面向高频交易场景的异常报价行为智能监测技术

蒋锐权

（上海金仕达软件科技股份有限公司，上海 201203）

摘　要： 为提升高频交易场景下异常报价行为的识别能力，基于逐笔委托、逐笔成交和盘口数据，构建多时间尺度异常报价监测方法。通过数据质量控制与时间对齐，从申报强度、撤单行为、订单存续、成交转化、价格偏离和盘口失衡6个维度提取18项报价特征，并对密集报价、快速撤单、价格偏离和盘口失衡等异常行为进行判别。监测结果表明，波动行情下的异常率明显高于平稳行情，且异常事件主要集中于少量高风险时间窗口。该方法能够反映不同交易状态下异常报价特征的变化，为异常报价实时筛查、重点时段复核和风险分级提供数据支撑。

关键词： 高频交易；异常报价；实时监测

中图分类号： TP391　　**文献标识码：** A

Intelligent monitoring technology for abnormal quotation behavior in high-frequency trading scenarios

JIANG Ruiquan

（Shanghai Kingstar Fintech Co., Ltd., Shanghai 201203, China）

Abstract: To improve the identification of abnormal quotation behavior in high-frequency trading scenarios, a multi-timescale abnormal quotation monitoring method is developed based on order-by-order data, trade-by-trade data, and order book data. Following data quality control and time alignment, 18 quotation features are extracted from six dimensions, including order submission intensity, order cancellation behavior, order lifetime, trade conversion, price deviation, and order book imbalance. Abnormal behaviors such as intensive quoting, rapid order cancellation, price deviation, and order book imbalance are then identified. Monitoring results show that the abnormality rate under volatile market conditions is significantly higher than that under stable market conditions, and abnormal events are mainly concentrated within a small number of high-risk time windows. The proposed method can capture changes in abnormal quotation characteristics under different trading conditions, providing data support for real-time screening of abnormal quotations, review of key time periods, and risk classification.

Key words: high-frequency trading, abnormal quotation, real-time monitoring

1　引言

高频交易具有报价密集、订单更新频繁等特征，传统基于低频统计的监测方式难以及时识别短时异常报价。异常申报、快速撤单等行为不仅增加了市场微观结构分析的难度，也对实时风险识别提出了更高要求。基于此，本文围绕逐笔委托、逐笔成交及盘口数据，构建多时间窗口特征体系和异常识别方法，并结合不同交易状态对监测结果进行分析，旨在为异常报价实时筛查、重点时段复核及风险分级提供佐证依据。

2　高频交易异常报价监测方案设计

2.1　监测指标体系与数据采集对象

监测对象包括逐笔委托、逐笔成交和盘口快照数据，核心指标有报价频率、撤单比例、订单存续时长、价格偏离度和成交转化率等。采集字段包括证券代码、订单编号、买卖方向、价格、数量、申报时间、撤单时间、成交数量及最优买卖价，时间戳精确至毫秒。在此基础上，采用不同时间窗口识别密集申报、连续撤单及报价偏离行为，并按照“申报—撤单—成交”关联事件链，统计委托、撤单和成交笔数，从而统一数据统计边界[1]。

2.2　数据采集组件与部署配置

数据采集组件按照高吞吐、低时延和连续接入要求进行配置。行情接入端采用多线程异步I/O方式接收逐笔委托、逐笔成交及盘口数据，服务器配置32核处理器、128 GB内存和双10 GbE网卡，原始数据存储于NVMe固态硬盘中[2]。采集节点采用双机并行部署，主节点负责实时数据接入，备用节点同步进行数据备份；通过独立接收

队列对不同证券数据进行分流，缓存容量在峰值消息量基础上预留30%以上。数据按照证券代码和交易时段分区写入，并同步记录数据接收时间和原始交易时间，从而支撑后续的质量校验、时间对齐及特征计算工作。

