ ，其 Pearson 中值偏度定义为 $\frac{3(\mu-m)}{\sigma}$ 。

(2) 成交量在价格上的分布反转因子 (v_p_reversal):

算得每组价格区间对应成交量后，每组成成交量与前后两组相加后（对于头尾情况，自身乘以 2 再分别加后一组与前一组）取值最大一组的标号作为 POC，这样做目的是寻找成交量最大的一个区域，避免出现某组异常高成交量但附近极低的情况。然后根据收盘价所处组别将其转化为标号 C ，记第 s 天的日内收益为 r_s 。

$$\sum_{s=t-\tau+1}^t \mathbb{I}_{r_s < 0} (POC - C + 1)r_s + \mathbb{I}_{r_s \geq 0} (C - POC + 1)r_s.$$

当 Pearson 中值偏度为负时，成交量在价格上分布的均值小于其中位数，分布左偏，即成交量更多分布在日内较高价格上，这可能对于未来价格走势是一种正向的推动。另外，POC 是日内多空博弈最为激烈且相对公允的价格，通常认为收盘价有向公允价格回归的趋势，所以 POC 与 C 的差值表示这种回归的强度。在日内收益为负时，我们希望用回归强度来放大反转效应，在日内收益为正时，我们希望用负向回归强度来增大正向的日内收益，同样提升反转效应。所以逻辑上这两个因子方向均为负，下面是因子测试结果。

表 20：日内成交量在价格上的分布相关因子测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空 Sharpe Ratio	多空最大回撤
成交量在价格上的分布偏移因子	2.93%	0.33	14.18%	13.58%	0.40	9.78%	7.51%	1.3	10.17%
成交量在价格上的分布反转因子	5.50%	0.73	14.00%	13.40%	0.82	30.88%	9.38%	3.29	9.10%

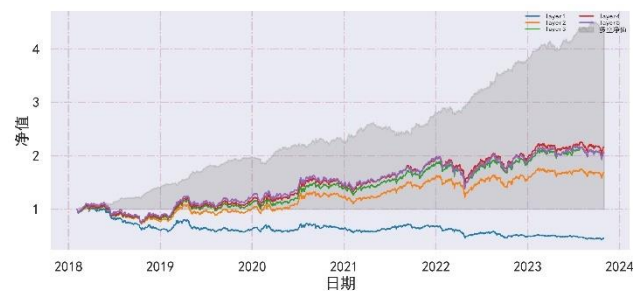
数据来源：东北证券，ricequant

图 38：成交量在价格上的分布偏移因子回测结果



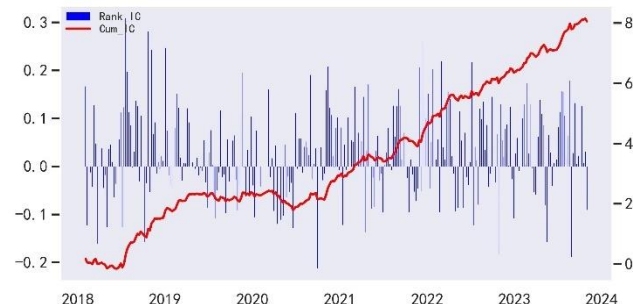
数据来源：东北证券，ricequant

图 40：成交量在价格上的分布反转因子回测结果



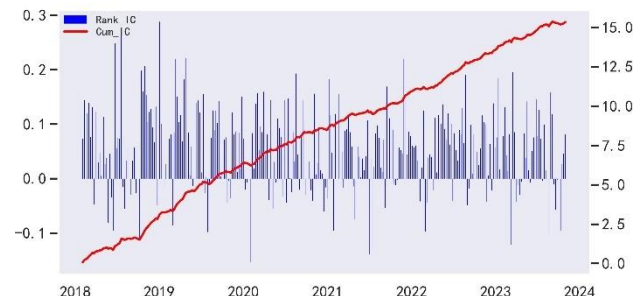
数据来源：东北证券，ricequant

图 39：成交量在价格上的分布偏移因子 Rank IC



数据来源：东北证券，ricequant

图 41：成交量在价格上的分布反转因子 Rank IC



数据来源：东北证券，ricequant

表 21：成交量在价格上的分布偏移因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空 Sharpe Ratio
2018	3.48%	0.33	-21.52%	14.48%	15.83%	7.43%	2.13
2019	1.74%	0.23	33.28%	-1.71%	5.52%	5.39%	1.03
2020	1.20%	0.14	20.00%	-3.72%	7.08%	8.00%	0.89
2021	3.41%	0.38	37.29%	27.41%	6.19%	8.22%	0.75
2022	3.68%	0.39	3.22%	27.71%	14.10%	8.85%	1.59
2023	4.25%	0.49	12.29%	19.43%	10.15%	6.42%	1.58
汇总	2.93%	0.33	14.18%	13.58%	9.78%	7.52%	1.30

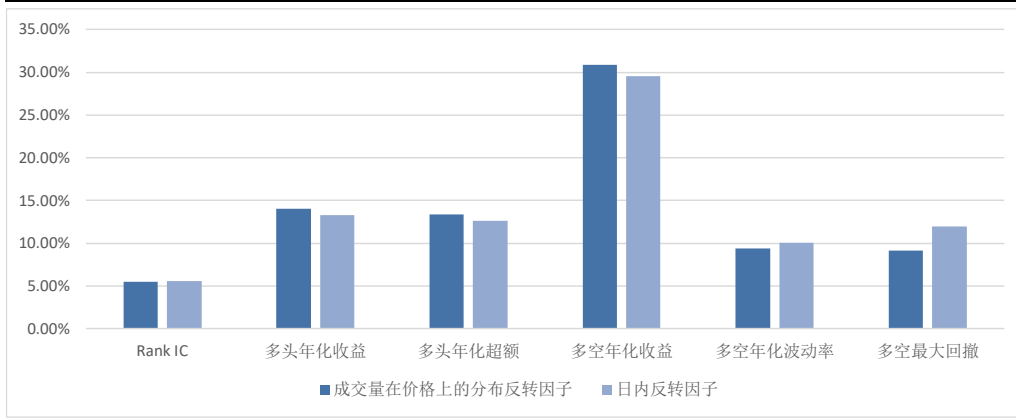
数据来源：东北证券，ricequant

表 22：成交量在价格上的分布反转因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	6.43%	0.73	-16.80%	21.36%	47.30%	8.08%	5.86
2019	7.15%	0.90	45.49%	7.30%	42.64%	7.38%	5.78
2020	4.39%	0.59	19.50%	-4.12%	16.78%	10.49%	1.60
2021	5.04%	0.68	32.67%	23.11%	24.75%	12.46%	1.99
2022	5.42%	0.91	-5.07%	17.45%	38.05%	8.99%	4.23
2023	4.45%	0.59	8.15%	15.02%	19.80%	7.42%	2.67
汇总	5.50%	0.73	14.00%	13.40%	30.88%	9.38%	3.29

