

打开高频交易的黑箱

项目要求

1. 根据项目描述, 使用 **Python** 完成下面的 10 道问题, 项目持续时间 2 天
2. 请在项目完成后, 提交测试报告, 格式: **策略_姓名.html**(jupyter notebook 导出为 html), 最终评价会从思路、代码、报告形式以及已有经验等多方面综合考虑
3. 项目过程中遇到任何问题请与面试负责人沟通

数据说明

./data/ 路径下有三份 csv, 分别: **0.csv**、**1.csv**、**2.csv**, 表示某只股票连续三日的 tick 级别交易行情数据。以 0.csv 为例, 每一行表示一个 tick 截面, 共包含 49 个字段, 具体说明见"字段说明.png"

为了简化分析, 本项目中提供 09:30:00~11:30:00、13:00:00~14:56:57 两个连续竞价时间段的 tick 行情

提示: 下面问题需分析多日行情, 提升结论的可信度

问题描述

1. 计算三个交易日中该股票的最高价、最低价、3 日累计涨幅、3 日累计成交金额、3 日平均振幅、3 日平均盘口变化次数(当前 tick 的 ap1 或 bp1 和上个 tick 不一样), 结果需要打印出来
2. 观察该股票日内相邻 tick 间成交金额的变化, 分析可能的原因
3. 理想情况下, tick 行情会 3s 推送一次, 但实际中一些因素的影响, 导致收到的行情是非均匀的甚至 tick 缺失, 请设计合适的清洗方式, 使得新数据以 09:30:00 起始、14:56:57 结束, tick 频率 3s 一次, 作为标准 tick 数据(每日行情 4741 行)

提示: 下面问题均使用清洗后的标准 tick 数据

4. 定义: 盘口中间价格 $mp = 0.5 * (ap1 + bp1)$ 、盘口价差比例 $spdr = (ap1 - bp1) / mp$, 观察

spdr 数据分布变化，总结有意义的现象，如某些 tick spdr 突然变大的原因

5. 选择你认为最优的方案，对于任意时刻 t ，定义过去 20 个 tick 的波动性，需考虑震荡上涨/下跌和快速上涨/下跌的区别，并分析该定义的优缺点
6. 在任意时间 t ，定义：订单薄的买卖不均衡 $\text{imbalance} = (\text{bv1} - \text{av1}) / (\text{bv1} + \text{av1})$ ，买方力量强时，因子值大；卖方力量强时，因子值小。我们认为： imbalance 与未来一段时间的价格走势相关。但该 imbalance 的定义过于简单，选择你认为更优的方案，重新定义 imbalance ，解释其背后的逻辑，并通过统计检验分析 imbalance 与未来价格走势的关系
7. 定义：相邻 tick 间的 成交量 = 主动买量 + 主动卖量
主动买 act_buy : 价格 高于或等于 卖方一档价格的买单
主动卖 act_sell : 价格 低于或等于 买方一档价格的卖单
请选择你认为最优的方案，估计相邻 tick 间的主动买卖量，并列出可能存在的问题
8. 根据主动买卖量，设计成交类型因子 trade_fc ，统计检验分析 trade_fc 与未来价格走势的关系
9. 定义撤单：相邻 tick 间，订单簿 orderbook 上每一档撤消的挂单量。根据 10 档 orderbook 的挂单量以及估算的主动买卖量，估计相邻 tick 间每一档价格上对应的撤单量
10. 结合对该股票的其他观察，统计和猜测，有什么可以盈利的策略？写出策略的思路、可以盈利的原因以及什么时候可能不盈利（提示：不计手续费；可以低买高卖/高卖低买）

最后结语

做到这里，想必你对高频交易已经有了一些了解：于细小处见甚微，于细小处见大势。我们从微小的盘口，流动性出发，推测市场上博弈的变化，激流勇进，其乐无穷。我们希望你可以从这个项目中得到不一样的启发，窥一斑而知豹；也希望你能提出可行的优化建议或者更期待的项目形式，期待你精彩的回答，我们下一轮面试见。