

高频因子计算的 GPU 加速

华泰研究

2023 年 10 月 16 日 | 中国内地

深度研究

研究员	林晓明
SAC No. S0570516010001	linxiaoming@htsc.com
SFC No. BPY421	+(86) 755 8208 0134
研究员	何康, PhD
SAC No. S0570520080004	hegang@htsc.com
SFC No. BRB318	+(86) 21 2897 2039

人工智能系列之 70: 使用 NVIDIA RAPIDS 对高频因子计算进行 GPU 加速

本文使用 NVIDIA RAPIDS 对高频因子计算进行 GPU 加速。量化因子计算场景下, RAPIDS 的直接助力是用 CuPy、cuDF 的 GPU 运算替代 NumPy、Pandas 的 CPU 运算。在 RTX 3090 和 i9-10980XE 测试环境下计算分钟线因子, CuPy 和 cuDF 替换库函数的提速效果约为 6 倍, 若同时将 for 循环替换为矩阵运算, 最终提速超 100 倍。预计 RTX 4090 和 A800 提速更显著。最大化 ICIR 法合成高频因子, 并与华泰金工神经网络多频率因子静态加权, 构建中证 1000 指数增强组合, 以 2016 年末至 2023 年 9 月为回测区间, 信息比率从 3.70 提升至 3.87, 超额收益 Calmar 比率从 2.41 提升至 3.96。

使用 CuPy/cuDF 替代 NumPy/Pandas, RTX 3090 实现约 6 倍性能提升
RAPIDS 的重要特性之一是将基于 CUDA 底层代码的优化以 Python 语言的形式体现。常用 API 如 CuPy (对标 NumPy)、cuDF (对标 Pandas)、cuML (对标 scikit-learn) 等。由于 API 语法几乎相同, 仅需轻量级代码修改, 即可实现 CPU 运算到 GPU 运算的迁移。针对全部因子使用 CuPy 和 cuDF 替换原函数, 部分因子使用矩阵运算替换 for 循环。结果表明: 不引入矩阵运算时, 单独替换库函数反而增加时间开销。若两步同时进行, 替换库函数带来的性能提升约 6 倍 (RTX 3090), 矩阵运算带来的性能提升约 18 倍。GPU 性能同型号和数据量相关, 单次运算数据量越大, 加速效果越显著。

分钟线选股因子: 价格全局特征类、价格局部特征类

本文测试 5 类共计 50 个分钟线选股因子。以下展示测试效果较好的因子及投资逻辑: (1)价格全局特征类: return_intraday、tp_diff、return_improved 的本质是不同形式的日内反转因子, return_var 的本质是日内低波动因子。(2)价格局部特征类: return_last_30min 和 return_skewness_last_30min 的本质是尾盘反转因子, return_upward_var 和 return_downward_var 的本质是日内低波动因子的精细化刻画。

分钟线选股因子: 成交量/额类、成交关联价格类、价量相关性类

(3)成交量/额类: volume_open_30min_ratio 给予开盘成交不活跃的股票风险溢价, 市场情绪一般在开盘释放, 该因子或反映理性交易者占比。amount_out_order_avg_ratio 给予单笔流出金额较大的股票风险溢价, 或对应快速下跌后的反转或者主力的操纵行为。(4)成交关联价格类: cum_return_top30_order 的本质是反转因子的精细化刻画, 大单推动的涨幅更具信息量。(5)价量相关性类: VP、VP_top33_volume、VR_1min_lag 的本质都是捕捉量价背离, 即缩量上涨或放量下跌。

最大化 ICIR 法合成分钟线因子, 与神经网络因子结合构建选股策略

进一步围绕分钟线因子构建选股策略。采用最大化 ICIR 法对前述 50 个因子进行合成, 以未来 5 日收益为预测目标, 合成因子 RankICIR (未年化) 1.50, 对冲组合夏普比率 3.59, Top 层信息比率 3.81。该合成因子和神经网络多频率因子相关系数为 0.25, 两者截面标准化后静态融合, 构建指数增强策略。结果显示, 结合分钟线合成因子后, 策略各项指标均有提升, 尤其体现在回撤控制。中证 500 增强信息比率从 2.78 提升至 3.01, 超额收益 Calmar 比率从 1.37 提升至 2.07, 中证 1000 增强信息比率从 3.70 提升至 3.87, 超额收益 Calmar 比率从 2.41 提升至 3.96。

风险提示: 人工智能挖掘市场规律是对历史的总结, 市场规律在未来可能失效。人工智能技术存在过拟合风险。深度学习模型受随机数影响较大。本文测试的选股模型调仓频率较高, 假定以 vwap 价格成交, 忽略其他交易层面因素影响。

正文目录

导言	4
NVIDIA RAPIDS 实践.....	5
RAPIDS 安装	5
操作系统和 WSL 安装	5
WSL Conda 和 RAPIDS 安装	5
高频因子代码优化	7
简单替换库函数	7
矩阵运算代替 for 循环，同时替换库函数	8
分钟线因子.....	11
价格全局特征	11
价格局部特征	11
成交量/额	12
成交关联价格	12
价量相关性	13
单因子测试方法与结果	14
价格全局特征	14
价格局部特征	15
成交量/额	17
成交关联价格	18
价量相关性	19
基于高频因子的指增策略.....	22
最大化 ICIR 法合成因子	22
指数增强组合	22
总结和讨论.....	25
来自 NVIDIA 的建议	26
风险提示	26

图表目录

图表 1: RAPIDS: 数据科学的 GPU 加速	4
图表 2: RAPIDS 常用库	4
图表 3: WSL 安装: 启动或关闭 Windows 功能	5
图表 4: WSL 安装: DNS 设置	5
图表 5: RAPIDS 安装: 选择版本	6
图表 6: RAPIDS 安装: 在 WSL 中运行	6
图表 7: 计算下行收益率方差 return_downward_var 代码: for 循环+CPU	7
图表 8: 计算下行收益率方差 return_downward_var 代码: for 循环+GPU	7

图表 9: 计算下行收益率方差 <code>return_downward_var</code> 代码: 矩阵运算+CPU	8
图表 10: 计算下行收益率方差 <code>return_downward_var</code> 代码: 矩阵运算+GPU	8
图表 11: 因子计算环节 CPU、GPU 性能对比 (明细)	9
图表 12: 因子计算环节 CPU、GPU 性能对比 (汇总)	10
图表 13: 数据读取、预处理及导出环节 CPU、GPU 性能对比	10
图表 14: 价格全局特征类因子	11
图表 15: 价格局部特征类因子	11
图表 16: 成交量/额类因子	12
图表 17: 成交关联价格类因子	12
图表 18: 价量相关性类因子	13
图表 19: 价格全局特征因子测试指标 (周频)	14
图表 20: <code>return_improved</code> 因子分层相对净值	14
图表 21: 价格全局特征因子测试指标 (双周频)	15
图表 22: 价格全局特征因子测试指标 (月频)	15
图表 23: 价格局部特征因子测试指标 (周频)	15
图表 24: <code>return_downward_var</code> 因子分层相对净值	16
图表 25: 价格局部特征因子测试指标 (双周频)	16
图表 26: 价格局部特征因子测试指标 (月频)	16
图表 27: 成交量/额因子测试指标 (周频)	17
图表 28: <code>amount_out_order_avg_ratio</code> 因子分层相对净值	17
图表 29: 成交量/额因子测试指标 (双周频)	18
图表 30: 成交量/额因子测试指标 (月频)	18
图表 31: 成交关联价格因子测试指标 (周频)	18
图表 32: <code>cum_return_top30_order</code> 因子分层相对净值	19
图表 33: 成交关联价格因子测试指标 (双周频)	19
图表 34: 成交关联价格因子测试指标 (月频)	19
图表 35: 价量相关性因子测试指标 (周频)	20
图表 36: <code>VP_top33_volume</code> 因子分层相对净值	20
图表 37: 价量相关性因子测试指标 (双周频)	20
图表 38: 价量相关性因子测试指标 (月频)	21
图表 39: 最大化 ICIR 法合成因子测试指标 (预测未来 5 日收益)	22
图表 40: 分钟线合成因子分层相对净值	22
图表 41: 指数增强组合构建方法	23
图表 42: 中证 500 指数增强组合回测绩效	23
图表 43: 中证 500 指数增强组合超额净值	23
图表 44: 中证 1000 指数增强组合回测绩效	24
图表 45: 中证 1000 指数增强组合超额净值	24

导言

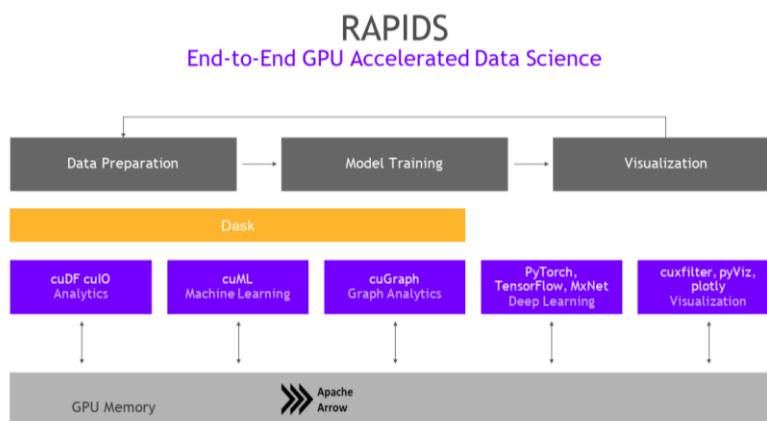
量化投资作为一种技术驱动的方法论，对效率的追求是其永恒的核心。近年来，AI 方法的运用显著提升了因子挖掘和合成环节的效率，然而因子计算环节似乎未充分享受技术红利。传统因子计算通常采用 CPU 运算，随着高频数据的广泛应用，CPU 性能面临上限，而 NVIDIA 引领的 GPU 软硬件生态日益成熟，使得高频因子计算的 GPU 加速成为可能。

RAPIDS 是 NVIDIA 针对数据科学与机器学习推出的 GPU 加速平台。RAPIDS 以 CUDA-X AI 为基础，由一系列开源软件库和 API 组成，支持完全在 GPU 上执行从数据加载和预处理到机器学习、图形分析和可视化的端到端数据科学工作流程。据 NVIDIA 官网，RAPIDS 能够为数据科学领域提供 50 倍的效率提升。

RAPIDS 的重要特性之一是将基于 CUDA 底层代码的优化以 Python 高级语言的形式体现。常用 API 包括 cuDF (对标 Pandas)、cuML (对标 scikit-learn)、cuGraph (对标 networkX) 等，由于 API 语法几乎相同，使用者仅需要轻量级的代码修改，即可实现 CPU 运算到 GPU 运算的迁移。量化投资因子计算场景下，RAPIDS 的直接助力是用 CuPy、cuDF 库替代 NumPy、Pandas 库，实现高频因子计算的 GPU 加速。

本研究使用 NVIDIA RAPIDS 对日内分钟线因子计算进行 GPU 加速。在 NVIDIA GeForce RTX 3090 和 Intel Core i9-10980XE 测试环境下，CuPy 和 cuDF 替换库函数的 GPU 提速效果约为 6 倍，若同时将 for 循环替换为矩阵运算，最终提速超过 100 倍。使用最大化 ICIR 法合成 50 个分钟线因子，并与华泰金工神经网络多频率因子静态加权，构建中证 1000 指数增强组合，以 2016 年末至 2023 年 9 月末为回测区间，组合信息比率从 3.70 提升至 3.87，超额收益 Calmar 比率从 2.41 提升至 3.96。

图表1：RAPIDS：数据科学的 GPU 加速



资料来源：NVIDIA，华泰研究

图表2：RAPIDS 常用库

RAPIDS 库	对标 Python 库	功能
cuDF	Pandas	数据集处理
CuPy	NumPy	科学计算
cuML	scikit-learn	机器学习
cuGraph	networkX	图结构处理
cuSpatial	geoPandas	空间分析
cuCIM	scikit-image	图像处理
cuSignal	scipy.signal	信号处理