2.3 数据预处理与特征计算

2.3.1 数据质量控制与时间对齐

高频交易数据预处理以事件完整性、时序一致性和字段有效性为质量控制重点，具体流程如图1所示。

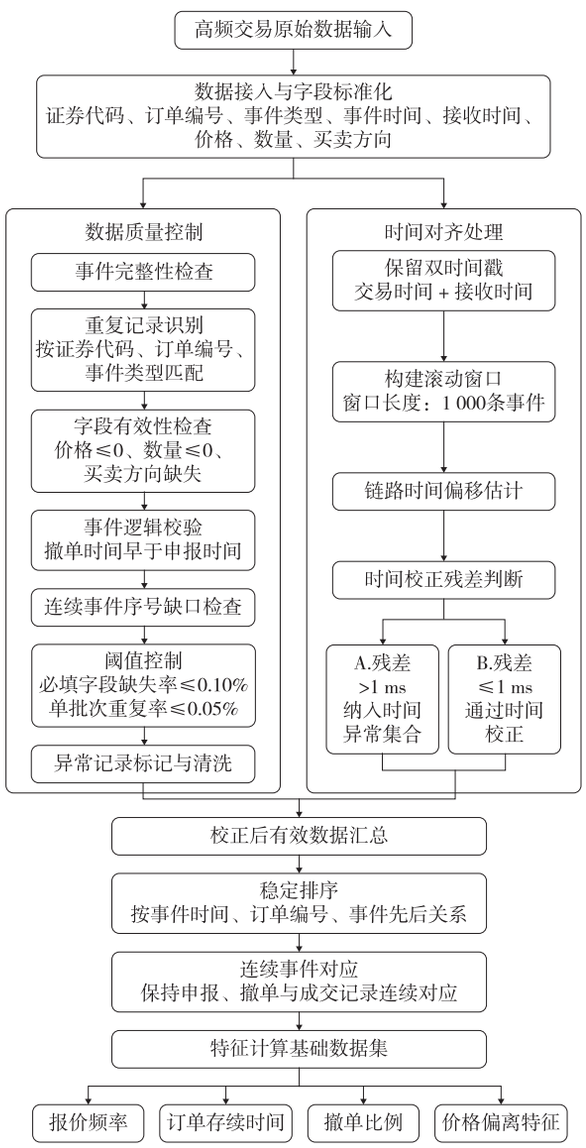

图1 高频交易数据质量控制与时间对齐流程

系统按照证券代码、订单编号和事件类型进行数据去重，并对价格≤0、数量≤0、买卖方向缺失以及撤单时间早于申报时间等异常情况进行标记；必填字段缺失率和单批次重复率阈值分别设置为0.10%与0.05%。在时间对齐过程中，同时保留交易时间和接收时间，并采用1 000条事件构成的滚动窗口估计链路时间偏移。当单条记录校正残差超过1 ms时，将其纳入时间异常集合[3]。时间校正量按下式计算：

$$\Delta t_i=(t_i^r-t_i^e)-\mathrm{median}_{j\in W_i}(t_j^r-t_j^e) \tag{1}$$

式中，Δt_i表示第i条事件的相对时间偏移；t_i^r表示数据接收时间；t_i^e表示交易事件原始时间；W_i表示包含第i条记录的滚动窗口；median表示窗口内时间差中位数。时间校正后的数据按照事件时间、订单编号及事件先后关系进行稳定排序，使申报、撤单与成交记录保持连续对应，并保证报价频率、订单存续时长、撤单比例及价格偏离特征计算的一致性。

