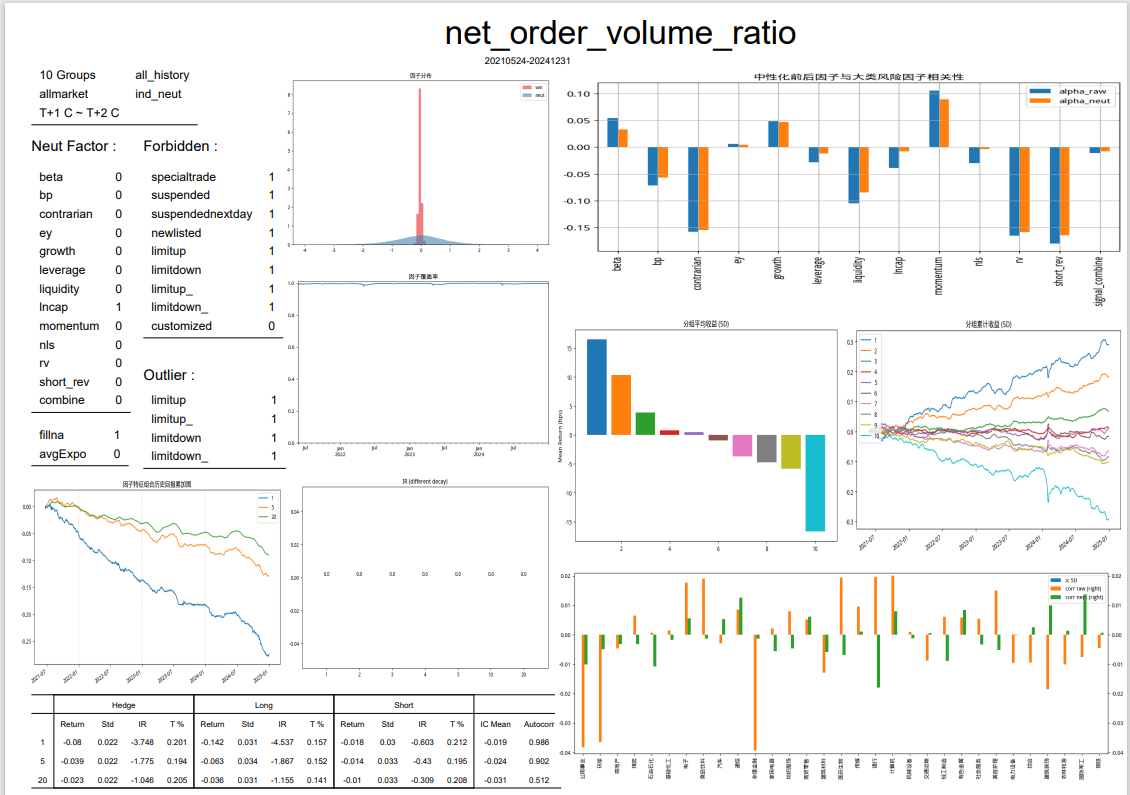
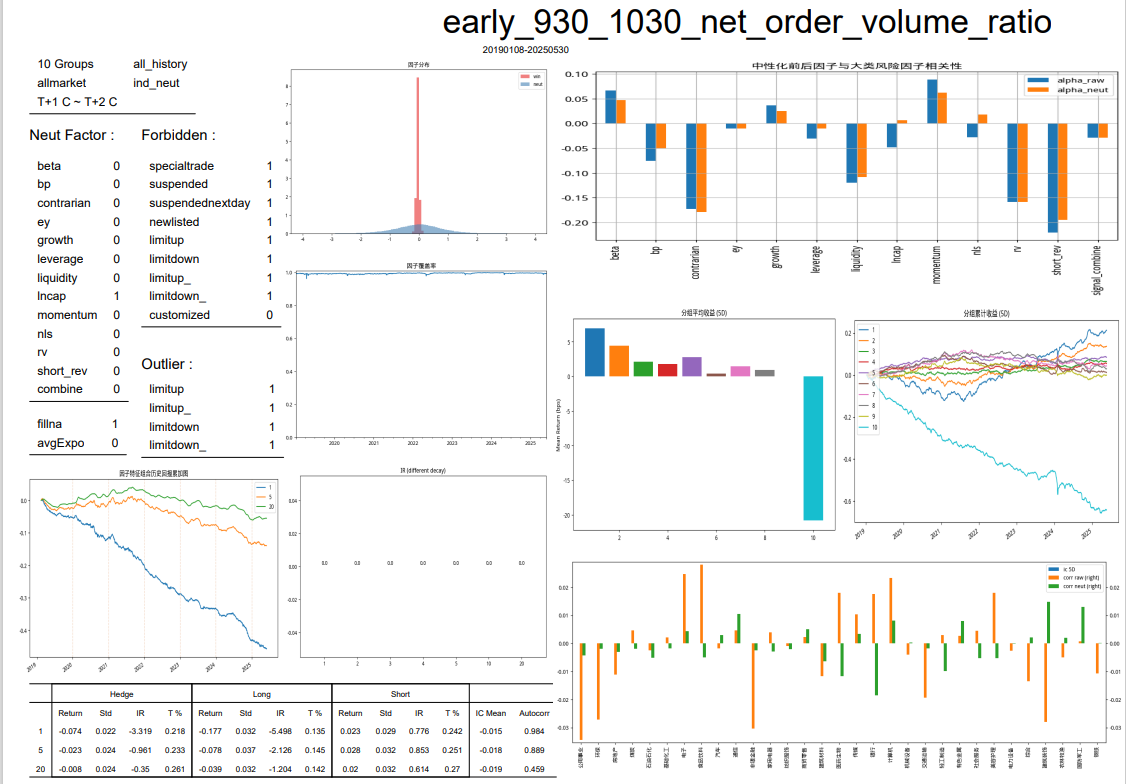