资料来源：NVIDIA，华泰研究

NVIDIA RAPIDS 实践

RAPIDS 安装

RAPIDS 官网提供了详细安装方法，可参考：<https://docs.rapids.ai/install>。实际安装过程中可能遇到较多问题，可参考本文经验。

操作系统和 WSL 安装

操作系统推荐 Ubuntu 20.04 和 CentOS 7。如使用 Windows 系统，需要通过 Linux 的 Windows 子系统 WSL2 (Windows Subsystem for Linux 2)。微软官网提供详细安装方法，参考：<https://learn.microsoft.com/en-us/windows/wsl/install>。

WSL 安装可能遇到较多问题，以下讲解要点：

1. 推荐使用 Window 11 系统。
2. 任务管理器—性能—CPU，确认 CPU 虚拟化已启用。
3. 启用或关闭 Windows 功能，勾选“适用于 Linux 的 Windows 系统”和“虚拟机平台”。
4. 管理员身份启动 Power Shell，运行 `bcdedit /set hypervisorlaunchtype auto`。
5. 网络和 Internet 设置—高级网络设置—属性，DNS 服务器分配设为“手动”；Ipv4 DNS 服务器，首选 DNS 设为 114.114.114.114，备用 DNS 设为 8.8.8.8。若此步不执行，安装 WSL 可能报错“无法解析服务器的名称或地址”。

图表3：WSL 安装：启动或关闭 Windows 功能



资料来源：华泰研究

图表4：WSL 安装：DNS 设置



资料来源：华泰研究

6. 管理员身份启动 CMD，运行 `wsl --update`。如此步不执行，启动 Ubuntu 可能报错。
7. 继续运行 `wsl --install -d Ubuntu`，安装 Ubuntu 子系统。
8. 启动“适用于 Linux 的 Window 子系统”（WSL），设置账户和密码，完成 WSL 安装。

WSL Conda 和 RAPIDS 安装

RAPIDS 官网提供了 WSL 中安装 Conda 和 RAPIDS 的方法，参考：<https://docs.rapids.ai/install#wsl2>。

原文档较简略，Conda 和 RAPIDS 安装可能遇到较多问题，以下讲解要点：

1. 启动 WSL，登录 Ubuntu 系统。
2. 依次运行下列指令，下载和安装 Miniconda3：

```
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
bash Miniconda3-latest-Linux-x86_64.sh
```


3. 运行 `export PATH=/home/username/miniconda3/bin:$PATH`，设置环境变量。
4. 在 RAPIDS 官网 (<https://docs.rapids.ai/install>) 选择需要安装的版本、方式、CUDA 版本、Python 版本、RAPIDS 库种类、附加扩展库，生成相应代码，在 WSL 中运行。以仅安装 cuDF 和 cuML 为例，运行代码 `conda create --solver=libmamba -n rapids-23.08 -c rapidsai -c conda-forge -c nvidia cudf=23.08 cuml=23.08 python=3.10 cuda-version=11.2`。该指令将自动创建名为 rapids-23.08 的 conda 环境。

图表5: RAPIDS 安装: 选择版本



RELEASE: Stable (23.08) | Nightly (23.10a)

METHOD: Conda | pip | Docker

ENV. CUDA: CUDA 11.2 | CUDA 11.8 | CUDA 12.0

PYTHON: Python 3.9 | Python 3.10

RAPIDS PACKAGES: Standard | Choose Specific Packages

cuDF | cuML | cuGraph | cuSpatial/cuProj | cuxfilter | cuSignal | cuCIM

ADDITIONAL PACKAGES: Dask-SQL | JupyterLab | Plotly Dash | Graphistry | TensorFlow | Xarray-Spatial

! RAPIDS currently doesn't support channel_priority: strict; use channel_priority: flexible instead

COMMAND: `conda create --solver=libmamba -n rapids-23.08 -c rapidsai -c conda-forge -c nvidia \ cudf=23.08 cuml=23.08 python=3.10 cuda-version=11.2`

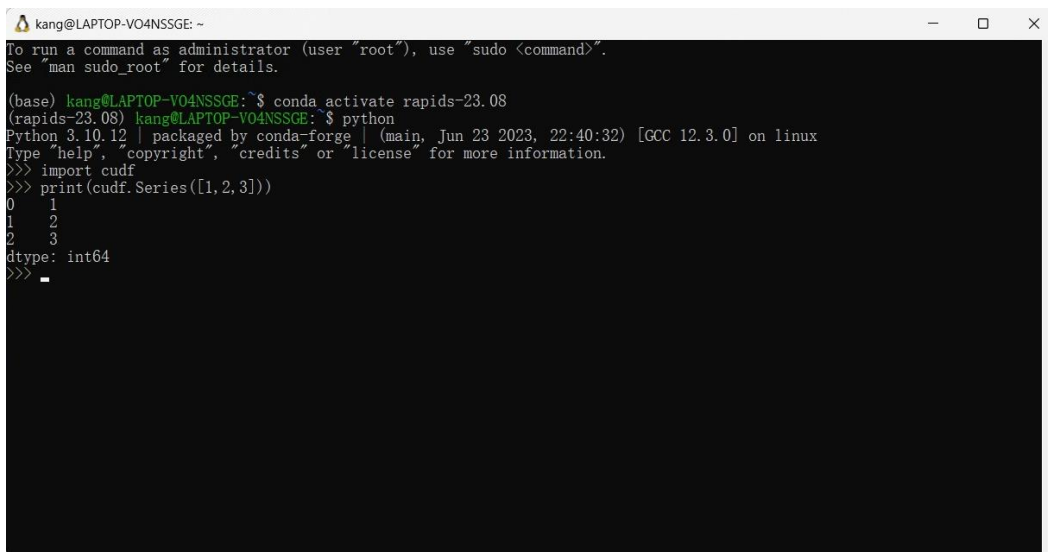
COPY COMMAND

资料来源: NVIDIA, 华泰研究

5. 运行 `conda activate rapids-23.08`，进入该环境。
6. 运行 `python`，进入 Python 环境，依次运行下列代码，若正常运行则代表安装成功。

```
import cudf
print(cudf.Series([1, 2, 3]))
```
7. Miniconda 环境仅安装少量库，可通过 `conda install` 安装 Pandas 等常用库。
8. 如需在代码中调用 Windows 系统文件，路径为 “/mnt/盘符/路径”，如 “/mnt/d/data”。

图表6: RAPIDS 安装: 在 WSL 中运行



```
kang@LAPTOP-VO4NSSGE: ~
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

(base) kang@LAPTOP-VO4NSSGE:~$ conda activate rapids-23.08
(rapids-23.08) kang@LAPTOP-VO4NSSGE:~$ python
Python 3.10.12 | packaged by conda-forge | (main, Jun 23 2023, 22:40:32) [GCC 12.3.0] on linux
Type "help", "copyright", "credits" or "license()" for more information.
>>> import cudf
>>> print(cudf.Series([1, 2, 3]))
0    1
1    2
2    3
dtype: int64
>>>
```

资料来源: 华泰研究

高频因子代码优化

本文测试环境为：CPU——Intel Core i9-10980XE，GPU——NVIDIA GeForce RTX 3090（以下分别简称 i9-10980XE 和 3090）。以 2023 年 4 月 24 日单日全 A 股分钟线数据为例测试时间开销。

CPU 因子计算代码改造为 GPU 版本过程中，采用两种优化方式：

1. 使用 CuPy 和 cuDF 替换 NumPy 和 Pandas 函数（针对全部 50 个因子）；
2. 使用矩阵运算替换 for 循环（针对 50 个因子中的 22 个因子）。

结果表明：

1. 单独使用(1)反而增加了时间开销。
2. 同时使用(1)(2)时间开销总体缩短约 117 倍，其中(1)贡献约 18.4 倍，(2)贡献约 6.4 倍。
3. GPU 性能表现主要取决于数据量，单次运算涉及的数据量越大，GPU 加速效果越显著。若数据量过小，反而不如 CPU 计算效率。

简单替换库函数

RAPIDS 的 CuPy、cuDF 对原始 CPU 代码的 NumPy、Pandas 具有同名接口性质，因此可以简单替换库名称简化工作量。如下图，以计算下行收益率方差 `return_downward_var` 代码为例，简单替换 `numpy.var` 为 `cupy.var` 即可。cuDF 具体指令集可参考官方文档：<https://docs.rapids.ai/api/cudf/stable/>。

图表7：计算下行收益率方差 `return_downward_var` 代码：for 循环+CPU

```

1. import numpy as np
2. def return_downward_var(_return):
3.     _return_down = _return * (_return<0)
4.     RDV = np.empty(_return_down.shape[1])
5.     for i_stk in range(_return_down.shape[1]):
6.         if (_return_down[_return_down[:,i_stk]!=0,i_stk]).shape[0] == 0:
7.             RDV[i_stk] = 0
8.         else:
9.             RDV[i_stk] = np.var(_return_down[_return_down[:,i_stk]!=0,i_stk])
10.    return RDV

```

资料来源：华泰研究

图表8：计算下行收益率方差 `return_downward_var` 代码：for 循环+GPU

```

1. import cupy as cp
2. def return_downward_var(_return):
3.     _return_down = _return * (_return<0)
4.     RDV = cp.empty(_return_down.shape[1])
5.     for i_stk in range(_return_down.shape[1]):
6.         if (_return_down[_return_down[:,i_stk]!=0,i_stk]).shape[0] == 0:
7.             RDV[i_stk] = 0
8.         else:
9.             RDV[i_stk] = cp.var(_return_down[_return_down[:,i_stk]!=0,i_stk])
10.    return RDV

```

资料来源：华泰研究

然而实证发现，CPU 版代码单日高频因子计算时间开销为 12.34 秒，简单替换后，GPU 版代码时间开销超过 100 秒。问题在于原因子计算中代码大量使用 for 循环。具体而言：(1)RAPIDS 的接口仅实现了接口内部的并行计算，对 for 循环本身没有优化作用；(2)由于数据在 CPU 和 GPU 间拷贝传递，带来耗时影响。GPU 更适合进行批量运算，如果仅简单替换库函数，for 循环逐轮小数据量地调用 GPU，反而造成拖累。

矩阵运算代替 for 循环，同时替换库函数

我们重新梳理 for 循环逻辑，改为矩阵运算，以 CuPy 接口进行批量处理。以下图计算下行收益率方差 return_downward_var 代码为例。

图表9：计算下行收益率方差 return_downward_var 代码：矩阵运算+CPU

```
1. import numpy as np
2. def return_downward_var(_return):
3.     _return_down = _return * (_return<0)
4.     mask = (_return_down!=0).sum(axis=0)==0
5.     _return_down[_return_down==0] = np.nan
6.     RDV = np.nanvar(_return_down, axis=0)
7.     RDV[mask]=0
8.     return RDV
```

资料来源：华泰研究

图表10：计算下行收益率方差 return_downward_var 代码：矩阵运算+GPU

```
1. import cupy as cp
2. def return_downward_var(_return):
3.     _return_down = _return * (_return<0)
4.     mask = (_return_down!=0).sum(axis=0)==0
5.     _return_down[_return_down==0] = cp.nan
6.     RDV = cp.nanvar(_return_down, axis=0)
7.     RDV[mask]=0
8.     return RDV
```

资料来源：华泰研究

因子计算时间开销明细和汇总结果如下表所示：

1. 对于 50 个替换库函数的因子，GPU 相比 CPU 总时间开销从 0.91 秒缩短至 0.16 秒，性能提升约 5.9 倍。
2. 对于 22 个使用矩阵运算代替 for 循环的因子，CPU 总时间开销从 12.34 秒缩短至 0.67 秒，性能提升约 18 倍。GPU 相比 CPU 总时间开销从 0.67 秒缩短至 0.11 秒，性能提升约 6.4 倍。两项改进总时间开销从 12.34 秒缩短至 0.11 秒，性能提升约 117 倍。