2.3.2 特征计算方法

完成交易事件清洗与时间校正后，按照账户、证券代码和时间片对数据进行分组，并从申报强度、撤单行为、订单存续、成交转化、价格偏离和盘口失衡6个维度提取18项连续特征[4]。申报强度分别根据10 ms，50 ms和200 ms时间窗口内的新增订单数量进行计算。当单个窗口内委托笔数超过该证券当日同期历史均值的3倍时，将该窗口标记为高密度报价样本。撤单特征主要统计撤单笔数、撤单数量和撤单间隔，在此基础上构建窗口撤单强度：

$$C_t=\frac{N_t^c}{N_t^o+\varepsilon}\times\frac{V_t^c}{V_t^o+\varepsilon}\times\ln\left(1+\frac{\tau_0}{\bar{\tau}_t+\varepsilon}\right) \tag{2}$$

式中，C_t表示撤单强度；N_t^c和N_t^o分别表示撤单笔数与申报笔数；V_t^c和V_t^o分别表示撤单数量与申报数量；$\bar{\tau}_t$表示窗口内撤单订单平均存续时间；τ_0设为1 000 ms；ε取10^{-8}。

订单存续特征分别统计存续时间在10 ms以内、10~100 ms、100~500 ms和500 ms以上的订单占比，用于区分快速撤单与正常挂单。价格偏离特征采用同期中间价进行标准化：

$$D_i=\frac{|p_i-m_t|}{s_t+\varepsilon}\times\ln\left(1+\frac{q_i}{\bar{q}_t+\varepsilon}\right) \tag{3}$$

式中，D_i表示第i笔申报的价格偏离值；p_i表示申报价格；m_t表示同期买一价与卖一价的均值；s_t表示最近50笔有效报价价格的标准差；q_i表示申报数量；$\bar{q}_i$表示同期平均申报数量。盘口特征读取前5档买卖挂单量，并计算失衡度：

$$I_t=\frac{\sum_{k=1}^{5}V_{k,t}^{\mathrm{bid}}-\sum_{k=1}^{5}V_{k,t}^{\mathrm{ask}}}{\sum_{k=1}^{5}V_{k,t}^{\mathrm{bid}}+\sum_{k=1}^{5}V_{k,t}^{\mathrm{ask}}+\varepsilon} \tag{4}$$

式中，I_t取值范围为$[-1,1]$；$V_{k,t}^{\mathrm{bid}}$和$V_{k,t}^{\mathrm{ask}}$分别表示第k档买方与卖方挂单量。上述18项特征按照时间窗口连续输出，将短时密集申报、高比例撤单、低成交转化、异常价格偏离和盘口单侧堆积等行为转化为统一的数值特征，从而支持后续异常报价判别[5]。

3 高频交易异常报价监测结果与应用

3.1 数据处理与异常判别

异常报价监测以连续竞价阶段的逐笔委托、逐笔成

交和5档盘口数据为分析对象，观测时段包括09:30—11:30和13:00—14:00，并根据价格波动幅度与订单流变化情况划分为平稳行情和波动行情两类交易状态。完成交易状态划分后，按照10 ms，50 ms和200 ms时间窗口对18项报价特征进行汇总，并以同一证券、同一交易时段的历史分布为参照开展异常判别。

具体判别规则如下：(1)密集报价根据窗口内申报笔数进行判别。例如，10 ms窗口内申报28笔，而同期均值为7笔，申报强度达到同期均值的4倍，则将该窗口判定为高强度报价。(2)快速撤单根据撤单率和订单存续时长进行判别。例如，50 ms窗口内共有40笔申报，其中34笔被撤回(撤单率为85.0%)，且其中26笔订单存续时间小于10 ms，则判定为集中快速撤单。(3)价格与盘口异常采用联合判别方式。当200 ms窗口内价格偏离程度超过同期3倍标准差、成交转化率低于20.0%，且5档盘口失衡度绝对值高于0.70时，将该时间窗口纳入异常报价集合，并保留订单编号、申报时间、撤单时间、价格和数量等事件链信息，用于后续行情比对与异常复核。

