因子说明

## 分钟 K 线类因子

说明：根据分钟 K 线包含的信息，可构建高频偏度、下行波动占比、尾盘成交占比、量价相关性、改进反转、平均单笔流出金额占比、大单推动涨幅，共七个因子。它们的月均 Rank IC 都在 3%以上，年化 ICIR 都大于 3。其中，尾盘成交占比因子表现最优，月均 Rank IC 为 4.86%，年化 ICIR为 3.59

因子效果

1. **高频偏度**

出处：海通证券（2017年5月）选股因子系列研究(十九)——高频因子之股票收益分布特征；Amaya et al.(2011)《Does Realized Skewness and Kurtosis Predict the Cross-Section of Equity Returns?》一文中发现，股票高阶矩与未来收益之间存在联系。并利用股票的日内分时数据，构建了高频方差、高频偏度和高频峰度三个指标，其中高频偏度具有较强的选股效果。

说明：Amaya et al.(2011)中发现，股票高阶矩与未来收益之间存在联系。并利用股票的日内分时数据，构建了高频方差、高频偏度和高频峰度三个指标，其中高频偏度具有较强的选股效果。高频偏度刻画了股票价格日内快速拉升或下跌的特征。假设有两只股票日内涨幅相同，其中一只股票的涨幅由持续稳定的小幅上涨累计而来，而另一只股票的上涨源自于短期的大幅拉升，那么后者在未来有较大概率出现收益反转。从风险溢价角度来看，日内经常快速下跌，或者下行风险大的股票具有更高的风险溢价。

因子计算公式如下：

其中，i、j、n 分别代表第 i 只股票、第 j 分钟和第 n 个交易日。月度选股下 T 取 20，周度选股下 T 取 5（下同）。

效果：高频偏度因子月均 Rank IC 为 3.41%，年化 ICIR 为 3.48。因子月均多空收益为1.07%，月度胜率为 83%。月均多头收益为 0.27%，月均空头收益为-0.80%。

1. **下行波动占比**

下行波动占比与高频偏度的逻辑基本一致，因子计算公式如下：

1. **尾盘成交占比**

一般来说，股票日内成交量呈现“U”型或者“W”型走势，即成交量在开盘和收盘阶段比其他交易时段更高，午间休市可能导致下午开盘时成交量也出现高点。各个时点的成交量分布能够反映投资者的行为特征，蕴含额外信息。以半小时为间隔划分成八个区间，计算每个区间成交量占比，并使用股票过去T日指标均值作为因子值。检验发现，10:00 之前，14:30 之后的成交量占比因子和股票下月收益负相关；10:00-11:00 的成交量占比因子和股票下月收益显著正相关。其中，尾盘（14:30 之后）成交占比因子的选股效果最显著。

在每个月的月末(t)，按照月末 A 股流通市值将所有A 股股票等分为 10 组。其中， A 股流通市值最小的 10%的股票在第一组，A 股流通市值最大的 10%的股票在第十组。然后，计算下一个月（t+1）各组流通市值加权的股票组合的月收益率及第十组与第一组收益率之差（即为因子收益率）。以此类推，在每个月月末计算 A 股流通市值并重新分组、计算组合收益率。 因子计算公式如下所示：

尾盘成交占比 =

尾盘成交占比因子具有较好的效果可能源于，（1）尾盘投机度高，容易出现价格操纵行为；（2）非知情交易者（散户）不愿承担日内波动，更倾向于尾盘交易，而知情交易者（机构）则倾向于在早盘交易。

尾盘成交占比因子月均 Rank IC 为 4.86%，年化 ICIR 为 3.59。因子月均多空收益为 1.62%，月度胜率为 82%。月均多头收益为 0.52%，月均空头收益为-1.10%。

1. **K 线最短路径非流动性因子**

出处：光大证券（2017） 基于 K 线最短路径构造的非流动性因子 ——多因子系列报告之七

说明：这是一种利用了高频数据的信息，构造的表示精度更高，效果更强的非流动性因子

效果：最终定义的 TS 非流动性因子选股表现最为优秀。多空收益高达 17.3%，年化波动率 11.8%，夏普比率 1.405，最大回撤 20.3%。以该因子构造的选股组合在 2010 年至 2017 年间，年化收益 18.2%，年化波动率 24.8%，夏普比率 0.8，最大回撤 48.3%，月度换手率 31.4%。