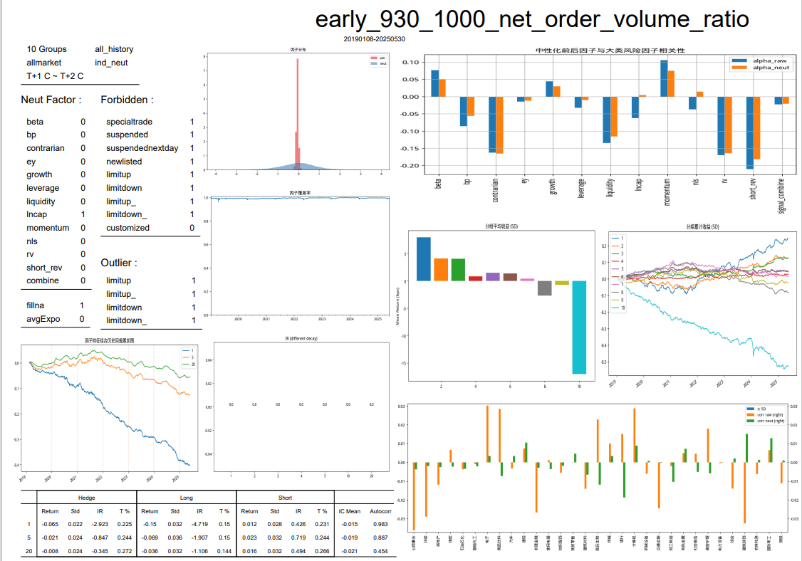
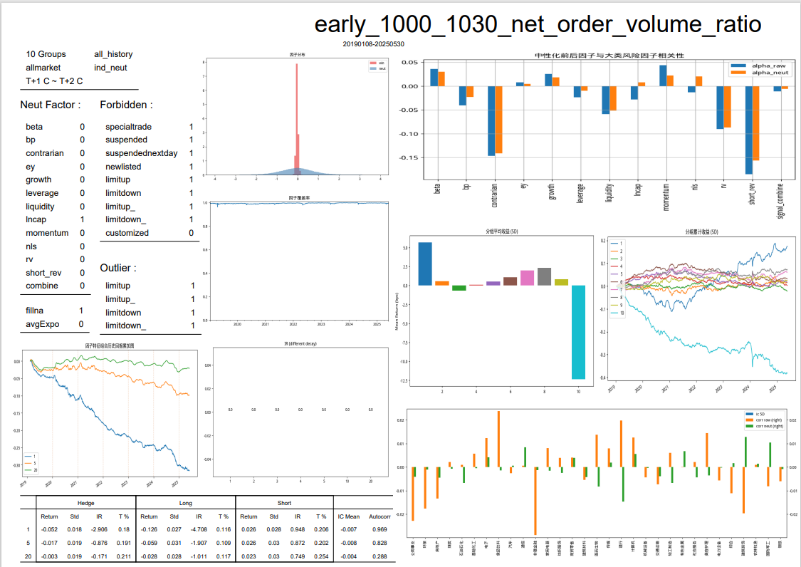
