第3卷 第3期 计 量 经 济 学 报 Vol. 3, No. 3
2023年7月 China Journal of Econometrics Jul., 2023

**doi:** 10.12012/CJoE2023-0001

# 基于集成神经网络模型的知情交易股票价格异象研究

张学勇, 李沛然

(中央财经大学金融学院, 北京 100081)

**摘 要** 本文使用中国 A 股的日内高频交易数据, 采用集成神经网络算法实现了针对知情交易行为的精准识别并证实 A 股市场中存在与股票知情交易程度相关的定价异象. 研究发现: 知情交易者的交易手法主要包括首尾盘操纵和策略化下单, 具体表现为首尾盘时段内量价指标的异常变化和日内买卖价差、订单簿深度的短期突变, 上述特征均可被本文建立的模型所捕捉. 进一步研究发现, 由于信息不对称所导致的流动性风险使得具有高知情交易倾向的股票需提供额外的风险补偿以吸引普通投资者的进入, 基于本文计算的知情交易指数所构建的多空组合每月可获得 1.38% 的等权收益率. 此外在市值规模较大、流动性较高、机构投资者和大股东持股比例较高的股票中组合收益更加显著. 本文的研究对于完善金融市场监管、提升资本市场定价效率具有一定的启示意义.

**关键词** 知情交易; 集成学习; 高频数据; 金融科技

# The "Informed Trading Anomaly" in China's A-Share Market — Research Based on Ensemble Neural Network

ZHANG Xueyong, LI Peiran

(School of Finance, Central University of Finance and Economics, Beijing 100081, China)

**Abstract** Based on the high-frequency trading data of China's A shares, we used ensemble learning algorithm to achieve accurate identification of informed trading activity, and found corresponding pricing anomaly in China's stock market. We found that: The trading methods of insiders including manipulation at the beginning and end of the market and quick order placement during the trading hours, which is manifested in abnormal changes in volume-price indicators, bid-ask spreads and order book

**收稿日期:** 2023-01-02
**基金项目:** 国家哲学社会科学基金重大项目 (19ZDA098)
**Supported by** Key Program of National Social Science Fund of China (19ZDA098)
**作者简介:** 张学勇, 中央财经大学金融学院院长, 教授, 博士生导师, 研究方向: 金融科技与金融市场, E-mail: zhangxueyong@cufe.edu.cn; 李沛然, 博士研究生, 研究方向: 深度学习与资产定价, E-mail: lipeiran156@163.com.

depth. Further research found that due to the liquidity risk caused by information asymmetry, stocks with high informed trading tendency need to provide additional risk compensation to attract ordinary investors to enter. The monthly long-short portfolio constructed based on our informed trading index can achieve significant excess returns. Besides this anomaly is more prevalent in large-sized stocks, high-liquidity stocks and stocks with high institutional ownership. Our paper is of great value for strengthening the financial market supervision and increasing the pricing efficiency of China's A-share market.

**Keywords** informed trading; ensemble learning; high-frequency data; fintech

## 1 引言

市场微观结构理论指出在股票市场中投资者依据信息产生交易决策并形成证券价格, 而信息在投资者间的不对称性则将投资者分为了知情交易者和非知情交易者两类 (O'Hara (1997)). 知情交易者利用自身的信息优势进行择时交易 (蔡宁 (2012)) 并获得 "信息租金" (沈冰等 (2012)), 或通过与市场中的信息劣势投资者进行交易谋求获利 (陈国进等 (2019)) 的交易行为则构成了市场中的知情交易现象.

