

关联信息中的Alpha

高频研究系列八

首席分析师：郑兆磊

分析师：陈子轩

目录 CATALOGUE

01

高频研究回顾与关联信息因子研究框架

02

关联信息基础型因子

03

关联信息中的全局特征：非线性因果关系

04

关联信息中的局部特征：“交易时间”和持续期特征

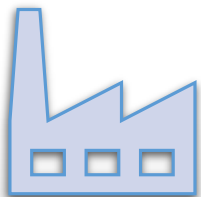
05

关联信息中的典型特征：再论“闪电崩盘”

06

总结

1.1 引言 – 高频研究系列回顾



高频因子构造方法

- 高频指标：我们将高频数据中的指标信息分为四类，分别是分布信息、时间信息、关联信息与另类信息。
- 高频因子：高频指标通过加权平均或标准差时序操作生成高频因子。



高频因子有效性分析

- 回测结果判断因子有效性：因子值截面正态分布、多空组合构建方式等；
- 基于时序相关性的特异性判断标准：时序相关性 v.s. 截面相关性



分布信息、时序信息因子（目前共计35个因子）

- 收益率分布：收益率噪音nos因子；极值理论exRtn因子；混合高斯分布模型gmm因子；
- 成交量分布：不稳定性与成交增量因子；同价成交量分布因子；
- 时序相关性：非同步交易程度因子；股价自相关性下的 VaR；“闪电崩盘”概率因子。

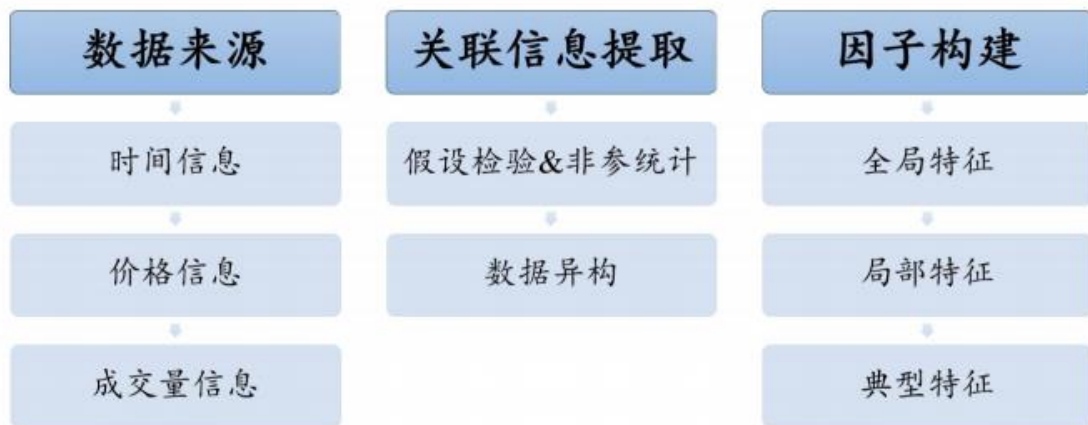
1.2 关联信息简析与因子回测说明

✓ 在高频漫谈中，我们详细阐述了提取日内数据总体分布信息的四种方法。其中，我们定义**关联信息**需要满足下述定义：

$$g = f(data_x, Reorder(data_y)) \neq f(data_x, data_y)$$

- ✓ 其中， $data_x, data_y$ 为**日内的行情数据**，如分钟级收盘价、成交量等。指标 g 对于数据 $data_x, data_y$ 的一一对应关系敏感，改变 $data_x, data_y$ 的对应关系会改变指标 g 的取值。典型的例子如量价相关性 $\rho(r, volume)$ ：改变对应收益率水平下的成交量会极大地改变量价相关性的取值。在此基础上，我们衍生出关联信息相关因子。在本文中，我们引入**时间、价格和成交量**这三类数据，由浅入深地构造因子。
- ✓ 本文引用的数据包括数据的类型（量、价）和数据的特征（时序和分布），数据复杂度较高。因此，本文做了适当的**范围限定和数据去噪**，以提升因子有效性和特异性。本文“由浅入深”地展开研究：从**关联信息的整体特征**，至**信息的局部特征**，最后聚焦于**日内某一典型关联特征**，从三个角度刻画一系列特异性高频因子。

图 1、关联信息高频因子研究方法论



资料来源：兴业证券经济与金融研究院整理

不同日内数据的一阶相关性特征

基础性因子

日内量价信息的全局特征

因果关系 —— 非线性因果性

日内量价信息的局部特征

“自然时间”至“交易时间”
交易聚集性

日内量价信息的典型特征

日内“V型”特征
新的“闪电崩盘”概率因子

目录 CATALOGUE

01

高频研究回顾与关联信息因子研究框架

02

关联信息基础型因子

03

关联信息中的全局特征：非线性因果关系

04

关联信息中的局部特征：“交易时间”和持续期特征

05

关联信息中的典型特征：再论“闪电崩盘”

06

总结

2.1 关联信息基础型因子构建

- 金融领域的时间序列通常存在着较强的自相关性。其中，**一阶自相关性是最为基础和直观的统计量之一**。在上一篇研究报告《时序信息中的Alpha》中，我们便使用了多种统计量构建了时序基础性因子。在本章中，我们再次聚焦于**一阶的相关性信息**，通过计算**相关系数**来刻画多类数据的相关性。
- 在计算维度，我们引入了多个维度的数据，包括**分钟级收盘价、最高价、最低价、成交量**等。由于个股在开盘和收盘阶段的量价信息波动较大，因此，我们均剔除了开盘和收盘10分钟数据。我们尝试下述几种构建方式，作为基础型关联信息因子。

表 2、一阶相关系数基础型因子定义

因子名	因子构造方法
rvc_cor	分钟收益率与成交量占比相关系数
rhl_cor	分钟收益率与分钟最高价最低价变化率相关系数
voh_cor	正分钟收益率样本下分钟收益率与成交量占比相关系数
vol_cor	负分钟收益率样本下分钟收益率与成交量占比相关系数

资料来源：兴业证券经济与金融研究院整理

2.2 因子表现

- 我们首先测试上述 4 个因子的表现。首先从日度 IC 测试结果上看，大多数的常见关联信息因子 IC 均值在 3% 左右，表现出较好的股价预测能力。
- 从日度组合测试结果上看，我们构建的一阶相关系数基础型因子的表现均相对优秀，一半因子的多空夏普大于 4。具体来看，从多空组合测试上看，voh_cor 因子的多空收益率在 34% 左右，多头收益率高达 14%，夏普比率在 5.2 左右。

表 3、一阶相关系数基础型因子定义 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
rvc_cor	0.027	0.062	0.44	20.58
rhl_cor	0.041	0.065	0.63	29.55
voh_cor	0.039	0.070	0.55	26.11
vol_cor	0.009	0.056	0.16	7.73

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 4、一阶相关系数基础型因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
rvc_cor	15.58%	5.93%	-8.55%	6.16%	2.53	-7.83%	18.15%
rhl_cor	31.80%	11.59%	-15.80%	6.31%	5.04	-6.86%	14.20%
voh_cor	33.79%	13.80%	-15.81%	6.53%	5.17	-5.68%	13.34%
vol_cor	4.93%	0.14%	-4.36%	5.92%	0.83	-28.27%	16.42%

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

目录 CATALOGUE

01

高频研究回顾与关联信息因子研究框架

02

关联信息基础型因子

03

关联信息中的全局特征：非线性因果关系

04

关联信息中的局部特征：“交易时间”和持续期特征

05

关联信息中的典型特征：再论“闪电崩盘”

06

总结

3.1 量价因果性的由来和因果关系的刻画

- ✓ 在前文中，我们从数据相关性的角度出发，构建了常见的关联信息因子，其重点主要聚焦在对日内信息的一阶相关性的刻画。结果表明，各类因子具有较好的选股能力。
- ✓ 上文更多地是从数据统计的角度出发。在本章中，我们将展开**特异性因子**的第一个部分，也就是通过**因果关系**刻画关联信息中全局特征：**量价因果性**。

“量”推动“价”

在弱式有效市场中，**股价**和**交易信息**往往是投资者最为关心的两类数据：

- 以成交量为代表的交易信息代表着流动性资产的供求关系；
- 股价的变动则代表着由供求关系导致的价格变动。

在上述逻辑之中，**供求关系的变动往往代表着“因”，股价变动则作为由“因”导致的“果”**：市场投资者根据外部信息决定供求关系的变化，进而推动股价变动。

“价”推动“量”

在非弱式有效的市场环境中，过去一段时间内的量价信息并不完全反映所有的过往信息。

因此，大量投资者通过市场的过往交易行为尝试分析出“内幕信息”，进而决定未来的交易行为，导致“因果倒置”的现象，如“跟风投资”等。在此情形之下，**股价的变动则变成为未来供求关系发生变化的“因”**。

- ✓ 因此，对于非弱式有效的股票市场，价格变动和交易信息往往存在着复杂的因果关系，不同股票短期内的因果关系往往也存在着较大差异。由此，**不同股票的量价因果差异能够一定程度上反映投资者对于该股的看法，成为 Alpha 信息的重要来源**。
- ✓ 在学术界和投资领域，研究者们对于股票量价的因果性有着大量的研究，其中主要分为**线性相关性**和**非线性相关性**。在实际操作中，部分研究已经表明成交量变化率和收益率不存在明显的双向线性相关性。因此，**想要以噪音程度较高的日内分钟数据测算股票自身的量价因果性时，需要考虑更为合适的非线性关系**。