数据来源：东北证券，ricequant

图 42：成交量在价格上的分布反转因子与其日内因子的绩效比较



数据来源：东北证券，ricequant

成交量在价格上的分布偏移因子多头近些年来表现较为突出，另外成交量在价格上的分布反转因子 Rank IC 达到了 5.50%，两者多头年化收益均达到约 14%，并且两个因子的相关性仅为 0.07，处于较低的水平。成交量在价格上的分布反转因子相比于日内反转因子在多头收益与多空表现上均有提升，这说明成交量分布信息对日内反转因子可以起到加强作用。

3. Logsig-Alpha 因子生成

不同于前一章人工构造成量分布相关因子，本章利用成交量序列的 log-signature 进行因子生成，生成的因子含有原始序列绝大部分信息，且具有优异的选股效果。

3.1. 成交量序列的 log-signature

Log-signature 方法可以全面提取成交量序列中的信息，蕴含着成交量分布中的有效成分。具体来说，我们将窗口期为 20 天的 5min 对数成交量序列进行 lead-lag 变换得到 (V^{lead}, V^{lag}) ，然后计算其对应的 log-signature 特征集 $(S^1, S^2, S^{1,1}, S^{1,2}, S^{2,1}, S^{2,2}, \dots)$ ，为了更好地提取成交量分布特征，我们将截断阶数定为 10 阶。

这里简单介绍一下 lead-lag 变换。对于原始长度为 $N + 1$ 的 d 维序列 X ，其 lead-lag 变换包含两个长度为 $2N + 1$ 的 d 维序列 X^{lead} 和 X^{lag} ：

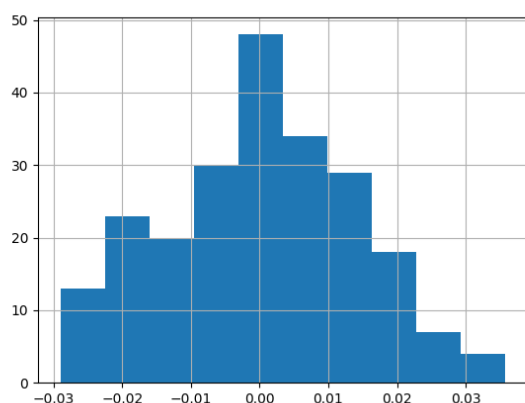
$$X_{t_j}^{lead} = \begin{cases} X_{t_i} & , \text{ if } j = 2i \\ X_{t_i} & , \text{ if } j = 2i - 1 \end{cases}$$

$$X_{t_j}^{lag} = \begin{cases} X_{t_i} & , \text{ if } j = 2i \\ X_{t_i} & , \text{ if } j = 2i + 1 \end{cases}$$

使用成交量序列的 lead-lag 变换来计算具有两方面的原因：其一，log-signature 是对于路径定义的，至少两个维度才能计算；其二，经过 lead-lag 变换计算 log-signature 可以显性地计算出序列的一些变差 (Variation) 特征，更有效地挖掘序列中的信息。

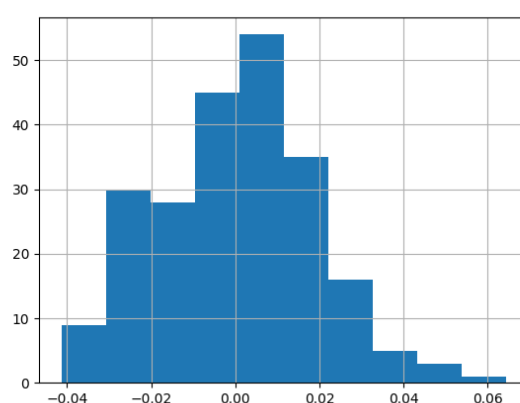
对于得到的 log-signature 特征直接进行 Rank IC 测试，结果如下：

图 43：成交量序列 log-signature 周度 Rank IC 分布



数据来源：东北证券，ricequant

图 44：成交量序列 log-signature 月度 Rank IC 分布



数据来源：东北证券，ricequant

直接计算得到的 log-signature 特征集含有原始成交量序列几乎所有的信息，其中部分因子在周度和月度均具有较好的选股效果，但另一部分因子不具有较好的线性性，下节对这些因子进行进一步的转化融合。

3.2. Logsig-Alpha：由序列到选股因子

本节提出一类新的从序列到因子的因子生成器 Logsig-Alpha，它可以端到端的将序列或数据流转化为因子。Logsig-Alpha 分为两个模块：其一为 log-signature 计算模块，将原始的序列转化为一个特征集，再进行标准化操作后得到基础因子；其二为因子正交化模块，它是一个 MLP 架构，旨在保留因子选股能力的同时降低因子之间的相关性，将正交转换后的因子进行标准化后输出，或是等权合成。架构简单高效，易于操作。

架构中有两个标准化的操作，第一个是对 log-signature 特征的标准化，原因是特征之间可能存在着较大的数量级差异，标准化让模型的训练更加稳定。第二个是对正交转化后的因子进行标准化，目的是平衡各因子的数量级，便于合成以及计算损失。

在实际应用中，我们使用上一节中 lead-lag 变换后的成交量序列，截断阶数为 10。我们从 2018 年开始到 2023 年 11 月按年滚动训练更新模型，每次考虑前 4 年的数据，前 3 年作训练集，后 1 年作验证集，训练标签为 $t+1$ 到 $t+N$ 的收益率（对于周频因子 $N=5$ ，月频因子 $N=20$ ），为了更好的训练效果，我们将标签也进行了标准化。在训练时，随机从训练集中采样日期，将当日内处理后全市场数据作为一个 batch 输入模型优化参数。