总的看，GPU (3090) 相比 CPU (i9-10980XE) 计算因子带来的性能提升在 5.9~6.4 倍。当然约 6 倍的加速并不是 GPU 的上限，加速效果受到显卡性能和数据量影响，4090、A100 的加速效果应更强，同时单次运算的数据量越大，加速效果也越显著。据 NVIDIA 官网，RAPIDS 的提速可达 50 倍 (<https://www.nvidia.cn/deep-learning-ai/software/rapids/>)。

同时我们观察没有 GPU 加速效果的因子：

1. 以 return_intraday、return_improved、return_last_30min 为例，仅使用日内两个时间点数据，使用数据量过少。
2. 以 intraday_maxdrawdown 为例，CuPy 还未实现 numpy.minimum.accumulate 接口，调用 GPU 反而耗时增加。

图表11：因子计算环节 CPU、GPU 性能对比（明细）

因子类别	因子名称	是否替换为 矩阵运算	矩阵运算+GPU 时间 开销（秒）	矩阵运算+CPU 时间 开销（秒）	for 循环+CPU 时间 开销（秒）
价格全局特征	return_intraday		0.0011	0.0001	0.0001
	return_var		0.0016	0.0030	0.0030
	return_skew		0.0024	0.0152	0.0152
	return_kurtosis		0.0035	0.0128	0.0128
	tp_diff		0.0022	0.0038	0.0041
	price_trend_ratio		0.0016	0.0025	0.0025
	intraday_maxdrawdown	是	0.0163	0.0060	1.8093
	max_price_pos		0.0031	0.0069	0.0072
价格局部特征	return_improved	是	0.0001	0.0000	0.3610
	return_last_30min		0.0002	0.0000	0.0000
	return_var_last_30min		0.0002	0.0004	0.0004
	return_skew_last_30min		0.0011	0.0025	0.0025
	return_kurtosis_last_30min		0.0005	0.0022	0.0022
	return_downward_var	是	0.0073	0.0237	0.1135
	return_upward_var	是	0.0021	0.0236	0.1145
	return_downward_var_ratio	是	0.0018	0.0256	0.1132
	return_upward_var_ratio	是	0.0018	0.0253	0.1158
	return_downward_volatility_ratio	是	0.0021	0.0039	1.3882
成交量/额	cum_return_top_10		0.0049	0.0321	0.0335
	volume_ratio_per_30min_1		0.0002	0.0009	0.0009
	volume_ratio_per_30min_2		0.0001	0.0009	0.0008
	volume_ratio_per_30min_3		0.0001	0.0006	0.0005
	volume_ratio_per_30min_4		0.0001	0.0005	0.0005
	volume_ratio_per_30min_5		0.0001	0.0005	0.0005
	volume_ratio_per_30min_6		0.0001	0.0005	0.0005
	volume_ratio_per_30min_7		0.0001	0.0005	0.0005
	volume_ratio_per_30min_8		0.0001	0.0005	0.0005
	volume_variance_ratio		0.0021	0.0120	0.0112
	volume_open_30min_ratio		0.0002	0.0002	0.0002
	amount_out_order_avg_ratio		0.0020	0.0080	0.0079
	amount_net_top30_order_ratio		0.0095	0.0609	0.0608
成交关联价格	WVAD		0.0029	0.0140	0.0135
	Amihud		0.0019	0.0073	0.0072
	PVI		0.0008	0.0026	0.0026
	cum_return_top30_order		0.0076	0.0507	0.0917
	return_var_top33_volume	是	0.0045	0.0455	0.1200
	return_kurtosis_top33_volume	是	0.0077	0.0607	1.8881
	return_skew_top33_volume	是	0.0066	0.0574	1.8552
价量相关性	VP	是	0.0052	0.0325	0.3756
	VR	是	0.0044	0.0326	0.3741
	VR_1min_lead	是	0.0044	0.0326	0.3770
	VR_1min_lag	是	0.0044	0.0329	0.3723
	VP_last_30min	是	0.0023	0.0075	0.3234
	VR_last_30min	是	0.0024	0.0070	0.3242
	VR_lead_last_30min	是	0.0024	0.0067	0.3237
	VR_lag_last_30min	是	0.0024	0.0067	0.3228
	VP_top33_volume	是	0.0065	0.0595	0.4074
	VR_top33_volume	是	0.0069	0.0609	0.4060
	VR_lead_top33_volume	是	0.0069	0.0602	0.4298
	VR_lag_top33_volume	是	0.0069	0.0595	0.4290

注：使用 20230424 分钟 K 线数据

资料来源：华泰数智中台，华泰研究

图表12：因子计算环节 CPU、GPU 性能对比（汇总）

改进方法	因子 for 循环+CPU 矩阵运算+CPU 矩阵运算+GPU 矩阵运算时间开销 替换库函数时间开销 总体时间开销						
	数量	(秒)	(秒)	(秒)	缩短(倍)	缩短(倍)	短(倍)
替换库函数	50	12.63	0.91	0.16	13.84	5.86	81.17
使用矩阵运算+替换库函数	22	12.34	0.67	0.11	18.42	6.35	117.01

注：使用 20230424 分钟 K 线数据

资料来源：华泰数智中台，华泰研究

除了因子计算环节，我们还使用 cudf 库对 pandas 库进行替换，对数据读取、预处理环节也进行了优化，如下表。cudf 库仅有数据导出环节的 to_csv 函数性能更优（约 5 倍加速）。

不过测试发现，数据读取和预处理部分性能不佳的主要原因是数据量不足。单日分钟 K 线数据的维度约为 $5000 \times 242 \times 11$ （全 A 股*分钟*字段）。我们尝试构建虚拟数据集，发现当数据量扩大 4 倍后，GPU 性能超过 CPU。

图表13：数据读取、预处理及导出环节 CPU、GPU 性能对比

步骤	代码	GPU (3090) 时间开销(秒)	CPU (i9-10980XE) 时间开销(秒)	CPU/GPU
数据读取	read_csv	0.9390	0.8006	0.8526
数据预处理	pivot	3.2265	2.1685	0.6721
	cal_return	3.1892	0.0114	0.0036
数据导出	to_csv	0.0425	0.2282	5.3753

注：cal_return 为 $((df_close - df_close.shift(1)) / df_close).fillna(0)$ ；使用 20230424 分钟 K 线数据

资料来源：华泰数智中台，华泰研究

分钟线因子

本文测试 50 个分钟线构建的选股因子，按构建逻辑分为如下 5 类：

1. 价格全局特征（9 个因子）；
2. 价格局部特征（10 个因子）；
3. 成交量/额（12 个因子）；
4. 成交关联价格（7 个因子）；
5. 价量相关性（12 个因子）。

价格全局特征

价格全局特征类包括收益率/价格的分布特征，如方差、偏度、峰度等，以及最高价、最大回撤等。

图表14：价格全局特征类因子

因子名称	含义	计算公式	因子方向
return_intraday	日内收益率	$close_{15.00} - open_{9.30}$	-1
return_var	收益率方差	$\sum_{i=1}^n (r_i - \bar{r})^2$	-1
return_skew	收益率的偏度	$\frac{\sum_{i=1}^n (r_i - \bar{r})^3}{n(\sum_{i=1}^n (r_i - \bar{r})^2)^{1.5}}$	-1
return_kurtosis	收益率的峰度	$\frac{\sum_{i=1}^n (r_i - \bar{r})^4}{n(\sum_{i=1}^n (r_i - \bar{r})^2)^2} - 3$	-1
tp_diff	典型价格差值	$\frac{(high_{15.00} + low_{15.00} + close_{15.00})}{3} - \frac{(high_{9.30} + low_{9.30} + close_{9.30})}{3}$	-1
price_trend_ratio	股价变动趋势占比	$\frac{close_{15.00} - close_{9.30}}{\sum_{i=2}^n (close_i - close_{i-1}) }$	-1
intraday_maxdrawdown	日内最大回撤	$\max \left(\frac{close_i}{\min_{i' < i} (close_{i'})} - 1 \right)$	-1
max_price_pos	日内最高价出现时间	$argmax(close_i)$	-1
return_improved	改进日内涨跌幅	$close_{15.00} / close_{10.00}$	-1

资料来源：华泰研究

价格局部特征

价格局部特征类包括收益率/价格分布的局部特征，如收盘前半小时、价格上行或下行阶段收益率/价格方差、偏度、峰度等因子。

图表15：价格局部特征类因子

因子名称	含义	计算公式	因子方向
return_last_30min	收盘前半小时收益率	$close_{15.00} / close_{14.31}$	-1
return_var_last_30min	收盘前半小时的收益率方差	$\sum_{t=14.31}^{15.00} (close_t - \overline{close})^2$	-1
return_skew_last_30min	收盘前半小时的收益率偏度	$kurtosis(r_t), \quad 14:31 \leq t \leq 15:00$	-1
return_kurtosis_last_30min	收盘前半小时的收益率峰度	$skew(r_t), \quad 14:31 \leq t \leq 15:00$	-1
return_downward_var	下行收益率方差	$var(r_t), \quad r_t < 0$	-1
return_upward_var	上行收益率方差	$var(r_t), \quad r_t > 0$	-1
return_downward_var_ratio	下行收益率方差占比	$upward_var(r_t) / var(r_t)$	-1
return_upward_var_ratio	上行收益率方差占比	$downward_var(r_t) / var(r_t)$	-1
return_downward_volatility_ratio	下行收益率波动占比	$\frac{\sqrt{n} \sum_{i=1}^n (r_i \cdot I_{r_i > 0})^2}{\sum_{i=1}^n r_i^2}$	-1
cum_return_top_10	前 10% 最大累计涨幅	$\prod (1 + r_i \cdot I_{y_i \geq y_{90\%}})$	-1

资料来源：华泰研究

成交量/额

成交量/额类包括成交额、成交量、成交笔数所构建的因子，如成交量的方差比率、开盘前 30 分钟成交量占比、平均每单流出、大单净流入占比等。

图表16：成交量/额类因子

因子名称	含义	计算公式	因子方向
volume_ratio_per_30min	每半小时（共八组）成交量占总成交量比例	$\frac{\sum_{j=30(j-1)+1}^{30j} volume_i}{\sum_{i=1}^{240} volume_i}$; j 为分组 id	-1/1/1/1/ 1/1/1/-1
volume_variance_ratio	方差比率，衡量成交量 5 分钟和 10 分钟自相关性	$\frac{var(cusum_volume_{5k})}{var(cusum_volume_{10k})} / j$; $cusum_volume_{i,5k} = \sum_{l=k}^i volume_l, k=5, j=10$	1
volume_open_30min_ratio	前 30min 和午间休市后开盘 30min 交易量比值	$\frac{\sum_{t=9.30}^{10.00} volume_t}{\sum_{t=13.30}^{13.00} volume_t}$	-1
amount_out_order_avg_ratio	平均每笔流出金额占比	$\frac{\sum_{i=1}^N (amount_i \cdot I_{r_i < 0}) / \sum_{i=1}^N (items \cdot I_{r_i < 0})}{\sum_{i=1}^N amount_i / \sum_{i=1}^N items}$	1
amount_top30_order_net_ratio	大单净流入占比	$\frac{\sum_{i \in top30} amount_i \cdot I_{r_i > 0} - \sum_{i \in top30} amount_i \cdot I_{r_i < 0}}{\sum_{i \in top30} items}$; $top30 = \left(percentile \left(\frac{amount_i}{items_i} \right) > 70\% \right)$	-1