为实现维护市场公平、增进资源配置效率等目的, 知情交易的识别与监测一直受到众多学者的广泛关注. 部分文献从投资者的身份出发, 指出机构投资者 (Czech et al. (2021))、公司股东 (Dang et al. (2021), 蔡宁 (2012)) 等传统意义上的信息优势投资者能够在业绩预告 (蔡宁 (2012))、借壳重组 (邵新建等 (2014)) 等公司重大事项公布之前掌握相关的内幕信息并通过交易行为获得显著的投资获利. 近年来伴随数据可获得性的不断提高, 日内交易数据逐步被应用于知情交易识别相关研究. 代表性方法包括使用买卖订单数量估计基于内幕信息产生的订单比例, 以 PIN (probability of informed trading) 指标度量市场中的知情交易程度 (Easley et al. (1996, 2008), Chen and Zhao (2012), 陈国进等 (2019)); 基于收盘前时段的高频交易数据进行尾盘操纵识别 (Aitken et al. (2018), Cumming et al. (2020), 李志辉等 (2018), 孙广宇等 (2021)), 和使用量价指标的异常波动判定知情交易事件 (李志辉和孙广宇 (2020)). 然而已有文献大多从订单均衡性、尾盘操纵等特定角度进行研究, 尚未有学者利用深度学习的高维数据处理能力提出针对知情交易的综合监管框架. 基于此本文利用中国 A 股的日内高频交易数据, 创新性地使用基于 Stacking 算法 (Wolpert (1992), Breiman (1996)) 的集成神经网络 (ensemble neural network), 在全面刻画市场交易行为特征的基础上充分发挥深度学习在大数据处理中的优势, 对市场中的知情交易行为进行逐股、逐日的精准识别. 具体而言, 本文综合考虑了包括首尾盘操纵、策略化下单、日内异常量价波动在内的多种知情交易潜在特征, 在大幅提升识别准确率的情况下对多种监测指标进行了充分的横向对比, 厘清了知情交易监测重点的同时进一步为多角度、全方位的综合监控提供了全新思路与方法指导.

定价异象的存在不仅对有效市场假说提出了挑战同时也扭曲了资产的价格发现过程, 因而受到众多学者的广泛讨论 (De Bondt and Thaler (1985), Jegadeesh and Titman (1993), George and Hwang (2010), Novy-Marx (2013), 朱红兵和张兵 (2020), 何诚颖等 (2021), 许泳昊等 (2022)). 明晰市场中的定价异象对完善资产定价理论、提升市场资源配置效率具有

重要意义. 目前关于知情交易因素定价效果的相关研究主要基于 PIN 指标进行展开. Easley et al. (2002), Chen and Zhao (2012), 杨之曙和姚松瑶 (2004), 韩立岩等 (2008), 陈国进等 (2019) 均证实了 PIN 类指标在中美市场具有产生风险溢价的能力, 然而其对收益的影响方向不同学者的结论却尚未统一. 同时由于中国市场所特有庄家与散户博弈的赢利模式 (韩立岩等 (2008)) 和 PIN 值需要依托交易活跃股票进行估算的限制 (陈国进等 (2019)) 使得 PIN 值在中国市场的定价效果出现下降. 综上本文基于深度学习算法构建了股票知情交易程度的全新度量指标, 突破了 PIN 类指标仅从订单数量角度进行估算的局限性, 为知情交易因素是否构成了市场的系统性风险并参与到股票的定价过程这一核心问题提供了更加精确的解答.

具体而言本文证实中国 A 股市场中存在与股票知情交易程度相关的定价异象, 股票当月的知情交易程度与其下月的预期收益率存在显著的负相关关系, 多空组合可以获得约 1.38% 的等权月度收益. 同时本文证实基于深度学习算法构建的股票知情交易程度指标并非 PIN 类指标的替代变量, 相较之下本文指标在中国 A 股市场具有更强的定价效果. 进一步的机制检验证实 Easley et al. (2011) 的交易流毒性 (flow toxicity) 理论在中国股票市场中成立, 知情交易现象导致普通投资者的市场参与意愿下降, 促使买卖价差扩大并引发了股票的流动性风险, 进而产生了显著的风险溢价.

本文的创新与贡献在于: 首先本文结合高频交易数据与深度学习算法构建了针对知情交易的全新识别框架, 为知情交易的精准识别与监管提供了有效的技术支持. 其次本文证实了知情交易因素能够导致股票的价格异象并详细论证了其背后的传导机制, 补充定价异象相关理论研究的同时具有极强的实用价值.

本文剩余部分安排如下: 第二部分为理论分析与研究假说; 第三部分为基于高频数据和深度学习算法的知情交易识别, 该部分详述了识别模型的原理与构建方法; 第四部分为实证结果, 详细验证了股票价格异象的存在性、异质性以及形成机制; 第五部分为稳健性检验; 最后为本文结论与启示.