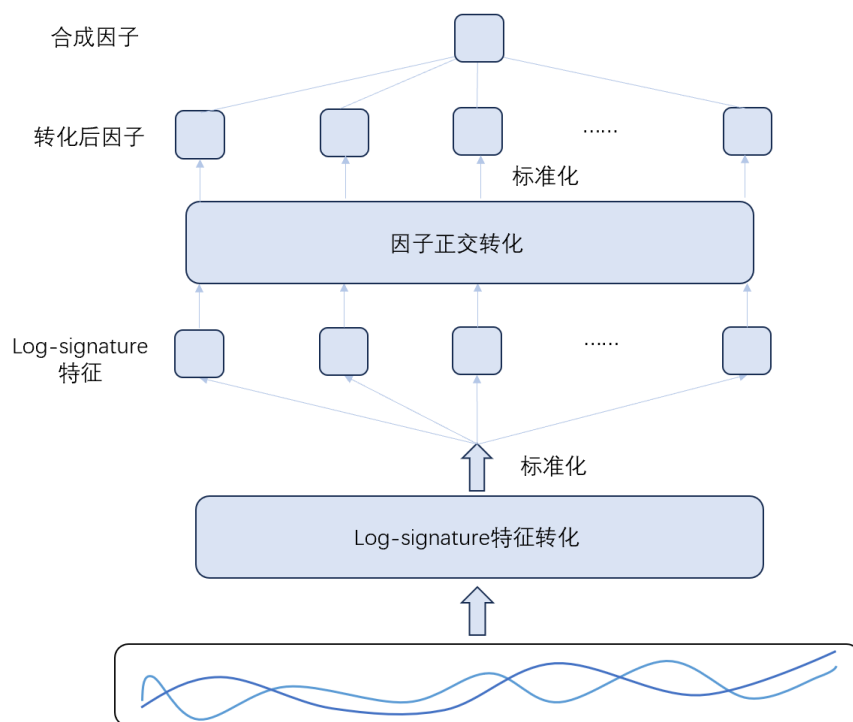
优化目标为最大化合成因子的 IC，但需要约束合成前各基础因子的相关性，所以损失函数 (Loss function) 中需要加入惩罚项，即合成前各基础因子相关系数矩阵的 L^2 范数，规范定义如下：

$$L(X_t, Y_t) = -(\rho(X_t, Y_t) - \lambda \|M_t\|_{L^2}),$$

其中, $X_t = (X_t^1, \dots, X_t^n)$ 为 t 日 n 个股票的合成因子值, $Y_t = (Y_t^1, \dots, Y_t^n)$ 为对应的标签, M_t 为合成前各因子的相关系数矩阵, λ 为对相关性的厌恶系数, 来控制生成因子相关性的强弱。

损失函数中的惩罚部分限制了合成前基础因子的相关性, 而优化合成因子 IC 部分也保证了基础因子具有一定的选股能力。这是因为当基础因子经正交化和标准化后, 优化合成因子的 IC 等价于优化基础因子的 IC。

图 45: Logsig-Alpha 因子生成器的架构



数据来源: 东北证券

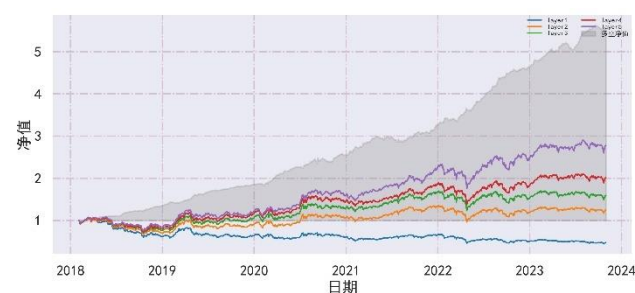
将生成的因子行业市值中性化后进行测试, 回测设定跟上一章相同。

表 23: Logsig-Alpha-v 周度因子和月度因子测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
Logsig-Alpha-v周度因子	7.52%	1.24	20.30%	19.67%	1.04	36.39%	7.11%	5.12	4.90%
Logsig-Alpha-v月度因子	9.12%	1.34	17.43%	16.06%	1.35	26.18%	6.84%	3.83	5.93%