相对中证全指年化收益 12.5%， 相对波动率 8.3%，相对最大回撤仅8.35%。

因子定义：

其中：

表示日内频率下第 j根K 线的最短路径

：表示日内频率下第j 根K线内的成交额

p:表示日内频率分段个数

d：表示日间移动平均的周期参数

1. **集合竞价阶段的成交量占比**

出处：光大证券（2017） 见微知著：成交量占比高频因子解析 ——多因子系列报告之五

说明：

1. **成交量比值因子 OCVP**

出处：光大证券 高频因子：日内分时成交量蕴藏玄机 ——多因子系列报告之八

说明：集合竞价阶段是反映投资者行为信息的重要时点，集合竞价阶段成交量反映了多空双方对个股开盘价格的认同度。而一般量在价先，集合竞价成交量占比更有意义。

效果：采用简单移动平均（MA）构造开盘集合竞价成交量占比因子OCVP。同时考虑信息的时间衰减效用，引入具有时变效用的权重修正OCVP 因子。经过检验 OCVP 因子具备良好的预测能力和单调性，其IC均值为-5.6%，IR 绝对值为 0.83；引入尾盘前5 分钟成交量占比因子，通过加权方式构造复合因子 OBCVP。最优权重组合下，复合因子的预测性和单调性显著提升，IC 均值为-7%，IR 绝对值大于 1。根据因子值将股票等分 5 组的多空对冲组合 8 年年化收益为 15.10%，夏普比率达3.03，最大回撤为 10.2%

因子定义：

其中：𝑉𝑂𝐿𝑐𝑎𝑙𝑙：表示每日开盘前集合竞价阶段成交量

𝑉𝑂𝐿𝑡𝑜𝑡𝑎𝑙：表示日内个股总成交量

𝑑：表示第t日向前移动平均的交易日个数

**：**引入时变效用的权重因子：

其中，表示每日集合竞价成交量占比

α：信息的衰减强度，α = 2/(1+𝑑)；

,为时间权重因子

**收盘前成交量占比因子BCVP（before closing volume percent）**

𝑉𝑂𝐿𝑐𝑙𝑜𝑠𝑒：表示收盘前 5 分钟内个股成交量

𝑉𝑂𝐿𝑡𝑜𝑡𝑎𝑙：表示日内个股总成交量

𝑑：表示第 t 日向前移动平均的交易日个数

1. **一致交易因子（**Total Consistent Volume**）**

出处：光大证券 一致交易因子：挖掘集体行为背后的收益 ——多因子系列报告之九

说明：如果一支股票在一段时间内大部分的交易属于集体一致交易行为（带有显著趋势性），那么很可能这支股票在这段时间内有新信息进入，正在经历市场消化（price-in）的过程。因此这类股票未来股价更可能产生大幅变化，为交易者创造出了交易机会。

效果：在一致参数为 0.95 时，一致交易因子、一致买入交易因子与一致卖出交易因子的 IR 值都达到十分显著的水平：0.84，0.64 和0.91。综合比较 IC 序列整体与近期的表现，**一致卖出交易因子更具有预测能力，其 IC 均值 7.7%，IR 值为 0.91。**

因子定义：

其中：

𝑑：表示移动平均的周期参数

𝐶𝑜𝑛𝑠𝑖𝑠𝑡𝑒𝑛𝑡\_𝑉𝑜𝑙𝑢𝑚𝑒：表示当日全部 5 分钟频率实体 K 线对应成交量总和

𝑉𝑜𝑙𝑢𝑚𝑒：表示当日总成交量

实体K线定义：，其中 α 是一致参数，取值在 0 到 1 之间，取 1 时被称为严格实体 K 线

1. **东吴证券𝐌𝐢𝐱𝐋𝐈𝐐因子**

出处：

说明：传统的流动性因子忽略了 A 股市场中特定的日内交易结构信息，东吴证券给出了一个针对传统流动性因子的改进方案。改进后的流动性因子在不同样本空间下均具有显著的选股能力。