资料来源：华泰研究

成交关联价格

成交关联价格类包括收益率/价格关联对应范围的成交量所构建的因子，如大成交量对应的收益率方差、偏度、峰度，以及放量时收益率、大单涨幅等。

图表17：成交关联价格类因子

因子名称	含义	计算公式	因子方向
WVAD	威廉变异离散量	$\sum_{i=1}^N \frac{close_i - open_i}{high_i - low_i} \cdot volume_i$	-1
Amihud	收益率除以总交易金额	$mean \left(\frac{ r_i }{volume_i \cdot close_i} \right)$	1
PVI	正成交量指标，放量时收益率之和	$\sum_{i=1}^N r_i \cdot I_{volume_i > volume_{i-1}}$	-1
cum_return_top30_order	前 30% 的大单推动涨幅	$\prod_{i \in top30} (1 + r_i)$; $top30 = rank \left(\frac{amount_i}{items_i} \right)$	-1
return_var_top33_volume	前 1/3 成交量对应的收益率方差	$var_{i \in top_{\frac{1}{3}}} (1 + r_i)$; $top_{\frac{1}{3}} = rank(volume_i)$	-1
return_kurtosis_top33_volume	前 1/3 成交量对应的收益率峰度	$kurtosis_{i \in top_{\frac{1}{3}}} (1 + r_i)$; $top_{\frac{1}{3}} = rank(volume_i)$	-1
return_skew_top33_volume	前 1/3 成交量对应的收益率偏度	$skew_{i \in top_{\frac{1}{3}}} (1 + r_i)$; $top_{\frac{1}{3}} = rank(volume_i)$	-1

资料来源：华泰研究

量价相关性

量价相关性类包括成交量/额和收益率/价格的相关系数所构建的因子，如量价相关性、收益率与量相关性、领先/滞后收益率与量相关性等。

图表18：量价相关性类因子

因子名称	含义	计算公式	因子方向
VP	量价相关性	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}, y = \text{close}$	-1
VP_last_30min	收盘前半小时的 VP	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_{14.30 \leq t \leq 15.00},$ $y = \text{close}_{14.30 \leq t \leq 15.00}$	-1
VP_top33_volume	前 1/3 成交量对应的 VP	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_{\text{top}_{\frac{1}{3}}},$ $y = \text{close}_{i \in \text{volume}_{\text{top}_{\frac{1}{3}}}}$	-1
VR	收益率与量相关性	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_i, y = \text{return}_i$	-1
VR_1min_lead	领先一分钟 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_i, y = \text{return}_{i+1}$	1
VR_1min_lag	滞后一分钟 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_i, y = \text{return}_{i-1}$	-1
VR_last_30min	收盘前半小时的 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_t, 14.30 \leq t \leq 15.00,$ $y = \text{return}_{t+1}, 14.30 \leq t \leq 15.00$	-1
VR_lead_last_30min	收盘前半小时的 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_t, 14.30 \leq t \leq 15.00,$ $y = \text{return}_{t+1}, 14.30 \leq t \leq 15.00$	-1
VR_lag_last_30min	收盘前半小时的滞后 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_t, 14.30 \leq t \leq 15.00,$ $y = \text{return}_{t-1}, 14.30 \leq t \leq 15.00$	-1
VR_top33_volume	前 1/3 成交量对应的 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_i, i \in \text{volume}_{\text{top}_{\frac{1}{3}}},$ $y = \text{return}_i, i \in \text{volume}_{\text{top}_{\frac{1}{3}}}$	-1
VR_lead_top33_volume	前 1/3 成交量对应的领先 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_i, i \in \text{volume}_{\text{top}_{\frac{1}{3}}},$ $y = \text{return}_{i+1}, i \in \text{volume}_{\text{top}_{\frac{1}{3}}}$	-1
VR_lag_top33_volume	前 1/3 成交量对应的领先 VR	$\frac{\text{cov}(x,y)}{\text{std}(x)\cdot\text{std}(y)};$ $x = \text{volume}_i, i \in \text{volume}_{\text{top}_{\frac{1}{3}}},$ $y = \text{return}_{i-1}, i \in \text{volume}_{\text{top}_{\frac{1}{3}}}$	-1

资料来源：华泰研究

单因子测试方法与结果

单因子测试方法如下：

1. 回测区间：2013/01/31~2023/09/28；
2. 回测样本：中证全指成分股；
3. 因子频率（采样频率）：日频；
4. 预测收益率区间（调仓频率）：5日（周频）/10日（双周频）/20日（月频）；
5. 因子处理：过去 5/10/20 日取均值或累计值，5 倍中位数去极值，行业市值中性化，Z-score 标准化，缺失值填充；反向因子则乘以-1；
6. 交易费率：均不考虑。

价格全局特征

预测周频收益的单因子测试指标如下表。return_intraday、return_var、tp_diff、return_improved 因子表现出色。其中 return_intraday、tp_diff、return_improved 的本质是不同形式的日内反转因子，return_var 的本质是日内低波动因子。

图表19：价格全局特征因子测试指标（周频）

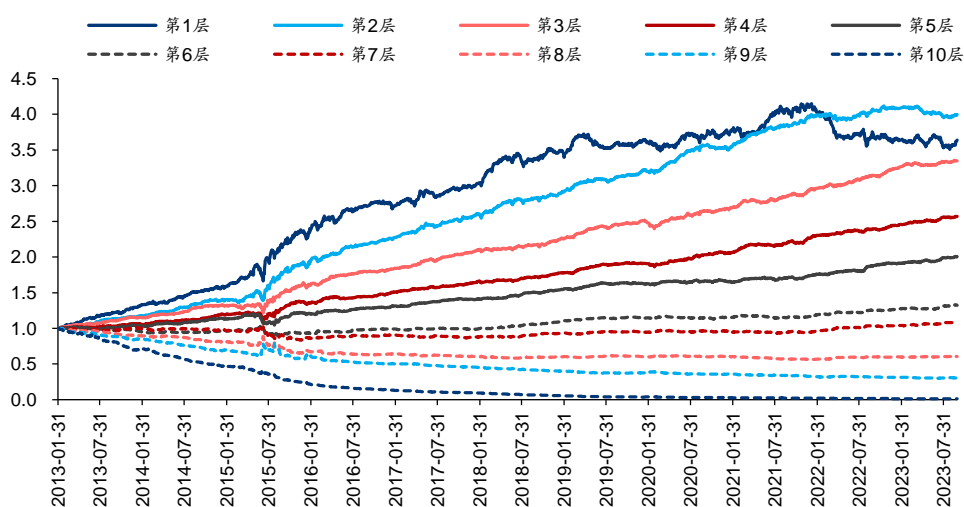
因子	RankIC 均值	RankICIR 胜率	RankIC 胜率	T 均值	Top 层年化超 额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信息 比率	Top 层年化 换手率
return_intraday	5.9%	0.63	75.1%	2.49	25.3%	-19.8%	22.6%	0.88	-0.71	1.54	1.19	88.7
return_var	5.9%	0.54	75.9%	2.05	24.5%	-14.3%	19.4%	1.38	-0.52	0.92	0.55	38.7
return_skewness	4.2%	0.98	83.6%	1.32	17.8%	-3.5%	10.6%	0.64	-0.13	1.62	0.55	86.6
return_kurtosis	4.2%	0.64	73.5%	1.46	26.0%	-2.0%	14.0%	0.92	-0.10	1.92	1.77	58.1
tp_diff	5.8%	0.65	75.3%	2.33	25.3%	-17.5%	21.4%	0.89	-0.70	1.64	1.09	81.0
price_trend_ratio	4.9%	0.59	72.4%	1.76	20.1%	-11.2%	15.7%	0.68	-0.43	1.39	0.70	88.9
intraday_maxdrawdown	5.1%	0.35	67.3%	1.38	22.3%	-2.1%	12.2%	1.42	-0.10	0.52	0.41	46.0
max_price_pos	1.2%	0.15	56.8%	0.92	19.8%	-0.6%	10.2%	1.07	-0.02	0.60	0.33	67.5
return_improved	6.3%	0.68	77.5%	2.99	28.5%	-25.9%	27.2%	0.97	-0.99	2.17	1.75	88.4

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+5 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

下图展示 return_improved 因子分层相对净值。Top 层 2022 年以来出现回撤，第 2 层 2023 年以来出现回撤，因子有效性整体下滑。

图表20：return_improved 因子分层相对净值



注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

同时展示双周频、月频单因子测试结果。

图表21：价格全局特征因子测试指标（双周频）

因子	RankIC 均值	RankICIR 胜率	RankIC 胜率	T 均值	Top 层年化超 额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信息 比率	Top 层年 化换手率
return_intraday	6.0%	0.64	75.1%	2.75	23.8%	-13.0%	18.4%	0.88	-0.46	1.30	1.10	44.4
return_var	6.2%	0.56	75.5%	2.35	22.4%	-7.5%	14.9%	1.26	-0.29	0.75	0.47	21.5
return_skewness	5.1%	1.12	86.9%	1.84	19.6%	-2.2%	10.9%	0.70	-0.14	1.75	0.91	43.0
return_kurtosis	4.7%	0.71	74.6%	1.77	25.3%	1.2%	12.1%	0.88	0.03	1.59	1.50	28.0
tp_diff	5.9%	0.68	76.1%	2.44	22.6%	-8.7%	15.6%	0.75	-0.33	1.18	0.65	40.5
price_trend_ratio	4.7%	0.59	73.2%	1.82	16.7%	-5.9%	11.3%	0.63	-0.24	1.12	0.42	44.3
intraday_maxdrawdown	5.9%	0.40	67.5%	1.83	20.9%	-1.1%	11.0%	1.28	-0.07	0.49	0.34	23.0
max_price_pos	0.2%	0.02	51.8%	0.65	18.2%	6.7%	5.8%	0.95	0.20	0.37	0.17	33.7
return_improved	6.2%	0.72	78.7%	3.09	24.7%	-15.3%	20.0%	0.82	-0.51	1.70	1.29	44.2

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+10 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

图表22：价格全局特征因子测试指标（月频）

因子	RankIC 均值	RankICIR 胜率	RankIC 胜率	T 均值	Top 层年化超 额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信息 比率	Top 层年 化换手率
return_intraday	7.8%	0.84	80.4%	3.51	26.4%	-6.7%	16.5%	0.87	-0.26	1.29	1.43	22.1
return_var	6.4%	0.60	74.4%	2.54	21.2%	-0.1%	10.7%	1.11	-0.03	0.60	0.46	12.3
return_skewness	6.5%	1.33	90.2%	2.39	21.2%	1.2%	10.0%	0.77	-0.01	1.66	1.53	21.1
return_kurtosis	4.9%	0.72	75.0%	1.86	23.4%	5.6%	8.9%	0.85	0.18	1.20	1.40	13.6
tp_diff	7.2%	0.84	79.5%	2.91	24.0%	-2.2%	13.1%	0.75	-0.11	1.10	0.79	20.0
price_trend_ratio	5.8%	0.72	76.4%	2.27	18.3%	-1.8%	10.1%	0.67	-0.13	1.04	0.63	22.0
intraday_maxdrawdown	6.7%	0.45	69.3%	2.22	20.7%	2.9%	8.9%	1.18	0.04	0.44	0.39	12.1
max_price_pos	0.2%	0.02	49.9%	0.68	17.7%	8.8%	4.4%	0.87	0.31	0.28	0.18	17.0
return_improved	7.2%	0.89	82.1%	3.65	26.5%	-7.9%	17.2%	0.83	-0.30	1.69	1.50	22.1

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+20 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

价格局部特征

价格局部特征因子预测周频收益的单因子测试指标如下表。return_last_30min、return_skewness_last_30min、return_upward_var、return_downward_var 因子表现出色。其中 return_last_30min 和 return_skewness_last_30min 的本质是尾盘反转因子，return_upward_var 和 return_downward_var 的本质是日内低波动因子的精细化刻画。

