

Quant Evaluation - HFT

1. Calculate the following statistics of each stock.

a. 3-day highest/lowest price.

b. 3-day cumulative return/trading amount.

c. The total number of BBO prices changes in 3 days. Here BBO price change is defined as bp1 or ap1 changes.

stock_id	highest	lowest	cum_ret	cum_amount	bbo_change
0	17.18	16.2	0.040681303	269528402	5014
1	4.05	3.75	0.025732756	933473960	1791
2	56.72	52.53	-0.004143775	802933876	8215

2. Standard Dataset.

```
股票0-20250305共有：4741行  
股票0-20250306共有：4741行  
股票0-20250307共有：4741行  
股票1-20250305共有：4741行  
股票1-20250306共有：4741行  
股票1-20250307共有：4741行  
股票2-20250305共有：4741行  
股票2-20250306共有：4741行  
股票2-20250307共有：4741行
```

```
Process finished with exit code 0
```

3. The Distribution of Spread Data.

在所有样本中，前 10% 的 Spread 非常小，几乎稳定在 0.00018，说明绝大多数样本的价差极小；极端大价差情况集中在高分位（75%、90%、95%、99%）说明部分流动性较差的时点或股票价差拉大。

在时间分布中，高分位（95%、99%）在 03-07 有明显抬升，说明该日可能出现了极端价差/低流动性的情况。

在截面分布中，Group 1 的 Spread 整体显著高于 Group 0 和 Group 2，说明该组股票存在更大的买卖价差。

某些快照（snapshot）中出现价差突然增大的可能原因：

- 重大新闻、财报发布、政策公告等发生
- 出现较大的单
- 不同时段也会存在差异（开盘和临近收盘）

Table A: All Sample

Percentile	Spread
1%	0.000181
5%	0.000182
10%	0.000183
25%	0.000549
50%	0.000607
75%	0.002541
90%	0.002567
95%	0.002594
99%	0.002656

Table B: Cross Date

Percentile	20250305	20250306	20250307
1%	0.000181	0.000181	0.000184
5%	0.000182	0.000182	0.000185
10%	0.000183	0.000182	0.000188
25%	0.000546	0.000543	0.000564
50%	0.000608	0.000601	0.0006
75%	0.002528	0.002535	0.002561
90%	0.002567	0.002548	0.002614
95%	0.002574	0.002554	0.002642
99%	0.002601	0.002567	0.002663

Table C: Cross Section

Percentile	0	1	2
1%	0.000592	0.002516	0.000181
5%	0.000594	0.002522	0.000182
10%	0.000595	0.002522	0.000182
25%	0.000598	0.002541	0.000183
50%	0.0006	0.002554	0.000189
75%	0.000607	0.002581	0.000549
90%	0.001201	0.002614	0.000924
95%	0.001214	0.002649	0.001292
99%	0.001813	0.004975	0.002358

4. Choose your optimal solution to define the volatility of the past 20 snapshots

滚动20个快照计算：Spread标准差 × 均值变化率系数，具体地：

先计算 20 个Spread的标准差 std，再得到窗口的趋势斜率slope，最后计算波动率为：

$$\text{volatility} = \text{std} * (1 + \text{abs}(\text{slope}))$$

这样做的好处是既可以衡量波动的幅度，其中的斜率又可以区分缓慢起伏

但如果存在极端值可能会造成影响，并且也不能区分波动的方向

5. Estimate the Active Buy and Sell Volume

基本思路

1. 遍历两个快照时间之间的每一笔成交
2. 若价格 \geq ap1 算作主动买；若价格 \leq bp1 算作主动卖

6. Orders Cancelled

基本思路是：撤单量 = 上一档位挂单量 - 当前档位挂单量 - 在该档位成交的数量

1. 从两个快照中提取每个价格档位的price 和volume
2. 获取总成交量变化
3. 估算成交方向（主动买/主动卖）及对应数量
4. 对每个档位：计算挂单减少量减去估算成交量，剩下部分就是撤单量

7. Define the Price Center

以往的mid_price没有考虑挂单的深度（volume），如果买一和卖一的挂单量非常不对称，mid_price 可能会失效

因此可以用挂单加权中间价来完善，即：

$$\text{weighted_price} = (\text{bid_price} * \text{bid_volume} + \text{ask_price} * \text{ask_volume}) / (\text{bid_volume} + \text{ask_volume})$$

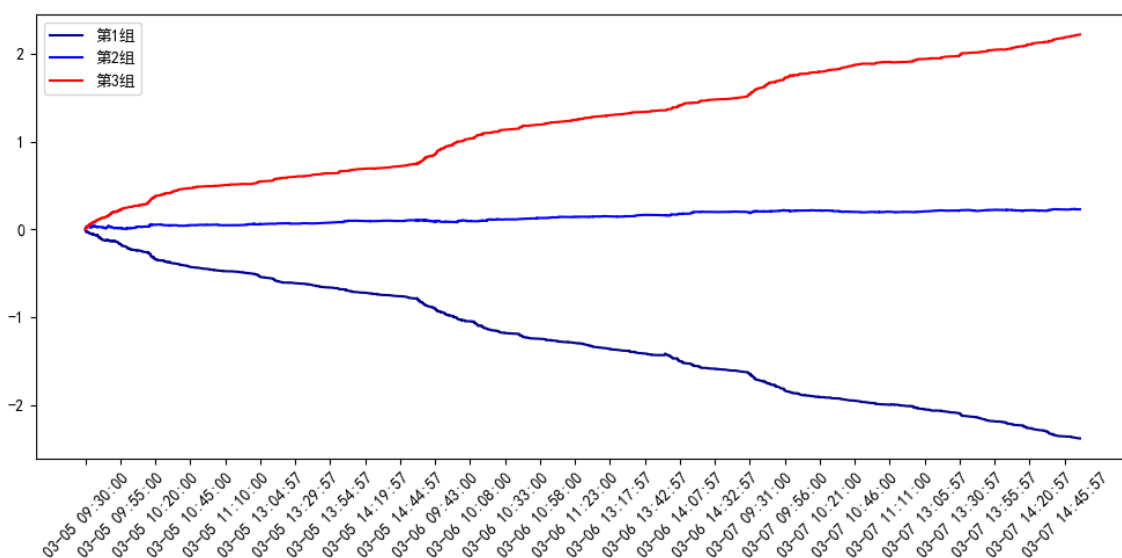
改公式表示，如果卖方挂单多，weighted_price 会偏向卖价；如果买方挂单多，weighted_price 会偏向买价。

8. Statistical Tests of Imbalance of Buy and Sell

只看一档（买一/卖一）可能不够稳定，容易被高频“瞬间挂单”扰动。更改的思路是计算多档累积的挂单不平衡因子，将买1-买N 和卖1-卖N 档位挂单量都考虑进去，即：

$$\text{Cumulative_Imbalance} = (\Sigma bq - \Sigma aq) / (\Sigma bq + \Sigma sq)$$

接下来还是先用 $(bq1 - aq1) / (bq1 + aq1)$ 计算不平衡指标imbalance看看和未来收益的相关性。由于本次检验的截面数据较少只涉及三只股票，因此因子最后呈现的IC均值为0.2453，ICIR为0.3714（当前bar的因子值与未来一个bar的收益）。从下图因子分组累计收益来看，因子值越高，累计收益越高。



不平衡因子分组的累计收益曲线