## 2 理论分析与研究假说

基于 Easley et al. (2011) 的交易流毒性 (flow toxicity) 理论并以 PIN、VPIN 等指标为代理变量, 已有文献针对知情交易因素在中美市场的定价效果进行了激烈的讨论 (Easley et al. (2002, 2012), Chen and Zhao (2012), 杨之曙和姚松瑶 (2004), 韩立岩等 (2008), 陈国进等 (2019)). 对于美国市场, Easley et al. (2002) 证实基于 PIN 指标构建的多空组合可以获得显著的月度超额收益, 知情交易因素显著参与了市场的定价过程. 基于此 Easley et al. (2012) 进一步利用高频数据改进了传统的 PIN 值估算方法, 提出以 VPIN (volume-synchronized probability of informed trading) 指标度量的知情交易程度的大幅上涨对由流动性蒸发导致的市场 “闪崩” 具有显著的预测作用, 且在横向对比 VIX 等指标后指出 VPIN 具有最佳的预测效果. 但 Andersen and Bondarenko (2014) 则对 VPIN 指标的预测能力提出了质疑, 指出投资者交易模式的变化直接干扰了 VPIN 指标的预测有效性, 并通过实证检验发现 VPIN 反而是崩盘现象的滞后指标. 国内学者中, 陈国进等 (2019) 以沪深 300 成分股为研究对象, 证实基于 VPIN 指标可以构建具有超额收益的投资组合并在机制检验中支持了 Easley et al. (2011) 的交易流毒性假说. 韩立岩等 (2008) 则发现虽然知情交易概率 (PIN) 在中国市场具

有风险定价能力, 但与之前研究相反其对收益产生的却是负效应. 因此韩立岩等 (2008) 认为是以散户为主的非知情交易者在交易过程中表现出的过度交易等行为偏误引发了股价的上升与反转, 从而导致了相应的价格异象.

虽然不同学者的观点与结论存在一定分歧, 但均指出知情交易者与非知情交易者之间存在的信息不对称是知情交易因素能够参与市场定价的理论基础. 因此从信息不对称的角度出发, 若某只股票近期发生了知情交易事件, 表明内部人已经使用了其所持有的内幕信息, 则对应股票的信息不对称程度获得了一定程度的释放, 投资风险下降; 反之若近期没有出现知情交易现象, 则内部人的信息优势在当期仍然维持高位, 对应股票具有更高的知情交易倾向并增大了投资者的风险承担. 基于此, 本文提出假说 1:

**假说 1** 中国 A 股市场中存在知情交易股票价格异象, 当月知情交易程度越高的股票, 其下月收益率越低.

由于公司内部人和机构投资者在信息获取能力和资金实力上所具有的显著优势, 大量文献探讨了内部人和机构投资者持股与知情交易现象之间的内在联系 (Benabou and Laroque (1992), Dittmar and Field (2015), Ali and Hirshleifer (2017), Czech et al. (2021), 曾庆生 (2008), 吴育辉和吴世农 (2010), 蔡宁 (2012), 邵新建等 (2014), 徐龙炳等 (2021)). 对于大股东和机构投资者, 持股比例的上涨将扩大其在公司的内部势力, 增强其对公司内幕信息的获取能力. 同时持股增多也使得大股东和机构投资者拥有了更多的交易筹码, 其股价操纵能力也获得了进一步的提升. 因而伴随大股东和机构投资者持股比例的上升, 投资者间的信息不对称程度和股票的知情交易程度均存在进一步增大的可能, 导致相应股票的价格异象更加明显. 据此本文提出假说 2:

**假说 2** 知情交易股票价格异象在截面维度上存在差异. 股权集中度更高、机构投资者持股比例更高的股票其价格异象更加明显.

交易流毒性理论 (flow toxicity, Easley et al. (2011)) 从风险补偿的视角为 A 股市场中知情交易股票价格异象的形成机制提供了一定的参考. 具体而言, Easley et al. (2011) 指出由于信息不对称的存在, 通常非知情交易者在与知情交易者的交易过程中会蒙受损失. 因此当股票知情交易程度上升时, 非知情交易者要求市场提供更低的买价和更高的卖价进行相应的风险补偿. 而买卖价差的扩大导致了股票流动性风险的产生, 极端情况下甚至将导致市场出现流动性蒸发并引发崩盘现象 (Easley et al. (2012)). 因此知情交易通过扩大买卖价差的方式引发股票的流动性风险进而参与市场定价是导致股票价格异象的潜在机制. 据此本文提出如下待验假说 3:

**假说 3** 知情交易通过扩大股票买卖价差所导致的流动性风险, 是知情交易股票价格异象存在的重要原因.

## 3 基于高频数据和深度学习算法的知情交易识别

### 3.1 数据来源与基本处理

本文使用 2013 年 1 月 1 日至 2020 年 12 月 31 日, 刷新频率约为 3 秒的 A 股日内高频交易数据, 统计包括开盘操纵、尾盘操纵、有效报价差、订单簿深度、日内交易量与日内价格波动等角度在内的 81 个量化指标作为知情交易行为识别的特征集合, 指标构建方法见附录.