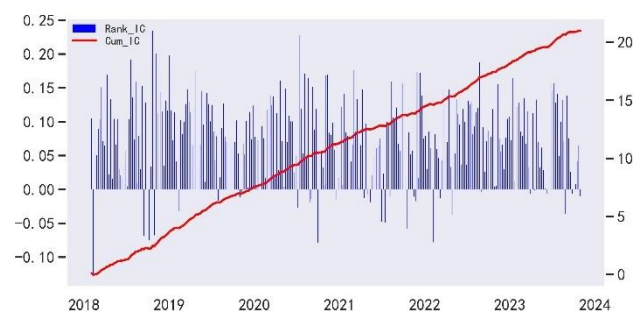
数据来源: 东北证券, ricequant

图 46: Logsig-Alpha-v 周度因子分层回测结果



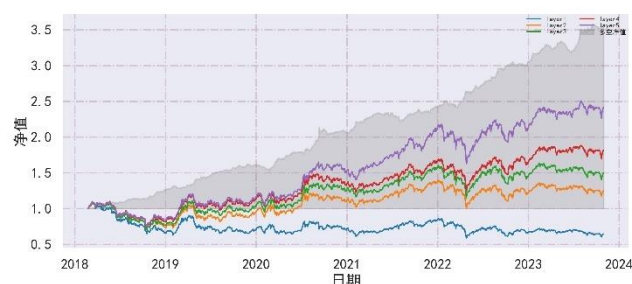
数据来源: 东北证券, ricequant

图 47: Logsig-Alpha-v 周度因子 Rank IC



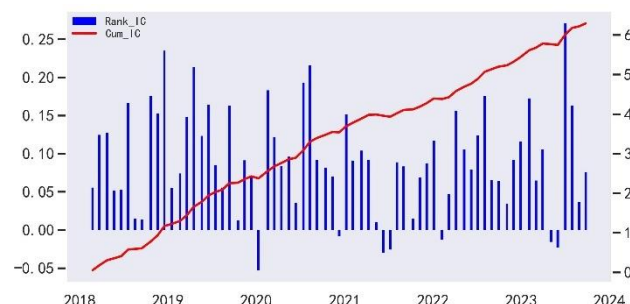
数据来源: 东北证券, ricequant

图 48: Logsig-Alpha-v 月度因子分层回测结果



数据来源：东北证券，ricequant

图 49: Logsig-Alpha-v 月度因子 Rank IC



数据来源：东北证券，ricequant

表 24: Logsig-Alpha-v 周度因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	7.91%	1.01	-17.13%	20.88%	41.62%	6.18%	6.73
2019	8.21%	1.65	45.38%	7.21%	39.53%	5.75%	6.87
2020	7.73%	1.22	32.12%	6.01%	39.69%	9.58%	4.15
2021	6.54%	1.04	40.99%	30.83%	29.93%	8.66%	3.46
2022	7.55%	1.44	6.60%	31.89%	43.85%	5.71%	7.68
2023	7.15%	1.25	13.63%	20.85%	24.90%	5.28%	4.71
汇总	7.52%	1.24	20.30%	19.67%	36.39%	7.11%	5.12

数据来源：东北证券，ricequant

表 25: Logsig-Alpha-v 月度因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	10.64%	1.46	-18.84%	15.51%	32.64%	5.33%	6.12
2019	10.49%	1.81	40.05%	3.28%	29.66%	6.11%	4.86
2020	9.27%	1.16	30.89%	5.02%	29.81%	9.47%	3.15
2021	6.14%	1.10	40.02%	29.93%	18.38%	6.90%	2.66
2022	8.96%	1.75	-2.80%	20.26%	26.74%	6.17%	4.33
2023	9.47%	0.99	15.06%	22.37%	21.95%	6.04%	3.64
汇总	9.12%	1.34	17.43%	16.06%	26.18%	6.85%	3.82

数据来源：东北证券，ricequant

Logsig-Alpha-v 周度因子与 Logsig-Alpha-v 月度因子的 Rank IC 分别达到 7.52%和 9.12%，表现良好且持续稳定；因子五分组多头年化收益分别高达 20.30%和 17.43%，且持续提供正向超额。周度因子和月度因子的优异表现说明 Logsig-Alpha 具有较强的序列信息提取能力。

正交转化后的基础因子的平均 Rank IC 为 0.02，两两相关系数的均值小于 5%。这里 Logsig-Alpha 仅使用成交量序列，但实现了较高的收益，原因可能在于：我们选择了较高的 log-signature 截断阶数（10 阶），这保留了高频序列中绝大部分的信息；在模型中加入了正交转化模块，使得模型可以生成低相关又具有选股能力的因子并进行合成，实际上每个基础因子表示原始 log-signature 特征集中某个独立的具有收益预测能力的部分；在每个交易日对标签也进行了标准化，这可能也在一定程度上提升了模型的表现。

4. 因子合成及稳健性测试

4.1. 相关性测试

由于同一类别中的因子有相对高的相关性，所以我们将类别中的细分因子进行合成，考虑大类因子的相关性及其表现。

表 26：大类合成因子的相关性分析

	Logsig-Alpha-v周度因子	日内对数成交量分布因子	日内成交量在价格上的分布因子	日内成交量在收益率上的分布因子	日内成交量在时间上的分布因子
Logsig-Alpha-v周度因子	1.00	-0.30	-0.05	-0.24	-0.35
日内对数成交量分布因子	-0.30	1.00	0.35	0.49	0.74
日内成交量在价格上的分布因子	-0.05	0.35	1.00	0.18	0.18
日内成交量在收益率上的分布因子	-0.24	0.49	0.18	1.00	0.41
日内成交量在时间上的分布因子	-0.35	0.74	0.18	0.41	1.00

数据来源：东北证券，ricequant

表 27：大类合成因子与 Barra 风格因子的相关性分析

	beta	btop	earyi	growth	lever	liqui	lncap	nlsize	resivol	rstr
Logsig-Alpha-v周度因子	-0.10	0.19	0.09	0.00	0.06	-0.29	-0.05	-0.02	-0.20	-0.06
日内对数成交量分布因子	-0.04	-0.04	-0.10	-0.08	0.06	0.27	-0.13	-0.05	0.31	-0.01
日内成交量在价格上的分布因子	-0.16	0.17	0.05	-0.03	0.17	-0.08	0.09	0.09	0.11	-0.03
日内成交量在收益率上的分布因子	0.07	-0.15	-0.15	-0.06	-0.06	0.24	-0.21	-0.10	0.22	-0.05
日内成交量在时间上的分布因子	0.02	-0.07	-0.10	-0.05	-0.01	0.20	-0.19	-0.11	0.29	-0.05

数据来源：东北证券，ricequant

由相关系数矩阵可以看出，日内对数成交量分布因子与日内成交量在时间上和收益率上的分布因子有较强的相关性，原因在于前者对于分布是整体考查的，而后者是局部考查的。Logsig-Alpha-v 周度因子与人工构造的因子相关性不高，这表明其可以更全面地提取成交量序列信息，为人工构造因子提供较强的信息增益。在 Barra 风格因子相关性上因子略偏向于流动性与特异性波动率。

4.2. 合成因子及其表现

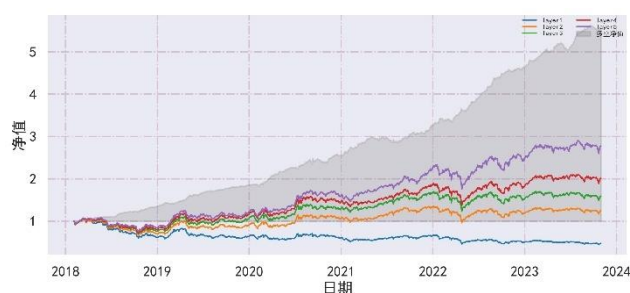
本节测试各个大类合成因子的表现，回测的设置跟之前保持一致。各类合成因子与细分因子相比更加稳健，因子表现均显著，其中 Logsig-Alpha-v 大类因子表现最为突出。