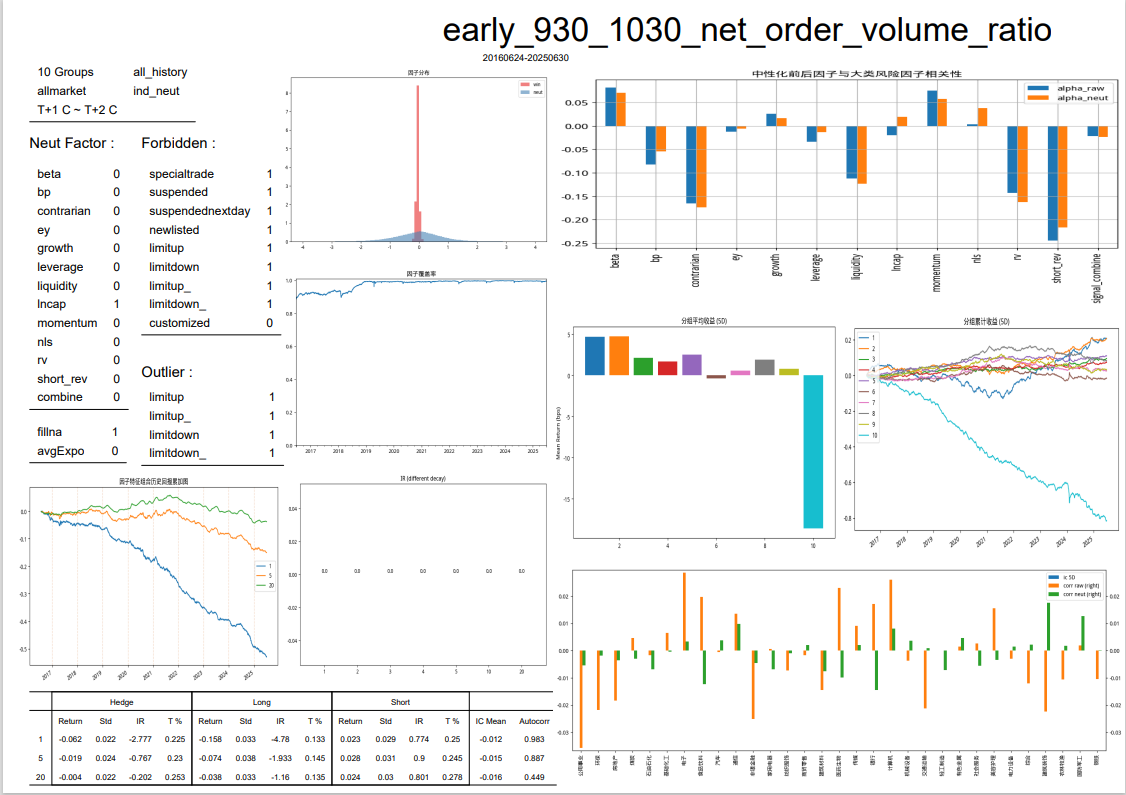
**2021年前后上交所、深交所主笔委托数据结构化差异分析**