效果：在股票5分组多空对冲情形下，其全样本年化超额收益达到 32.4%，信息 比率达 3.98。在中证 500 成分股内选股，多空对冲年化收益达到 25.2%，信息比率为 2.46。在沪深 300 成分股内选股，多空对冲年化 收益达到 21.7%，信息比率为 1.61。

因子定义：

说明：、、、、的计算方式：

横截面回归得到的残差项即为所要最终得到的“纯净的局部流动性因子”

、、、、，股票的局部流动性因子：分别代表股票在过去 20 个交易日在隔夜时段、9:30-10:30、10:30-11:30、13:00-14:00、14:00-15:00 这 5 个时段的流动性

的计算方式：

其中，at、bt为回归系数，截面回归的残差项即为我们所需要的得到，即LIQ(4)因子。

将 t-19 至 t 日共 20 个交易日的 14:00-15:00 时段的换手率相加，而后取对数，

为股票s在t − i日 14:00-15:00 时段的换手率，因子在横截面上关于对数流通市值因子做回归:

其中，at、bt为回归系数，截面回归的残差项即为我们所需要的得到，即LIQ(4)因子。

1. **APM 因子**

出处：方正证券（2016年10月） 凤鸣朝阳：股价日内模式中蕴藏的选股因子

说明：在A股市场中，市场微观结构领域的实证研究显示，知情交易概率（probabiliy of informed trading）在日内呈现"倒J形"曲线。关于这个现象的简单理解是∶由于隔夜时段的交易暂停，每个交易日开盘后，市场累积的大量私有信息，将通过交易迅速释放，因此知情交易概率在开盘后呈现快速下降的形态.换言之，知情交易者更加倾向于在每日上午进行交易，上午的价格行为蕴藏了更多可用于选股的信息量。通过模型分析上午和下午行情数据所蕴含信息量的差别，是构建APM因子模型的核心思想。

定义：**原始APM因子**的计算步骤∶

第一步∶对选定股票，回溯取其过去20日数据，记逐日上午的股票收益率为，指数收益率为;逐日下午的股票收益率为，指数收益率为;

第二步∶将得到的40组上午与下午（r，R）的收益率数据进行回归∶

，得到残差项

第三步∶ 以上得到的40个残差中，上午残差记为，下午残差记为，进一步计算每日上午与下午残差的差值.

第四步∶ 构造统计量stat来衡量上午与下午残差的差异程度，计算公式如下（μ为均值，a为标准差）;

第五步∶为了消除动量因子的影响，将统计量stat对动量因子进行横截面回归∶ stat; =bRet20;＋E;，其中Ret20为股票过去20日的收益率，代表动量因子;

第六步∶ 将回归得到的残差值作为原始的APM因子。

图1给出了原始APM因子在剔除市值和行业因子后的表现，可以看到，该因子在样本内（2013年5月-2016年10月）表现优异，多空对冲年化收益达到了15.90%，期间最大回撤为2.52%，信息比率为2.84，月度胜率为78.6%;样本外（2016年11月-2019年12月）的表现略逊于样本内，多空对冲年化收益为6.41%，最大回撤为5.72%。尤其是2019年，因子出现了连续的回撤。

**改进的APM因子**：

在第一步中，用隔夜收益替代上午收益，（指数收益也相应用隔夜的指数收益替代），其余步骤不变，回归所得到的因子记为 。

1. **聪明钱因子**

出处：方正证券（2016年7月）跟踪聪明钱：从分钟行情数据到选股因子

说明：先利用聪明度指标 S，从分钟数据中筛选出属于“聪明钱”的交易。在此基础上，我们构造了聪明钱情绪因子 Q，该因子实际上反映了聪明钱参与交易的相对价位。因子 Q 的值越大，表明聪明钱的交易越倾向于出现在价格较高处，这是逢高出货的表现，反映其悲观态度；因子 Q 的值越小，则表明聪明钱的交易多出现在价格较低处，这是逢低吸筹的表现，反映其乐观情绪。