3.2 非线性量价因果性：Granger 因果关系和转移熵

- 在对因果刻画的研究中，**Granger 因果关系**是最为常见的统计方式，用于分析两个时间序列之间的相互影响关系。但传统的 Granger 因果检验更接近于考察变量间的**线性关系**，从而可能导致研究结论出现明显偏差。
- 具体来说，如果时间序列 X 能够提供统计上显著的信息来预测时间序列 Y ，并且同时使用 X 和 Y 的历史数据对 Y 进行预测得到的效果要优于单独使用 Y 的历史数据进行的预测时，那么称 X 是 Y 的 Granger 原因。
- 如今，学术界将非线性因果检验的前沿方法归纳为三类：基于**回归模型**、基于**非参函数**和基于**信息理论**。其中，信息理论检验能从状态转移的角度出发刻画因果性，对于数据噪音的适应性较强，更加适合用于金融时序数据。
- 信息理论算法之一便是**转移熵 (Transfer Entropy)**，是衡量两个随机过程之间有向信息传递量的非参数统计量，刻画在给定过去状态的情况下，时间序列 X 对于时间序列 Y 未来值不确定性的减少量，其具体计算公式如下：

$$TE_k^{X \rightarrow Y} = \sum p(x_{t-k}, y_t) \log \frac{p(y_t | x_{t-k})}{p(y_t)}$$

- 其中， k 为需要主观确定的滞后项， $p(x_{t-k}, y_t)$ 是 X 在时间 $t-k$ 的状态和 Y 在时间 t 的状态的联合概率， $p(y_t | x_{t-k})$ 是给定 X 在时间 $t-k$ 的状态时 Y 在时间 t 的状态的条件概率，而 $p(y_t)$ 是 Y 在时间 t 的状态的边缘概率。

3.2 非线性量价因果性：Granger 因果关系和转移熵

转移熵的值越大，说明 X 对于 Y 的因果性更显著。与 Granger 因果关系不同，转移熵是针对**非线性因果关系的有向刻画**，可以捕捉更复杂的动态关系，**同时两组序列的异向转移熵可能不同，这允许我们确定信息流动的方向**。

考虑到转移熵的计算特点以及日内信息的高阶相关性（详见《时序信息中的Alpha》）。因此，我们类比于之前的 LBQ 因子刻画方式，采用**多期滞后**项的方式计算量价层面的转移熵，进而刻画出量价的**全局非线性因果性**。具体计算方式如下：

确定状态序列

- 使用五分钟滚动收益率序列和五分钟滚动成交量之和作为收益率和成交量序列。
- 由于转移熵本质是条件概率，我们**将成交量与收益率序列修改为状态序列**：以 30% 和 70% 分位点作为切分线，将该股日内收益率/成交量大于 70% 的样本点定义为 2，小于 30% 定义为 0，其余为 1。例如：

$$rtnStatus = [1, 2, 0, 1, \dots, 0]$$

计算转移熵序列

- 由于转移熵需要设置滞后长度和方向，我们分别计算**收益率→成交量**和**成交量→收益率**两类转移熵，并设置多期滞后长度：即对于收益率→成交量而言，计算滞后项为 1-5 共五期转移熵。
- 最终，单日单只股票可以得到两组转移熵序列： $TE_{[1-5]}^{rtn \rightarrow vol}$ 和 $TE_{[1-5]}^{vol \rightarrow rtn}$ 。以 $TE_{[1-5]}^{rtn \rightarrow vol}$ 为例，可以得到：

$$TE_{[1-5]}^{rtn \rightarrow vol} = \left[\sum p(rtnStatus_{t-k}, volStatus_t) \log \frac{p(volStatus_t | rtnStatus_{t-k})}{p(volStatus_t)} \right]_{k=1-5}$$

3.3 因子表现

- 从日度 IC 测试结果上看，量价非线性因果因子表现较好，其中**收益率→成交量方向**刻画的因子更为出色，Rank IC 均值为 0.045，ICIR 大于 0.5，表现出较好的股价预测能力。
- 从日度组合测试结果上看，量价非线性因果因子的表现十分优秀，多空夏普均较高，且无明显回撤，多头收益率较高，收益率→成交量的因子表现更优。具体来看，从多空组合测试上看，te_r2v 的多空收益率在 47% 左右，多头收益率约为 17%，夏普比率在 5 左右。
- 从 Rank IC 时序特征、多空净值曲线以及分位数组合测试结果上看，te_r2v 因子的多空净值长期呈现上升趋势，且无明显回撤，表现十分稳定。

表 5、量价非线性因果因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
te v2r	0.032	0.084	0.38	17.96
te r2v	0.045	0.080	0.56	26.61

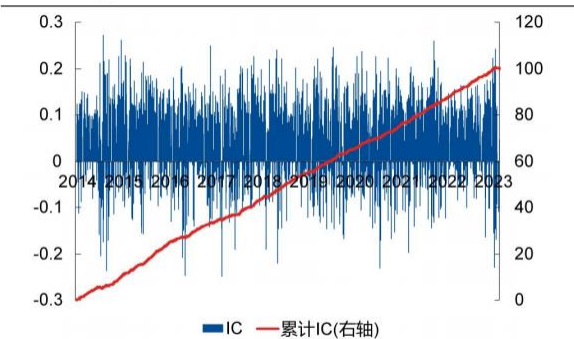
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 6、量价非线性因果因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
te_v2r	37.42%	14.53%	-17.74%	8.43%	4.44	-6.96%	8.91%
te_r2v	47.28%	17.48%	-21.47%	8.62%	5.49	-6.72%	12.35%

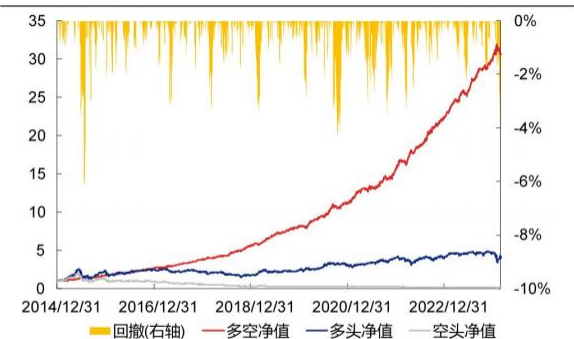
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 2、te_r2v 因子 Rank IC 和累计 IC



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 3、te_r2v 因子多空组合净值



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

3.4 因子相关性检验和正交化处理

- 由于两者在逻辑和相关性上接近，我们最终将两者等权合成得到 te_Comb 因子，并将其与时序相关性最高的已有高频因子进行正交化处理，得到 te_CombN 因子。经过正交化后该因子与底层所有高频因子的时序相关性均值约为 27%，截面相关性均值约为 17%，表现出较高的特异性。
- 从日度 IC 测试结果上看，te_Comb 因子表现较好，正交化前 Rank IC 均值为 0.046，正交化后均值为 0.025，ICIR 分别约为 0.5 和 0.4，表现出较好的股价预测能力。
- 从日度组合测试结果上看，te_Comb 因子的表现十分优秀，多空夏普均较高，且无明显回撤，多头收益率较高，多空收益率在 49% 左右，多头收益率约 18%，夏普比率在 5 左右。经过正交化之后，表现并未出现明显下降。

表 7、te_Comb 因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
te_Comb	0.046	0.093	0.49	23.22
te_CombN	0.025	0.063	0.40	18.97

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 8、te_Comb 因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
te_Comb	48.72%	17.84%	-22.08%	9.43%	5.17	-6.39%	9.14%
te_CombN	28.20%	12.89%	-12.57%	6.65%	4.24	-8.84%	13.21%

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

目录 CATALOGUE

01 高频研究回顾与关联信息因子研究框架

02 关联信息基础型因子

03 关联信息中的全局特征：非线性因果关系

04 关联信息中的局部特征：“交易时间”和持续期特征

05 关联信息中的典型特征：再论“闪电崩盘”

06 总结

- ✓ 在前序章节中，我们从**关联信息的全局角度**出发，构建了量价**非线性因果关系**因子，该因子也侧面论证了日内量价的关联信息存在着丰富的 Alpha 信息。在本章中，我们将进一步探究**关联信息中的局部特征**，从“自然时间”到“交易时间”，再进一步衍生出最终的**个股的持续期特征**。

“自然时间”与“交易时间”

此前，大多数研究在讨论收益率特征时往往将“自然时间”默认为收益率的时序标签，如每个交易日收盘价的变动幅度等。

在 1973 年，Clark 首次在价格变动的随机过程中引入从属过程，**将价格变动从“自然时间”变化至交易活动强度加权下的“交易时间”，进而更好地表示波动率聚集现象以及收益率的“尖峰厚尾”特征。**

无论是日内还是日间，收益率的分布特征往往存在着明显的**“尖峰厚尾”**的现象。

近年来，随着日内交易数据的逐步完善，学术界针对日内收益率的分布特征提出了同样类似的观点：**“尖峰厚尾”是由于流动性资产在日内不同交易时间内的交易活跃度不同导致。**在分钟或更高频的观测维度之下，每一个等量的时间间隔对于价格的影响作用并不相同。**这本质上是“自然时间”和“交易程度”的不对等特征导致的。**

信息的透明程度与流动性

- 参考 Clark 等相关研究的思路，在本节中我们将以**成交量**作为交易活跃度的刻画指标，尝试将“**自然时间**”调整为“**交易时间**”，致力于在相对统一的“交易时间”内衡量“量”和“价”的变动特征，进而挖掘出日内 Alpha 信息。
- 具体来说，我们首先根据分钟级成交量和收盘价数据，刻画出统一的交易活跃度加权下的收盘价，即“交易时间”间隔下的收盘价数据。具体流程如下：

设置“交易时间”阈值

类比于“自然时间”中 60 秒等于 1 分钟的设置，我们首先需要设定一个“**交易时间间隔**”的**阈值**。

我们计算一只股票 T 日内分钟成交量总和，再进一步除以 50 得到成交量阈值 $tradingVol$ ：

$$tradingVol = \sum_{t=1}^{240} volume_t / 50$$

其中， $volume_k$ 是第 t 个自然时间下的成交量。

计算“交易时间”价格

基于“交易时间间隔”计算活跃度加权下的收盘价。

从第一分钟开始，我们**逐步累加成交量**，若成交量累加达到“交易时间间隔”时，说明当前“交易时间”达到“一个时刻”。此时，我们将囊括在内的所有分钟收盘价按照分钟成交量占比加和，得到当前“**交易时间**”的**收盘价**：

$$tradingClose = \left[\frac{\sum (volume_t \cdot close_t \cdot 1\{t \in i\})}{tradingVol} \right]_{i=1,2,\dots,50}$$