**21年之后表现较好**

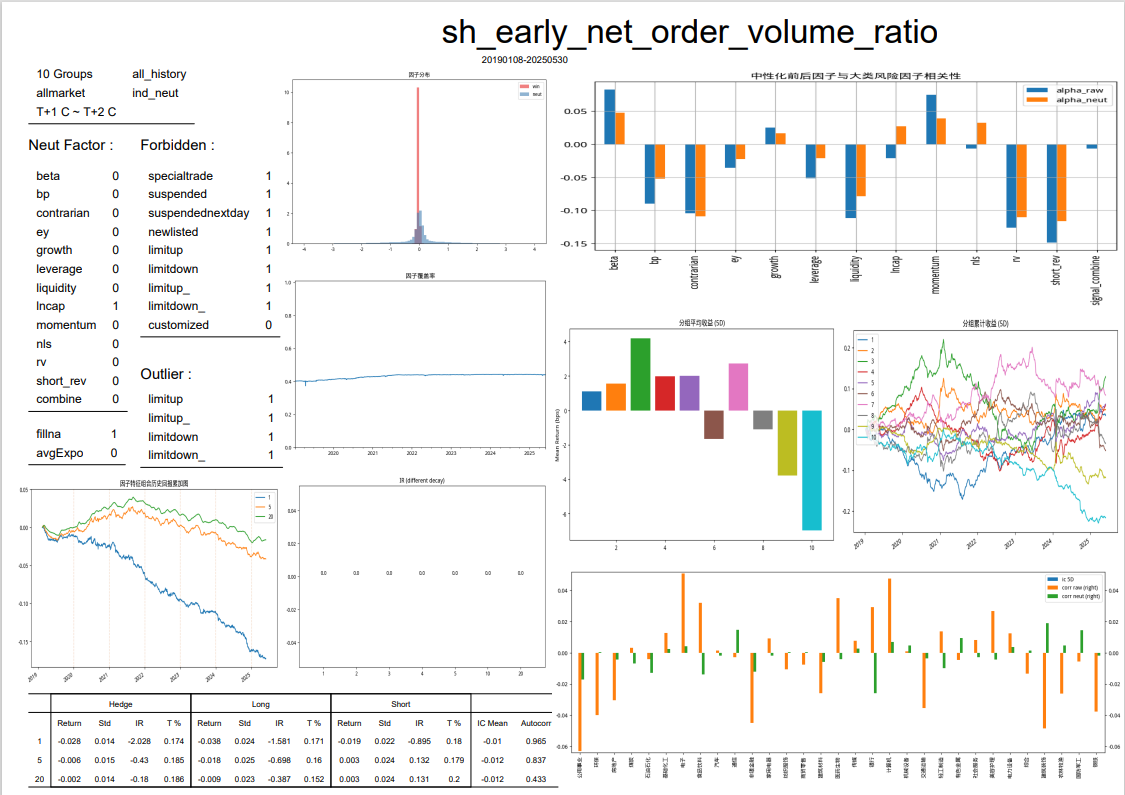
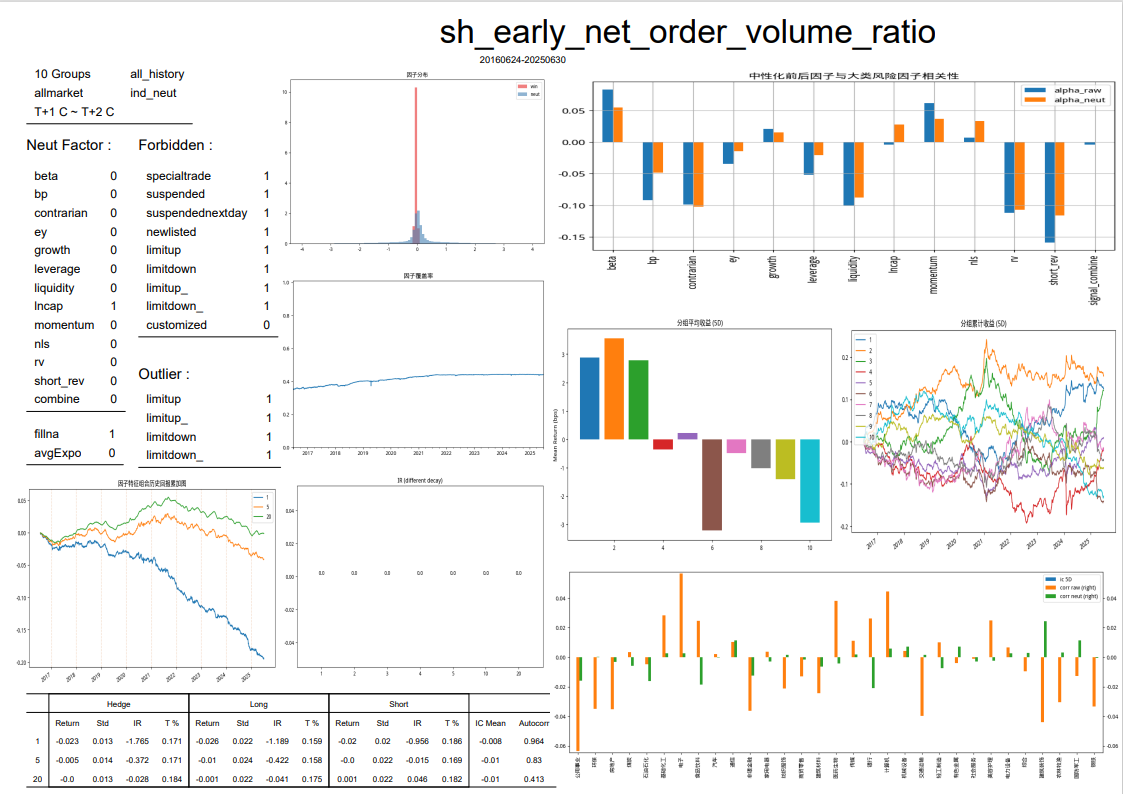
**19-21年表现一般**