同时本文选取了相同时段内 A 股上市公司的公告发布时间、财务数据等基础数据, 和中国 Fama-French 三、五因子 (Fama and French (1993, 2015), Carhart 四因子 (Carhart (1997)), 中国市场 (CH) 三、四因子 (Liu et al. (2019)) 数据. 数据来源包括 Resset 数据库、中国资产管理研究中心和 Stambaugh 学者主页.

### 3.2 集成神经网络模型概述

集成学习 (ensemble learning) 通过将多个学习器进行结合来完成学习任务, 核心目标是获得比单一学习器更加优越的泛化能力, 这一特性与当前金融业界对提升模型样本外预测能力的需求不谋而合. 对于学习器的结合策略大致可分为平均法、投票法和学习法三大类, 相对而言学习法是一种更加强大的结合策略, 是指通过训练另一个次级学习器实现对初级学习器的整合 (周志华 (2016)). 为充分利用深度学习算法的算力优越性, 本文以不同层次结构的前馈神经网络作为初级学习器, 以线性回归作为次级学习器从而避免模型复杂度的上升对其泛化能力的影响, 并采用 Stacking 算法 (Wolpert (1992), Breiman (1996)) 将具有不同特性的神经网络进行集成.

Stacking 算法的核心思想是: 首先使用初始训练集训练 N 个初级学习器. 其次以 N 个初级学习器的输出特征作为次级训练集的输入特征, 而初始样本的标记仍当作样例标记构建次级训练样本集合. 最终采用次级训练集训练次级学习器, 从而实现将多个初级学习器进行集成的目的. 模型组概览如图 1 所示, 本文分别采用隐藏层数目为 1、3、5 的前馈神经网络作为初级学习器, 若最终的集成神经网络泛化能力强于任何一个单独的初级学习器则证实了集成算法的有效性, 且一定程度上避免了对于神经网络层数设定变化是否会增强模型泛化能力的争论.

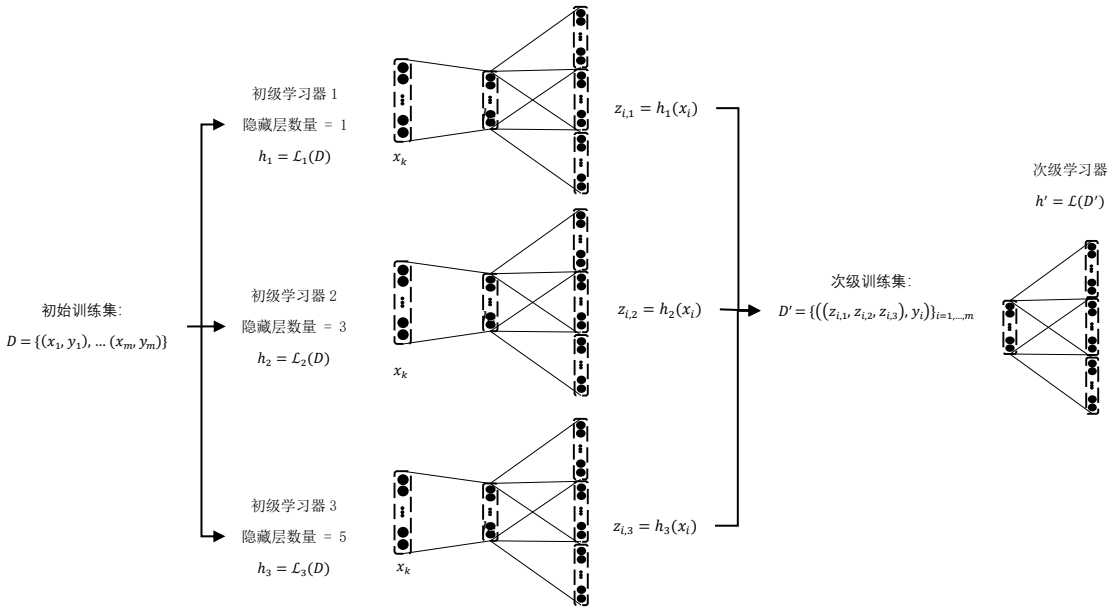



**图 1　基于 Stacking 算法的集成神经网络概览**

### 3.3 基于集成神经网络算法的知情交易识别框架

#### 3.3.1 集成神经网络训练方法与多模型横向比较

深度学习模型的训练依赖合理的训练集, 本文借鉴李志辉和孙广宇 (2020) 采用公司公告发布前一定时段内证券价格和交易量是否出现异常变化作为知情交易行为的识别依据, 所得到的知情交易发生日即作为后续模型的训练集数据. 训练集分为 3 类标签, 分别代表 “买入型” 知情交易、“卖出型” 知情交易和未发生知情交易的数据样本, 具体构建方法见附录.