其中， $tradingClose_i$ 为第 i 个“交易时间”下的收盘价， $volume_t$, $close_t$ 分别是第 t 个自然时间下的成交量和收盘价， $1\{k \in i\}$ 表示当前自然时间 k 是否属于第 i 个交易时间。

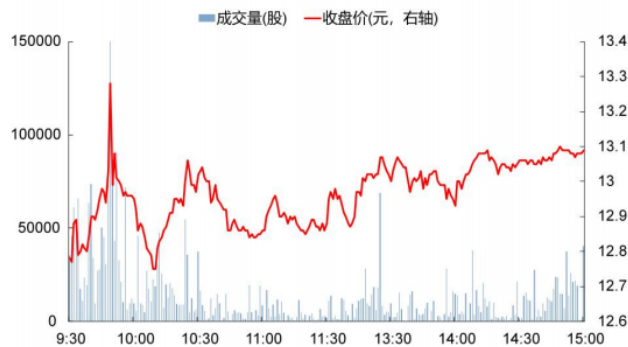
除收盘价，我们进一步统计**每个“交易时间”的分钟间隔数**，记为 $tradingTime$ ：

$$tradingTime = \left[\sum 1\{t \in i\} \right]_{i=1,2,\dots,50}$$

4.2 “交易时间”的刻画方式

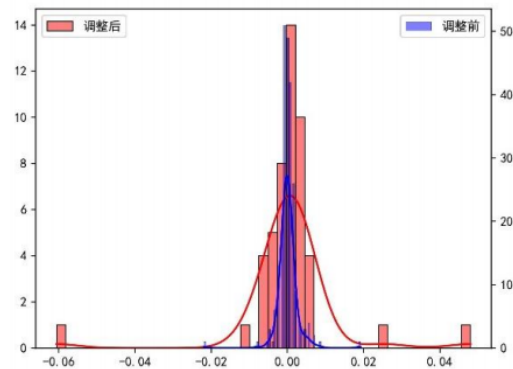
- 至此，我们根据每支股票每天的交易情况，将日内“自然时间”对应的数据刻画为“交易时间”，共计 50 个。其中，我们总共得到两组数据，一组代表**每个交易时间的股价**，另一组相当于**“交易时间”与“自然时间”的转换关系**。
- 转换为“交易时间”的目的之一是以交易活跃度作为衡量方式时，收益率的“尖峰厚尾”特征会有所改变，分布更接近正态分布。
- 以 2024 年 3 月 29 日某一只股票 (603303) 作为样例，展示转换前后的股价特征和收益率分布。下图左侧展示了该股当日的收盘价和成交量，可以明显看出：该股在 9 点 45 至 10 点左右有一次较大幅度的波动；下午该股整体走势相对平缓。从原始收益率分布特征上看 (右侧图蓝色部分)，该股呈现出相对明显的“尖峰厚尾”特征，其分钟收益率的峰度值为 26.78，偏度值为 -1.26，“尖峰”特征十分明显。之后我们通过上述算法将“自然时间”切换为“交易时间”后，其收益率分布的正态性得到了十分明显的改善，偏度值为 -0.53，峰度值为 16.94。从频率分布直方图和核密度图上看，调整后的分布 (右侧图中红色部分) 的正态性相对于蓝色也得到了明显的改善。这一对比说明：**经过“交易时间”转换后的确部分改善了收益率分布的正态性。**

图 4、2024 年 3 月 29 日 603303 分钟收盘价和成交量



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 5、2024 年 3 月 29 日 603303 收益率分布直方图-调整前后对比



注：调整前代表原始分钟收益率分布特征；调整后代表“交易时间”下的收益率分布特征；

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

4.3 “交易时间”下的基础性因子表现与分析

- 经过上述的计算，我们已经可以基于每支股票当日的交易特征，刻画出“交易时间”下的股价走势以及每个“交易时刻”的时间间隔。
- 在本节中，我们先尝试基于上述刻画出的两组数据构建因子。本质上，新的股价时序的数据特征和原始股价特征一致。我们首先统计“交易时间”下的收益率的均值和方差，作为“交易时间”下的**收益率均值因子**和**已实现方差因子**，记为 cutVol_rtnMean 和 cutVol_rtnVar。此外，我们进一步统计“交易时刻”的时间间隔序列的均值和方差，作为“交易时间”下的**时间特征均值因子**和**方差因子**记为 cutVol_timeMean 和 cutVol_timeVar。
- 从日度 IC 测试结果上看，除单个因子之外，“交易时间”下的基础性因子表现较好，其中 cutVol_rtnMean 和 cutVol_rtnVar 因子更为出色：以 cutVol_rtnVar 为例，该因子的 Rank IC 均值为 0.047，ICIR 为 0.59，表现出较好的股价预测能力。
- 从日度组合测试结果上看，“交易时间”下的基础性因子的表现十分优秀，多空夏普均较高，且无明显回撤，多头收益率较高，收益率→成交量的因子表现更优。具体来看，从多空组合测试上看，cutVol_rtnMean 的多空收益率在 43% 左右，多头收益率约为 16%，夏普比率在 5 左右。

表 9、“交易时间”基础性因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
cutVol_timeMean	0.008	0.051	0.15	7.18
cutVol_timeVar	0.026	0.072	0.37	17.27
cutVol_rtnMean	0.046	0.089	0.52	24.54
cutVol_rtnVar	0.047	0.080	0.59	27.84

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 10、“交易时间”基础性因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
cutVol_timeMean	8.68%	3.97%	-4.32%	4.81%	1.81	-19.93%	11.42%
cutVol_timeVar	27.76%	12.41%	-11.85%	6.47%	4.29	-5.74%	8.13%
cutVol_rtnMean	42.79%	16.17%	-18.88%	8.40%	5.10	-7.56%	15.36%
cutVol_rtnVar	35.24%	14.68%	-16.18%	8.08%	4.36	-6.00%	7.19%

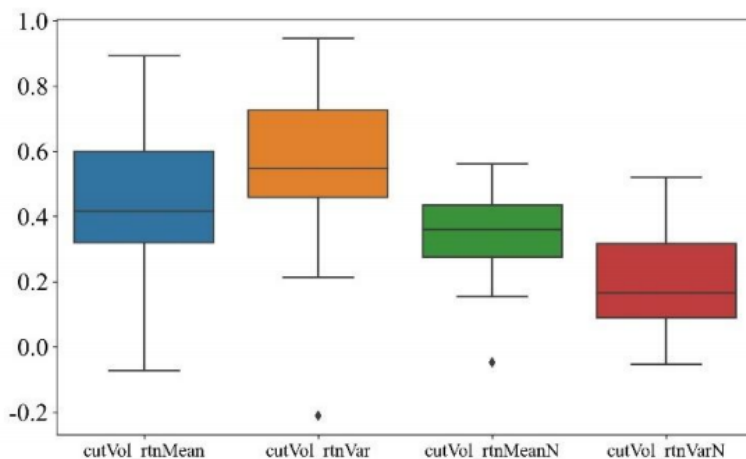
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

4.3

“交易时间”下的基础性因子表现与分析

- 与之前的收益率均值因子和已实现方差因子类似，这类因子都是通过多个角度刻画收益率分布特征偏离正态分布的距离，也就是**通过对异常交易行为的刻画来提取 Alpha 信息**。
- 但“交易时间”的刻画方面，我们加入交易活跃度作为从属特征，从一定程度上避免了诸如流动性、集中交易导致的收益率偏移。换个角度而言，如果在“交易时间”下收益率仍然明显偏移正态分布，其异常交易行为的特征将更加明显，由此得到的 Alpha 信息也应当更加纯粹。
- 因此，理论上 **cutVol_rtnMean** 和 **cutVol_rtnVar** 应当能够包含不同于普通日内收益率均值和方差因子的信息。
- 从时序相关性的检验上看，这两个因子对应的相关性最高的因子正是均值和方差因子。我们将其分别进行正交化之后，两个因子与底层所有高频因子的相关性均值分别约是 30% 和 20%。

图 6、“交易时间”部分基础性因子正交化前后时序相关性



注：左侧两个箱形图为正交化前因子时序相关性，右侧为正交化后；

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

4.3 “交易时间”下的基础性因子表现与分析

- 我们进一步展示两个因子正交化之后的回测结果，具体如下。
- 从日度 IC 测试结果上看，正交化之后“交易时间”下的基础性因子表现同样优秀，两个因子的 Rank IC 均值在 0.023 左右，ICIR 仍然在 0.50 左右，表现出较好的股价预测能力以及特异性。
- 从日度组合测试结果上看，“交易时间”下的基础性因子表现十分优秀，多空夏普均较高，且无明显回撤，多头收益率较高。具体来看，从多空组合测试上看，cutVol_rtnVarN 的多空收益率在 22% 左右，多头收益率约为 13%，夏普比率在 5.5 左右，表现出较好的稳定性和特异性。
- 从多空净值曲线以及分位数组合测试结果上看，两个因子的多空净值长期呈现上升趋势，且最近几年无明显回撤，表现十分稳定。
- 综上，“交易时间”的转化本质上是在统计个股日内异常特征的同时，通过标准化交易量的方式避免诸如流动性、集中交易导致的收益率异常值，进而提取相对更加纯粹的日内信息。从上述结果上看，即便是同样的统计量，“交易时间”和“自然时间”下构造的因子也存在明显的差异性，能够以此构造出更加丰富的高频因子。

表 11、“交易时间”部分基础性因子正交化后 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
cutVol_rtnMeanN	0.024	0.058	0.42	19.68
cutVol_rtnVarN	0.023	0.040	0.56	26.58