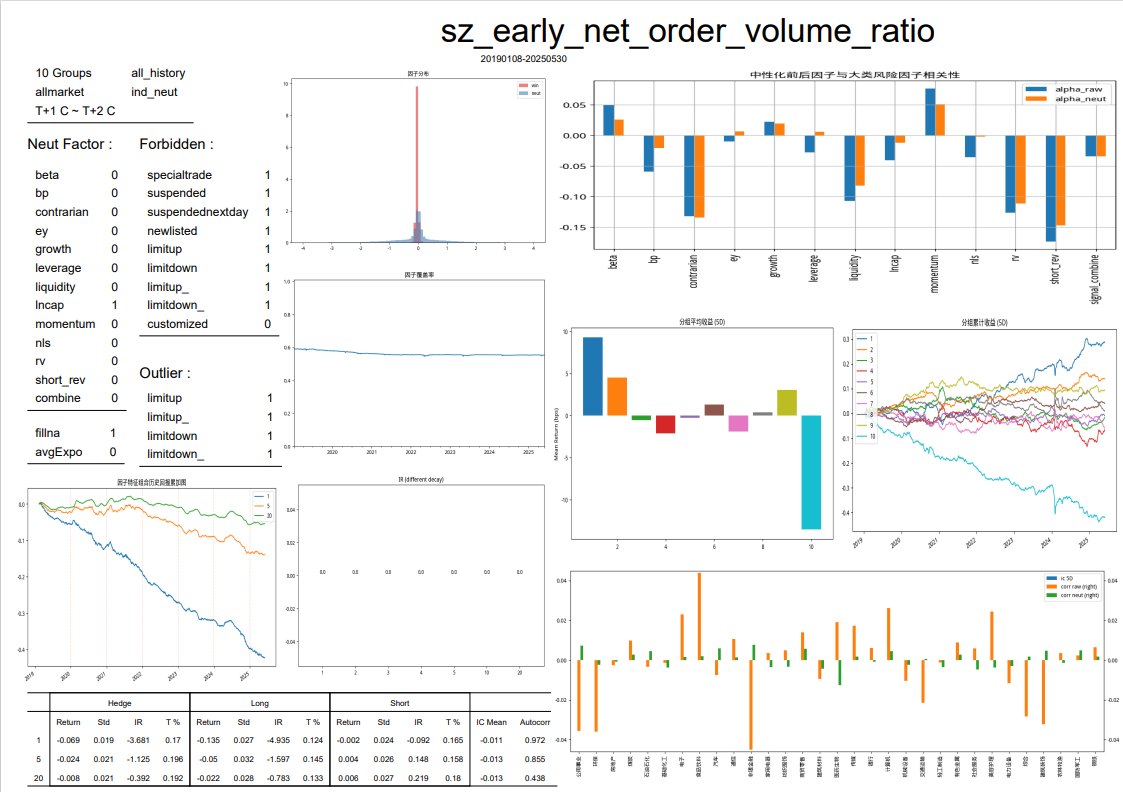
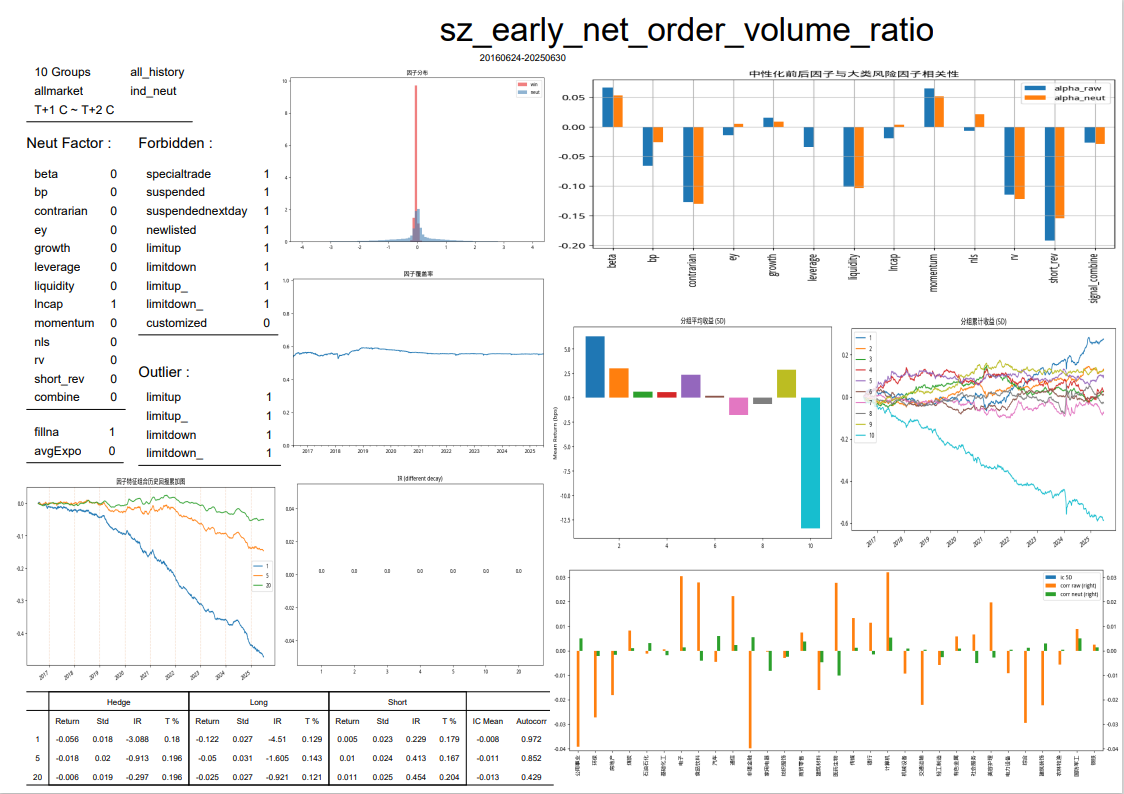
**16-19表现也比较差**

