

量化专题报告

“量价淘金”选股因子系列研究（十）

订单簿资金流因子簇的构建与生产加速

——基于 Memory Map 的分段读取性能优势

前言：本文为国盛金工《“量价淘金”选股因子系列研究》的第十篇报告，也是我们“因子生产加速”相关研究的再次探索，仍然围绕“数据存储与读写的技术”、“体系化的因子构建思路”这两个话题，在上一篇报告的基础上展开进一步讨论。

Memory Map 在数据分段读取方面的性能优势：在某些因子的构建过程中，经常需要分段读取数据。传统的 CSV、Parquet 等存储形式，需要将全天数据加载到内存中，再截取其中一部分用于后续的因子计算；而 Memory Map 则可以直接在硬盘上截取数据，只将需要用到的数据映射到内存中，大幅减少内存开销，提升因子的计算效率。以读取宽表存储格式下、某一交易日所有股票的第一小时分钟收盘价数据为例，Memory Map 的读取速度约为 Parquet 的 280 倍、CSV 的 350 倍。

因子的体系化构建与批量生产——订单簿资金流因子簇：将订单簿资金流因子的构建流程，拆分为“设计资金流指标”、“是否标准化”、“计算因子”这 3 个步骤，分别展开详细讨论。借助 Memory Map 分段读取数据的高性能优势，基于不同的经济学逻辑，最终构建了约 10 万个资金流因子，并通过进一步筛选，保留其中效果最优且相关性较低的 50 个因子，得到“订单簿资金流因子簇”。

订单簿资金流综合因子：在因子簇中选取样本内信息比率最高的 10 个因子、等权合成，得到订单簿资金流综合因子。回测期 2016/01/01-2024/12/31 内，在全体 A 股中，该因子的月度 RankIC 均值为 0.110，年化 RankICIR 为 4.24；10 分组多空对冲的年化收益为 41.37%，信息比率为 3.29，月度胜率为 87.85%，最大回撤为 13.26%。在剔除了市场常用风格和行业的影响后，纯净因子仍然有效，年化 RankICIR 达到 3.43，全市场 10 分组多空对冲的信息比率为 3.52。

指数增强组合的表现：基于订单簿资金流综合因子，构建月频调仓的指数增强组合：（1）沪深 300 指数增强组合的超额年化收益为 6.66%，跟踪误差为 3.27%，信息比率为 2.04，月度胜率为 71.03%，最大回撤为 3.65%；（2）中证 500 指数增强组合的超额年化收益为 10.72%，跟踪误差为 4.52%，信息比率为 2.37，月度胜率为 75.70%，最大回撤为 3.58%；（3）中证 1000 指数增强组合的超额年化收益为 15.97%，跟踪误差为 5.59%，信息比率为 2.86，月度胜率为 78.50%，最大回撤为 3.52%。

风险提示：以上结论均基于历史数据和统计模型的测算，如果未来市场环境发生明显改变，不排除模型失效的可能性。

作者

分析师 沈芷琦

执业证书编号：S0680521120005

邮箱：shenzhiqi@gszq.com

分析师 刘富兵

执业证书编号：S0680518030007

邮箱：liufubing@gszq.com

研究助理 阮俊烨

执业证书编号：S0680124070019

邮箱：ruanjunye@gszq.com

相关研究

- 1、《“量价淘金”选股因子系列研究（一）：如何将隔夜涨跌变为有效的选股因子？——基于对知情交易者信息优势的刻画》2022-04-26
- 2、《“量价淘金”选股因子系列研究（二）：不同交易者结构下的动量与反转》2022-10-24
- 3、《“量价淘金”选股因子系列研究（三）：如何基于 RSI 技术指标构建有效的选股因子》2023-03-06
- 4、《“量价淘金”选股因子系列研究（四）：高/低位放量：从事件驱动到选股因子》2023-12-18
- 5、《“量价淘金”选股因子系列研究（五）：基于趋势资金日内交易行为的选股因子》2024-05-28
- 6、《“量价淘金”选股因子系列研究（六）：创新高股票中的 Alpha》2024-06-26
- 7、《“量价淘金”选股因子系列研究（七）：盲目追随趋势资金的极端交易行为分析——羊群效应的识别与因子构建》2024-08-06
- 8、《“量价淘金”选股因子系列研究（八）：逐笔买卖差异中的选股信息——条件成交不平衡因子》2024-10-09
- 9、《“量价淘金”选股因子系列研究（九）：Memory Map 在因子生产加速上的应用——以构建羊群效应因子簇为例》2024-11-20

内容目录

一、前言	3
二、Memory Map 的分段读取性能优势	6
2.1 概念简介及数据存储代码展示	6
2.2 分段读取数据的性能优势及代码示例	7
2.3 分段读取数据与因子计算速度对比	8
三、订单簿资金流因子簇的构建	10
3.1 因子构建思路概览	10
3.2 设计资金流指标	10
3.3 是否标准化	12
3.4 计算因子	13
3.5 因子构建方法小结	14
四、订单簿资金流因子簇的表现	16
4.1 因子的批量生产与筛选	16
4.2 订单簿资金流综合因子示例	17
4.3 指数增强组合的表现	19
五、总结	22
风险提示	23

图表目录

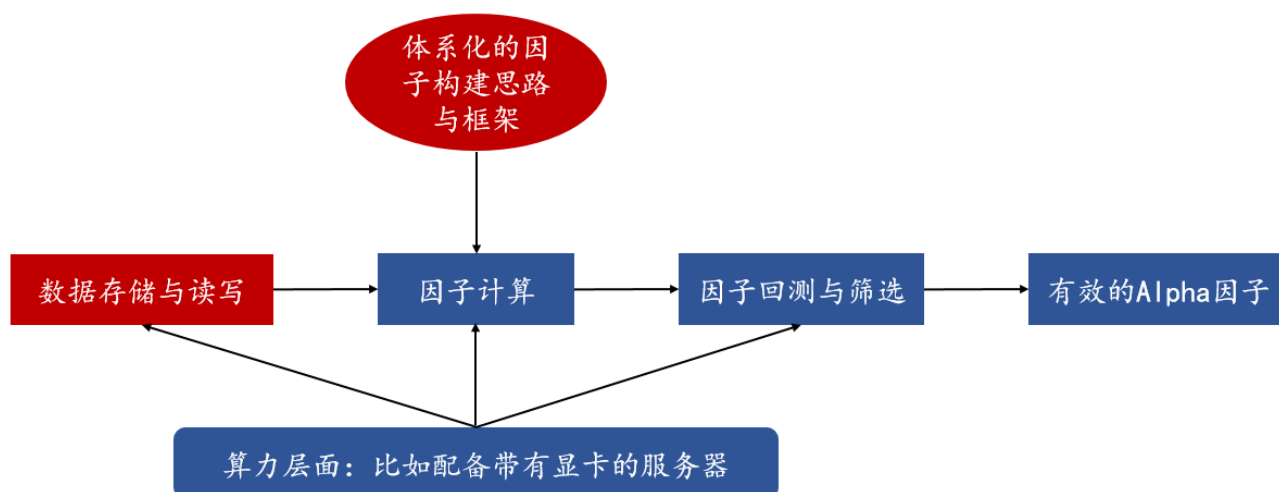
图表 1: 因子生产过程中的重要环节与因素	3
图表 2: 上一篇报告的主要内容	4
图表 3: 本篇报告的主要内容	5
图表 4: 宽表存储示例: 以某日个股分钟收盘价为例	6
图表 5: Memory Map 数据存储结构: 以某日个股分钟收盘价为例	6
图表 6: Memory Map 数据存储代码: 以某日个股分钟收盘价为例	7
图表 7: Memory Map 分段读取代码: 以某日个股分钟收盘价为例	8
图表 8: Memory Map、Parquet 和 CSV 存储文件系统的数据读取速度对比	9
图表 9: 资金流因子簇的构建思路	10
图表 10: wind 基于固定金额对大小单进行划分的标准	11
图表 11: 平安银行 (000001.SZ) 20240830 逐笔成交数据片段	11
图表 12: 资金流指标的设计框架	12
图表 13: 资金流指标的标准化处理	13
图表 14: 以小时为单位的时间段划分	13
图表 15: 将分钟资金流指标降频为日度因子值	14
图表 16: 订单簿资金流因子簇的构建方法汇总	15
图表 17: 订单簿资金流因子簇回测绩效示例	16
图表 18: 订单簿资金流综合因子 10 分组及多空对冲净值	17
图表 19: 订单簿资金流综合因子分年度表现	18
图表 20: 订单簿资金流综合因子与 Barra 风格因子相关系数	18
图表 21: 纯净订单簿资金流综合因子 10 分组及多空对冲净值	19
图表 22: 纯净订单簿资金流综合因子分年度表现	19
图表 23: 沪深 300 指数增强组合净值	20
图表 24: 中证 500 指数增强组合净值	20
图表 25: 中证 1000 指数增强组合净值	21