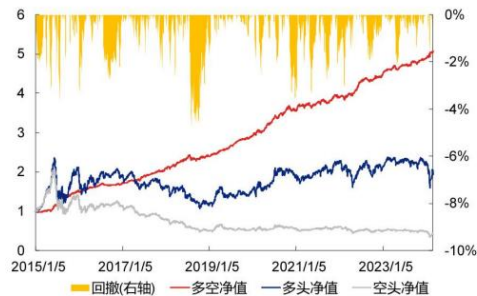
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 12、“交易时间”部分基础性因子正交化后组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
cutVol_rtnMeanN	18.97%	9.18%	-8.43%	5.83%	3.25	-11.91%	25.32%
cutVol_rtnVarN	21.64%	13.27%	-6.90%	3.92%	5.52	-4.07%	11.84%

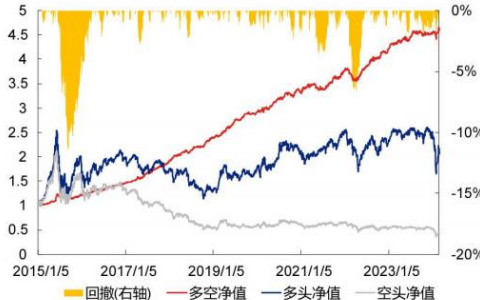
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 7、cutVol_rtnVarN 因子多空组合净值



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

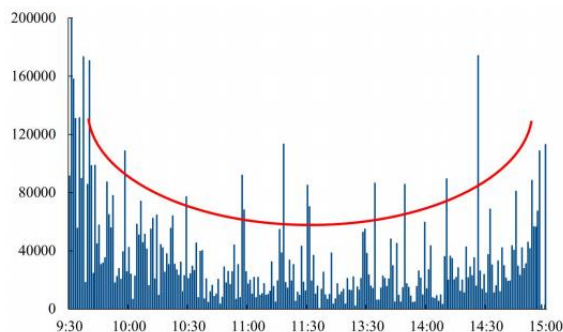
图 8、cutVol_rtnMeanN 因子多空组合净值



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

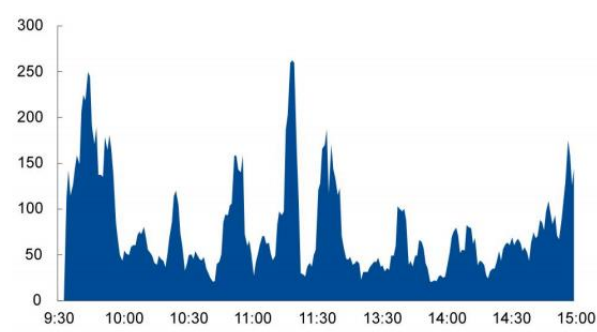
- ✓ 上文中，我们基于成交量将“自然时间”转换为“交易时间”，并集中于转换之后价格信息，刻画了分布特征因子，表现出较好的选股能力和特异性。上文的高频因子相对于偏向基础的统计。在本章中，我们尝试进一步深耕“交易时间”带来的增量信息——分钟统计下的条件持续期模型。
- ✓ 在现有的日内 Level-1 数据研究中，高频数据均是在细小的时间间隔上抽取的观测值，并将原始数据过滤产生新的固定频率的数据。这种处理方法能够高效地清理出体量更小、更适合直接应用的高质量高频数据，但同样也降低了数据频率，并丧失交易数据的某些特性，其中便包括了交易发生的间隔时间。
- ✓ 在上文中，我们通过自行设计的算法刻画出了“交易时间”，每一个“交易时刻”时序数据代表着一个单位交易量下的时间间隔和股价变动。那么，我们可以将“交易时间”视为一个简单的点过程：考虑 t_i 是第 i 个“交易时间”到达时对应的“自然时间”，从第 $i-1$ 到第 i 个“交易时刻”的时间间隔为 $\Delta t_i = t_i - t_{i-1}$ 。对于上述刻画出的 50 个日内“交易时刻”而言，我们同样可以得到 50 个间隔时间，记为交易到达的间隔时间，每一个间隔时间代表着下一笔同等交易水平的到达时间。
- ✓ 与股价的时序模型类似，交易到达的间隔时间也存在着部分异常特征，其中最为著名的便是交易的聚集性。交易聚集性类似于波动率聚集现象：较短的交易间隔时间后面也往往跟随着短的交易间隔时间，长的交易间隔时间后面也往往跟随着长的交易间隔时间。举例来说，A 股市场的成交量分布呈现出十分明显的“U 型”分布特征：即开盘和收盘阶段的交易强度较大，如下方左图。这便是交易聚集性的典型体现。此外，股票市场存在着大量的“跟风交易者”，这导致部分时间点成交量出现激增。如下方右图中，我们计算该股日内五分钟滚动成交量之和。可以明显看出，该股日内的成交量呈现出十分明显的聚集现象：部分时间段成交量集中增加。

图 9、某支股票日内分钟成交量时序上呈 U 型分布
(单位：股)



资料来源：上交所、深交所行情数据，兴业证券经济与金融研究院整理

图 10、某支股票日内五分钟滚动成交量和 (单位：万股)



资料来源：上交所、深交所行情数据，兴业证券经济与金融研究院整理

- ✓ 针对上述特征，学术界引入**持续期模型**，在考虑交易聚集性的情况下构建交易间隔时间的时序模型，如 **ACD (Autoregressive Conditional Duration)** 模型。ACD 模型采用极大似然估计法估计模型中的参数，进而根据样本数据刻画中交易间隔的聚集性现象的强度。然而，此类模型的计算复杂度相对较高，同时并未将股价涨跌特征纳入考虑。**本节中，我们从两个角度出发，参考 ACD 模型的假设刻画出日内股票的聚集现象。**

检验方式1：时序相关系数

对于一个时间序列，时序相关系数是检验聚集性最为有效且简单的方式之一。我们基于“交易时间”的时间间隔数据(共计 50个)，统计间隔数据的一阶自相关性，记为 cutVol_timeCor 因子。自相关性越大，说明时序上时间间隔存在着相对明显地聚集性。

检验方式2：分桶熵

在上述统计方式中，我们仅考虑了交易时间自身，并未引入股价涨跌特征。在本节中，我们在引入股价涨跌特征的同时，参考之前的方法，从整体性的角度出发刻画交易时间的分化特征，进而反向刻画出交易聚集性的强度。我们的计算方式如下：

1. 刻画考虑涨跌的交易时间间隔：基于上述算法得到的“交易时间”数据，我们将“交易时间”的分钟间隔数差乘以这一时刻的收益率正负号，得到一组带有正负的分钟间隔数：

$$tradingTimeSign = [tradingTime_i \cdot sign tradingRtn_i]_{i=2,3,\dots,50}$$

其中 $tradingRtn_i$ 为第 i 个交易时间的收益率，由此我们可以得到 49 个带正负号的时间间隔数据。

2. 计算分桶熵：将带有正负的时间间隔按照最小到最大进行值得五等分，并计算分桶熵，得到 cutVol_entropy 原始因子值。

$$cutVol_entropy = - \sum_{k=1}^{10} p_k \ln(p_k) \cdot 1_{(p_k > 0)}$$

分桶熵的思路本质上是**刻画一组数据的分化程度与随机性**：如果同号同大小的时间间隔数据大量出现，如大量出现上涨且间隔绝对值均为 3 的数据，说明该股出现了交易聚集现象，整体交易规则呈现出极强的一致性，反之我们可以认为交易的时间间隔相对随机。

- ✓ 综上，上述两个因子均为正向逻辑的因子，即因子值越大，股票的交易聚集现象更加强烈。**最终我们计算过去 20 个交易日的均值，作为最终的因子值进行测试。**

4.4 “交易时间”与交易聚集性

- 从日度 IC 测试结果上看，交易聚集性相关因子表现较好，其中 cutVol_timeCor 因子更为出色，Rank IC 均值为 0.045，ICIR 约为 0.5，表现出较好的股价预测能力。
- 从日度组合测试结果上看，交易聚集性相关因子的表现也十分优秀，多空夏普均较高，且无明显回撤，多头收益率较高。从多空组合的夏普比率上看，cutVol_entropy 的表现更为优秀，其多空收益率在 24% 左右，多头收益率约为 11%，夏普比率在 4 左右。
- 从 Rank IC 时序特征、多空净值曲线以及分位数组合测试结果上看，te_Comb 因子的多空净值长期呈现上升趋势，且最近几年无明显回撤，表现十分稳定。

表 13、交易聚集性相关因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
cutVol_timeCor	0.045	0.099	0.45	21.20
cutVol_entropy	0.020	0.066	0.31	14.41

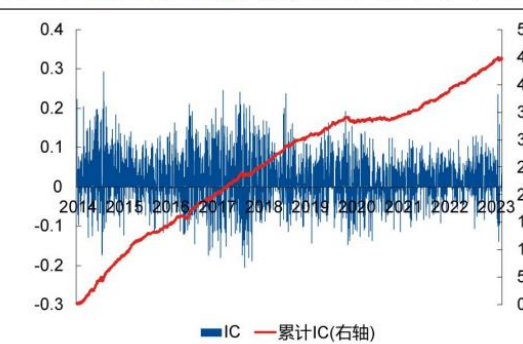
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 14、交易聚集性相关因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
cutVol_timeCor	33.81%	16.12%	-14.79%	8.62%	3.92	-7.48%	8.71%
cutVol_entropy	24.34%	10.95%	-10.41%	5.86%	4.16	-5.78%	12.19%

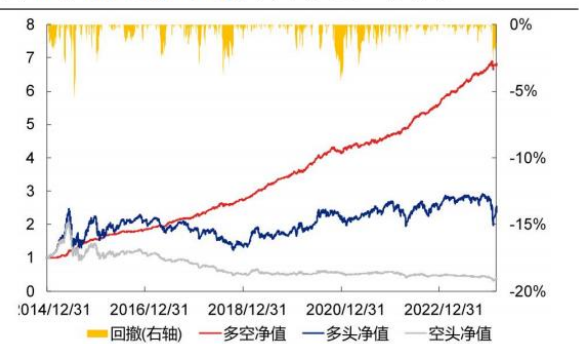
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 11、cutVol_entropy 因子 Rank IC 和累计 IC