表 28：大类合成因子测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
Logsig-Alpha-v周度因子	7.52%	1.24	20.30%	19.67%	1.04	36.39%	7.11%	5.12	4.90%
日内对数成交量分布因子	6.31%	0.97	15.35%	14.75%	0.59	30.72%	7.48%	4.11	9.27%
日内成交量在价格上的分布因子	2.81%	0.44	13.57%	12.97%	0.79	16.12%	6.53%	2.47	5.43%
日内成交量在收益率上的分布因子	5.82%	0.91	19.21%	18.58%	0.60	22.39%	6.42%	3.49	7.33%
日内成交量在时间上的分布因子	5.16%	0.98	14.69%	14.09%	0.64	23.70%	6.12%	3.87	6.72%

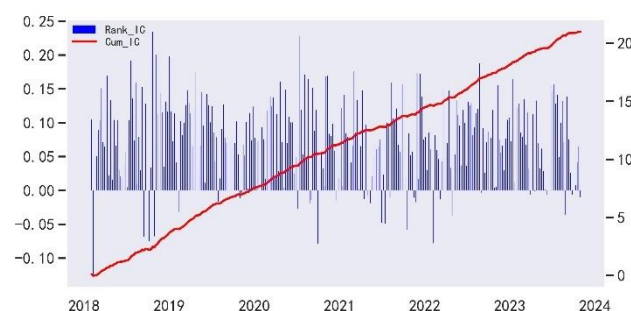
数据来源：东北证券，ricequant

图 50：Logsig-Alpha-v 大类因子分层回测结果



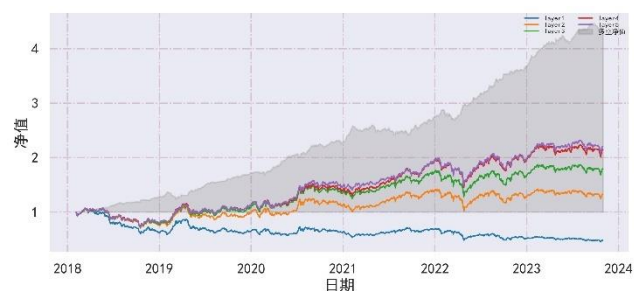
数据来源：东北证券，ricequant

图 51：Logsig-Alpha-v 大类因子 Rank IC



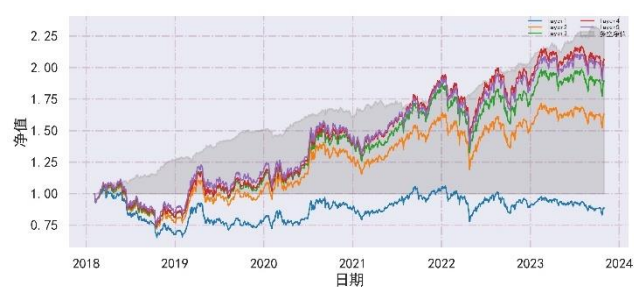
数据来源：东北证券，ricequant

图 52：日内对数成交量分布因子分层回测结果



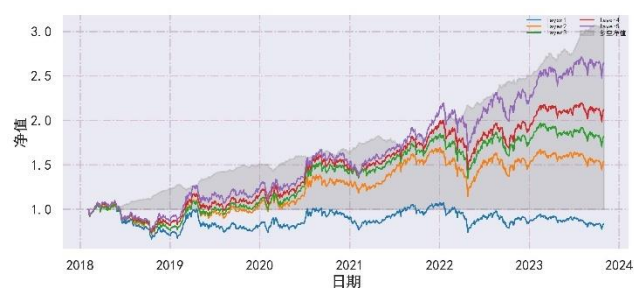
数据来源：东北证券，ricequant

图 54：日内成交量在价格上的分布因子回测结果



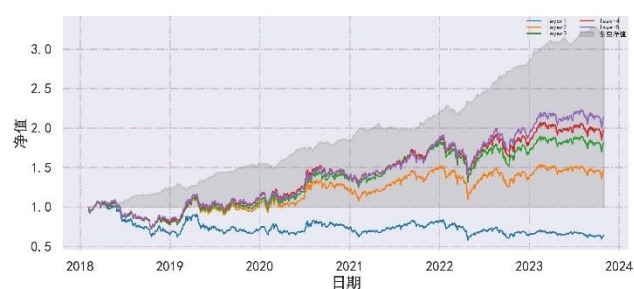
数据来源：东北证券，ricequant

图 56：日内成交量在收益率上的分布因子回测结果



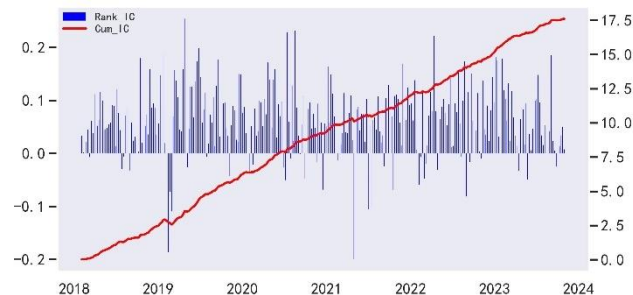
数据来源：东北证券，ricequant

图 58：日内成交量在时间上的分布因子回测结果



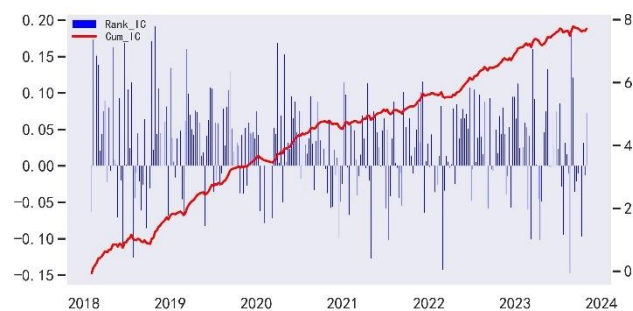
数据来源：东北证券，ricequant

图 53：日内对数成交量分布因子 Rank IC



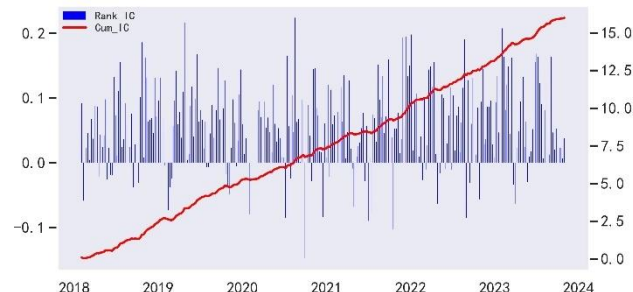
数据来源：东北证券，ricequant

图 55：日内成交量在价格上的分布因子 Rank IC



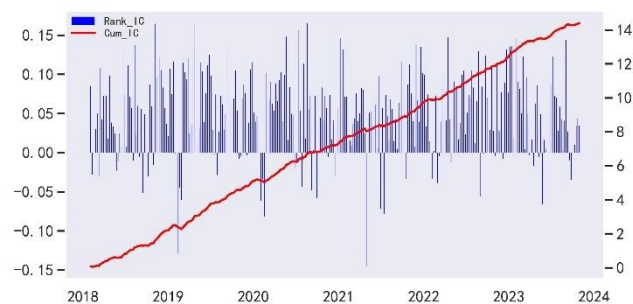
数据来源：东北证券，ricequant

图 57：日内成交量在收益率上的分布因子 Rank IC



数据来源：东北证券，ricequant

图 59：日内成交量在时间上的分布因子 Rank IC



数据来源：东北证券，ricequant

表 29: Logsig-Alpha-v 大类因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	7.91%	1.01	-17.13%	20.88%	41.62%	6.18%	6.73
2019	8.21%	1.65	45.38%	7.21%	39.53%	5.75%	6.87
2020	7.73%	1.22	32.12%	6.01%	39.69%	9.58%	4.15
2021	6.54%	1.04	40.99%	30.83%	29.93%	8.66%	3.46
2022	7.55%	1.44	6.60%	31.89%	43.85%	5.71%	7.68
2023	7.15%	1.25	13.63%	20.85%	24.90%	5.28%	4.71
汇总	7.52%	1.24	20.30%	19.67%	36.39%	7.11%	5.12

数据来源：东北证券，ricequant

表 30: 日内对数成交量分布因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	5.56%	1.24	-22.78%	12.64%	32.40%	5.93%	5.46
2019	7.66%	0.95	43.75%	6.01%	37.09%	7.86%	4.72
2020	6.38%	1.00	30.17%	4.45%	34.73%	8.28%	4.20
2021	5.49%	0.79	32.41%	22.88%	22.09%	8.99%	2.46