图表23：价格局部特征因子测试指标（周频）

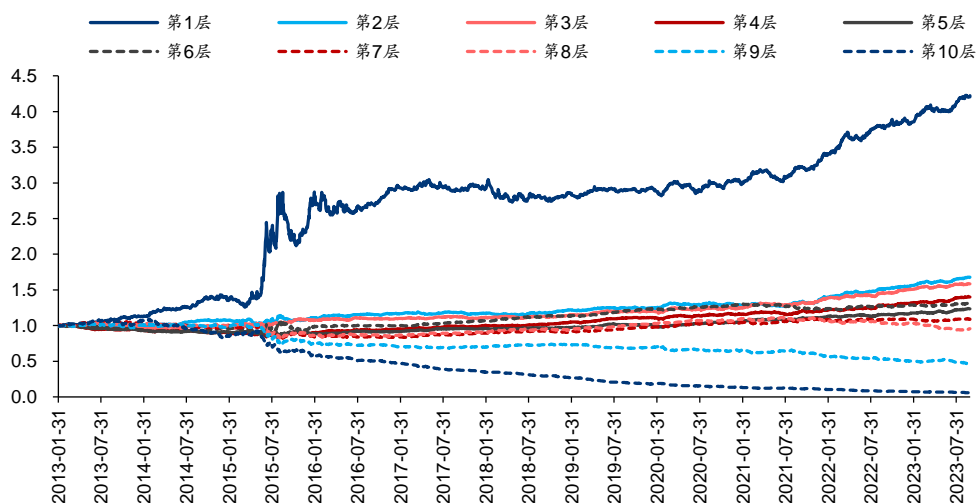
因子	RankIC 均值	RankICIR 胜率	RankIC 胜率	T 均值	Top 层年化超 额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信 息比率	Top 层年 化换手率
return_last_30min	4.3%	0.67	78.9%	2.31	24.4%	-19.7%	22.0%	0.84	-0.73	2.58	1.42	84.7
return_var_last_30min	3.4%	0.39	69.0%	1.55	22.0%	-9.5%	15.7%	1.21	-0.37	0.88	0.44	45.3
return_skewness_last_30min	3.3%	0.86	82.5%	1.56	22.9%	-5.5%	14.2%	0.78	-0.20	2.67	1.49	88.0
return_kurtosis_last_30min	1.3%	0.20	57.2%	0.65	17.5%	5.5%	6.0%	0.63	0.21	0.87	0.52	73.2
return_upward_var	6.7%	0.66	76.7%	2.33	31.1%	-14.3%	22.7%	1.61	-0.51	1.15	1.04	47.1
return_downward_var	5.3%	0.49	71.3%	1.86	28.7%	-9.0%	18.8%	1.52	-0.34	0.93	0.87	44.0
return_upward_var_ratio	4.4%	0.55	70.5%	1.50	25.3%	-1.3%	13.3%	0.91	-0.05	1.38	1.27	42.1
return_downward_var_ratio	1.0%	0.11	51.4%	0.39	20.6%	11.6%	4.5%	0.74	0.43	0.50	0.67	38.8
return_downward_volatility_ratio	5.5%	0.91	82.1%	1.81	20.1%	-10.3%	15.2%	0.71	-0.40	1.75	0.95	88.1
cum_return_top_10	5.5%	0.47	72.7%	1.86	22.1%	-13.7%	17.9%	1.28	-0.48	0.81	0.37	39.1

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+5 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

下图展示 return_downward_var 因子分层相对净值。Top 层表现较突出，2017 至 2020 年长期走平，2021 年后稳定上行；中间层区分度不高。

图表24：return_downward_var 因子分层相对净值



注：股票池：中证全指成分股；回溯区间：2013-01-31 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

图表25：价格局部特征因子测试指标（双周频）

因子	RankIC 均值	RankICIR	RankIC 胜率	T 均值	Top 层年化 超额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信 息比率	Top 层年 化换手率
return_last_30min	3.2%	0.52	74.2%	1.86	18.6%	-6.3%	12.5%	0.64	-0.23	1.64	0.56	42.1
return_var_last_30min	3.6%	0.41	69.8%	1.75	19.8%	-3.3%	11.5%	1.07	-0.15	0.69	0.31	22.8
return_skewness_last_30min	2.9%	0.73	78.0%	1.42	19.8%	0.9%	9.4%	0.69	0.03	1.84	0.93	44.0
return_kurtosis_last_30min	1.5%	0.19	56.5%	0.70	17.7%	8.6%	4.6%	0.59	0.31	0.55	0.29	33.4
return_upward_var	7.2%	0.70	77.2%	2.81	28.0%	-9.3%	18.6%	1.46	-0.35	1.02	0.91	25.0
return_downward_var	5.8%	0.53	71.0%	2.23	26.2%	-4.1%	15.1%	1.40	-0.15	0.79	0.78	23.1
return_upward_var_ratio	4.9%	0.58	71.4%	1.83	24.5%	0.8%	11.9%	0.89	-0.02	1.27	1.14	20.0
return_downward_var_ratio	1.4%	0.14	53.2%	0.52	21.3%	12.8%	4.3%	0.75	0.50	0.41	0.67	17.7
return_downward_volatility_ratio	6.2%	1.00	84.5%	2.20	20.7%	-5.9%	13.3%	0.75	-0.27	1.56	1.09	43.6
cum_return_top_10	5.6%	0.47	72.3%	2.12	20.3%	-6.3%	13.3%	1.19	-0.25	0.66	0.33	21.2

注：股票池：中证全指成分股；回溯区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+10 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

图表26：价格局部特征因子测试指标（月频）

因子	RankIC 均值	RankICIR	RankIC 胜率	T 均值	Top 层年化 超额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信 息比率	Top 层年 化换手率
return_last_30min	2.5%	0.44	67.3%	1.49	15.5%	2.4%	6.5%	0.47	0.04	0.88	-0.17	20.8
return_var_last_30min	3.8%	0.46	70.7%	1.86	19.2%	3.0%	8.1%	0.96	0.04	0.57	0.30	12.3
return_skewness_last_30min	2.9%	0.73	76.0%	1.43	19.4%	6.0%	6.7%	0.64	0.15	1.22	0.72	22.0
return_kurtosis_last_30min	1.6%	0.17	55.9%	0.55	16.8%	12.0%	2.4%	0.58	0.41	0.33	0.24	15.1
return_upward_var	7.6%	0.73	77.2%	3.09	25.8%	-1.6%	13.7%	1.28	-0.11	0.88	0.93	14.0
return_downward_var	6.1%	0.55	71.2%	2.40	24.0%	1.8%	11.1%	1.18	0.02	0.66	0.71	12.8
return_upward_var_ratio	5.3%	0.61	71.4%	1.99	22.6%	4.6%	9.0%	0.84	0.10	1.03	0.99	9.9
return_downward_var_ratio	1.2%	0.12	53.1%	0.42	19.9%	14.5%	2.7%	0.70	0.50	0.32	0.52	8.3
return_downward_volatility_ratio	7.9%	1.19	87.3%	2.85	21.5%	-2.4%	12.0%	0.81	-0.15	1.52	1.56	21.4
cum_return_top_10	5.9%	0.50	71.1%	2.34	19.8%	0.4%	9.7%	1.08	0.01	0.51	0.37	11.9

注：股票池：中证全指成分股；回溯区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+20 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

成交量/额

成交量/额因子预测周频收益的单因子测试指标如下表。volume_open_30min_ratio、amount_out_order_avg_ratio 因子表现出色。前者给予开盘成交不活跃的股票风险溢价，由于市场情绪一般在开盘释放，该因子可能反映理性交易者占比。后者给予单笔流出金额较大的股票风险溢价，可能对应快速下跌后的反转或者主力的操纵行为。

图表27：成交量/额因子测试指标（周频）

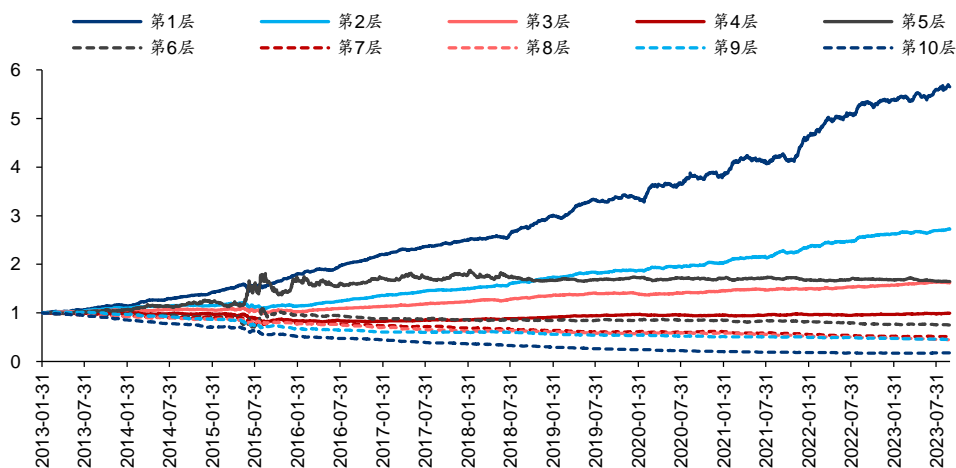
因子	RankIC 均值	RankICIR 胜率	RankIC 胜率	T 均值	Top 层年化 超额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信 息比率	Top 层年 化换手率
volume_ratio_per_30min_1	4.6%	0.78	80.2%	1.28	18.8%	-0.2%	9.5%	0.74	-0.06	0.98	0.71	75.4
volume_ratio_per_30min_2	0.7%	0.17	56.8%	0.40	16.8%	7.9%	4.4%	0.57	0.24	0.83	0.18	88.1
volume_ratio_per_30min_3	2.8%	0.74	78.0%	0.91	19.3%	2.2%	8.6%	0.69	0.00	1.68	0.88	88.7
volume_ratio_per_30min_4	3.5%	0.86	82.9%	0.98	19.3%	-0.4%	9.9%	0.69	-0.07	1.69	0.83	88.1
volume_ratio_per_30min_5	3.4%	0.80	80.9%	1.03	18.9%	-0.2%	9.5%	0.69	-0.03	1.49	0.75	86.9
volume_ratio_per_30min_6	3.1%	0.75	79.1%	0.78	17.8%	3.5%	7.2%	0.67	0.04	1.22	0.65	87.1
volume_ratio_per_30min_7	3.0%	0.69	77.3%	0.59	15.6%	5.2%	5.2%	0.56	0.12	0.76	-0.15	85.9
volume_ratio_per_30min_8	0.3%	0.05	51.7%	0.79	18.3%	3.1%	7.6%	0.58	0.11	0.77	0.22	78.6
volume_variance_ratio	5.2%	0.68	75.2%	1.36	26.5%	-1.6%	14.0%	1.01	-0.09	1.46	1.80	72.2
volume_open_30min_ratio	5.6%	1.12	88.1%	1.80	25.5%	-3.3%	14.4%	0.98	-0.16	1.75	2.65	80.8
amount_out_order_avg_ratio	4.8%	1.02	85.3%	1.91	32.2%	-1.1%	16.6%	1.25	-0.05	2.55	3.24	76.1
amount_top30_order_net_ratio	3.8%	0.65	74.5%	1.54	21.6%	-6.7%	14.2%	0.76	-0.28	1.68	1.18	81.2

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+5 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