效果：根据情绪因子 Q 对所有 A 股进行排序并等分五组，多空对冲的年化收益为26.0%，信息比率 2.70，最大回撤 7.97%，月度胜率 75.7%，收益回撤比 3.26。将 Q 值小视为聪明钱逢低买入的表现，因此可选取五分组中 Q 值最小的第一组，作为跟踪聪明钱动向的 SMART 组合。SMART 组合对冲中证 500 指数后，年化收益为 31.6%，信息比率 3.66，最大回撤 5.9%，月度胜率 79.0%，收益回撤比高达5.4。

因子计算公式：

指标S衡量每一个分钟交易的“聪明”程度：

其中，Rt为第 t 分钟的涨跌幅，Vt为第 t 分钟的成交量。指标 St的值越大，则表示该分钟的交易越“聪明。

对于特定股票、特定时段的分钟行情数据，按照上述方法划分出

聪明钱的交易之后，我们可以构造聪明钱的情绪因子 Q：

其中，是聪明钱的成交量加权平均价，是所有交易的成交量加权平均价。

## 基于 TICK 委托数据的高频因子

1. **开盘后净委买增额占比**

出处：

说明：盘口委托挂单数据刻画了投资者的买入意愿，开盘后 30 分钟内的委买增量越大， 投资者在这段时间内的买入意愿越强。

效果：开盘后净委买增额占比因子月均 Rank IC 为 4.19%，年化 ICIR 为 4.24。因子月均多空收益为 1.33%，月度胜率为 82%。月均多头收益为 0.68%，月均空头收益为-0.65%。

因子计算公式如下：

1. **基于订单簿的spread指标**

出处：天风证券（2019年9月）

说明：盘口信息将是判断个股短期走势的重要依据。根据每个Tick上的买卖挂单强度，研报构建了 Spread 指标以度量当前时间的盘口买卖挂单的强弱差异。

效果：从5个Tick到 200 个Tick的预测窗口，研报发现Spread 指标对股价短期走势具有显著预测能力，在买盘挂单相对越厚的时点股价短期上行概率将越大。 将指标调整为日频后，研报发现 Spread 指标在日频率与 Tick 频率窗口下的含义具有明显差异。在日频率上，盘口委买挂单相对越厚的股票在当日实际承受了越强的主动卖压，二者之间相关性显著，因此这部分股票在截面上的相对收益应该越低。从"高频时序"到"低频 alpha" 研报将高频数据低频化后构建了月度的 Spread 选股因子，从 2010 年至 2019 年8月 30 日的回溯区间内，因子表现出显著的选股能力。因子中性化后多空组合年化收益为 14.6%，多空组合 IR达到 2.01;因子 IC 均值-4.0%， ICIR 达到-2.10，IC 胜率为 75.7%。 研报分别在沪深 300 指数、中证 500 指数、中证 800指数以及中证 1000指数成分股中回溯了因子的选股能力，各股票池中因子IC均值均超过4.0%，因子选股能力呈现出较好的稳健性。此外，Spread 因子相对于反转、波动、换手、流动性等量价指标具有较高的独立性，截面相关系数均值分别为 13.6%，-10.4%、11.7%、-15.1%，因子能带来较显著的增量信息。

计算公式：

## 基于逐笔成交数据的高频因子

1. **开盘后净主买占比**

出处：

说明：逐笔成交数据中包含投资者主动买入和主动卖出的信息，将两者的成交金额相减即可得到净主动买入的金额（简称“净主买”）。开盘后净主买占比因子刻画了投资者在开盘后 30 分钟内，净买入行为的强度。开盘后净主买占比越高，投资者的主动买入行为越强。

效果：开盘后净主买占比因子月均 Rank IC 为 3.31%，年化 ICIR 为 2.51。因子月均多空收益为 0.98%，月度胜率为 73%。月均多头收益为 0.55%，月均空头收益为-0.42%。

计算公式：

1. **开盘后净主买强度**

出处：

说明：开盘后净主买强度因子刻画了投资者在开盘后 30 分钟内，净买入行为的稳健性。开盘后净主买强度越高，投资者的主动买入行为越稳健。

效果：开盘后净主买强度因子月均 Rank IC 为 4.46%，年化 ICIR 为 3.29。因子月均多空收益为 1.34%，月度胜率为 72%。月均多头收益为 0.44%，月均空头收益为-0.91%。