****

**简单处理：区分上交所与深交所股票数据，分别构建因子。回填至2016-06-30再次进行回测，发现效果依然不佳**

**可以看出，在因子本身效果不好的同时，上交所和深交所数据确实有差异。因子效果21年前后的差异应该不是两交易所数据的结构性差异导致的。**

****

****

**# 上交所历史数据接口变动与逐笔委托数据差异说明**

**## 1. 历史数据接口变动说明**

**| 日期 | 版本 | 说明 |**

**| ---- | ---- | ---- |**

**| 2020-02-19 | 1.0.0 | 首次发布 |**

**| 2020-09-23 | 1.0.1 | 1. Level-1 行情快照增加均价字段<br>2. 期权快照和 K 线增加均价、昨天结算价和今日结算价字段 |**

**| 2021-04-23 | 1.0.2 | 1. Level-1 行情快照增加收盘价、消息序号、发送时间字段<br>2. Level-1 行情 K 线增加 K 线数量、交易日字段<br>3. Level-2 行情竞价逐笔成交增加成交序号、频道代码、内外盘标志、业务序列号字段<br>4. \*\*Level-2 行情增加竞价逐笔委托数据\*\*<br>5. 数据文件及字段说明中补充启用时间 |**

**| 2021-11-25 | 1.0.3 | 1. Level-2 行情增加债券逐笔类数据<br>2. Level-2 行情快照新增部分字段<br>3. Level-2 竞价逐笔成交新增发送时间字段 |**

**| 2022-10-27 | 1.1.0 | 1. 调整文档结构，对各行情数据进行归类<br>2. 调整 Level-2 行情快照及 K 线、Level-1 行情快照及 K线的 IOPV 字段精度 |**

**| 2023-05-31 | 1.1.1 | 1. 调整 Level-1 行情快照、盘后固定价格交易快照及 K 线的成交金额精度<br>2. 调整 Level-2 行情快照、盘后固定价格交易快照、盘后固定价格交易逐笔成交及 K 线的成交金额精度<br>3. 调整股票期权行情快照及 K 线的成交金额精度 |**

**| 2023-12-13 | 1.1.2 | \*\*Level-2 行情中增加竞价逐笔合并数据\*\* |**

**| 2024-06-21 | 1.1.3 | 1. Level-2 行情的逐笔类数据中，竞价逐笔成交（Tick.csv）和逐笔委托（Entrust.csv）文件自 2024 年 7 月 8 日起不再提供；之后竞价逐笔类数据将通过竞价逐笔合并数据文件 sh2\yyyymmdd\StockTick.csv 发布 <br>2. 部分描述调整 |**

**## 2. Level-2行情数据重点信息**

**Level-2 行情包括快照类数据、逐笔类数据和 K 线数据。**

**### 2.1 快照类数据重点信息**

**- 开盘集合竞价数据仅提供 2018 年 12 月 24 日前全部证券的开盘集合竞价数据，之后通过行情快照文件 sh2\yyyymmdd\Snapshot.csv 发布。**

**- Level-2 行情快照数据的快照间隔为 3 秒或 5 秒（其中，部分指数快照间隔为 5 秒，个股及部分指数快照间隔为 3 秒，相邻快照如果完全相同就保留第一幅），将全天所有证券的快照保存在同一个文件中。收盘集合竞价数据包含在 Level-2 行情快照中。**