一、前言

传统量化策略在当前市场环境下正面临一系列挑战，其中最显著的问题是边际效益递减。随着市场的发展与竞争的加剧，信息快速传播且投资者行为趋同，基于公开信息的策略越来越难以获得显著的 Alpha，现有模型和因子的有效性在逐渐衰减；另一方面，靠人力挖掘 Alpha 新因子的难度也越来越大，提取增量信息的效率快速下降，需要持续投入大量资源对现有模型进行不断的升级和迭代，以维持未来获取超额收益的能力。

因此，如何系统性、快速、批量挖掘 Alpha 新因子，对传统模型进行迭代，就成为一项至关重要的技术。在此背景下，国盛金工推出“因子生产加速”相关研究，尝试在数据存储、因子计算、因子构建框架等各个层面做出努力，提升因子的生产效率。

图表1：因子生产过程中的重要环节与因素



资料来源：国盛证券研究所整理

上一篇报告《“量价淘金”选股因子系列研究（九）：Memory Map 在因子生产加速上的应用——以构建羊群效应因子簇为例》，是我们“因子生产加速”系列研究的首次尝试，主要包含以下 3 点贡献：

- （1）聚焦数据存储端，重点介绍了 Memory Map 这一数据存储技术，并展示了其相比于 CSV、Parquet 等传统数据存储形式的优势；
- （2）在因子构建思路，将事件研究的想法融入到选股因子的构建中，具体以羊群效应因子为例，详细讨论了如何相对体系化地批量构建因子；
- （3）最终提出“因子簇”的概念，得到羊群效应话题相关的一簇有效因子。

图表2: 上一篇报告的主要内容

数据存储与读写

Memory Map数据存储
技术的优势

相比于传统的CSV、Parquet等存储形式，Memory Map减少了显式的文件I/O操作，将数据直接映射到内存中，可大幅提升访问速度

以读取宽表存储形式下、某一交易日所有股票的分钟收盘价数据为例，Memory Map的读取速度约为Parquet的260倍、CSV的330倍

体系化的因子构建思路

将事件研究的想法融入到选股因子的构建中，以“羊群效应因子”为例

事件识别

根据价格走势、成交量变化、成交量内部结构的变化等，对趋势资金是否行动做出判断

因子定义

考察趋势资金行动之后的量价特征，比如价格、成交量、价量相关性、大小单资金流等等，衡量羊群效应的强弱

数据频率

在日频、分钟、逐笔等不同数据频率上，识别趋势资金、定义羊群效应因子

提出“因子簇”的概念

基于因子计算速度的提升&体系化的因子构建框架，系统性、快速、批量生产因子，得到羊群效应话题相关的一簇有效因子

资料来源：国盛证券研究所整理

本文为国盛金工《“量价淘金”选股因子系列研究》的第十篇报告，继续围绕“因子生产加速”展开研究，仍然以“数据存储与读写的技术”、“体系化的因子构建思路”这两个角度为切入点，进行更深入的讨论。那么本文与上一篇报告的异同在于：

（1）数据存储与读写方面，上一篇报告重点讨论了如何利用 Memory Map 存储数据，但只是粗略展示了它的读写性能；本篇报告将结合“分段读取数据”的具体应用场景，分析 Memory Map 的读写性能优势，并与 CSV、Parquet 等传统数据存储形式的读取速度进行对比，属于对上一篇报告的延伸与具象化；

（2）因子构建思路方面，上一篇报告给出的“事件识别-因子定义-数据频率”框架，并不能囊括所有的因子研究思路；本篇报告跳出上述框架，给出另一种体系化的思考方式，具体以“订单簿资金流因子”为例，展示在另一套研究框架下，如何系统性地批量构建因子；在这一方面，本篇报告与上一篇报告为并行关系，即基于不同底层逻辑、讨论了不同的因子研究思路；

（3）最终的研究成果展示形式，本篇报告与上一篇报告一致，通过批量生产与筛选，保留一簇有效且相关性较低的因子，称之为“订单簿资金流因子簇”。

图表3: 本篇报告的主要内容

数据存储与读写

是对上一篇报告的延伸

结合“分段读取数据”的具体应用场景，分析Memory Map的读写性能优势，并与CSV、Parquet等传统数据存储形式进行对比

体系化的因子构建思路

与上一篇报告为并行关系，即基于不同底层逻辑、讨论不同的因子研究思路

跳出上一篇报告“事件识别-因子定义-数据频率”的框架，给出另一种体系化的思考方式，具体以“订单簿资金流因子”为例，展示在另一套研究框架下，如何系统性地批量构建因子

延续“因子簇”的概念

研究成果展示形式，与上一篇报告一致

通过批量生产与筛选，保留一簇有效且相关性较低的因子，得到“订单簿资金流因子簇”

资料来源：国盛证券研究所整理

二、Memory Map 的分段读取性能优势

2.1 概念简介及数据存储代码展示

关于 Memory Map 的概念及数据存储方法，我们在上一篇报告《“量价淘金”选股因子系列研究(九): Memory Map 在因子生产加速上的应用——以构建羊群效应因子簇为例》中已进行详细描述，此处做简单回顾。

Memory Map（内存映射）技术是一种将文件或设备的内容直接映射到进程虚拟地址空间中的方法，通过这种方式，进程可以像访问内存一样访问文件或设备，而不需要显式地进行读写操作。相比于传统的 CSV、Parquet 等存储形式，Memory Map 减少了显式的文件 I/O 操作，由于数据直接映射到内存中，因此提高了访问速度。在需要频繁访问大文件的场景中，Memory Map 更能发挥其独特优势，如读取所有 A 股的高频量价数据。

在具体存储 A 股行情数据时，我们采用宽表形式。以分钟行情数据为例，将每个分钟特征设置为二维矩阵，并保存为单独的文件，比如对于某一交易日所有股票的分钟收盘价数据，我们将其存为大小为 $242 \times N$ （242 为每日分钟数据的行数，N 为当日股票数量）的矩阵文件。

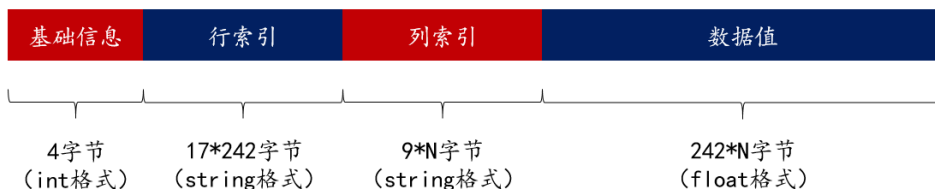
图表4: 宽表存储示例: 以某日个股分钟收盘价为例

	股票1	股票2	股票3	股票4	股票5	股票6		股票N
分钟1								
分钟2								
分钟3	每个格子里，存放的都是对应股票对应分钟的收盘价数据							
分钟4								
分钟5								
.....								
分钟242								

资料来源: Wind, 国盛证券研究所

在利用 Memory Map 格式进行数据存储时，首先需要向系统申请一块连续的内存空间，包括基础信息（行索引个数、行索引最大长度、列索引个数、列索引最大长度）、行索引（分钟时间戳）、列索引（股票代码）、数据值，具体的存储结构如图表 5 所示。

图表5: Memory Map 数据存储结构: 以某日个股分钟收盘价为例



资料来源: Wind, 国盛证券研究所

以 2024/10/31 所有股票的分钟收盘价数据对应的宽表文件 close_20241031.mmap 为例，图表 6 展示了具体的存储代码。对于 Data Frame 文件，我们首先计算出行索引、列索引的最大元素长度和元素数量（代码的第 4-7 行），然后定义内存空间的数据类型及写入的数据（代码的第 10、11 行），最后调用 Numpy 的 memmap 方法申请一块内存空间（代码的第 14 行）并写入数据（代码的第 15 行）。

图表6: Memory Map 数据存储代码: 以某日个股分钟收盘价为例

```
1 # df.shape (242,5082) DataFrame格式
2 index = df.index.map(str)
3 columns = df.columns.astype(str)
4 idxlen = index.map(lambda x: len(x)).max() # 行索引最大长度
5 clmlen = columns.map(lambda x: len(x)).max() # 列索引最大长度
6 idxnum = len(index) # 行索引元素个数 (比如每日分钟数据行数242)
7 clmnum = len(columns) # 列索引元素个数 (即个股数量)
8 dtype = []
9 data = []
10 dtype += ['<i4i4', '%s%d'%(idxnum, idxlen), '%s%d'%(clmnum, clmlen), '%d%s'%(idxnum*clmnum, 'float64')] # 定义数据类型
11 data += [[idxnum, clmnum, idxlen, clmlen], index, columns] # 定义写入数据
12 data += [df.values.flatten()]
13 dtype = ','.join(dtype)
14 fb = np.memmap('close_20241031.mmap', dtype=dtype, mode='w+', shape=(1)) # 申请内存空间
15 fb[0] = tuple(data) # 数据写入
```

资料来源: Wind, 国盛证券研究所

2.2 分段读取数据的性能优势及代码示例

在某些因子的构建过程中，我们经常需要分段读取数据。比如由于 A 股日内不同时间段的投资者结构不同，导致利用不同时段数据计算的因子逻辑和效果存在差异，因此我们经常需要截取日内特定时段的数据来构建因子。