2022	6.49%	1.00	-1.40%	21.99%	35.47%	7.24%	4.90
2023	6.18%	1.03	12.65%	19.81%	25.08%	5.61%	4.47
汇总	6.31%	0.97	15.35%	14.75%	30.72%	7.49%	4.10

数据来源：东北证券，ricequant

表 31: 日内成交量在价格上的分布因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	3.75%	0.46	-16.77%	21.41%	30.91%	6.64%	4.65
2019	3.89%	0.75	40.53%	3.64%	20.54%	5.13%	4.00
2020	2.16%	0.36	20.04%	-3.68%	7.85%	6.75%	1.16
2021	2.29%	0.40	32.03%	22.52%	12.94%	8.24%	1.57
2022	2.62%	0.48	-4.73%	17.87%	13.58%	6.12%	2.22
2023	2.03%	0.27	9.88%	16.87%	14.14%	5.79%	2.44
汇总	2.81%	0.44	13.57%	12.97%	16.12%	6.54%	2.47

数据来源：东北证券，ricequant

表 32: 日内成交量在收益率上的分布因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	5.36%	0.95	-14.43%	24.83%	26.78%	4.08%	6.57
2019	5.82%	0.96	43.39%	5.75%	23.80%	5.27%	4.51
2020	4.71%	0.69	21.56%	-2.47%	10.22%	7.41%	1.38
2021	5.86%	0.86	38.45%	28.48%	17.84%	8.36%	2.13
2022	6.10%	0.92	2.30%	26.56%	30.00%	7.02%	4.28
2023	7.16%	1.09	21.91%	29.65%	28.71%	4.83%	5.95
汇总	5.82%	0.91	19.21%	18.58%	22.39%	6.43%	3.48

数据来源：东北证券，ricequant

表 33: 日内成交量在时间上的分布因子分年度测试结果

year	Rank IC	ICIR	多头年化收益	多头年化超额	多空年化收益	多空年化波动率	多空Sharpe Ratio
2018	4.97%	0.97	-19.78%	17.01%	28.61%	5.23%	5.47
2019	5.86%	1.02	38.46%	2.12%	26.51%	6.32%	4.20
2020	4.54%	0.82	22.23%	-1.92%	19.71%	6.74%	2.93
2021	5.01%	0.90	34.06%	24.41%	20.85%	7.16%	2.91
2022	5.55%	1.15	0.33%	24.13%	30.90%	5.74%	5.38
2023	4.98%	0.98	12.47%	19.61%	17.39%	5.03%	3.46
汇总	5.16%	0.98	14.69%	14.09%	23.70%	6.12%	3.87

数据来源：东北证券，ricequant

另外对大类因子进一步等权合成，最终合成周度因子 Rank IC 为 6.50%，五分组多头年化收益为 20.17%，多空年化收益为 29.69%，周度调仓年化换手率约 20 倍。

4.3. 不同选股域的测试结果

本节测试各大类因子在沪深 300，中证 500 以及中证 1000 中的表现。

表 34：大类合成因子在沪深 300 中的测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
Logsig-Alpha周度因子	1.41%	0.12	2.97%	6.36%	0.87	6.88%	10.27%	0.67	18.71%
日内对数成交量分布因子	2.69%	0.26	3.75%	7.16%	0.53	7.61%	10.02%	0.76	20.01%
日内成交量在价格上的分布因子	0.56%	0.06	1.52%	4.86%	0.73	3.68%	9.96%	0.37	18.97%
日内成交量在收益率上的分布因子	2.57%	0.26	0.65%	3.96%	0.62	7.38%	10.31%	0.72	14.23%
日内成交量在时间上的分布因子	1.56%	0.16	-0.27%	3.01%	0.59	2.47%	9.74%	0.25	19.23%