因子计算公式：

1. **大单成交占比**

出处：

说明：基于逐笔成交数据中的单号，可将逐笔数据合成为单成交数据，并从单的角度区分大小单。再根据对应的“买”、“卖”标记，可进一步计算大买单占全天成交金额的比例（简称“大买成交占比”）。大买成交占比越高，大单买入行为越强。

效果：大买成交占比因子月均 Rank IC 为 4.35%，年化 ICIR 为 1.96。因子月均多空收益为 1.35%，月度胜率为 66%。月均多头收益为 0.39%，月均空头收益为-0.96%。

因子计算公式：

在界定大小单时，可从单的成交量分布出发。例如，某一单的成交量处于当日成交

量分布的均值+1 倍标准差之外，则可被认为是大单。

1. **逐笔大单因子**

出处：海通证券（2020年）选股因子系列研究（七十一）——逐笔大单因子与大资金行为

说明：异动股票中统计得到的大资金行为与大单因子有较高关联度，交易异动股票往往具有长期负向收益，剔除异动效应后大单因子选股效果明显增强。基金重仓股票与基金增持股票在基金定期报告披露前有很好的表现。利用非流动性，剔除大买的大卖，正交后的大买三个因子对于上期基金重仓组合进行增强，可以在基金定期报告期披露前两个月获得显著正向收益。

**交易异动与大单因子**

说明：当股票出现超高换手等异常交易情况时，会被定义为交易异动股票，同时需要披露参与该股票交易的买卖前五大营业部买卖金额情况。如果研报所构建的大单因子 与异动股票前五大席位总成交金额有较高相关性，则可以证明研报所构建的大单因子确实有捕捉大资金交易行为的特点。

效果：基于交易异动股票显著的长期负向选股效果，研报中尝试通过将过去 10 个交易日内，发生过交易异动的股票从股票池中剔除。无论是剔除股票还是正交哑变量的方法，在剔除异动效应之后大单因子的选股效果均有显著的提升，选股效果的稳定性也显著增强。

因子计算公式：

大单：即参与成交买卖订单有一方成交额超过全天所有订单成交额均值的成交

金额占比。

大买+大卖成交占比：下简称“大买+大卖”，即参与成交买卖订单双方成交额

均超过全天所有订单成交额均值的成交金额占比。

剔除大买的大卖：即参与成交买卖订单中，卖单成交额均超过全天所有订单成

交额均值，而买单成交金额没有超过全天所有订单成交金额均值的成交金额占

比。

大买：即参与成交买卖订单中，买单成交额均超过全天所有订单成交额均值的

成交金额占比。

正交大单的大买：利用上述大买因子对上述大单因子进行线性回归的残差。

**基金重仓与大单因子**

说明：公募基金建仓行为也是一种大资金交易动向，公募基金建仓过程中，“机构专用”账户大概率参与这些股票的交易过程当中。根据异动股票中大资金行为与大单因子的关联分析，因此我们有理由相信大单因子与公募基金建仓行为也有一定关联。

效果：基金重仓股票与基金增持股票在基金定期报告披露前有很好的表现。重仓组合在报告期前收益情况显著优于报告期之后，报告期前无论是原始收益还是市值中性

化后的收益，相对于市场均有一定的超额收益，基金增持组合也具有同样的特征。

因此，如果我们可以在报告期前 2 个月对于基金持仓量，基金持仓增量进行预测，

那么该预测指标大概率可以帮助我们找到公募基金更倾向于持有或增持的股票，从而获取在报告期前不错的股票收益。

因子计算公式：

1. 筛选基金：只保留主观型公募基金，剔除被动型基金和指数增强基金
2. 构建组合：每个季末保留前十大重仓股，计算基金持仓市值占总市值比例作为组合权重
3. 计算指标：以当期持仓占比作为持仓量指标，以每个有增持该股票的基金增持市值占比合计作为持仓增量指标
4. **开盘后知情主卖占比**