传统的 CSV、Parquet 等存储形式，需要将全天数据都加载到内存中，再截取其中一部分来计算因子；而 Memory Map 则可以直接在硬盘上截取数据，只将需要用到的数据映射到内存中，大幅减少了内存开销，提升因子的计算效率。

如上一小节的图表 5 所示，在存储数据时，对于数据值部分，我们将当日所有股票的分钟收盘价按行展平成 $(242 \times N, 1)$ 的一维向量，并利用连续的空间进行存储。基于 Memory Map 分段读取数据的特性，我们可以通过计算出数据开始位置的索引、数据截取长度，对需要的数据进行自由截取和读取。例如，我们需要读取所有股票第一小时的分钟数据，那么数据起始位置便是 0，截取长度为 $60 \times N$ ，即可得到一个 $(60 \times N, 1)$ 的一维向量；获取这部分数据后，我们再对其进行 reshape，即可得到维度为 $(60, N)$ 的第一个小时分钟收盘价数据矩阵。

图表 7 展示了具体的分段读取代码。对于 Memory Map 文件，我们首先调用 Numpy 中的 memmap 方法读取基础信息，获取数据的行索引个数、最大行索引长度、列索引个数和最大列索引长度。memmap 方法内置了 offset 和 shape 这两个参数，可以指定所需要数据的开始位置和截取长度。如代码第 13 行、第 15 行所示，我们只需设置 shape=（结束分钟位置-开始分钟位置，当日股票数量），offset=数据值开始位置+开始分钟位置*当日股票数量，便可获取日内任意时间段的股票行情数据。

图表7: Memory Map 分段读取代码: 以某日个股分钟收盘价为例

```

1 def from_mmap(filename, parse_dates = True, datatype = 'float32', start_min=0, end_min=242):
2     offset = 0
3
4     idxnum, clmnum, idxlen, clmlen = np.memmap(filename, dtype='int32', mode='r', shape=(4), offset=offset)
5
6     offset += 4*4
7     index = np.memmap(filename, dtype='S%d'%idxlen, mode='r', shape=(end_min - start_min), offset=offset + start_min)
8
9     offset += idxlen * idxnum
10    columns = np.memmap(filename, dtype='S%d'%clmlen, mode='r', shape=(clmnum), offset=offset)
11
12    offset += clmlen * clmnum
13    _shape = (end_min - start_min, clmnum)
14
15    data = np.memmap(filename, dtype=datatype, mode='r', shape=_shape, offset=offset + start_min * clmnum)
16
17    if parse_dates:
18        index = pd.to_datetime(index.astype(str))
19    else:
20        index = index.astype(str)
21    columns = columns.astype(str)
22
23    df = pd.DataFrame(data, index=index, columns=columns)
24    return df

```

资料来源: Wind, 国盛证券研究所

2.3 分段读取数据与因子计算速度对比

本节内容测试对比不同存储文件系统的性能差异以及因子的计算耗时, 通过实例展示 Memory Map 在分段读取数据方面的性能优势。本文的测试系统环境为 Ubuntu 22.04 LTS, 硬件配置为: CPU, AMD 7T83 (64 核, 主频 2.5GHz); 显卡, NVIDIA RTX4090; 存储, 三星 PM9A3 15.36T NVME 固态硬盘 raid0 阵列 (软 raid)。

我们通过 Python 的 timeit 方法测试了 Memory Map、Parquet 和 CSV 存储文件系统的数据读取速度, 读取的数据即为前文所述的宽表存储形式下、某一交易日所有股票的第一小时分钟收盘价数据。其中 Memory Map 格式的文件通过分段读取机制来完成, Parquet 和 CSV 格式文件则先将全天的分钟数据读到内存中、再进行截取。

对比发现, Memory Map 的分段读取性能显著优于另外两种存储文件系统。如图表 8 所示, Memory Map 的读取耗时不到 800 μ s, 而 Parquet、CSV 的读取耗时分别约为 216ms、269ms, 约为 Memory Map 的 280 倍、350 倍。

图表8: Memory Map、Parquet 和 CSV 存储文件系统的数据读取速度对比

<pre># mmap %timeit df = mpd.from_mmap('data/20231229.mmap',parse_dates=False,start_min=0,end_min=60) ✓ 6.2s</pre>	765 μ s $\pm$ 7.09 μ s per loop (mean $\pm$ std. dev. of 7 runs, 1,000 loops each)
<pre># parquet %timeit df = parquet_test() ✓ 1.7s</pre>	216 ms $\pm$ 9.99 ms per loop (mean $\pm$ std. dev. of 7 runs, 1 loop each)
<pre># csv %timeit df = parquet_csv() ✓ 2.1s</pre>	269 ms $\pm$ 6.64 ms per loop (mean $\pm$ std. dev. of 7 runs, 1 loop each)

资料来源: Wind, 国盛证券研究所

在因子计算上，我们也通过一个具体的案例进行对比说明。测试案例为：

- (1) 在 CPU 多进程并行下，读取某只股票某日第一小时的分钟收盘价、分钟成交量数据，计算两个序列的相关系数，得到该股当日的的第一小时高频量价相关性；
- (2) 计算全体 A 股 2021/01/01-2024/12/31 的上述日度指标。

计算结果显示，若原始数据的存储格式为 CSV，该过程大约需要 40s；若为 Parquet，大约需要 25s；若为 Memory Map，只需 10s 左右。

三、订单簿资金流因子簇的构建

3.1 因子构建思路概览

本文基于高频订单簿数据构建资金流因子，可以分为以下 3 个步骤：

(1) 设计资金流指标：具体利用哪些订单簿数据、提取关于资金流的何种信息。比如从数据来源上，订单簿数据可分为逐笔成交、逐笔委托；从交易者结构上，订单簿数据可被拆分为大单、小单；从买卖方向上，订单簿数据又有主动买入、主动卖出之分。不同细分数据蕴含的信息量通常存在较大差异，我们需要基于不同的经济学逻辑，设计不同的资金流指标。在本文中，通过这一步骤，我们将原始的逐笔数据降频为分钟级别的资金流指标。

(2) 是否标准化：是否排除市场、行业等截面因素的影响。个股的资金流数据往往会受到整体市场、或者所在行业的影响，比如某只股票某日的成交金额突然变多，可能是由于它自身因素的刺激，也有可能是所在行业受到了投资者的关注、同行业内的股票普遍放量，亦或仅仅是因为整个市场的交易突然火热了起来。因此我们需要考虑，是否对个股的资金流指标进行截面标准化处理。

(3) 计算因子：如何利用资金流指标构建最终的因子。我们可以关注资金流本身的分布情况，比如计算某段时间内资金流指标的均值、标准差、偏度、峰度；或者也可以考察资金流与收益率的交互作用，比如计算资金流指标与收益率数据的相关系数，得到最终的选股因子。在本文中，通过这一步骤，我们将分钟级别的资金流指标降频为日度因子值。

图表9：资金流因子簇的构建思路



资料来源：国盛证券研究所整理

后文将逐一对上述 3 个步骤进行详细描述，展示完整的因子构建流程。

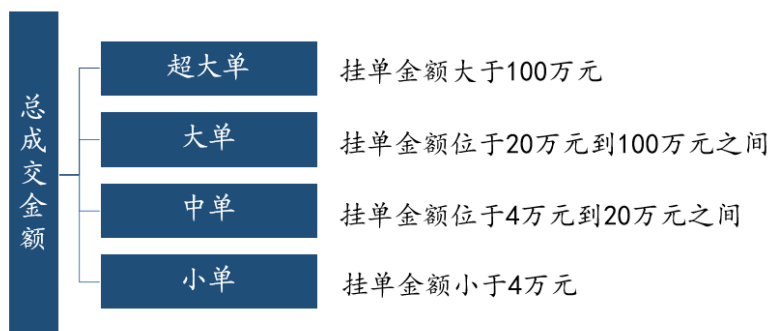
3.2 设计资金流指标

本小节内容，我们讨论如何根据不同的数据来源、交易者结构、买卖方向等信息，设计不同的资金流指标。此环节的关键因素，可用一句话概括，即“基于何种数据，关注何人、何方向的何种量”，涉及的关键因素共有 4 个：何种数据、何人、何方向、何种量。

首先，关于“何种数据”，我们考察订单簿中的 3 种数据来源：成交、委托、撤单。即在构建资金流指标时，我们可以用逐笔成交的数据来计算，也可以用所有逐笔委托的数据来计算，或者只用委托单中的撤单数据来计算。

其次，关于“何人”，我们通过“大小单”进行细分，既可以是不区分大小单的整体的资金流，也可以是超大单、大单、中单、小单的资金流。关于超大单、大单、中单、小单的定义，可以仿照 wind 数据库基于固定金额进行划分的方式（如图表 10），也可以采用浮动金额进行划分，比如基于逐笔成交数据，对于每只股票，将当日的逐笔成交数据与过去 20 日的所有逐笔成交数据进行对比，处于过去 20 日数据 99%分位数以上的定义为当日的超大单，处于 90%-99%分位数的定义为大单，处于 70%-90%分位数的定义为中单，处于 70%分位数以下的定义为小单。

