

2021 年之后中国学者的研究

2024.1.11-2024.1.18

● 日内高频数据因子

● 尾市交易操纵：当天交易结束前 15 分钟内监测指标出现异常变化

当天交易结束前 15 分钟内监测指标出现异常变化，即： $|\Delta EOD_{it} - \Delta EOD_i| >$

$3\sigma_i$ ，其中  $\Delta EOD_{it} = (X_{eod, it} - X_{eod-15ms, it}) / X_{eod-15ms, it}$ ，表示股票  $i$  在交易日  $t$  的尾市监测指标相对收盘前 15 分钟的变化率； $\Delta EOD_i = 1/30 \times \sum_{t=-30}^{t=-1} \Delta EOD_{it}$ ，表示交易日  $t$  前 30 个交易日的滚动窗口下交易结束前 15 分钟监测指标变化率的均值， $\sigma_i$  为相同时间窗口下  $\Delta EOD_{it}$  的标准差。

[1]孙广宇,李志辉,杜阳等.市场操纵降低了中国股票市场的信息效率吗——来自沪市 A 股高频交易数据的经验证据[J].金融研究,2021(09):151-169.

● 股票流动性：相对报价价差  $SPREAD_{i,t}$

采用张峥等（2014）的计算方法。根据中国股市交易机制的特点，用买卖价格的绝对报价价差除以买卖价格的中点价格得出百分比，再对日内实时的相对报价价差作时间加权平均，得到  $t$  日相对报价价差  $SPREAD_{i,t}$ ，数值越大，表示交易成本越高，即流动性越差。具体见（4）式与（5）式。买卖价差是以“报价”为基础，不一定最终成交，能更好地反映投资者的参与度和关注度，而信息机制满足该特征。

$$SPREAD_{i,t} = \sum w_{\Delta t} \times SPREAD_{i,\Delta t} \quad (4)$$

$$SPREAD_{i,\Delta t} = \frac{(\text{卖价} - \text{买价})}{(\text{卖价} + \text{买价})/2} \quad (5)$$

[2]杨何灿,吴隽豪,杨咸月.北向资金与境内股票市场流动性——基于高频数据的传导机制[J].经济研究,2023,58(05):190-208.

● 买卖不平衡指标：

Lee&Ready(1991)提出日内分笔交易中的主动性买单与主动性卖单识别方法，高于买卖报价

中点的成交单定义为主动性买单，低于报价中点的成交单则为主动性卖单。当成交单报价等于买卖报价中点时，高于前一笔成交价定义为主动性买单，反之为主动性卖单。买卖不平衡指标定义为主动性买单减去主动性卖单的占比，记为

$OIB_{i,t}$ ，正值表示主动性买入的力量强于主动性卖出的力量。

$$OIB_{i,t} = \frac{\text{主动性买入成交额} - \text{主动性卖出额}}{\text{股票流通盘市值}}$$

[2] 杨何灿,吴隽豪,杨咸月.北向资金与境内股票市场流动性——基于高频数据的传导机制[J].经济研究,2023,58(05):190-208.

2024.1.18-2024.1.25

● 高阶矩与资产价格的关系

Campbell and MacKinlay (1997)

不仅证实资产收益的分布具有非对称性和尖峰厚尾特征，并事实表明越是高频的数据，收益率数据越偏离正态分布。

Merton (1980)

第一次指出，随着采样频率的增加，波动率可以任意精确地测量，构建了根据日内收益的平方和计算的已实现波动率的估计方法，进而得到已实现偏度和峰度，该方法被广泛应用于日内高阶矩的测度。

Amaya(2015)

基于美股长达 20 年的高频数据构建了已实现高阶矩指标，实证表明已实现偏度与未来收益率存在显著负相关。

[3]Diego Amaya, Peter Christoffersen, Kris Jacobs & Aurelio Vasquez. (2015). Does realized skewness predict the cross-section of equity returns?. *Journal of Financial Economics* (1), 135-167.

陈坚、张轶凡(2018)

当前较低的已实现偏度可以显著预测下个月中国股票市场较高的超额收益率  
基于日内高频数据估计已实现高阶矩（已实现波动率、已实现偏度、已实现峰度）

本文参考 Amaya 等 (2015)<sup>[20]</sup>，给出已实现二至四阶矩的计算方法：

$$RDVar_t = \sum_{i=1}^N r_{t,i}^2, \quad (1)$$

$$RSDkew_t = \frac{\sqrt{N} \sum_{i=1}^N r_{t,i}^3}{RDVar_t^{\frac{3}{2}}}, \quad (2)$$

$$RDKurt_t = \frac{N \sum_{i=1}^N r_{t,i}^4}{RDVar_t^2}, \quad (3)$$

其中，式(1)-(3)中 $r_{t,i}$ 为相邻样本的对数差价，即收益率序列，N 为每日的采样的价格频数。文本基于中国 A 股市场每个交易日 5 分钟高频数据，从 9:30-11:30 到 13:00-15:00 两个交易时段，共 240 分钟，得到 48（N=48）条日内收益率。基于式(1)-(3)得到日内高阶矩数据。

（可进一步通过周标准化处理，得到周度的已实现波动率、已实现偏度、已实现峰度。）

[4]潘娜,李子洋,周勇.已实现波动、偏度和峰度与股票未来收益——来自中国 A 股市场的证据[J/OL].数理统计与管

理,2023(06):1127-1140[2024-01-23].<https://doi.org/10.13860/j.cnki.sltj.20231030-007>.