**- Level-2 盘后固定价格交易行情快照针对科创板股票，自 2019 年 7 月 22 日启用, 快照间隔为 3 秒。**

**### 2.2 逐笔类数据重点信息**

**- Level-2 逐笔类数据包括竞价逐笔成交、盘后固定价格交易逐笔成交、竞价逐笔委托和债券逐笔类数据。**

**- 2021-04-23起，上交所Level-2 行情增加竞价逐笔委托数据**

**- 竞价逐笔委托数据与竞价逐笔成交数据将于 2024 年 7 月 8 日起不再提供，合并为竞价逐笔合并数据文件 sh2\yyyymmdd\StockTick.csv 发布**

**- 竞价逐笔合并数据包括新增委托订单和删除委托订单（撤单）、产品状态订单及成交，自 2023年 11 月 27 日启用**

**### 2.3 K 线数据重点信息**

**- K 线数据包括日 K 线和分钟 K 线两种，分别保存在独立的文件中。**

**- 科创板（非必有）日 K线中自 2019 年 7 月22 日启用。**

**## 3. 上交所与深交所逐笔委托数据差异说明**

**上交所的逐笔委托数据中的委托数量不同于深交所。它代表的是经过一次\*\*成交撮合后剩余数量\*\*，具体来说当某一个订单经过一次撮合全部成交的话，是不会有对应逐笔委托数据的，只会有逐笔成交数据。**

**### 3.1 一次撮合成交定义**

**一次撮合成交：订单发送到交易所的撮合平台能立即成交，不需要在订单队列中排队。比如【市价单】，限价单的【主动买入】（大于等于当前盘口的卖一价的买入）或【主动卖出】（小于等于买一价的卖出）。