图表10: wind 基于固定金额对大小单进行划分的标准



资料来源: Wind, 国盛证券研究所

接着，关于“何方向”，我们通过逐笔成交数据中的“主买”、“主卖”信息进行区分。“主买”、“主卖”方向是逐笔成交数据中的一项重要细节，如图表 11 展示了平安银行（000001.SZ）在 2024 年 8 月 30 日的逐笔成交数据片段，其中有一列“方向”数据，列明了每一笔成交的买卖方向，标签为“B”代表这一笔成交为主动买入单，即对应着在这一笔成交中，买单是主动吃单方，卖单是挂单方、等待成交；标签为“S”则代表该笔成交为主动卖出单，即买单是挂单方、等待成交，而卖单是主动吃单方。这一“主买”或“主卖”的方向差异，在一定程度上反应了交易双方的激进程度以及对市场未来走势的预期差异。

在讨论“何方向”时，我们既可以计算不区分主买、主卖方向的整体资金流，也可以只计算主买、主卖方向的资金流，或者计算主买与主卖资金流的差值。

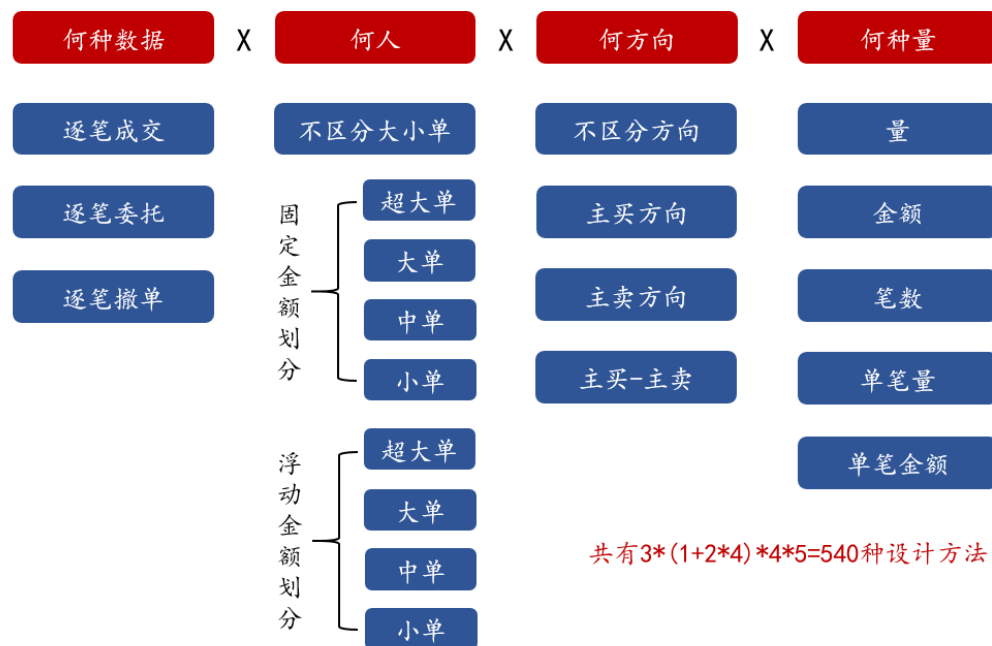
图表11: 平安银行（000001.SZ）20240830 逐笔成交数据片段

时间戳	成交价格	成交量	成交状态	方向	买单ID	卖单ID	成交ID
10:08:00:560	10.16	100	1	B	12145278	12143550	12145290
10:08:00:560	10.16	200	1	B	12145278	12143611	12145291
10:08:00:560	10.16	100	1	B	12145278	12143843	12145292
10:08:00:630	10.17	600	1	B	12145353	11513896	12145354
10:08:00:860	10.16	1700	1	S	12145278	12145819	12145820
10:08:02:050	10.16	200	1	S	12145278	12149488	12149489
10:08:02:050	10.16	100	1	S	12145278	12149504	12149505
10:08:02:050	10.16	400	1	S	12145278	12149506	12149507
10:08:02:050	10.16	200	1	S	12145278	12149508	12149509
10:08:02:090	10.16	300	1	S	12145549	12149510	12149511
10:08:03:130	10.15	1100	1	S	4581621	12153063	12153064
10:08:03:530	10.16	100	1	B	12154146	12149510	12154147
10:08:04:020	10.16	100	1	B	12155727	12149510	12155728
10:08:04:020	10.16	300	1	B	12155727	12149514	12155729
10:08:04:020	10.16	200	1	B	12155727	12149527	12155730
10:08:05:010	10.16	200	1	B	12158627	12149527	12158628

资料来源: Wind, 通联数据, 国盛证券研究所

最后，关于“何种量”，我们考察 5 个指标：量、金额、笔数、单笔量（即量除以笔数）、单笔金额（即金额除以笔数）。

图表12: 资金流指标的设计框架



资料来源: 国盛证券研究所整理

在数据处理上, 本文关于资金流指标的构建, 都是将原始的逐笔数据降频为 1 分钟级别的资金流指标, 比如基于逐笔成交数据, 计算某只股票某一分钟内、浮动金额划分方式下大单主动买入的成交金额。

如图表 12 所示, 4 个因素中, “何数据” 有 3 种数据来源, “何人” 有 9 种选择, “何方向” 有 4 种区分方法, “何种量” 可以考察 5 个指标, 它们互相搭配, 一共可以产生 540 种资金流指标的计算方法。

另外, 关于上文中涉及的参数 (比如浮动金额划分大小单环节中设定的历史分位数), 我们在此重点申明: 所有参数都不会进行遍历测试与样本内优化, 我们会先验地测试参数的敏感性, 若某一参数非常敏感、则舍弃对应的定义方式, 只在参数不敏感的前提下, 选取一个较为合理的取值, 进入上述框架中。即对于两种不同的设计方法, 我们要求它们的区别必须来源于底层逻辑上稍有差异, 而不是简单修改了同一逻辑下的参数取值。

3.3 是否标准化

个股的资金流数据, 往往会受到整体市场、或者所在行业的影响, 比如某只股票某日的成交量突然放大, 可能是受到它自身信息的刺激, 也有可能是所在行业受到了投资者的关注、同行业内的股票普遍放量, 或者也有可能仅仅是因为整个市场的交易量突然放大。

在构建选股因子时, 我们通常希望因子值的变化, 能够尽量反映股票个体因素层面的变化, 因此我们需要考虑, 是否对个股的资金流指标进行截面标准化处理, 剔除市场、行业的影响。

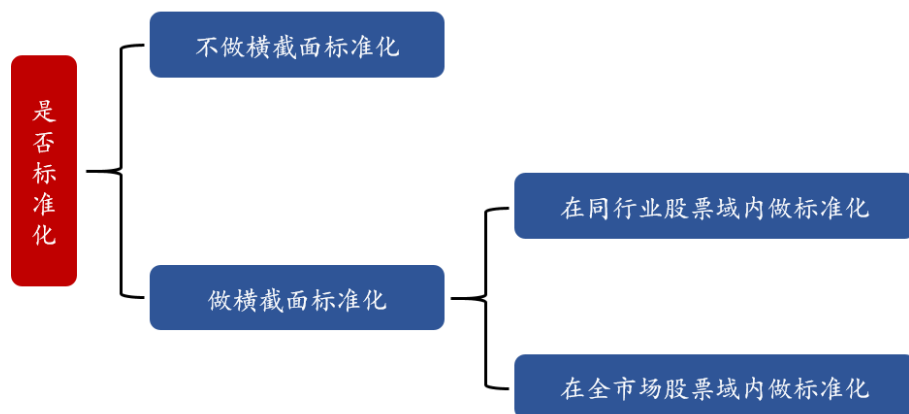
根据是否标准化、以及标准化所选取的样本域, 可对资金流指标做 3 种不同方式的处理。以计算股票某一分钟的大单主动买入成交金额为例:

(1) 不做横截面标准化处理, 计算个股某一分钟的大单主动买入成交金额, 用于后续的因子计算;

(2) 相对于个股所在中信一级行业成分股、做横截面标准化处理：将个股这一分钟的大单主动买入成交金额，在同行业内所有成分股这一分钟的大单主动买入成交金额中排序，取排序值，用于后续的因子计算；

(3) 相对于全市场所有股票、做横截面标准化处理：将个股这一分钟的大单主动买入成交金额，在全市场所有股票这一分钟的大单主动买入成交金额中排序，取排序值，用于后续的因子计算。