3.2　监测结果分析

典型交易日的平稳行情与波动行情监测结果分别如表1、表2所列。

表1　场景A(平稳行情)监测结果

时段	总报价笔数	异常标记数	异常率/%	最高异常评分
09:30—10:30	285 432	127	0.044	0.72
10:30—11:30	302 187	156	0.052	0.78
13:00—14:00	267 894	103	0.038	0.69

表1数据显示，场景A共记录855 513笔报价，其中386笔被标记为异常，整体异常率为0.045%。3个时段的异常率分布为0.038%~0.052%，最高异常评分为0.78。由此可见，平稳行情下异常记录数量较少，各时段之间差异较小，短时密集申报、快速撤单和价格偏离等特征未出现明显集中叠加。

表2　场景B(波动行情)监测结果

时段	总报价笔数	异常标记数	异常率/%	最高异常评分
09:30—10:30	412 567	1 843	0.447	0.94
10:30—11:30	398 234	2 106	0.529	0.96
13:00—14:00	435 678	1 967	0.451	0.93

表2数据显示，场景B共记录1 246 479笔报价，其中5 916笔被标记为异常，整体异常率为0.475%，约为场景A的10.52倍。其中，10:30—11:30时段异常率最高(0.529%)，最高异常评分达到0.96，表明该时段的报价频率、撤单行为和盘口偏离等特征变化更为明显。此外，场景B 3个时段的最高异常评分均高于0.90，但异常报价占全部报价的比例仍低于0.60%，异常记录主要集中于少量高风险时间窗口，并未出现大范围连续触发。

对比两类行情可以看出，波动行情下的异常率和异常评分均明显高于平稳行情，说明该监测方法能够反映不同交易状态下异常报价特征的变化。然而，当前分析缺少人工复核标签或已确认异常样本作为对照，无法计算检出率和误报率，因此异常率只能反映异常标记情况，不能直接用于评价监测准确性。

3.3　监测结果应用

结合异常标记、异常评分、时间窗口特征和原始订单链，可对监测结果进行分层应用，并根据平稳行情与波动行情下的差异开展风险筛查和业务复核。

(1)异常报价实时筛查。按照10 ms，50 ms和200 ms时间窗口汇总监测结果，对异常评分大于0.90，且同时出现高频申报、快速撤单或盘口失衡特征的时间片进行优先标记，并进一步关联对应订单信息。

(2)重点账户与异常时段复核。按照账户、证券和交易时段归集订单链，对异常率较高或异常评分集中的时间段进行重点核查。例如，场景B中10:30—11:30时段异常率达到0.529%，最高异常评分为0.96，可结合撤单比例、订单存续时长和价格偏离程度进一步分析相关账户的报价行为。

(3)监测参数校准与风险分级。将平稳行情下0.038%~0.052%的异常率作为参考，并结合波动行情中异常事件的持续时长、发生频次和异常评分变化对监测参数进行动态调整，从而降低单一指标触发造成误判的可能性，同时依据异常程度划分不同风险等级。

4　结束语

本文基于逐笔委托、逐笔成交和盘口数据，构建了面向高频交易场景的异常报价监测方法，通过多时间尺度特征计算、事件链关联和异常阈值判别，实现对密集报价、快速撤单、价格偏离和盘口失衡等行为的持续监测。应用结果表明，波动行情下的异常率和异常评分明显高于平稳行情，异常事件主要集中于少量高风险时间窗口，表明该方法能够反映不同交易状态下异常报价特征的变化。后续可进一步引入人工复核标签、已确认异常样本及跨交易日数据，持续优化判别阈值与特征组合，并结合检出率、误报率等指标完善监测效果评价，从而提升复杂市场状态下异常报价监测的稳定性和准确性。

参考文献：

[1] 李佳熠，陈石，孔傲，等.中国股市谁在提供流动性？——基于订单簿下单行为的博弈分析[J].管理科学学报，2024，27(12)：103-115.

[2] 张小日，孙芳芳，叶强.A股程序化交易与盈余公告前股价

(下转第127页)