下图展示 amount_out_order_avg_ratio 因子分层相对净值。Top 层表现较突出，并且长期收益稳定。

图表28：amount_out_order_avg_ratio 因子分层相对净值



注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

图表29：成交量/额因子测试指标（双周频）

因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
volume_ratio_per_30min_1	6.1%	0.96	84.6%	1.87	19.3%	-0.3%	9.8%	0.76	-0.02	0.89	0.77	36.0
volume_ratio_per_30min_2	0.4%	0.10	53.8%	0.39	15.2%	9.9%	2.6%	0.51	0.36	0.43	-0.18	43.7
volume_ratio_per_30min_3	3.4%	0.91	81.5%	1.30	20.7%	3.5%	8.6%	0.71	0.06	1.60	1.10	44.1
volume_ratio_per_30min_4	4.2%	1.09	86.8%	1.42	20.1%	0.9%	9.6%	0.69	-0.03	1.64	0.89	43.6
volume_ratio_per_30min_5	4.4%	1.01	85.3%	1.57	20.2%	0.0%	10.1%	0.72	-0.07	1.67	1.01	42.6
volume_ratio_per_30min_6	4.0%	0.98	84.5%	1.29	19.2%	2.9%	8.1%	0.69	0.05	1.17	0.77	42.8
volume_ratio_per_30min_7	3.9%	0.83	81.7%	0.98	16.9%	5.5%	5.7%	0.62	0.15	0.72	0.20	42.1
volume_ratio_per_30min_8	0.0%	0.01	50.2%	0.72	16.1%	6.3%	4.9%	0.52	0.21	0.53	-0.02	38.2
volume_variance_ratio	6.3%	0.72	76.6%	1.70	23.0%	0.6%	11.2%	0.92	-0.03	1.12	1.26	33.4
volume_open_30min_ratio	7.1%	1.30	91.3%	2.48	25.8%	-3.1%	14.5%	0.96	-0.16	1.69	2.59	39.3
amount_out_order_avg_ratio	6.3%	1.13	87.4%	2.49	29.6%	-1.0%	15.3%	1.12	-0.11	2.33	2.61	36.8
amount_top30_order_net_ratio	3.9%	0.70	76.8%	1.61	18.0%	-2.2%	10.1%	0.62	-0.13	1.24	0.29	39.7

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+10 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

图表30：成交量/额因子测试指标（月频）

因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
volume_ratio_per_30min_1	7.1%	0.97	83.7%	2.27	18.4%	1.8%	8.3%	0.74	-0.01	0.80	0.65	17.2
volume_ratio_per_30min_2	0.2%	0.04	50.6%	0.41	16.0%	11.2%	2.4%	0.55	0.38	0.48	0.17	21.4
volume_ratio_per_30min_3	3.7%	0.96	84.0%	1.46	19.8%	6.3%	6.7%	0.67	0.15	1.33	0.95	21.8
volume_ratio_per_30min_4	4.9%	1.16	88.4%	1.78	19.3%	3.5%	7.9%	0.69	0.06	1.34	0.96	21.4
volume_ratio_per_30min_5	5.5%	1.15	86.1%	2.09	20.8%	3.0%	8.9%	0.77	0.04	1.36	1.31	20.8
volume_ratio_per_30min_6	4.9%	1.02	85.9%	1.71	19.0%	4.3%	7.3%	0.70	0.08	1.03	0.83	20.9
volume_ratio_per_30min_7	4.8%	0.83	81.1%	1.39	17.2%	6.0%	5.6%	0.65	0.16	0.66	0.33	20.4
volume_ratio_per_30min_8	-0.1%	-0.01	50.1%	0.68	15.6%	8.4%	3.6%	0.49	0.34	0.28	-0.11	18.5
volume_variance_ratio	7.5%	0.79	77.2%	2.18	20.9%	2.2%	9.4%	0.82	0.01	0.91	0.83	15.4
volume_open_30min_ratio	8.1%	1.30	89.2%	2.88	24.1%	1.0%	11.5%	0.91	-0.06	1.33	2.52	19.1
amount_out_order_avg_ratio	7.4%	1.11	85.9%	2.91	27.2%	2.1%	12.5%	1.09	0.02	1.75	2.38	17.7
amount_top30_order_net_ratio	4.3%	0.77	78.8%	1.74	17.0%	1.8%	7.6%	0.61	-0.01	0.96	0.28	19.4

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+20 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

成交关联价格

成交关联价格因子预测周频收益的单因子测试指标如下表。cum_return_top30_order 因子表现出色。该因子的本质是反转因子的精细化刻画，大单推动的涨幅更具信息量。

图表31：成交关联价格因子测试指标（周频）

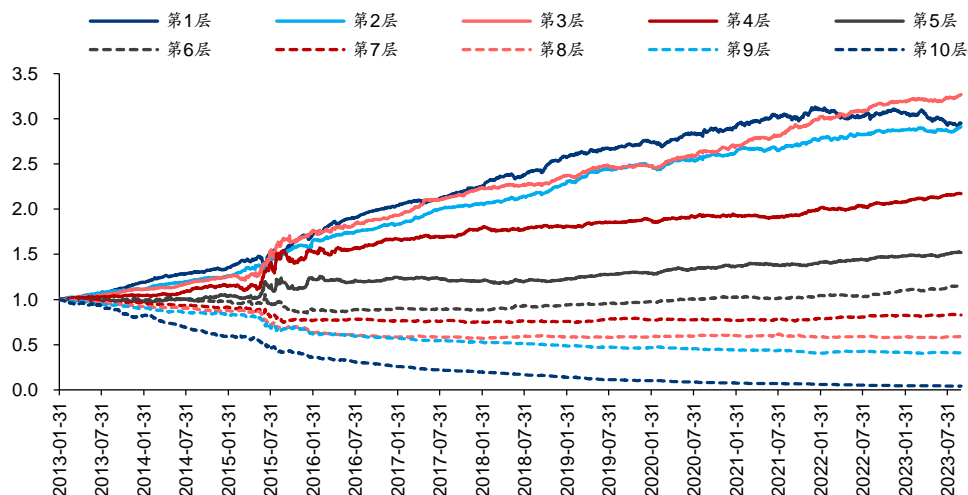
因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
WVAD	2.3%	0.27	57.9%	0.71	4.1%	-6.5%	5.3%	0.07	-0.27	0.40	-1.91	75.2
Amihud	6.4%	0.75	78.2%	1.85	32.0%	2.9%	14.5%	1.18	0.09	0.99	2.62	38.8
PVI	4.1%	0.47	68.4%	1.46	17.8%	-7.5%	12.6%	0.59	-0.31	1.04	0.37	86.4
cum_return_top30_order	5.2%	0.97	85.4%	2.33	27.8%	-13.7%	20.8%	0.95	-0.51	2.46	2.67	85.9
return_var_top33_volume	6.8%	0.63	78.0%	2.35	26.5%	-17.6%	22.0%	1.44	-0.60	1.03	0.68	42.5
return_kurtosis_top33_volume	2.8%	0.52	69.6%	1.03	22.5%	3.6%	9.5%	0.81	0.12	1.61	1.44	70.0
return_skewness_top33_volume	3.5%	0.95	83.1%	1.15	18.7%	0.1%	9.3%	0.69	-0.02	1.66	0.84	86.4

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+5 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

下图展示 cum_return_top30_order 因子分层相对净值。Top 层 2022 年以来出现回撤，第 3 层表现较好。

图表32: cum_return_top30_order 因子分层相对净值



注：股票池：中证全指成分股；回溯区间：2013-01-31 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

图表33：成交关联价格因子测试指标（双周频）

因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
WVAD	1.2%	0.16	54.3%	0.30	1.3%	-1.3%	1.3%	0.00	-0.07	0.09	-2.32	37.6
Amihud	7.3%	0.84	79.1%	2.16	29.4%	6.5%	11.4%	1.07	0.28	0.79	2.14	19.9
PVI	3.4%	0.40	65.6%	1.19	14.3%	-0.9%	7.6%	0.43	0.01	0.55	-0.43	43.3
cum_return_top30_order	6.0%	1.01	86.7%	2.73	25.8%	-8.4%	17.1%	0.93	-0.31	2.00	2.74	42.3
return_var_top33_volume	7.2%	0.65	78.0%	2.77	24.0%	-10.9%	17.5%	1.32	-0.40	0.88	0.60	23.2
return_kurtosis_top33_volume	3.0%	0.54	68.2%	1.25	22.3%	5.3%	8.5%	0.79	0.17	1.38	1.30	33.0
return_skewness_top33_volume	4.3%	1.01	84.3%	1.62	20.4%	0.2%	10.1%	0.72	-0.03	1.75	1.04	42.7

注：股票池：中证全指成分股；回溯区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+10 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

图表34：成交关联价格因子测试指标（月频）

因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
WVAD	1.6%	0.22	55.6%	0.46	5.6%	0.6%	2.5%	0.16	0.06	0.16	-1.68	18.9
Amihud	7.9%	0.93	81.3%	2.40	25.8%	9.9%	8.0%	0.93	0.35	0.64	1.57	10.5
PVI	3.9%	0.45	66.3%	1.36	16.0%	2.9%	6.6%	0.53	0.16	0.49	0.09	21.8
cum_return_top30_order	7.1%	1.11	87.6%	3.21	24.6%	-3.3%	14.0%	0.90	-0.18	1.59	2.46	20.6
return_var_top33_volume	7.4%	0.68	76.9%	3.03	22.5%	-3.5%	13.0%	1.17	-0.16	0.74	0.60	13.1
return_kurtosis_top33_volume	2.9%	0.50	68.9%	1.21	20.8%	8.8%	6.0%	0.76	0.29	1.00	1.19	15.5
return_skewness_top33_volume	5.3%	1.20	86.6%	2.05	20.6%	3.3%	8.7%	0.76	0.09	1.56	1.40	21.1

注：股票池：中证全指成分股；回溯区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+20 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

价量相关性

价量相关性因子预测周频收益的单因子测试指标如下表。VP、VP_top33_volume、VR_1min_lag 因子表现出色。这些因子的本质都是捕捉量价背离，即缩量上涨或放量下跌。

图表35：价量相关性因子测试指标（周频）

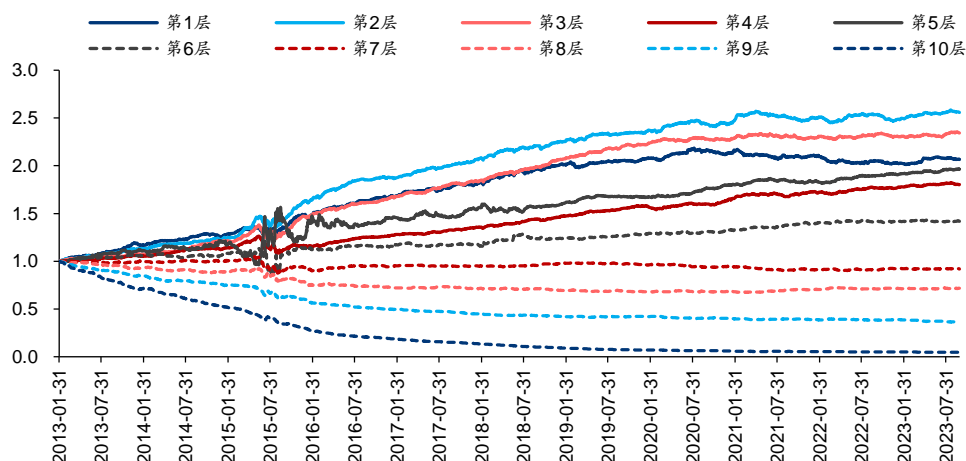
因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
VP	5.3%	0.92	82.9%	1.87	22.0%	-12.5%	17.2%	0.78	-0.50	2.20	1.23	86.8
VP_last_30min	2.7%	0.56	69.8%	1.31	19.7%	-5.4%	12.5%	0.67	-0.22	1.93	0.83	88.6
VP_top33_volume	5.6%	1.02	84.9%	1.98	23.6%	-12.9%	18.2%	0.83	-0.51	2.64	1.76	87.6
VR	3.9%	0.58	71.7%	1.21	16.5%	-5.2%	10.8%	0.55	-0.23	1.26	0.01	86.3
VR_1min_lead	0.5%	0.16	57.2%	0.25	12.3%	8.6%	1.9%	0.44	0.27	0.43	-0.86	87.6
VR_1min_lag	5.7%	0.89	81.8%	2.03	18.5%	-16.9%	17.7%	0.64	-0.62	1.76	0.55	89.3
VR_last_30min	2.5%	0.58	72.0%	1.08	21.9%	1.9%	10.0%	0.78	0.04	1.94	1.64	88.3
VR_lead_last_30min	0.3%	0.12	55.3%	0.11	15.4%	13.2%	1.1%	0.56	0.46	0.32	0.00	89.7
VR_lag_last_30min	2.3%	0.67	75.1%	1.13	20.6%	-0.9%	10.8%	0.75	-0.02	2.23	1.48	90.2
VR_top33_volume	3.6%	0.60	72.3%	1.12	17.6%	-3.3%	10.4%	0.62	-0.16	1.39	0.39	87.3
VR_lead_top33_volume	3.7%	0.61	72.7%	1.16	17.7%	-3.8%	10.7%	0.59	-0.17	1.32	0.24	87.3
VR_lag_top33_volume	3.1%	0.56	71.2%	0.97	17.1%	-0.3%	8.7%	0.58	-0.03	1.14	0.18	88.2