图表13: 资金流指标的标准化处理



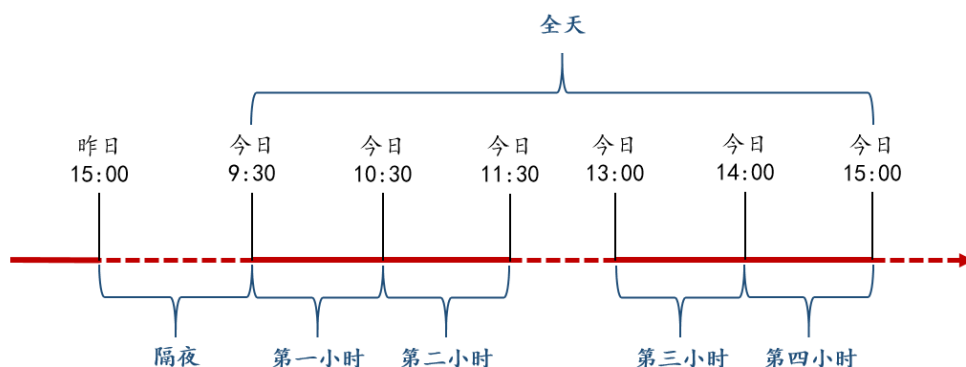
资料来源：国盛证券研究所整理

3.4 计算因子

本小节内容讨论如何利用前文的资金流指标构建最终的因子，涉及到的关键因素有：采用何时的数据，如何将分钟数据降频、得到日度因子值。

关于“何时”，我们可对日内数据进行分时段划分，比如以小时为单位，将每个交易日划分为9:30-10:30(第一小时)、10:30-11:30(第二小时)、13:00-14:00(第三小时)、14:00-15:00(第四小时)这4个时间段(本文暂时舍弃了开盘集合竞价部分的数据)。我们既可以用全天的分钟资金流指标构建因子，也可以只关注某个细分时间段内的资金流情况。

图表14: 以小时为单位的时间段划分

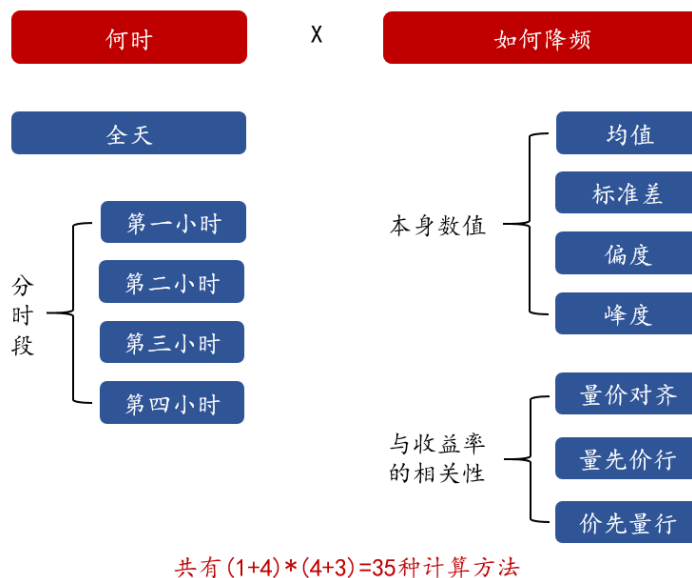


附注：本文暂未考虑开盘集合竞价（即隔夜）部分的数据

资料来源：国盛证券研究所整理

关于“如何将分钟数据降频、得到日度因子值”，我们可以关注资金流本身的分布情况，比如计算某段时间内分钟资金流指标的均值、标准差、偏度、峰度；或者也可以考察资金流与收益率的相互关系，比如计算分钟资金流与分钟收益率的相关系数，具体在计算相关系数时，根据两个序列是否错位，又可分为量价对齐、量先价行、价先量行这 3 种情况。

图表15: 将分钟资金流指标降频为日度因子值



资料来源：国盛证券研究所整理

如图表 15 所示，“何时”有 5 种选择，“如何降频”有 7 种方式，因此一共可以产生 35 种不同的因子计算方法。

3.5 因子构建方法小结

至此，订单簿资金流因子簇的构建方法，已经讨论完毕。我们通过“设计资金流指标”、“是否标准化”、“计算因子”这 3 个步骤，一共可构建 5 万多个资金流因子：

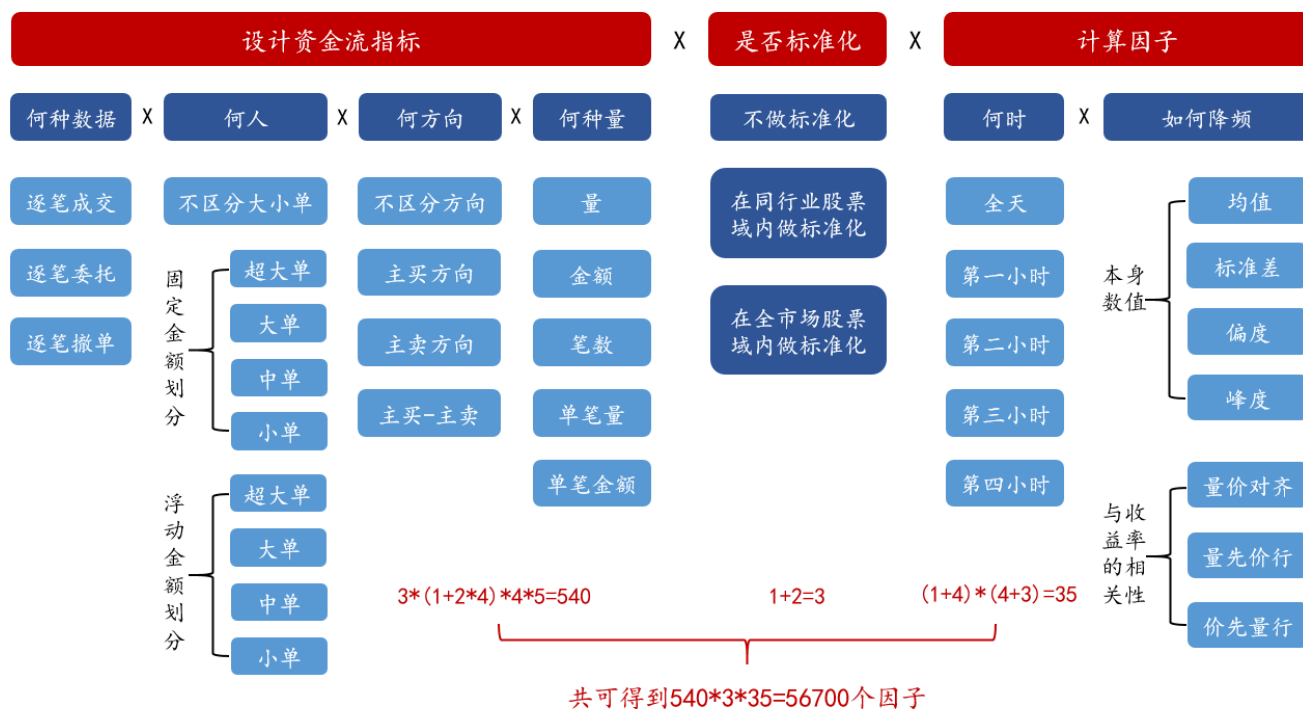
(1) 设计资金流指标：从数据来源、交易者结构、买卖方向等多个角度出发，提取订单簿数据中不同维度的信息，将逐笔数据降频为 1 分钟级别的资金流指标；

(2) 是否标准化：可以不对资金流指标做标准化，也可以将个股的资金流指标在同行业股票域内或者全市场股票域内进行截面标准化处理、剔除行业或者市场因素的影响；

(3) 计算因子：观测全天、或者聚焦日内不同时间段，考察资金流本身的分布情况、或者计算资金流与收益率的相关系数，将 1 分钟资金流指标降频为日度因子值。

实际上，我们能够想到的方法还有更多，比如在因子定义层面，还可以计算不同资金流指标之间的相互关系，或者将资金流指标对收益率做正交化处理、剔除收益率对资金流的干扰，本文不再展开讨论。本节内容的主要目的，在于展示上述体系化的思维与框架，各位读者可根据实际情况，基于各种衍生的经济学逻辑，在框架中修改、删减或增加具体的操作方法。

图表16: 订单簿资金流因子簇的构建方法汇总



资料来源: 国盛证券研究所整理

四、订单簿资金流因子簇的表现

4.1 因子的批量生产与筛选

在实际生产过程中，考虑到有些定义的逻辑相关性较高（比如成交量与成交金额往往高度正相关），因此我们对图表 16 中提到的“资金流指标”做了精简；另外，在“计算因子”层面，除了图表 16 展示的方法，我们还加入了“不同资金流指标之间的相互关系”、“将资金流指标对收益率做正交化处理”等计算方法，最终一共构建了约 10 万个资金流因子。通过进一步筛选，留下有效且相关性较低的因子，得到“订单簿资金流因子簇”：

“将资金流指标对收益率做正交化处理”等计算方法，最终一共构建了约 10 万个资金流因子。通过进一步筛选，留下有效且相关性较低的因子，得到“订单簿资金流因子簇”：

（1）在时间段 2016/01/01-2018/12/31 内，综合考虑因子的 IC、ICIR、10 分组多头超额收益、10 分组多头超额收益的信息比率以及因子互相之间的相关性，挑选效果最优且相关性较低的因子，大约有 2000 个；

（2）对于挑选出的约 2000 个因子，再考虑它们 2019/01/01-2021/12/31 的 IC、ICIR、10 分组多头超额收益、10 分组多头超额收益的信息比率以及因子互相之间的相关性，进一步挑选其中效果最优且相关性较低的因子，最终剩下大约 50 个，称之为基于高频订单簿数据构建的资金流因子簇（2022 年开始即为样本外）。