出处：

说明：结合逐笔成交数据中的主买、主卖和预期外收益，可定义知情主卖，用以刻画知情交易者的卖出行为。开盘后知情主卖占比越高，知情交易者卖出行为越强，投资者对股票未来的表现越悲观。

效果：开盘后知情主卖占比因子月均 Rank IC 为 2.86%，年化 ICIR 为 2.51。因子月均多空收益为 0.76%，月度胜率为 75%。月均多头收益为 0.46%，月均空头收益为-0.30%。

因子计算公式：

判断是不是知情主买和知情主卖主要看股票的预期外收益，而股票的预期外收益则由以下公式中的残差项得到；残差序列为股票的预期外收益。当预期外收益为正时，投资者的主动卖出行为可被认为是知情主卖；反之，则被认为是知情主买:

其中，为股票 i 在 n 日第 j 分钟的收益； 为虚拟变量（k=1, 2, 3, 4），分别表示周一至周四；为时间段虚拟变量（k=1, 2, 3），分别表示开盘后 30 分钟、盘中及收盘前 30 分钟；为分钟收益滞后项。

## 基于 TICK 委托与逐笔成交数据的高频因子

1. **开盘后买入意愿占比**

出处：

说明：委托挂单体现投资者尚未释放的交易意愿，而主买/主卖则代表已进行的交易行为。 若将两者结合，可以得到广义的投资者主动买入意愿。开盘后 30 分钟内买入意愿占比越高，投资者的买入意愿越强。

效果：开盘后买入意愿占比因子月均 Rank IC 为 4.34%，年化 ICIR 为 3.43。因子月均多空收益为 1.57%，月度胜率为 85%。月均多头收益为 0.89%，月均空头收益为-0.68%。

因子计算公式：

1. **开盘后买入意愿强度**

出处：

说明：开盘后 30 分钟内的买入意愿强度越高，投资者的买入意愿越稳健。

效果：开盘后买入意愿强度因子月均 Rank IC 为 5.07%，年化 ICIR 为 3.93。因子月均多空收益为 1.82%，月度胜率达 84%。月均多头收益为 0.68%，月均空头收益为-1.14%。

因子计算公式：

附录：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 因子名称 | 数据类型 | 月均 Rank IC | 年化ICIR | 月胜率 | 年化收益 | 最大回撤 |
| 高频偏度 | 分钟 K 线 | 3.41% | 3.48 | 83% |  |  |
| 下行波动占比 | 分钟 K 线 | 3.03% | 3.02 | 78% |  |  |
| 尾盘成交占比 | 分钟 K 线 | 0.0486 | 3.59 |  |  |  |
| 量价相关性 | 分钟 K 线 | >3% | >3 |  |  |  |
| 改进反转 | 分钟 K 线 | >3% | >3 |  |  |  |
| 平均单笔流出金额占比 | 分钟 K 线 | >3% | >3 |  |  |  |
| 大单推动涨幅 | 分钟 K 线 | >3% | >3 |  |  |  |
| 一致交易因子 |  |  |  |  |  |  |
| 开盘后净委买增额占比 | 委托队列 | 0.0419 | 4.24 |  |  |  |
| 开盘后净主买占比 | 逐笔成交 | 0.0331 | 2.51 |  |  |  |
| 开盘后净主买强度 | 逐笔成交 | 0.0446 | 3.29 |  |  |  |
| 大买成交占比 | 逐笔成交 | 0.0435 | 1.96 |  |  |  |
| 开盘后买入意愿占比 | TICK 委托与逐笔成交 | 0.0434 | 3.43 |  |  |  |
| 开盘后买入意愿强度 | TICK 委托与逐笔成交 | 0.0507 | 3.93 |  |  |  |
| 开盘后知情主卖占比 | 逐笔成交 | 0.0286 | 2.51 |  |  |  |
| APM | 分钟 K 线 |  | 2.07 | 0.7 | 0.1099 | 0.0572 |
| 开盘集合竞价成交量占比因子（OCVP） | 分钟 K 线 |  | 3.17 | 0.8125 | 0.1281 | 0.0232 |
| 改进的流动性因子（𝐌𝐢𝐱𝐋𝐈𝐐） | 分钟 K 线 |  | 4.44 | 0.889 |  |  |