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+5 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

下图展示 VP_top33_volume 因子分层相对净值。Top 层表现不突出，第 2、3 层相对较好，但 2021 年以来收益走平。

图表36：VP_top33_volume 因子分层相对净值



注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

图表37：价量相关性因子测试指标（双周频）

因子	RankIC	RankICIR	RankIC	T 均值	Top 层年化	Bottom 层年	对冲组合年	Top 层夏	Bottom 层	对冲组合	Top 层信	Top 层年
	均值		胜率		超额收益	化超额收益	化超额收益	普比率	夏普比率	夏普比率	息比率	化换手率
VP	4.6%	0.80	79.4%	1.57	17.0%	-2.6%	9.8%	0.64	-0.11	1.34	0.43	43.3
VP_last_30min	1.3%	0.28	60.1%	0.77	14.3%	4.2%	5.0%	0.50	0.13	0.87	-0.34	43.9
VP_top33_volume	5.3%	0.95	83.6%	1.82	19.1%	-3.8%	11.5%	0.71	-0.14	1.78	1.01	43.6
VR	3.4%	0.54	70.3%	1.05	13.8%	1.5%	6.2%	0.51	0.04	0.79	-0.33	42.7
VR_1min_lead	0.9%	0.26	60.2%	0.32	12.9%	9.4%	1.8%	0.42	0.35	0.18	-1.06	43.3
VR_1min_lag	5.8%	0.94	83.8%	2.10	17.0%	-8.3%	12.6%	0.63	-0.34	1.35	0.40	44.4
VR_last_30min	1.8%	0.43	66.1%	0.84	18.4%	7.4%	5.5%	0.65	0.30	0.93	0.68	44.1
VR_lead_last_30min	0.6%	0.26	60.7%	0.23	16.4%	13.7%	1.4%	0.62	0.54	0.26	0.47	44.6
VR_lag_last_30min	1.7%	0.51	70.4%	0.84	17.6%	5.9%	5.9%	0.62	0.22	1.18	0.54	44.9
VR_top33_volume	3.0%	0.54	70.6%	0.95	14.5%	2.8%	5.9%	0.53	0.09	0.85	-0.22	43.2
VR_lead_top33_volume	3.1%	0.55	71.2%	0.98	14.5%	2.3%	6.1%	0.55	0.05	0.94	-0.13	43.2
VR_lag_top33_volume	2.6%	0.49	69.0%	0.71	14.4%	5.8%	4.3%	0.54	0.15	0.75	-0.17	43.8

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+10 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

图表38： 价量相关性因子测试指标（月频）

因子	RankIC 均值	RankICIR 0.90	RankIC 胜率	T 均值	Top 层年化 超额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信 息比率	Top 层年 化换手率
VP	5.0%	0.90	81.9%	1.69	16.5%	2.3%	7.1%	0.63	0.04	1.13	0.39	21.7
VP_last_30min	0.8%	0.16	54.8%	0.55	13.0%	7.8%	2.6%	0.41	0.23	0.46	-0.93	21.6
VP_top33_volume	6.0%	1.11	87.1%	2.11	19.3%	0.5%	9.4%	0.69	-0.05	1.52	0.85	21.7
VR	3.9%	0.63	73.0%	1.21	15.3%	5.1%	5.1%	0.57	0.10	0.82	0.06	21.4
VR_1min_lead	0.9%	0.25	58.4%	0.28	13.1%	11.0%	1.1%	0.45	0.40	0.11	-0.91	21.4
VR_1min_lag	7.5%	1.20	87.8%	2.75	18.7%	-4.3%	11.5%	0.72	-0.20	1.29	0.94	22.0
VR_last_30min	2.0%	0.45	65.5%	0.85	18.3%	9.8%	4.2%	0.62	0.32	0.74	0.52	21.9
VR_lead_last_30min	1.0%	0.48	68.9%	0.45	16.6%	13.2%	1.7%	0.60	0.44	0.55	0.35	22.2
VR_lag_last_30min	1.7%	0.51	69.6%	0.82	16.2%	8.6%	3.8%	0.53	0.27	0.78	-0.25	22.5
VR_top33_volume	3.6%	0.65	73.1%	1.10	15.6%	6.2%	4.7%	0.56	0.17	0.77	-0.07	21.6
VR_lead_top33_volume	3.7%	0.65	73.8%	1.12	15.7%	6.0%	4.8%	0.57	0.16	0.79	0.02	21.5
VR_lag_top33_volume	3.0%	0.55	71.6%	0.83	15.3%	8.0%	3.7%	0.56	0.23	0.62	-0.01	21.8

注：股票池：中证全指成分股；回测区间：2013-01-31 至 2023-09-28；因子进行行业市值中性；预测收益使用 T 至 T+20 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

基于高频因子的指增策略

最大化 ICIR 法合成因子

基于前述 50 个高频因子,采用最大化 ICIR 法进行因子合成。每个交易日选取 T-260 至 T-20 日因子未来 20 日 RankIC 值,并结合 Ledoit-Wolf 压缩估计求解因子权重,最终得到分钟线合成因子。

分钟线合成因子 IC 测试、回归测试、分层测试指标及分层相对净值如下图表所示。以未来 5 日收益为预测目标,因子 RankICIR (未年化) 1.50, 对冲组合夏普比率 3.59, Top 层信息比率 3.81。最大化 ICIR 法合成后,因子单调性提升,但仍然表现出空头端筛选效果明显优于多头端的特点,这也是大多数高频因子的典型特征。

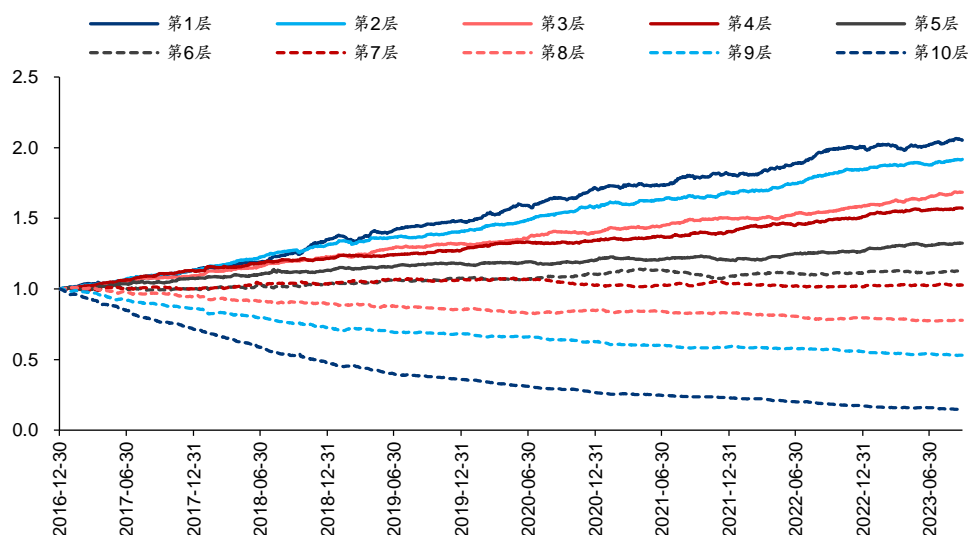
图表39：最大化 ICIR 法合成因子测试指标（预测未来 5 日收益）

因子	RankIC 均值	RankICIR 胜率	RankIC 胜率	T 均值	Top 层年化超 额收益	Bottom 层年 化超额收益	对冲组合年 化超额收益	Top 层夏 普比率	Bottom 层 夏普比率	对冲组合 夏普比率	Top 层信 息比率	Top 层年 化换手率
分钟线合成因子	7.0%	1.50	93.5%	3.26	14.7%	-25.6%	20.2%	0.69	-1.20	3.59	3.81	42.4

注：股票池：中证全指成分股；回测区间：2016-12-30 至 2023-09-28；因子进行行业市值中性；预测收益统一使用 T 至 T+5 日收盘价区间收益率

资料来源：Wind，华泰数智中台，华泰研究

图表40：分钟线合成因子分层相对净值



注：股票池：中证全指成分股；回测区间：2016-12-30 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

指数增强组合

华泰金工研报《神经网络多频率因子挖掘模型》(2023-05-11) 使用 GRU 网络学习日 K 线和 15 分钟 K 线数据,得到神经网络多频率因子。原始因子和上述分钟线合成因子相关系数为 0.248。将两者截面标准化后,采用 9:1 静态比例线性结合,分别构建中证 500 和中证 1000 指数增强策略。其余细节如下表所示。

图表41：指数增强组合构建方法

步骤	参数	参数值
股票池和回测	股票池	中证全指成分股
区间	回测区间	2016/12/30-2023/09/28
构建组合	基准	中证 500 指数/中证 1000 指数
	优化目标	最大化预期收益
	组合仓位	1
	个股权重下限	0
	个股偏离权重约束	[-1%, 1%]
	行业偏离权重约束	[-2%, 2%]
	风格偏离标准差约束	[-0.2, 0.2]
	风格因子	对数流通市值（预处理：5 倍 MAD 缩尾，zscore 标准化）
	调仓周期	每 5 个交易日
	单次调仓单边换手率上限	25%
	成分股权重约束	80%
回测	单边费率	0.002
	交易价格	vwap
	特殊处理	停牌不买入/卖出；一字板涨停不买入；一字板跌停不卖出；其余股票重新分配权重

资料来源：华泰研究

中证 500 指数增强回测绩效和超额净值如下所示。结合分钟线合成因子后，各项指标均有提升，2023 年上半年回撤大幅降低。信息比率从 2.78 提升至 3.01，超额收益 Calmar 比率从 1.37 提升至 2.07。