数据来源：东北证券，ricequant

表 35：大类合成因子在中证 500 中的测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
Logsig-Alpha周度因子	3.83%	0.43	8.68%	10.32%	0.92	15.06%	8.67%	1.74	8.24%
日内对数成交量分布因子	3.99%	0.45	6.16%	7.76%	0.55	12.45%	9.86%	1.26	18.95%
日内成交量在价格上的分布因子	0.55%	0.07	3.16%	4.72%	0.76	2.56%	8.46%	0.3	10.03%
日内成交量在收益率上的分布因子	3.61%	0.40	5.18%	6.77%	0.59	5.70%	9.03%	0.63	13.05%
日内成交量在时间上的分布因子	2.84%	0.36	3.99%	5.56%	0.61	7.80%	8.69%	0.9	15.90%

数据来源：东北证券，ricequant

表 36：大类合成因子在中证 1000 中的测试结果

因子名称	Rank IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
Logsig-Alpha周度因子	6.19%	0.89	13.29%	15.02%	1.03	31.74%	8.13%	3.9	7.90%
日内对数成交量分布因子	6.13%	0.84	11.93%	13.63%	0.59	30.79%	8.35%	3.69	10.38%
日内成交量在价格上的分布因子	2.71%	0.38	8.78%	10.44%	0.80	14.81%	7.96%	1.86	10.97%
日内成交量在收益率上的分布因子	5.40%	0.76	12.21%	13.92%	0.62	22.57%	7.64%	2.95	10.58%
日内成交量在时间上的分布因子	4.87%	0.76	8.69%	10.34%	0.65	20.64%	7.34%	2.81	7.96%

数据来源：东北证券，ricequant

从测试结果可以看出，因子普遍在中小市值的股票池中表现更好。

5. 总结

本报告一方面从日内成交量分布角度出发，分四个维度解析了成交量分布相关因子。由于成交量序列存在稀疏性和周期性特点，我们应用了日内 5min 的成交量序列，并且采用分段在日间对比的思想进行因子构造。日内对数成交量分布因子主要描述对数成交量在日内的离散程度和偏移程度；日内成交量在时间上的分布因子主要描述日内 U 形形态，以及日内开盘收盘时段的成交量异动与稳定性；日内成交量在收益率上的分布因子主要关注在日内价格显著上涨时段，上行的波动与推动价格上行的成交量占比的稳定性或是成交量的稳定性；日内成交量在价格上的分布因子主要关注价格区间成交量分布的偏移，这反映了日内公允价格区域偏低或是偏高，另外可以依据成交量在价格上的分布对日内反转因子做出改进。

另一方面，利用 log-signature 全面解析成交量的分布，对成交量序列进行特征提取，直接提取到的因子部分具有一定的选股能力，但另一部分线性性较弱。所构建的 Logsig-Alpha 因子生成器将序列简便高效地转化为选股因子，使用成交量序列构建的 Logsig-Alpha-v 周度因子和月度因子表现优异。Logsig-Alpha 首先计算序列的 log-signature 特征集，标准化后用因子正交转化模块将其转化为一批具有选股能力但两两相关性较低的基础因子，标准化后等权合成进行输出。高截断阶数的 log-signature 保留了序列中的大部分信息，使得模型具有优异的表现。

将每类细分因子进行合成，合成因子总体上表现良好且稳定，Logsig-Alpha 构造的成交量因子表现优异，整体上因子间相关性不高，另外构造的因子在 Barra 风险因子中与流动性与特异性波动率相关性略高。因子总体上更契合于中小市值股票池。

6. 参考文献

- [1] Panas, E. Generalized beta distributions for describing and analysing intraday stock market data: testing the U-shape pattern. Applied Economics 37, 191–199 (2005).
- [2] Takada, H. H. & Stern, J. M. Intraday trading volume and non-negative matrix factorization in (2016).

[3] Liao, S., Lyons, T., Yang, W. & Ni, H. Learning stochastic differential equations using RNN with log signature features (Aug. 2019).

[4] Liao, S. Log signatures in machine learning PhD thesis (Sept. 2022).

[5] Chevyrev, I. & Kormilitzin, A. A Primer on the Signature Method in Machine Learning (2016).

[6] Cohen, S. N. et al. Nowcasting with signature methods (2023).

[7] De Curtò, J., de Zarzà, I., Yan, H. & Calafate, C. T. Signature and Log-signature for the Study of Empirical Distributions Generated with GANs (2022).

[8] Kidger, P. & Lyons, T. Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU in International Conference on Learning Representations (2021).