回测结果显示，在 2016/01/01-2024/12/31 期间，全体 A 股中，50 个因子的年化 ICIR 绝对值都能达到 2 以上。图表 17 罗列了其中 20 个因子的回测绩效，为了方便展示，因子名称以序号的方式做了简化。特别地，这些因子都是未经行业、市值中性化操作的原始因子，大部分因子中性化后的稳定性可进一步提升。

图表17: 订单簿资金流因子簇回测绩效示例

因子名称	IC 信息				10 分组多空对冲绩效指标				
	月度 IC 均值	年化 ICIR	月度 RankIC 均值	年化 RankICIR	年化收益率	年化波动率	信息比率	月度胜率	最大回撤率
因子 1	-0.065	-2.42	-0.095	-3.30	30.61%	12.53%	2.44	75.70%	11.48%
因子 2	-0.058	-2.33	-0.079	-2.74	27.91%	11.43%	2.44	75.70%	11.04%
因子 3	-0.057	-2.26	-0.081	-2.82	26.26%	11.25%	2.33	79.44%	11.53%
因子 4	-0.057	-2.54	-0.076	-3.04	27.50%	10.34%	2.66	80.37%	9.94%
因子 5	-0.056	-2.20	-0.077	-2.66	26.94%	11.38%	2.37	78.50%	7.61%
因子 6	-0.054	-2.35	-0.073	-2.79	25.38%	10.74%	2.36	79.44%	10.54%
因子 7	-0.054	-3.30	-0.065	-3.51	27.38%	6.96%	3.93	85.98%	4.91%
因子 8	-0.054	-2.28	-0.072	-2.74	25.62%	10.93%	2.34	74.77%	10.00%
因子 9	-0.053	-2.44	-0.073	-2.94	26.36%	9.84%	2.68	78.50%	7.37%
因子 10	-0.052	-2.35	-0.072	-2.91	24.70%	10.47%	2.36	77.57%	8.96%
因子 11	-0.052	-2.32	-0.073	-2.93	25.14%	10.76%	2.34	75.70%	10.09%
因子 12	-0.048	-2.53	-0.074	-3.50	23.31%	9.28%	2.51	74.77%	10.91%
因子 13	-0.047	-2.28	-0.066	-2.97	22.72%	9.46%	2.40	76.64%	4.99%
因子 14	-0.047	-2.28	-0.066	-2.96	23.31%	9.68%	2.41	75.70%	5.53%
因子 15	-0.046	-2.39	-0.065	-2.95	21.76%	9.07%	2.40	70.09%	7.19%
因子 16	-0.046	-2.44	-0.061	-2.95	22.36%	9.26%	2.41	79.44%	11.53%
因子 17	-0.044	-2.41	-0.059	-2.90	20.83%	8.83%	2.36	76.64%	11.27%
因子 18	-0.040	-2.53	-0.051	-2.74	19.43%	7.51%	2.59	79.44%	3.88%
因子 19	-0.034	-2.71	-0.046	-3.40	16.53%	6.38%	2.59	82.24%	3.66%
因子 20	0.035	2.08	0.069	2.93	19.41%	8.32%	2.33	76.64%	7.75%

资料来源：Wind，通联数据，国盛证券研究所

以图表 17 中的因子 7 为例，我们详细介绍该因子是如何构建的：

- (1) 从逐笔委托数据中提取撤单信息，得到股票每分钟的撤单笔数，计算撤单笔数的变化率，即“这一分钟撤单笔数/上一分钟撤单笔数-1”；
- (2) 计算股票每分钟以开盘价衡量的收益率，即“这一分钟开盘价/上一分钟开盘价-1”；
- (3) 每个交易日，计算当日日内“分钟撤单笔数变化率”序列与“分钟收益率”序列的相关系数；
- (4) 滚动回看 20 个交易日，求 20 日相关系数的算术平均值，得到当日因子值（本文展示的回测效果，均为月度调仓，取每月月底最后一个交易日对应的因子值即为月度因子值，用于后续回测）。

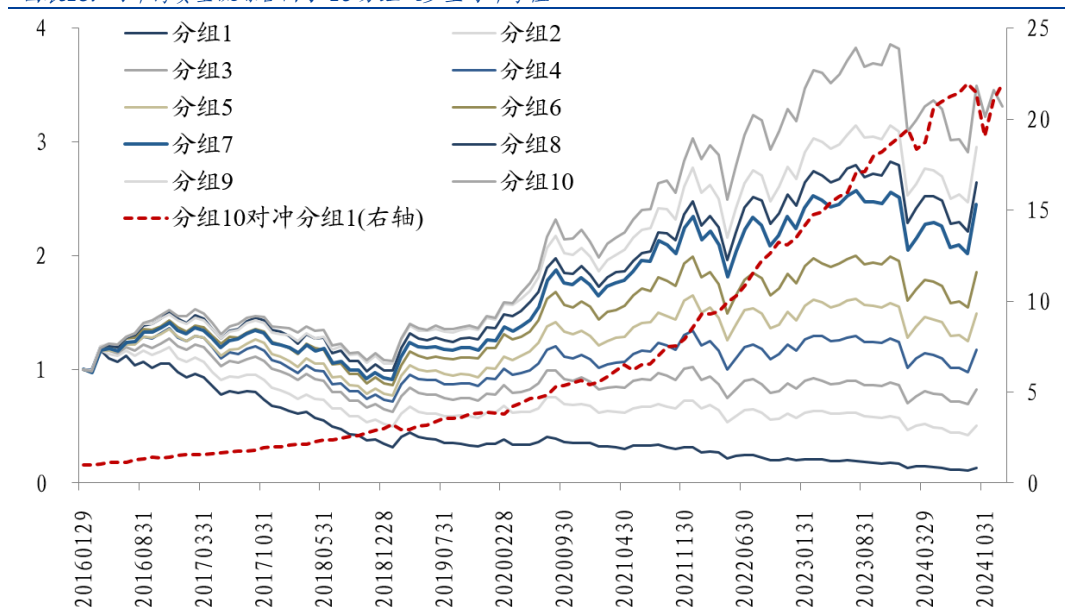
以上述因子为例，对比不同原始数据存储格式下的因子计算速度。回测期 2016/01/01-2024/12/31 内，在 CPU 单进程下，从读取个股分钟数据开始至计算得到所有 A 股的日度因子值，若原始数据的存储格式为 CSV，该过程大约需要 40min；若为 Parquet，大约需要 20min；若为 Memory Map，只需要 5min 左右（服务器配置为：CPU，AMD 7T83（64 核，主频 2.5GHz）；显卡，NVIDIA RTX4090；存储，三星 PM9A3 15.36T NVME 固态硬盘 raid0 阵列（软 raid））。可见 Memory Map 的性能优势确实大幅提升了因子的计算速度。

4.2 订单簿资金流综合因子示例

以选取样本内信息比率最高的 10 个因子等权合成为例，展示订单簿资金流综合因子的表现。回测时间段 2016/01/01-2024/12/31 内，以全体 A 股为研究样本，订单簿资金流综合因子的月度 IC 均值为 0.087，年化 ICIR 为 3.39，月度 RankIC 均值为 0.110，年化 RankICIR 为 4.24；10 分组多空对冲的年化收益为 41.37%，年化波动为 12.57%，信息比率为 3.29，月度胜率为 87.85%，最大回撤为 13.26%。

图表 18 展示了订单簿资金流综合因子的 10 分组及多空对冲净值走势，图表 19 则汇报了其各年度表现情况。

图表18: 订单簿资金流综合因子 10 分组及多空对冲净值



资料来源：Wind，通联数据，国盛证券研究所

图表19: 订单簿资金流综合因子分年度表现

年份	年化收益率			分组 10 对冲分组 1 绩效指标			
	分组 1	分组 10	分组 10 对冲分组 1	年化波动率	信息比率	月度胜率	最大回撤率
2016	-3.09%	52.37%	54.77%	11.28%	4.86	81.82%	0.88%
2017	-31.53%	-6.85%	34.17%	6.63%	5.16	100.00%	0.00%
2018	-47.96%	-21.38%	47.97%	6.77%	7.09	91.67%	0.87%
2019	0.23%	36.43%	30.41%	14.46%	2.10	83.33%	8.78%
2020	1.46%	44.09%	39.63%	15.60%	2.54	75.00%	4.42%
2021	-10.75%	43.20%	58.64%	12.33%	4.76	91.67%	4.40%
2022	-35.42%	5.05%	57.50%	7.99%	7.20	91.67%	1.19%
2023	-15.59%	19.83%	41.23%	6.67%	6.18	100.00%	0.00%
2024	-27.42%	-13.30%	15.07%	20.71%	0.73	75.00%	13.26%

资料来源: Wind, 通联数据, 国盛证券研究所

得到了选股效果不错的综合因子后, 我们考察其与市场常用风格因子的相关性。图表 20 展示了订单簿资金流综合因子与 10 个 Barra 风格因子的相关系数, 可以看到, 综合因子与波动率、流动性因子的相关性相对较高, 相关系数绝对值在 0.30 左右, 与其他因子的相关性较低。