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 12、cutVol_entropy 因子多空组合净值



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

4.5 因子相关性检验和正交化处理

- 我们进一步进行时序和截面相关性检验。其中，cutVol_entropy 表现出极强的特异性，其与底层所有高频因子的时序相关性低于 35%，均值仅 16%。对于 cutVol_timeCor，我们将其与时序相关性最高的已有高频因子进行正交化处理，得到 cutVol_timeCorN 因子。经过正交化之后，该因子与底层所有高频因子的时序相关性均值约为 35%。
- 从日度 IC 测试结果上看，cutVol_timeCorN 因子表现较好，正交化后均值为 0.025，ICIR 接分别约为 0.5 和 0.4，表现出较好的股价预测能力。
- 从日度组合测试结果上看，cutVol_timeCorN 因子的表现同样优秀，多空夏普均较高，且无明显回撤，多头收益率较高，多空收益率在 16% 左右，多头收益率约为 8%，夏普比率在 3 左右。

表 15、cutVol_timeCorN 因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
cutVol_timeCorN	0.022	0.056	0.39	23.22

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 16、cutVol_timeCorN 因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
cutVol_timeCorN	16.03%	8.10%	-7.36%	5.41%	2.96	-6.72%	13.56%

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

目录 CATALOGUE

01

高频研究回顾与关联信息因子研究框架

02

关联信息基础型因子

03

关联信息中的全局特征：非线性因果关系

04

关联信息中的局部特征：“交易时间”和持续期特征

05

关联信息中的典型特征：再论“闪电崩盘”

06

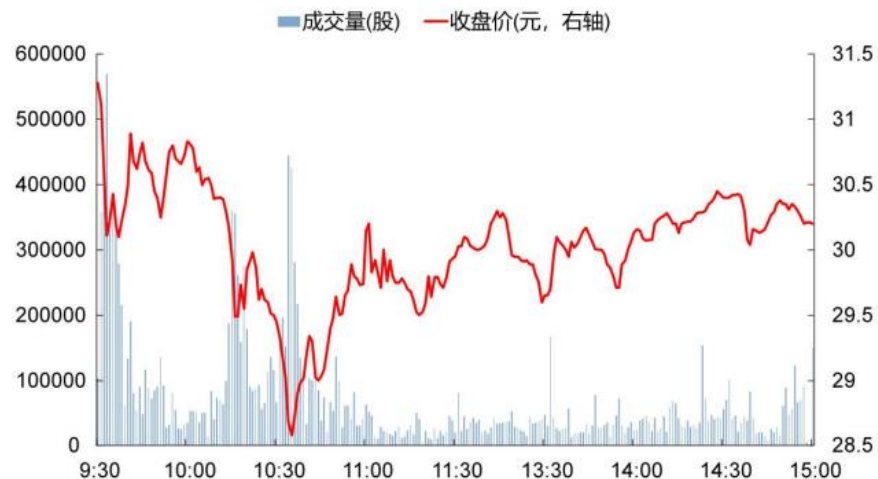
总结

5.1

“闪电崩盘”因子回顾与反思

- ✓ 在前序章节中，我们从**关联信息的全局角度和局部特征**出发，构建了**量价非线性因果关系**和**“交易时间”**相关因子。
- ✓ 在本章中，我们将角度进一步缩小，从某一具体的量价特征出发，构建带入量价信息的“V型”走势特征与最终的改进后“闪电崩盘”因子。
- ✓ 在往期报告《高频研究系列六——时序信息中的 Alpha》中，我们首次提出**“闪电崩盘”**这一概念。
- ✓ 具体来说，我们将“闪电崩盘”通指为流动性资产的价格在日内短时间出现大幅超跌，同时后续允许价格出现恢复，即在日内出现深“V”的价格走势，如下图所示：

图 13、300659.SZ 日内价格走势



资料来源：上交所、深交所行情数据，兴业证券经济与金融研究院整理

注：数据日期为 2023 年 3 月 31 日

5.1 “闪电崩盘”因子回顾与反思

- 在前序的研究中，我们以连续下跌次数作为出发点，通过刻画日内分钟级收益率连续下跌次数序列引出“快速下跌”这一特征，并进一步引入泊松分布，以表示未来一段时间内，能够出现“快速下跌”的概率，进而构建“闪电崩盘”flashCrashProb 因子。无论是在截面选股层面还是时序层面，该因子均表现出了较好的预测能力，并能够基于概率表达动态剔除预期可能会下跌的股票。具体算法可参考发布于2023年6月29日的报告《高频研究系列六—时序信息中的Alpha》。
- 截至2024年3月底，该因子已在样本外跟踪近1年。从回测全时段上看，该因子 Rank IC 均值为 0.014，多空夏普比率为 2.68。
- 从 Rank IC 时序特征、多空净值曲线以及分位数组合测试结果上看，flashCrashProb 因子的多空净值长期呈现上升趋势，且最近几年无明显回撤，尤其是 2024 年初并无明显回撤，表现十分稳定。

表 17、flashCrashProb 因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
flashCrash	0.014	0.052	0.27	12.93

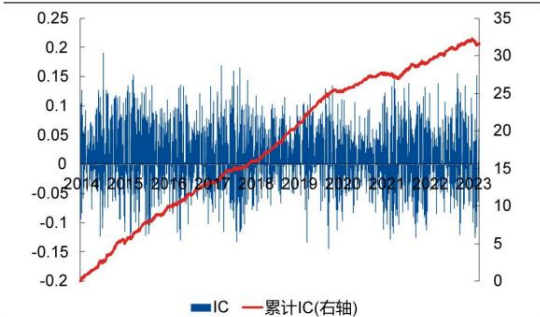
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 18、flashCrashProb 因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
flashCrashProb	21.61%	8.74%	-10.26%	4.62%	2.68	-4.84%	10.99%

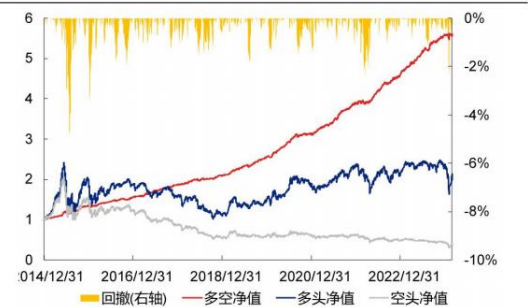
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 14、flashCrashProb 因子 Rank IC 和累计 IC



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 15、flashCrashProb 因子多空组合净值

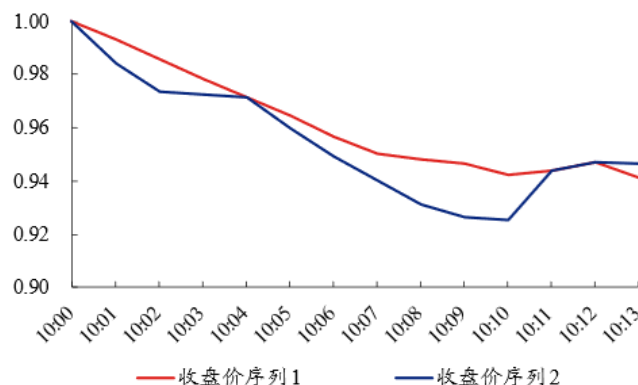


资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

5.1 “闪电崩盘”因子回顾与反思

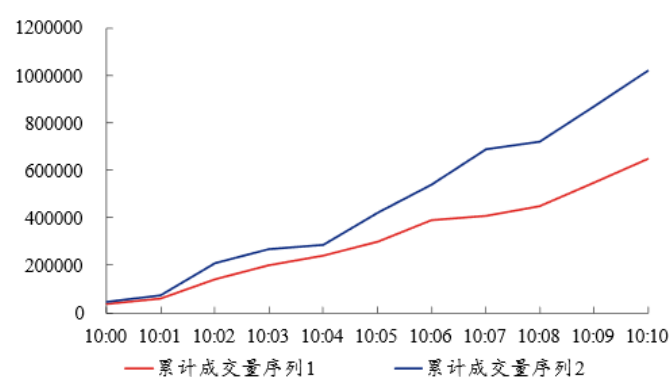
- 在往期的研究中，我们在刻画“快速下跌”特征时仅单一地考虑了下跌的分钟累加数，也就是时间间隔的概念。从本质上说，日内股价的下跌特征能够从多个方面进行刻画，包括下跌幅度、下跌中放量情况等。此外，将一次下跌和一次上涨合并刻画也更加符合“V型”特征。
- 如下图样例所示，左侧两个收益率序列在下跌次数计算中均为连续10次下跌(10点至10点10分)，但从幅度上看，蓝线的下跌幅度更大。在后续走势中，两者呈现出不同的走势：蓝线在下跌后出现一定幅度的上涨，红线则相对震荡，最终价格相对接近。在现有的统计中，两者在连续下跌和连续上涨次数统计中结果相对一致，但显然蓝线走势更加符合“V型”下跌的特征。此外，除收盘价序列之外，蓝线在下跌过程的放量特征更加明显(下方右图)。**因此，现有的全局次数统计法难以精准地刻画出类似于“V型走势”这样的典型特征。**
- 综上，尽管该因子在样本外仍保持了一定的预测能力，但其刻画的事件相对宽泛，且对于“V型”这一典型特征的刻画精准度相对一般，且并未考虑除下跌次数以外的额外信息。因此，我们将在此前的研究基础之上，对该因子进行改进。具体来说，**我们将改进分为两个步骤：第一步我们将基于量、价和时间三维度的信息，刻画股票的“V型特征”；第二步我们将基于“V型特征”，刻画新的“闪电崩盘”概率因子，并衡量其有效性。**