7. 风险提示

分析基于历史数据与模型，存在模型失效风险，历史数据回测结果不代表未来表现。

8. 附录

表 37：成交量相关细分因子汇总

因子名称	Rank	IC	ICIR	多头年化收益	多头年化超额	多头换手率	多空年化收益	多空年化波动率	多空Sharpe Ratio	多空最大回撤
对数成交量日内波动因子	1.55%	0.22	9.93%	9.35%	0.44	10.66%	6.35%	1.68	7.40%	
对数成交量日内偏度因子	5.94%	0.72	16.43%	15.82%	0.55	26.29%	8.33%	3.16	7.90%	
早盘成交量占比因子	4.61%	0.54	9.25%	8.68%	0.54	13.40%	8.91%	1.5	12.53%	
早盘成交量占比稳定性因子	4.89%	1.06	13.95%	13.35%	0.73	22.62%	5.31%	4.26	5.14%	
早盘成交量波动因子	4.33%	0.86	12.87%	12.27%	0.62	19.21%	5.93%	3.24	6.32%	
早盘成交量波动稳定性因子	4.03%	0.85	15.61%	15.01%	0.67	22.74%	4.89%	4.65	5.22%	
早午开盘成交量比例因子	5.56%	0.92	15.93%	15.32%	0.60	27.38%	6.69%	4.09	6.93%	
早午开盘成交量比例稳定性因子	4.35%	0.78	17.09%	16.47%	0.59	27.45%	5.39%	5.09	4.34%	
晚午收盘成交量比例因子	2.58%	0.53	14.78%	14.18%	0.63	19.58%	5.02%	3.9	4.47%	
晚午收盘成交量比例稳定性因子	1.70%	0.36	13.99%	13.39%	0.62	13.74%	4.50%	3.05	2.95%	
价格显著上行价稳量比因子	6.30%	0.83	18.01%	17.39%	0.59	23.14%	7.85%	2.95	10.68%	
价格显著上行价稳量比稳定性因子	5.20%	0.93	20.59%	19.95%	0.63	21.89%	5.36%	4.08	5.40%	
价格显著上行价量波动因子	5.93%	0.90	18.46%	17.84%	0.60	22.50%	6.73%	3.34	8.75%	
价格显著上行价量波动稳定性因子	4.84%	0.93	20.39%	19.75%	0.63	19.62%	4.86%	4.04	5.34%	
成交量在价格上的分布偏移因子	2.93%	0.33	14.18%	13.58%	0.40	9.78%	7.51%	1.3	10.17%	
成交量在价格上的分布反转因子	5.50%	0.73	14.00%	13.40%	0.82	30.88%	9.38%	3.29	9.10%	
Logsig-Alpha-v周度因子	7.52%	1.24	20.30%	19.67%	1.04	36.39%	7.11%	5.12	4.90%	
Logsig-Alpha-v月度因子	9.12%	1.34	17.43%	16.06%	1.35	26.18%	6.84%	3.83	5.93%	

数据来源：东北证券，ricequant

研究团队简介：

王琦：帝国理工学院数学与金融荣誉硕士，南开大学统计学学士。2021 年加入东北证券上海证券研究咨询分公司任金融工程首席分析师，研究方向为金融工程。曾任职于兴业财富资产管理有限公司，任 FOF 投资经理。

贾英：伦敦大学学院金融数学荣誉硕士，厦门大学数学与应用数学本科。2022 年加入东北证券，研究方向为金融工程、因子选股。现任东北证券金融工程组研究助理。。

张栋梁：复旦大学金融硕士，南京大学金融学本科。2022 年加入东北证券，研究方向为因子选股，现任东北证券上海证券研究咨询分公司金融工程研究助理。

王国鑫：伦敦国王学院金融数学荣誉硕士，武汉大学金融学本科。2023 年加入东北证券，研究方向为量化择时、行业轮动，现任东北证券上海证券研究咨询分公司金融工程组研究人员。

江雨航：加州大学洛杉矶分校金融工程硕士，南开大学理学/经济学学士。2023 年加入东北证券，研究方向为量化固收策略，现任东北证券上海证券研究咨询分公司金融工程组研究人员。

田靖航：北京大学金融硕士，上海财经大学经济学学士。2023 年加入东北证券，研究方向为基金研究，现任东北证券上海证券研究咨询分公司金融工程组研究人员。

刘昱亨：北京大学计算机硕士，北京航空航天大学工学学士。2023 年加入东北证券，研究方向为机器学习与衍生品量化研究，现任东北证券上海证券研究咨询分公司金融工程组研究人员。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，并在中国证券业协会注册登记为证券分析师。本报告遵循合规、客观、专业、审慎的制作原则，所采用数据、资料的来源合法合规，文字阐述反映了作者的真实观点，报告结论未受任何第三方的授意或影响，特此声明。

投资评级说明

股票 投资 评级 说明	买入	未来 6 个月内，股价涨幅超越市场基准 15%以上。	投资评级中所涉及的市场基准： A 股市场以沪深 300 指数为市场基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为市场基准；香港市场以摩根士丹利中国指数为市场基准；美国市场以纳斯达克综合指数或标普 500 指数为市场基准。
	增持	未来 6 个月内，股价涨幅超越市场基准 5%至 15%之间。	
	中性	未来 6 个月内，股价涨幅介于市场基准-5%至 5%之间。	
	减持	未来 6 个月内，股价涨幅落后市场基准 5%至 15%之间。	
	卖出	未来 6 个月内，股价涨幅落后市场基准 15%以上。	
行业 投资 评级 说明	优于大势	未来 6 个月内，行业指数的收益超越市场基准。	
	同步大势	未来 6 个月内，行业指数的收益与市场基准持平。	
	落后大势	未来 6 个月内，行业指数的收益落后于市场基准。	

重要声明

本报告由东北证券股份有限公司（以下称“本公司”）制作并仅向本公司客户发布，本公司不会因任何机构或个人接收到本报告而视其为本公司的当然客户。

本公司具有中国证监会核准的证券投资咨询业务资格。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅反映本公司于发布本报告当日的判断，不保证所包含的内容和意见不发生变化。

本报告仅供参考，并不构成对所述证券买卖的出价或征价。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的证券买卖建议。本公司及其雇员不承诺投资者一定获利，不与投资者分享投资收益，在任何情况下，我公司及其雇员对任何人使用本报告及其内容所引发的任何直接或间接损失概不负责。

本公司或其关联机构可能会持有本报告中涉及到的公司所发行的证券头寸并进行交易，并在法律许可的情况下不进行披露；可能为这些公司提供或争取提供投资银行业务、财务顾问等相关服务。

本报告版权归本公司所有。未经本公司书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的，须在本公司允许的范围内使用，并注明本报告的发布人和发布日期，提示使用本报告的风险。

若本公司客户（以下称“该客户”）向第三方发送本报告，则由该客户独自为此发送行为负责。提醒通过此途径获得本报告的投资者注意，本公司不对通过此种途径获得本报告所引起的任何损失承担任何责任。

东北证券股份有限公司

网址：<http://www.nesc.cn> 电话：95360,400-600-0686 研究所公众号：dbzqyanjiusuo

地址	邮编
中国吉林省长春市生态大街 6666 号	130119
中国北京市西城区锦什坊街 28 号恒奥中心 D 座	100033
中国上海市浦东新区杨高南路 799 号	200127
中国深圳市福田区福中三路 1006 号诺德中心 34D	518038
中国广东省广州市天河区冼村街道黄埔大道西 122 号之二星辉中心 15 楼	510630