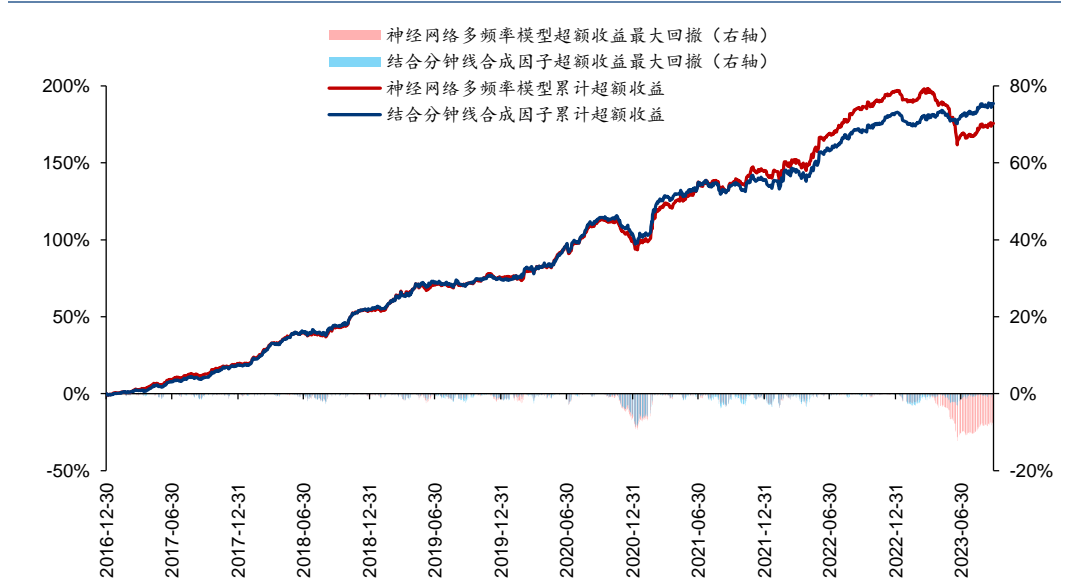
图表42：中证 500 指数增强组合回测绩效

	年化收益	年化波动	夏普比率	最大回撤	Calmar 比率	年化超额收益率	年化跟踪误差	信息比率	超额收益最大回撤	超额收益 Calmar 比率	相对基准月胜率	年化双边换手率
神经网络多频率模型	15.2%	21.4%	0.71	31.0%	0.49	16.8%	6.1%	2.78	12.3%	1.37	72.8%	25.2
结合分钟线合成因子	16.0%	21.4%	0.75	30.4%	0.53	17.7%	5.9%	3.01	8.6%	2.07	76.5%	25.2

注：回测区间：2016-12-30 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

图表43：中证 500 指数增强组合超额净值



注：回测区间：2016-12-30 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

中证 1000 指数增强回测绩效和超额净值如下所示。结合分钟线合成因子后，各项指标同样均有提升，并且 2023 年上半年回撤大幅降低。信息比率从 3.70 提升至 3.87，超额收益 Calmar 比率从 2.41 提升至 3.96。

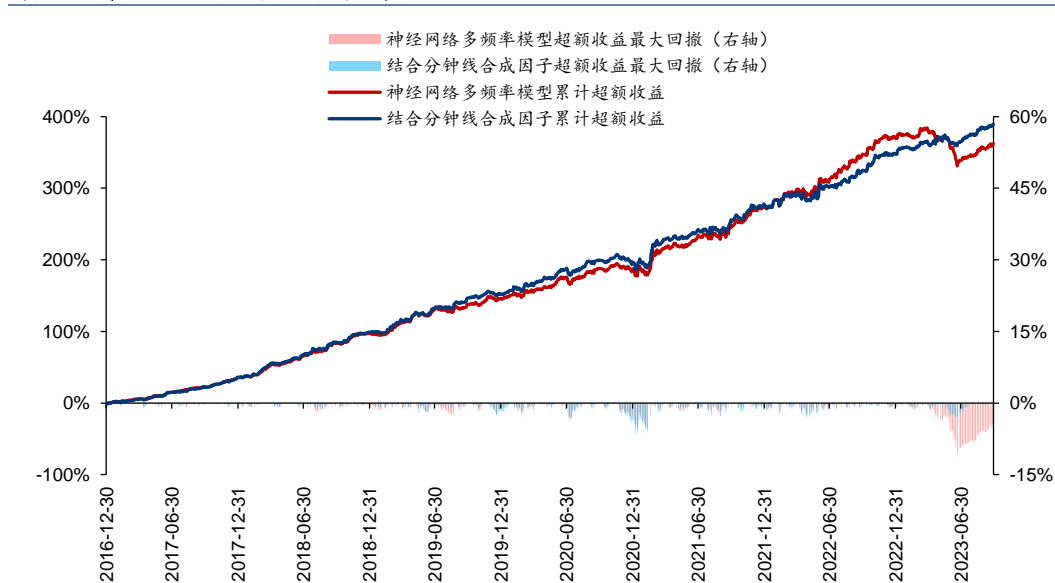
图表44：中证 1000 指数增强组合回测绩效

	年化收益 率	年化波动 率	夏普比率	最大回撤	Calmar 比率	年化超额 收益率	年化跟踪 误差	信息比率	超额收益 最大回撤	超额收益 Calmar 比率	相对基准 月胜率	年化双边 换手率
神经网络多频率模型	20.3%	23.4%	0.87	31.8%	0.64	26.5%	7.2%	3.70	11.0%	2.41	77.8%	25.1
结合分钟线合成因子	21.2%	23.8%	0.89	33.7%	0.63	27.6%	7.1%	3.87	7.0%	3.96	84.0%	25.3

注：回测区间：2016-12-30 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

图表45：中证 1000 指数增强组合超额净值



注：回测区间：2016-12-30 至 2023-09-28

资料来源：Wind，华泰数智中台，华泰研究

总结和讨论

本文使用 NVIDIA RAPIDS 对高频因子计算进行 GPU 加速。量化因子计算场景下，RAPIDS 的直接助力是用 CuPy、cuDF 的 GPU 运算替代 NumPy、Pandas 的 CPU 运算。在 NVIDIA GeForce RTX 3090 和 Intel Core i9-10980XE 测试环境下计算分钟线因子，CuPy 和 cuDF 替换库函数的提速效果约为 6 倍，若同时将 for 循环替换为矩阵运算，最终提速超 100 倍。预计 RTX 4090 和 A800 提速更显著。最大化 ICIR 法合成因子，并与华泰金工神经网络多频率因子静态加权，构建中证 1000 指数增强组合，以 2016 年末至 2023 年 9 月末为回测区间，信息比率从 3.70 提升至 3.87，超额收益 Calmar 比率从 2.41 提升至 3.96。

RAPIDS 的重要特性之一是将基于 CUDA 底层代码的优化以 Python 高级语言的形式体现。常用 API 如 CuPy（对标 NumPy）、cuDF（对标 Pandas）、cuML（对标 scikit-learn）等。由于 API 语法几乎相同，仅需要轻量级代码修改，即可实现 CPU 运算到 GPU 运算的迁移。针对全部因子使用 CuPy 和 cuDF 替换原函数；针对部分因子使用矩阵运算替换 for 循环。结果表明：若不引入矩阵运算，单独替换库函数反而增加时间开销。若两步同时进行，替换库函数带来的性能提升约在 6 倍（RTX 3090），矩阵运算带来的性能提升约在 18 倍。GPU 性能同型号和数据量相关，单次运算数据量越大，加速效果越显著。

本文测试 5 类共计 50 个分钟线选股因子。以下展示测试效果较好的因子及投资逻辑。

1. 价格全局特征类：return_intraday、tp_diff、return_improved 的本质是不同形式的日内反转因子，return_var 的本质是日内低波动因子。
2. 价格局部特征类：return_last_30min 和 return_skewness_last_30min 的本质是尾盘反转因子，return_upward_var 和 return_downward_var 的本质是日内低波动因子的精细化刻画。
3. 成交量/额类：volume_open_30min_ratio 给予开盘成交不活跃的股票风险溢价，市场情绪一般在开盘释放，该因子或反映理性交易者占比。amount_out_order_avg_ratio 给予单笔流出金额较大的股票风险溢价，或对应快速下跌后的反转或者主力的操纵行为。
4. 成交关联价格类：cum_return_top30_order 的本质是反转因子的精细化刻画，大单推动的涨幅更具信息量。
5. 价量相关性类：VP、VP_top33_volume、VR_1min_lag 的本质都是捕捉量价背离，即缩量上涨或放量下跌。

进一步围绕分钟线因子构建选股策略。采用最大化 ICIR 法对前述 50 个因子进行合成，以未来 5 日收益为预测目标，合成因子 RankICIR（未年化）1.50，对冲组合夏普比率 3.59，Top 层信息比率 3.81。该合成因子和神经网络多频率因子相关系数为 0.25，两者截面标准化后静态融合，构建指数增强策略。结果显示，结合分钟线合成因子后，策略各项指标均有提升，尤其体现在回撤控制。中证 500 增强信息比率从 2.78 提升至 3.01，超额收益 Calmar 比率从 1.37 提升至 2.07，中证 1000 增强信息比率从 3.70 提升至 3.87，超额收益 Calmar 比率从 2.41 提升至 3.96。

本文有如下未尽之处：

1. 本文采集的数据均基于 RTX 3090 显卡，对目前行业里主流的 A800、RTX 4090 等显卡未进行测试。
2. RAPIDS 除 CuPy 和 cuDF 外，cuML（对标 scikit-learn）在量化投资行业可能也有实用意义，如使用 cuML 进行遗传规划因子挖掘，值得后续探索。
3. 测试的分钟线因子具有非线性特征，部分因子 Top 层不如第 2、3 层，进行非线性变换后或能提升因子表现；
4. 从分钟线因子和神经网络因子结合的方式上看，本文先对分钟线因子进行最大化 ICIR 加权，再与神经网络因子进行静态合成，若采用动态或 AI 方法或有进一步提升空间。

来自 NVIDIA 的建议

以下是 NVIDIA 对 GPU 加速因子计算的建议(资料来源: NVIDIA 官网、RAPIDS 官网):

1. NumPy、Pandas 的数据处理中浮点数的默认格式是 float64, 而 GPU 比较擅长于 float32 的计算(RTX 3090 中, FP64 的运算能力仅为 FP32 的 1/64)。因此, 在保证精度的情况下, 可以考虑尽量使用 FP32 来进行计算, 性能还会有进一步的提升。
2. 由于 cuDF 的 IO 读写函数采用了 GDS 技术(不经 CPU 和内存, 直接从 GPU 读写 NVME 硬盘), 在性能方面有较大的提升, 所以建议直接把数据读到 GPU 中, 在 GPU 里进行完整的数据处理。
3. 对于目前 cuDF 暂不支持的函数, 可以通过 NUMBA-CUDA 来进行手工实现, 可以进一步获得较大的性能收益(比如上千倍的性能提升)。

风险提示

人工智能挖掘市场规律是对历史的总结, 市场规律在未来可能失效。人工智能技术存在过拟合风险。深度学习模型受随机数影响较大。本文测试的选股模型调仓频率较高, 假定以 vwap 价格成交, 忽略其他交易层面因素影响。

免责声明

分析师声明

本人，林晓明、何康，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构（以下统称为“华泰”）对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。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。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或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报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、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的意见或建议不一致的投资决策。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人（无论整份或部分）等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作，在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题，请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师林晓明、何康本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15% 以上

增持：预计股价超越基准 5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准 15% 以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国: 华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格, 经营许可证编号为: 91320000704041011J

香港: 华泰金融控股(香港)有限公司具有香港证监会核准的“就证券提供意见”业务资格, 经营许可证编号为: AOK809

美国: 华泰证券(美国)有限公司为美国金融业监管局(FINRA)成员, 具有在美国开展经纪交易商业业务的资格, 经营业务许可编号为: CRD#:298809/SEC#:8-70231

华泰证券股份有限公司**南京**

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码: 210019

电话: 86 25 83389999/传真: 86 25 83387521

电子邮件: ht-rd@htsc.com

深圳

深圳市福田区益田路5999号基金大厦10楼/邮政编码: 518017

电话: 86 755 82493932/传真: 86 755 82492062

电子邮件: ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层/

邮政编码: 100032

电话: 86 10 63211166/传真: 86 10 63211275

电子邮件: ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码: 200120

电话: 86 21 28972098/传真: 86 21 28972068

电子邮件: ht-rd@htsc.com

华泰金融控股(香港)有限公司

香港中环皇后大道中99号中环中心58楼5808-12室

电话: +852-3658-6000/传真: +852-2169-0770

电子邮件: research@htsc.com

<http://www.htsc.com.hk>

华泰证券(美国)有限公司

美国纽约公园大道280号21楼东(纽约10017)

电话: +212-763-8160/传真: +917-725-9702

电子邮件: Huatai@htsc-us.com

<http://www.htsc-us.com>

©版权所有2023年华泰证券股份有限公司