图 16、日内某两只股票分钟收盘价对应净值



资料来源：上交所、深交所行情数据，兴业证券经济与金融研究院整理

图 17、日内某两只股票分钟累计成交量（单位：股）



资料来源：上交所、深交所行情数据，兴业证券经济与金融研究院整理

5.2 改进步骤 1: “V 型特征” 的刻画

- 前文提及，日内股价的“V 型特征”可能包含三维度的信息：**股价走势的幅度、持续时间和成交量特征**。
- 沿用此前的想法，我们在此前的基础之上，我们加入幅度、成交量和间隔时间等多个维度的信息，刻画股票日内“V 型特征”，具体算法如下。

1、计算分钟量价幅度序列

- 在每一分钟，首先计算**分钟级对数收益率**和**分钟级成交量占比**数据，相乘得到分钟量价幅度数据序列 rv :

$$rv = \left[rtn_t \cdot \frac{volume_t}{\sum_t volume_t} \cdot 10^4 \right]_{t=1,2,\dots,240}$$

- 其中， rtn_t 为分钟收益率， $volume_t$ 为分钟成交量， rv 代表着考虑了量价共同作用的股票走势，记为量价幅度。**更大的成交量、更大的振幅将会得到更大的量价幅度。**

2、计算累计分钟量价幅度序列

- ✓ 计算量价幅度序列同号累加值：若一个量价幅度为负数，则开始统计连续量价幅度为负数之和，直至下一个量价幅度非负，该个数则为一个累计量价幅度样本点；

$$rvCum_i = \sum_{t=k}^K rv_t \cdot 1\{condition\}$$

- ✓ 其中， $condition$ 表明前序累加的量价幅度同号。最终我们由此得到多个连续下跌量价幅度样本点，该数据应当为正负号相接的一组序列。

3、刻画“V型”特征

- ✓ 在“V型”特征中，股价应该为大幅度下跌，小幅度上涨，且下跌幅度大于上涨幅度的一段走势。根据之前得到的累计量价幅度，我们按要求计算 $rvCum_{\{i-1\}}$ 为负数、 $rvCum_i$ 为正数的两者之和，并筛选和为负数的数据，取绝对值作为最终的“V 型”走势特征 rvV 。

$$rvV = [|rvCum_{i-1} + rvCum_i| \cdot 1\{condition\}]$$

$$condition = rvCum_{i-1} < 0,$$

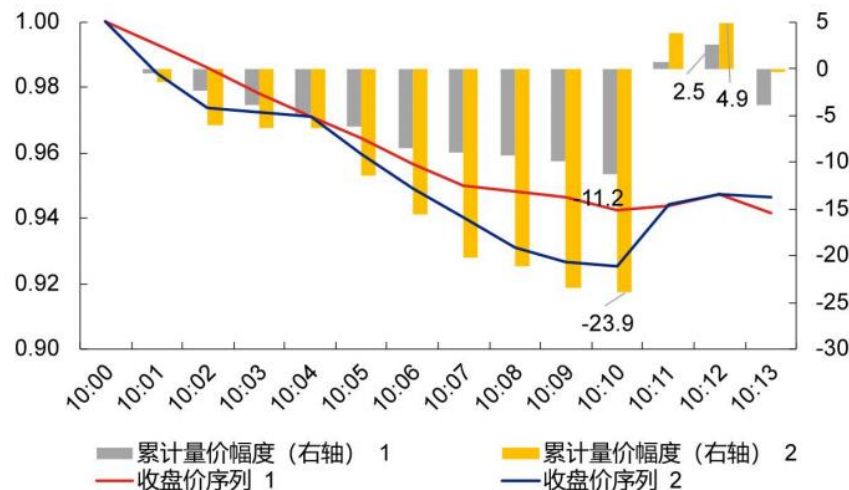
$$rvCum_i > 0, rvCum_{i-1} + rvCum_i < 0$$

- ✓ 还可以得到每个“V 型”走势的持续分钟数，记为 $timeV$ 。

5.2 改进步骤 1：“V 型特征”的刻画

- 综上，在每日收盘后，根据股票自身的日内走势，**将收盘价、成交量和时序特征三者进行综合考虑**，并在最终刻画时同时考虑下跌和上涨的连续性，最终得到每只股票日内的“V 型”走势特征。
- 我们以上文中提及的两个样例进行展示计算。在目前的计算中，首先根据价格振幅和成交量占比计算出量价幅度，再累加得到 10 点 10 分的累计量价幅度，分别约为 -11.2 和 23.9。在考虑了下跌幅度和交易情况之和，明显第 2 个序列的下跌特征更加明显，对应幅度累加值也同样更大。此后两者均上涨两分钟，得到累计幅度为 2.5 和 4.9。之后两股均下跌，由此我们得到一个完整的“V 型”走势。按照上述算法，我们得到两只股票的“V 型”走势特征分别是 8.68 和 19.05，两者的“V 型”走势持续分钟数均为 12。
- 显然，在目前的算法中，我们对于“V 型”走势特征的刻画能够清晰地区分出下述两种股票的走势差异，进而找出“V 型”走势特征更加明显的 2 号股票。

图 18、日内某两只股票分钟收盘价对应净值与累计量价幅度



资料来源：上交所、深交所行情数据，兴业证券经济与金融研究院整理

5.3 改进步骤 2：“V 型”特征基础因子构造与表现展示

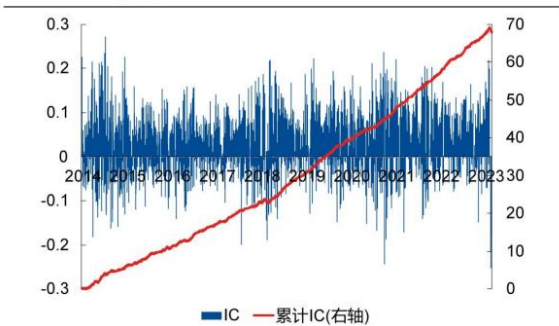
- 在本节中，我们将首先基于个股的“V 型”走势幅度和时间数构建“V 型”特征基础因子。
- 根据每只股票的“V 型”走势特征序列，计算均值和最大值，记为 negV_mean 和 negV_max。值越大，说明“V 型”特征走势越明显，股价日内风险较大，预期收益较低；将“V 型”走势特征除以对应的持续分钟数，刻画时间加权下的“V 型”走势特征，持续时间越短且原始走势特征越大，“V 型”走势特征越强烈，对应数值越大。同样计算时间加权下“V 型”走势特征的均值和最小值，逻辑与上述一致，记为 negVwgt_mean 和 negVwgt_max。最终在日度时序上取过去 20 日均值，得到最终的因子值。
- 从日度 IC 测试结果上看，“V 型”走势基础因子表现较好，其中原始“V 型”走势特征的相关因子表现更为出色，最小值计算得到因子表现也更为出色，以 negV_max 为例，该因子 Rank IC 均值为 0.031，ICIR 约为 0.5，表现出较好的股价预测能力。
- 从日度组合测试结果上看，“V 型”走势基础性相关因子的表现也十分优秀，多空夏普均较高，且无明显回撤，多头收益率较高。从多空组合的夏普比率上看，negV_max 的表现更为优秀，其多空收益率在 20% 左右，多头收益率约为 10%，夏普比率在 3 左右。

表 19、“V 型”走势基础性因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
negV_mean	0.028	0.074	0.38	17.70
negV_max	0.031	0.067	0.46	21.49
negVwgt_mean	0.021	0.073	0.29	13.64
negVwgt_max	0.028	0.063	0.44	20.76

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 19、negV_max 因子 Rank IC 和累计 IC



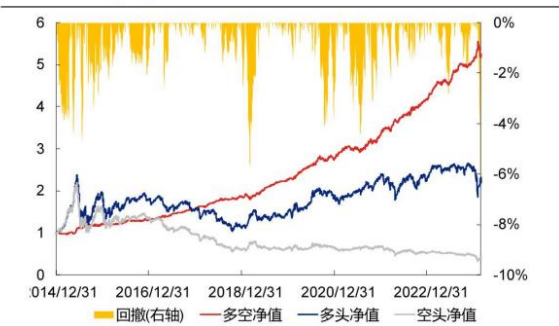
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 20、“V 型”走势基础性因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
negV_mean	19.93%	8.79%	-10.12%	7.55%	2.64	-7.88%	7.73%
negV_max	20.62%	9.89%	-9.51%	6.75%	3.06	-6.85%	9.49%
negVwgt_mean	16.68%	7.42%	-8.51%	7.11%	2.35	-7.86%	7.46%
negVwgt_max	20.36%	9.52%	-9.46%	6.28%	3.24	-6.18%	9.40%

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 20、negV_max 因子多空组合净值



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

5.4 改进步骤 3：“V 型”特征下的“闪电崩盘”概率

- 在前序研究中，我们引入泊松分布刻画概率，发现“闪电崩盘”概率因子本身概率化的含义**不仅能够截面上对股票进行排序，还能在时序上刻画流动性资产的未来风险**。在本章中，我们同样尝试将“V 型”特征通过类似的想法进行概率化，进而得到最终的“闪电崩盘”概率因子。
- 在上述的计算中，我们发现：**日内出现过“V型”特征的股票，其预期收益更低**。然而，在现有的结果中，“V 型”特征本身由此前的正整数变为了实数值，并不满足泊松分布的实际含义和分布特征。在目前的情况，我们需要刻画的是：**在未来一天内，发生“V 型”走势特征数值小等于某一阈值的概率是多少**。因此在本节中，我们将改变思路，使用更为贴切的连续性分布刻画“闪电崩盘”概率。具体如下：



引用分布：类比于泊松分布，**指数分布**同样是对事件发生相关信息进行刻画，且有着明显的长尾特征。不同于泊松分布对于次数的刻画，指数分布是在确定时间区间的情况下，事件发生的时间间隔对应概率，或事件严重程度对应概率；