图表20: 订单簿资金流综合因子与 Barra 风格因子相关系数

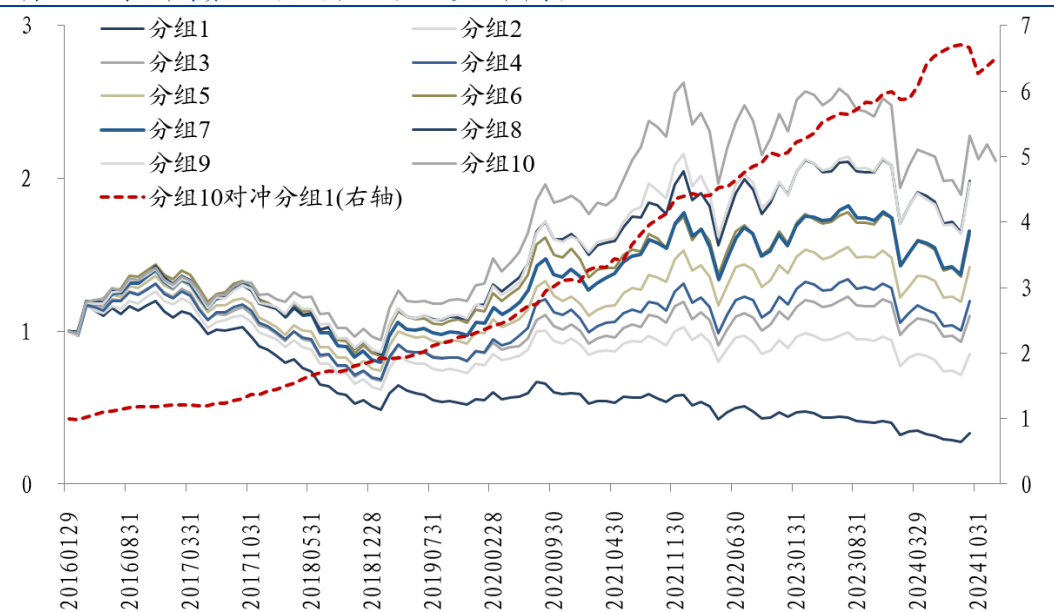
订单簿资金流综合因子		订单簿资金流综合因子	
Beta	-0.061	NonLinearSize	-0.028
BooktoPrice	0.088	Size	-0.102
EarningsYield	0.061	Momentum	-0.077
Growth	0.006	Liquidity	-0.276
Leverage	-0.011	ResidualVolatility	-0.293

资料来源: Wind, 通联数据, 国盛证券研究所

为了剔除市场常用风格和行业的干扰, 我们每月月底将订单簿资金流综合因子对 Barra 风格因子和中信一级行业虚拟变量进行回归, 取残差作为纯净因子, 检验其选股效果。图表 21 展示了纯净因子在全体 A 股中的 10 分组及多空对冲净值走势, 图表 22 则汇报了其分年度的表现情况。

剔除常用风格与行业的影响后, 因子仍然有效。纯净资金流综合因子的月度 IC 均值为 0.048, 年化 ICIR 为 3.58, 月度 RankIC 均值为 0.046, 年化 RankICIR 为 3.43; 10 分组多空对冲的年化收益为 23.33%, 年化波动为 6.64%, 信息比率为 3.52, 月度胜率为 85.05%, 最大回撤为 6.47%。

图表21: 纯净订单簿资金流综合因子 10 分组及多空对冲净值



资料来源: Wind, 通联数据, 国盛证券研究所

图表22: 纯净订单簿资金流综合因子分年度表现

年份	年化收益率			分组 10 对冲分组 1 绩效指标			
	分组 1	分组 10	分组 10 对冲分组 1	年化波动率	信息比率	月度胜率	最大回撤率
2016	14.00%	37.17%	20.82%	5.16%	4.03	90.91%	1.80%
2017	-22.01%	-7.06%	18.84%	6.74%	2.80	75.00%	1.03%
2018	-42.21%	-22.43%	32.43%	5.10%	6.36	91.67%	1.19%
2019	8.87%	35.20%	23.56%	4.33%	5.44	91.67%	0.43%
2020	5.87%	41.59%	33.69%	6.97%	4.83	91.67%	0.74%
2021	-0.10%	42.45%	41.97%	8.20%	5.12	83.33%	0.91%
2022	-24.42%	-12.13%	15.30%	4.51%	3.39	83.33%	0.97%
2023	-9.42%	7.33%	18.35%	3.66%	5.02	83.33%	0.26%
2024	-21.34%	-14.74%	8.31%	9.25%	0.90	75.00%	6.47%

资料来源: Wind, 通联数据, 国盛证券研究所

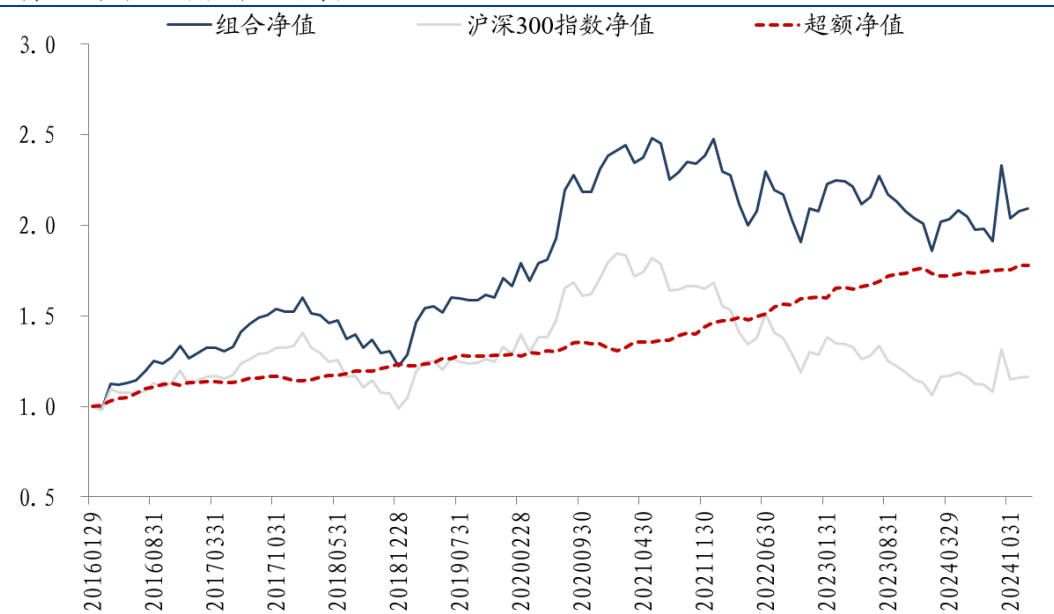
4.3 指数增强组合的表现

本小节内容基于上一小节的订单簿资金流综合因子，以最大化预期收益为优化目标，构建沪深 300、中证 500、中证 1000 指数增强组合，组合的调仓频率为月度。

沪深 300 指数增强组合的约束条件为：（1）指数成分股权重不低于 85%；（2）中信一级行业权重偏离不超过 0.5%；（3）个股权重偏离不超过 0.3%。

2016/01/01-2024/12/31，沪深 300 指数增强组合的年化收益为 8.64%，相对沪深 300 指数的超额年化收益为 6.66%，跟踪误差为 3.27%，信息比率为 2.04，月度胜率为 71.03%，最大回撤为 3.65%。

图表23: 沪深 300 指数增强组合净值

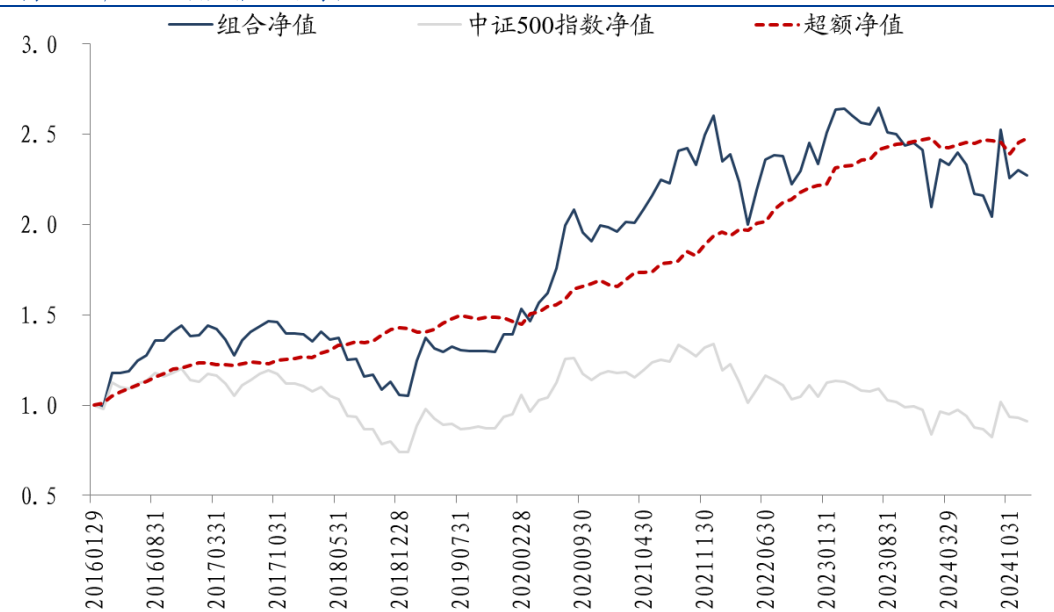


资料来源: Wind, 通联数据, 国盛证券研究所

中证 500 指数增强组合的约束条件为: (1) 指数成分股权重不低于 85%; (2) 中信一级行业权重偏离不超过 0.5%; (3) 个股权重偏离不超过 0.5%。