**

**### 3.2 数据识别方法**

**一般很容易在逐笔成交中查找到这类订单，连续竞价阶段：**

**- 若主动买【买方订单号=卖方订单号】或【买卖标志='B'】，将无法在逐笔委托中查找到买方的原始委托；**

**- 若主动卖【买方订单号=卖方订单号】或【买卖标志='S'】，将无法在逐笔委托中查找到卖方的原始委托。**

**### 3.3 差异的持续性**

**尽管上交所在2023年12月13日新增了竞价逐笔合并数据，但上述数据差异仍然存在：**

**- \*\*上交所\*\*：**

**- 逐笔委托数据记录的是经过一次成交撮合后的剩余数量**

**- 一次撮合全部成交的订单不会在逐笔委托数据中出现**

**- 需要结合逐笔成交数据才能获得完整市场信息**

**- \*\*深交所\*\*：**

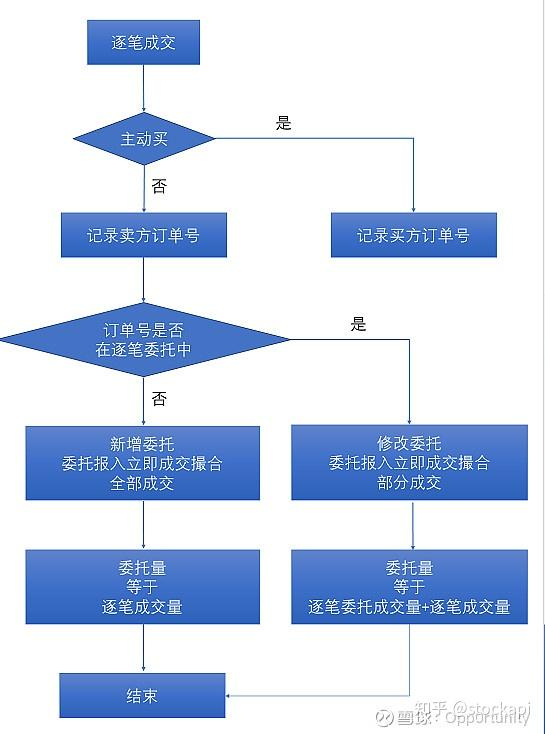
**- 记录的是原始委托数量**

**- 所有委托都会记录**

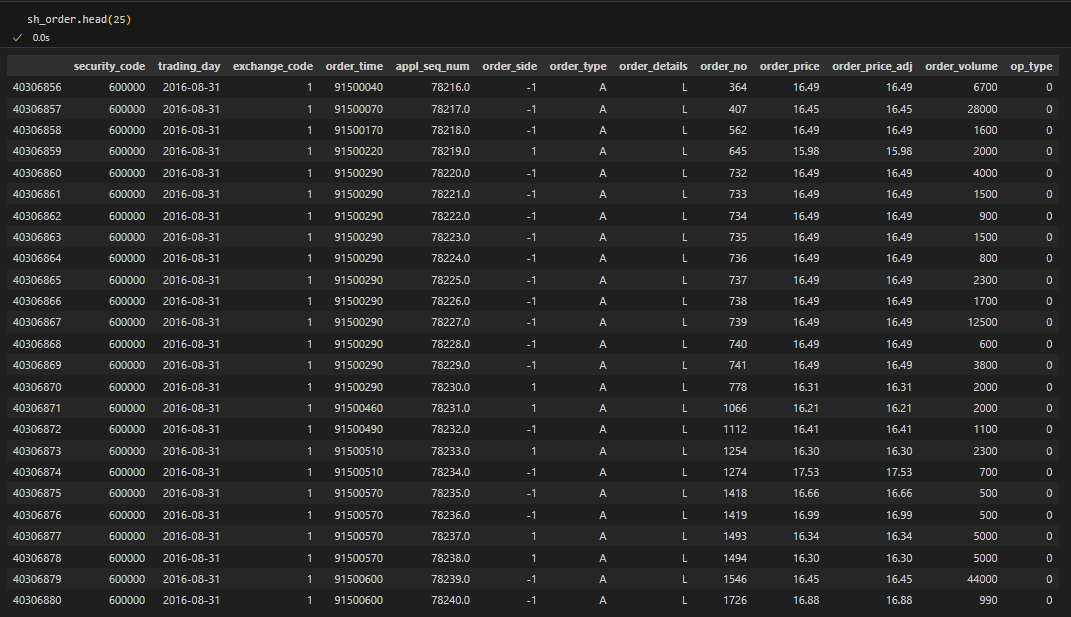
**### 3.4 数据处理建议**

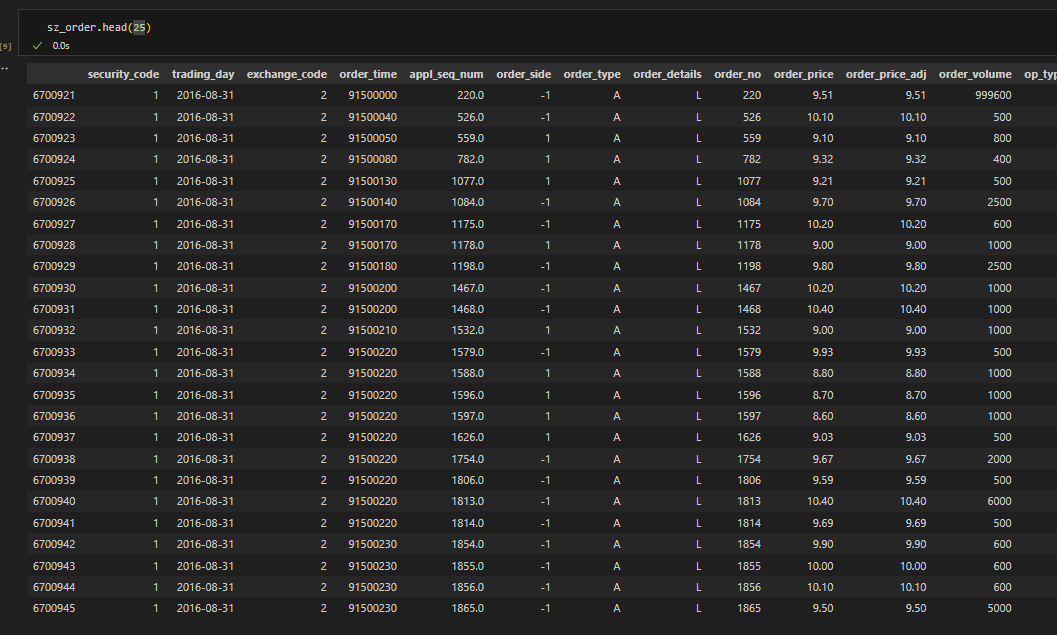
**通过上交所成交数据重构委托数据，可以获得更完整的市场信息。在分析上交所数据时，需要注意这一特殊性，并结合逐笔成交数据进行综合分析。**

**可参考知乎文章： https://zhuanlan.zhihu.com/p/1932952519044206797**

****

**数据概览：时间颗粒度上差距不大**

****

****