刻画事件：确定好分布后，我们将“V 型”走势类比于一个事件，即发生一次“V 型”走势，其最终的幅度值类比于事件发生的时间间隔，越大的幅度值对应更大的时间间隔，发生的难度越大；



估计分布：在上述测试，我们发现每日的最小幅度值，即最明显的“V 型”特征更具代表性。在 T 日，我们回望过去 20 日，得到一只股票过去 20 个“V 型”走势特征的最大值，以此作为该股指数分布的样本值 $VexpSample_T^i$ ，进而估计得到该股的指数分布函数；其中， rvV_t 为股票 i 在 t 日的“V 型”走势特征序列。

$$VexpSample_T^i = [\max(rvV_t^i)]_{t=T-20, T-19, \dots, T}$$



计算概率：类比于前序报告，我们以当日市场上所有股票的“V 型”走势特征均值的前 25% 分位数作为阈值，刻画该股在指数分布下 V 型走势特征比阈值的概率为多少，其中， λ_T^i 为股票 i 通过样本数据估计得到的对应指数分布参数 λ 。该概率是在表达：对于某一只股票而言，最为明显的“V 型”走势特征能够超过全市场前 25% 的概率为多少，数值越大则说明未来发生“闪电崩盘”的概率越大。

$$flashCrashProbV_T^i = Prob(x \geq threshold) = 1 - (1 - e^{-\lambda_T^i \cdot threshold})$$

5.5 改进步骤 4：“闪电崩盘” 概率应用

- 计算因子值之前，我们首先根据闪电崩盘概率，统计其对于预期风险的规避能力，即：如果当期某些股票的“闪电崩盘”概率较高，其未来短期内是否会出现相对明显的负向收益。
- 在测算层面，我们设置发出警示信号的阈值分别是 60%、75% 和 85%，即概率超过阈值时，该股被给予警示信号，我们统计其未来 1 日、3 日和 5 日的区间收益率和区间超额收益率，并进一步统计时序上所有被给予警示信号的股票的收益率均值、超额收益率均值以及负向收益率样本数占比。此外，由于小市值股票的“崩盘”概率从逻辑层面上概率较高，我们进一步以流通市值 25% 分位点作为范围，单独统计对应阈值下对小市值股票的预测能力。统计结果展示如下图。
- 我们的测试时间区间和全文设置的回测区间一致，且为每日判断。从测试结果上看，所有测试维度中的收益率和超额收益率均值均为负数，且负向收益率占比值基本在 55% 以上，这说明“闪电崩盘”概率在时序上具备一定的风险预测能力。以 75% 作为阈值统计，未来 3 和 5 日的负向收益率占比约为 52%，在小市值中达到 56%；负向超额的占比达到约 58%，在小市值中达到 60%，具备较强的预测能力。此外，随着阈值的逐步提升，指标的预测能力逐步提升，呈现出明显的单调性。

图 21、“闪电崩盘”概率时序预测能力统计

	60%	60%-后 25%市值	75%	75%-后 25%市值	85%	85%-后 25%市值
1D收益率均值	-0.2%	-0.3%	-0.4%	-0.6%	-0.6%	-1.0%
3D收益率均值	-0.4%	-0.4%	-0.9%	-1.1%	-1.5%	-1.9%
5D收益率均值	-0.5%	-0.5%	-1.3%	-1.3%	-2.3%	-2.6%
1D超额收益率均值	-0.2%	-0.3%	-0.4%	-0.6%	-0.7%	-1.0%
3D超额收益率均值	-0.5%	-0.5%	-1.0%	-1.1%	-1.6%	-2.0%
5D超额收益率均值	-0.7%	-0.7%	-1.5%	-1.5%	-2.5%	-2.8%
1D负收益率概率	49.7%	50.0%	51.5%	52.5%	53.1%	55.3%
3D负收益率概率	52.8%	52.2%	56.0%	55.8%	58.2%	59.4%
5D负收益率概率	53.3%	52.4%	56.8%	56.2%	60.0%	60.7%
1D负超额收益率概率	57.2%	57.2%	58.4%	59.5%	59.1%	61.4%
3D负超额收益率概率	58.3%	58.2%	60.9%	61.5%	62.4%	63.7%
5D负超额收益率概率	58.9%	58.7%	61.4%	61.5%	64.4%	64.7%

资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

5.5 改进步骤 4：“闪电崩盘” 概率应用

- 进一步，我们在日度层面取过去 20 日均值，作为最终的“V 型”特征下的“闪电崩盘”概率因子，记为 flashCrashProbV。
- 从日度 IC 测试结果上看，该因子表现较好，Rank IC 均值为 0.031 (同期此前闪电崩盘因子均值为 0.014)，ICIR 约为 0.5，表现出较好的股价预测能力。
- 从日度组合测试结果上看，该因子的表现也十分优秀，多空夏普均较高，且无明显回撤，多空收益率在 21% 左右，多头收益率为 10%，夏普比率在 3 左右。
- 从 Rank IC 时序特征、多空净值曲线以及分位数组合测试结果上看，flashCrashProb 因子的多空净值长期呈现上升趋势。

表 21、flashCrashProbV 因子 Rank IC 测试结果

因子名称	IC 均值	IC 标准差	ICIR	T 统计量
flashCrashProbV	0.031	0.067	0.46	21.56

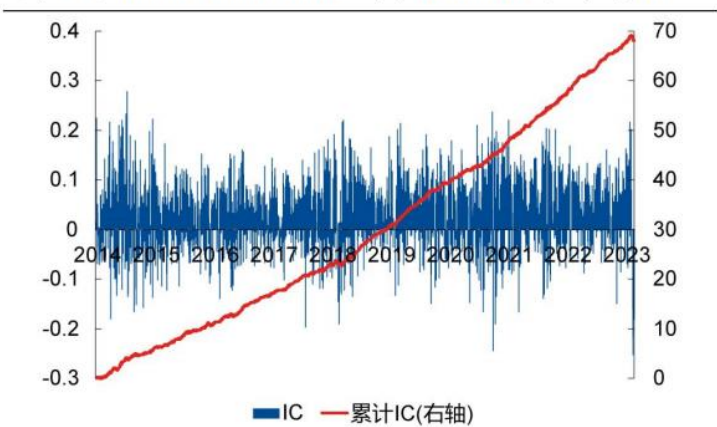
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

表 22、flashCrashProbV 因子组合测试结果

因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
flashCrashProbV	20.72%	9.97%	-9.51%	6.73%	3.08	-6.73%	9.77%

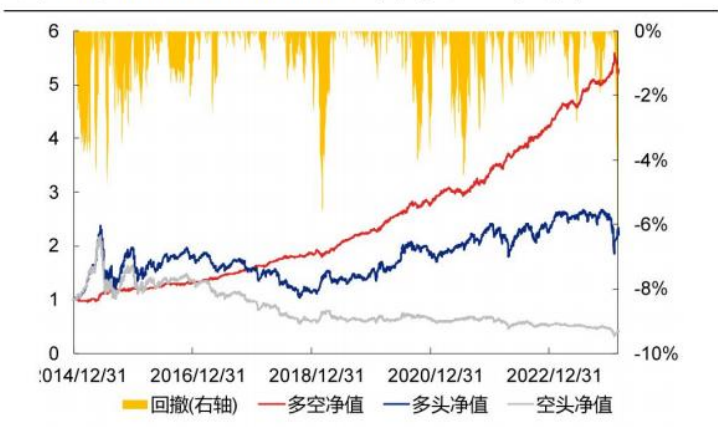
资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 22、flashCrashProbV 因子 Rank IC 和累计 IC



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

图 23、flashCrashProbV 因子多空组合净值



资料来源：上交所、深交所行情数据，Wind，聚源，兴业证券经济与金融研究院整理

目录 CATALOGUE

01

高频研究回顾与关联信息因子研究框架

02

关联信息基础型因子

03

关联信息中的全局特征：非线性因果关系

04

关联信息中的局部特征：“交易时间”和持续期特征

05

关联信息中的典型特征：再论“闪电崩盘”

06

回测表现更新

6

回测表现更新：单因子日度回测-截至20240531

因子名称	2014年底至今				2021年底至今			
	IC均值	IC标准差	ICIR	T统计量	IC均值	IC标准差	ICIR	T统计量
te_v2r	0.031	0.083	0.379	18.122	0.036	0.082	0.438	10.568
te_r2v	0.045	0.079	0.571	27.294	0.049	0.077	0.637	15.370
cutVol_timeVar	0.027	0.072	0.374	17.911	0.027	0.053	0.509	12.291
cutVol_rtnMean	0.046	0.089	0.520	24.896	0.042	0.091	0.462	11.151
cutVol_rtnVar	0.048	0.080	0.594	28.419	0.052	0.075	0.701	16.912
cutVol_timeCor	0.045	0.099	0.450	21.533	0.049	0.096	0.513	12.370
cutVol_entropy	0.020	0.066	0.311	14.888	0.021	0.047	0.450	10.856
negV_mean	0.028	0.075	0.376	17.983	0.035	0.069	0.504	12.167
negV_max	0.031	0.068	0.454	21.706	0.040	0.070	0.562	13.561
flashCrashV	0.031	0.068	0.456	21.795	0.040	0.070	0.566	13.654