2016/01/01-2024/12/31, 中证 500 指数增强组合的年化收益为 9.65%, 相对中证 500 指数的超额年化收益为 10.72%, 跟踪误差为 4.52%, 信息比率为 2.37, 月度胜率为 75.70%, 最大回撤为 3.58%。

图表24: 中证 500 指数增强组合净值

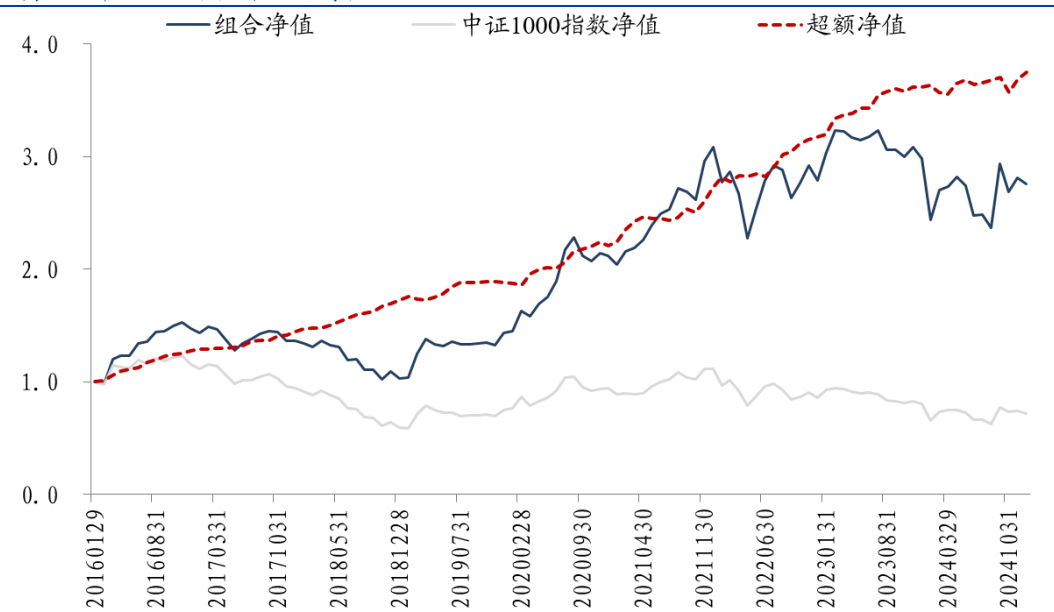


资料来源: Wind, 通联数据, 国盛证券研究所

中证 1000 指数增强组合的约束条件为: (1) 指数成分股权重不低于 85%; (2) 中信一级行业权重偏离不超过 0.5%; (3) 个股权重偏离不超过 0.5%。

2016/01/01-2024/12/31, 中证 1000 指数增强组合的年化收益为 12.04%, 相对中证 1000 指数的超额年化收益为 15.97%, 跟踪误差为 5.59%, 信息比率为 2.86, 月度胜率为 78.50%, 最大回撤为 3.52%。

图表25: 中证 1000 指数增强组合净值



资料来源: Wind, 通联数据, 国盛证券研究所

五、总结

本文为国盛金工《“量价淘金”选股因子系列研究》的第十篇报告，也是我们“因子生产加速”相关研究的再次探索，仍然围绕“数据存储与读写技术”、“体系化的因子构建思路”这两个话题，在上一篇报告的基础上展开进一步深入讨论。

在数据存储与读写方面，本篇报告结合“分段读取数据”的具体应用场景，分析了 Memory Map 的读写性能优势。在展示了 Memory Map 分段读取数据的代码案例后，我们测试对比了不同存储文件系统的数据读取速度。以读取宽表存储格式下、某一交易日所有股票的第一小时分钟收盘价数据为例，Memory Map 的读取速度约为 Parquet 的 280 倍、CSV 的 350 倍。

在因子构建框架层面，我们以“订单簿资金流因子”为例，通过“设计资金流指标”、“是否标准化”、“计算因子”这 3 个步骤，展示了一套相对体系化的因子构建思路：

(1) 设计资金流指标：从数据来源、交易者结构、买卖方向等多个角度出发，提取订单簿数据中不同维度的信息，将逐笔数据降频为 1 分钟级别的资金流指标；

(2) 是否标准化：可以不对资金流指标做标准化，也可以将个股的资金流指标在同行业股票域内或者全市场股票域内进行截面标准化处理、剔除行业或者市场因素的影响；

(3) 计算因子：观测全天、或者聚焦日内不同时间段，通过考察资金流本身的分布情况、计算资金流与收益率的相关系数、计算不同资金流指标之间的相关系数、将资金流指标对收益率做正交化处理等方式，将 1 分钟资金流指标降频为日度因子值。

在 Memory Map 分段读取性能的助力下，本文构建了约 10 万个资金流因子，并通过进一步筛选，保留其中效果最优且相关性较低的 50 个因子，称之为“订单簿资金流因子簇”。回测结果显示，50 个资金流因子的年化 ICIR 绝对值都能达到 2 以上。

为了展示综合因子的表现，我们在“订单簿资金流因子簇”中选取样本内信息比率最高的 10 个因子、等权合成，得到订单簿资金流综合因子。回测期 2016/01/01-2024/12/31 内，在全体 A 股中，订单簿资金流综合因子的月度 IC 均值为 0.087，年化 ICIR 为 3.39，月度 RankIC 均值为 0.110，年化 RankICIR 为 4.24；10 分组多空对冲的年化收益为 41.37%，年化波动为 12.57%，信息比率为 3.29，月度胜率为 87.85%，最大回撤为 13.26%。在剔除了市场常用风格和行业的影响后，纯净因子仍然有效，年化 ICIR 达到 3.58，全市场 10 分组多空对冲的信息比率为 3.52。

更进一步，我们基于订单簿资金流综合因子，构建月频调仓的指数增强组合：

(1) 沪深 300 指数增强组合的超额年化收益为 6.66%，跟踪误差为 3.27%，信息比率为 2.04，月度胜率为 71.03%，最大回撤为 3.65%；

(2) 中证 500 指数增强组合的超额年化收益为 10.72%，跟踪误差为 4.52%，信息比率为 2.37，月度胜率为 75.70%，最大回撤为 3.58%；

(3) 中证 1000 指数增强组合的超额年化收益为 15.97%，跟踪误差为 5.59%，信息比率为 2.86，月度胜率为 78.50%，最大回撤为 3.52%。

风险提示

以上结论均基于历史数据和统计模型的测算，如果未来市场环境发生明显改变，不排除模型失效的可能性。

免责声明

国盛证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，但本公司及其研究人员对该等信息的准确性及完整性不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，可能会随时调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，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。

本公司力求报告内容客观、公正，但本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。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。

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。

本报告版权归“国盛证券有限责任公司”所有。未经事先本公司书面授权，任何机构或个人不得对本报告进行任何形式的发布、复制。任何机构或个人如引用、刊发本报告，需注明出处为“国盛证券研究所”，且不得对本报告进行有悖原意的删节或修改。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的任何观点均精准地反映了我们对标的证券和发行人的个人看法，结论不受任何第三方的授意或影响。我们所得报酬的任何部分无论是在过去、现在及将来均不会与本报告中的具体投资建议或观点有直接或间接联系。

投资评级说明

投资建议的评级标准		评级	说明
评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。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（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准。	股票评级	买入	相对同期基准指数涨幅在 15%以上
		增持	相对同期基准指数涨幅在 5%~15%之间
		持有	相对同期基准指数涨幅在 -5%~+5%之间
		减持	相对同期基准指数跌幅在 5%以上
	行业评级	增持	相对同期基准指数涨幅在 10%以上
		中性	相对同期基准指数涨幅在 -10%~+10%之间
		减持	相对同期基准指数跌幅在 10%以上

国盛证券研究所

北京

地址：北京市东城区永定门西滨河路 8 号院 7 楼中海地产广场东塔 7 层

邮编：100077

邮箱：gsresearch@gszq.com

南昌

地址：南昌市红谷滩新区凤凰中大道 1115 号北京银行大厦

邮编：330038

传真：0791-86281485

邮箱：gsresearch@gszq.com

上海

地址：上海市浦东新区南洋泾路 555 号陆家嘴金融街区 22 栋

邮编：200120

电话：021-38124100

邮箱：gsresearch@gszq.com

深圳

地址：深圳市福田区福华三路 100 号鼎和大厦 24 楼

邮编：518033

邮箱：gsresearch@gszq.com