回测区间	因子名称	多空收益率	多头收益率	空头收益率	年化波动率	夏普比率	最大回撤	多头换手率
2014年底至今	te_v2r	36.76%	14.33%	-17.45%	8.36%	4.396	-6.96%	8.89%
	te_r2v	47.86%	17.95%	-21.47%	8.59%	5.572	-6.72%	12.33%
	cutVol_timeVar	28.31%	12.77%	-11.96%	6.44%	4.395	-5.74%	8.10%
	cutVol_rtnMean	43.15%	16.41%	-18.90%	8.40%	5.137	-7.56%	15.34%
	cutVol_rtnVar	35.74%	14.98%	-16.29%	8.08%	4.423	-6.00%	7.17%
	cutVol_timeCor	33.70%	16.11%	-14.73%	8.63%	3.906	-7.48%	8.68%
	cutVol_entropy	24.72%	11.22%	-10.47%	5.82%	4.248	-5.78%	12.18%
	negV_mean	20.52%	9.18%	-10.27%	7.61%	2.697	-7.88%	7.72%
	negV_max	20.95%	10.13%	-9.60%	6.80%	3.080	-7.18%	9.48%
	flashCrashV	21.06%	10.21%	-9.60%	6.78%	3.107	-7.13%	9.75%
2021年底至今	te_v2r	31.42%	0.57%	-23.76%	6.46%	4.864	-4.14%	7.38%
	te_r2v	41.21%	3.76%	-27.15%	7.18%	5.742	-4.72%	10.68%
	cutVol_timeVar	26.40%	0.61%	-20.27%	4.81%	5.490	-2.55%	6.39%
	cutVol_rtnMean	36.13%	0.15%	-26.12%	8.13%	4.443	-7.56%	13.78%
	cutVol_rtnVar	34.16%	2.52%	-24.14%	6.80%	5.026	-4.56%	5.66%
	cutVol_timeCor	29.07%	1.69%	-22.23%	7.72%	3.766	-6.12%	6.67%
	cutVol_entropy	22.25%	-1.23%	-18.86%	4.22%	5.268	-3.65%	10.72%
	negV_mean	26.50%	0.04%	-21.48%	6.68%	3.970	-6.07%	6.56%
	negV_max	25.55%	0.49%	-20.59%	6.65%	3.841	-7.18%	7.92%
	flashCrashV	25.68%	0.43%	-20.72%	6.65%	3.863	-7.13%	8.17%

回测表现更新：等权复合因子周度测试结果-截至20240531

股票池	组合	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率
沪深300	Top	11.1%	21.1%	0.53	44.5%	32.7%	8.0%
	1	5.2%	20.7%	0.25	86.5%	39.5%	2.3%
	2	2.4%	21.4%	0.11	92.8%	45.5%	-0.2%
	3	0.0%	22.3%	0.00	82.3%	47.2%	-2.4%
	Bottom	-5.4%	23.9%	-0.22	43.9%	64.8%	-7.5%
	市场	2.7%	20.8%	0.13		45.8%	
	L-S	15.7%	13.0%	1.21		28.2%	
股票池	组合	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率
中证500	Top	13.3%	24.5%	0.54	46.2%	37.9%	7.9%
	1	10.5%	25.1%	0.42	90.0%	45.9%	5.4%
	2	8.2%	25.3%	0.32	96.9%	41.5%	3.2%
	3	2.7%	26.3%	0.10	86.3%	51.1%	-1.7%
	Bottom	-9.5%	27.6%	-0.35	47.7%	82.0%	-13.3%
	市场	5.0%	24.2%	0.21		51.5%	
	L-S	23.4%	10.9%	2.14		15.9%	
股票池	组合	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率
全市场	Top	20.8%	26.4%	0.79	59.7%	39.2%	12.5%
	1	19.6%	26.8%	0.73	114.0%	37.8%	11.6%
	2	15.8%	26.8%	0.59	130.4%	41.0%	8.0%
	3	15.9%	26.9%	0.59	137.0%	38.1%	8.2%
	5	14.7%	27.3%	0.54	139.3%	39.8%	7.2%
	6	10.8%	27.6%	0.39	138.3%	47.2%	3.6%
	7	9.1%	27.8%	0.33	134.4%	48.5%	2.1%
	8	2.4%	27.9%	0.09	124.9%	61.0%	-4.2%
	9	-4.2%	28.4%	-0.17	105.5%	74.3%	-10.8%
	Bottom	-25.7%	29.9%	-0.86	63.0%	97.1%	-30.3%
	市场	7.4%	25.8%	0.29		50.3%	
	L-S	59.4%	11.6%	5.13		8.9%	
股票池	组合	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率
中证800	Top	12.6%	23.4%	0.54	43.7%	37.6%	8.1%
	1	9.3%	23.1%	0.41	86.6%	42.3%	5.0%
	2	5.9%	22.7%	0.26	92.8%	41.3%	1.6%
	3	1.2%	23.8%	0.05	82.6%	50.2%	-2.7%
	Bottom	-7.4%	25.4%	-0.29	45.4%	75.2%	-10.8%
	市场	4.2%	22.5%	0.19		49.1%	
	L-S	20.3%	10.6%	1.91		16.8%	
股票池	组合	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率
中证1000	Top	17.8%	27.9%	0.64	50.3%	42.8%	11.6%
	1	14.2%	28.6%	0.50	95.1%	45.7%	8.5%
	2	10.9%	29.2%	0.37	102.1%	50.5%	5.6%
	3	2.6%	29.7%	0.09	91.7%	63.6%	-2.2%
	Bottom	-15.4%	29.8%	-0.52	53.5%	91.3%	-19.4%
	市场	5.6%	27.0%	0.21		56.4%	
	L-S	37.5%	10.2%	3.67		10.4%	
股票池	组合	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率
国证2000	Top	22.3%	27.1%	0.82	61.8%	39.7%	11.4%
	1	22.7%	28.6%	0.79	117.0%	39.3%	12.3%
	2	19.6%	29.0%	0.68	133.1%	45.3%	9.6%
	3	19.5%	28.9%	0.67	139.9%	44.2%	9.4%
	5	18.0%	29.8%	0.61	142.0%	43.9%	8.4%
	6	16.8%	29.5%	0.57	141.1%	45.8%	7.2%
	7	10.3%	29.4%	0.35	137.1%	51.4%	1.1%
	8	3.6%	29.4%	0.12	127.5%	59.8%	-5.0%
	9	-6.5%	29.4%	-0.22	108.8%	80.1%	-14.3%
	Bottom	-24.3%	30.6%	-0.79	65.2%	96.8%	-30.5%
	市场	9.7%	26.9%	0.36		50.3%	
	L-S	58.2%	12.3%	4.74		8.7%	

回测表现更新：单因子周度测试结果-2024年以来

因子名称	Top超额收益率	多空夏普比率	最大回撤	因子名称	Top超额收益率	多空夏普比率	最大回撤
rtn_condVaR	13.1%	6.76	2.0%	cumsumvol_mean	4.2%	2.21	5.7%
exRtn_minFre	12.1%	5.03	2.2%	negV_mean	4.0%	3.77	4.2%
vsa_high2min	11.7%	4.44	3.5%	highStdRtn_mean	3.6%	3.85	7.7%
exRtn_maxVal	9.9%	8.22	1.9%	gmm_meandif2wgt	3.5%	6.52	12.1%
logvol_skew	9.7%	5.82	3.3%	cutVol_timeVar	3.2%	3.00	9.6%
gmm_wgt	8.4%	3.67	1.2%	cutVol_rtnVar	2.9%	5.45	5.2%
logvol_90tail	8.2%	4.52	3.1%	cutVol_timeCor	2.8%	2.19	5.1%
nos_gs	8.2%	2.74	3.1%	gmm_mean	2.6%	6.21	14.5%
vsa_low2max	8.0%	0.97	8.3%	cutVol_entropy	2.1%	2.50	8.8%
real_kurt	7.6%	2.41	4.9%	flashCrashV	1.9%	3.27	3.8%
foc_Comb	7.2%	4.65	4.3%	negV_max	1.9%	3.36	3.8%
volroc_skew	7.2%	2.70	2.2%	rv_up	1.6%	4.12	5.8%
DW_Comb	6.5%	1.36	6.2%	flashCrashProb	1.0%	-0.16	3.4%
LBQ_Comb	6.5%	2.29	3.8%	cutVol_rtnMean	0.6%	5.17	11.3%
vol_entropy	5.8%	2.52	1.8%	rho_Comb	0.0%	3.46	12.4%
rv_down	5.7%	3.23	6.0%	real_skew	-0.2%	4.49	14.7%
real_var	5.7%	4.36	4.5%	te_v2r	-1.3%	0.51	7.3%
logvol_10tail	5.7%	3.32	4.1%	gmm_mean2wgt	-2.8%	3.73	10.0%
vol_maxstd	5.4%	4.40	2.8%	rtn5_mean	-2.9%	2.81	14.3%
vol_maxmean	4.9%	4.74	3.6%	rv_umd	-3.4%	3.43	16.0%
te_r2v	4.7%	4.99	3.8%	gmm_meandif	-4.6%	1.45	12.5%
volroc_kurt	4.6%	2.14	3.3%	vsa_ratio	-8.8%	-1.89	13.5%
cumsumvol_std	4.5%	2.30	3.8%				

免责声明

DISCLAIMER

投资评级说明

投资建议的评级标准	类别	评级	说明
材料中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为材料发布日后的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。其中：A股市场以沪深300指数为基准；新三板市场以三板成指为基准；香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		增持	相对同期相关证券市场代表性指数涨幅在5%~15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%~5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

使用本研究报告的风险提示及法律声明

本公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本材料仅供本公司的客户使用，本公司不会因接收人收到本材料而视其为客户。本材料中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。投资者自主作出投资决策并自行承担投资风险，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，任何有关本材料的摘要或节选都不代表本材料正式完整的观点，一切须以本公司向客户发布的相关报告完整版本为准。该等信息、意见并未考虑到获取本材料人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本材料中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本材料所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本材料所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本材料所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本材料所载资料、意见及推测不一致的报告；本公司不保证本材料所含信息保持在最新状态。

除非另行说明，本材料中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本材料意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向材料所有接收者进行更新的义务。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材料中的意见或建议不一致的投资决策。

特别声明

在法律许可的情况下，本公司可能会持有本材料中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，委托人应当考虑到本公司及/或其相关人员可能存在影响本材料观点客观性的潜在利益冲突。投资者请勿将本材料视为投资或其他决定的唯一信赖依据。

THANKS

—— 研究发现价值 研究创造价值 ——