进一步本文采用筛选所得的知情交易发生日作为训练样本, 以基于高频数据计算的 81 个交易行为指标作为特征集合进行模型训练. 特征维度的增多一方面增强了对于交易行为刻画的细致程度但也使得传统的回归方法不再适用, 而本文选用的深度学习算法在克服维度灾难的同时也解决了潜在的特征非线性问题.

在训练方法上, 为避免前视偏差 (looking forward bias) 并增强结论的样本外稳健性, 本文采用 2013 年至 $T-1$ 年内的知情交易事件样本以 7:3 的比例随机划分为训练集 (train set) 和验证集 (validation set), 采用 SMOTE 算法解决样本标签不均衡问题后进行训练. 对于集成神经网络算法模型组, 本文采用分层训练的方法, 首先单独训练作为初级学习器的各个前馈神经网络; 在训练完成并生成次级训练集数据后, 冻结初级学习器的相关参数并进一步优化次级学习器, 最终获得训练完成的集成学习模型组. 训练过程采用自适应的梯度下降算法 (adam) 和交叉熵损失函数 (cross entropy loss), 并设定学习率 $\alpha = 0.001$. 在获得当前最优模型后以第 $T$ 年的知情交易事件样本为测试集 (test set) 检验模型的样本外判别能力, 训练集与测试集样本区间逐年向前滚动, 具体如图 2 所示.

多种机器学习模型的识别准确率横向对比结果如表 1 所示, 旨在证实集成神经网络模型 (ensemble neural network) 在样本外预测能力上的优越性并侧面说明本文选用集成神经网络的必要性. 结果可见传统机器学习模型当中 XGBoost 作为梯度提升技术的代表性算法在样

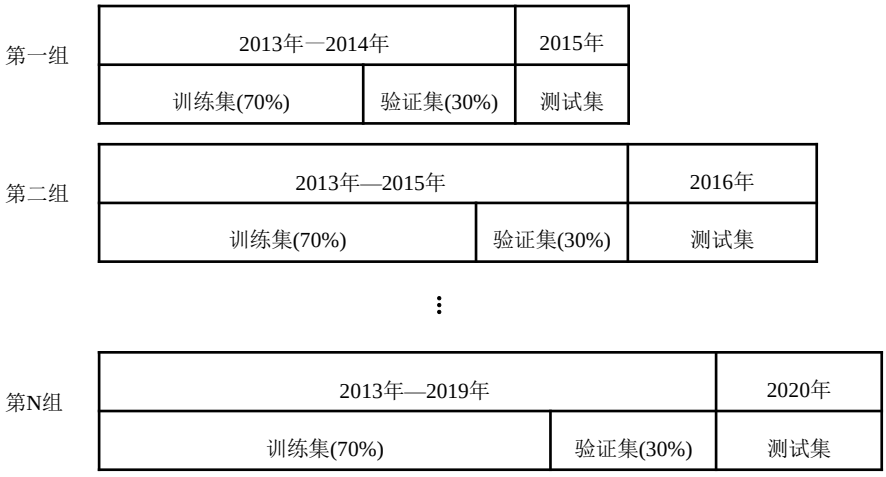



**图 2 训练集、验证集与测试集划分方法**

本内表现出了极高的分类准确性, 但在样本外的预测准确率上有所欠缺, 机器学习模型的样本外预测能力普遍低于神经网络算法. 而进一步对比多种神经网络模型可见本文构建的集成神经网络模型具有最高的样本外判别准确率. 此外值得关注的是具有 5 层隐藏层的前馈神经网络表现出了更高的样本内准确性, 但在样本外的预测准确率上低于本文采用的集成神经网络, 这一结果直接回答了是否应当继续叠加网络深度的疑问. 网络层次越复杂, 其对于训练样本的拟合程度将逐渐上升, 但过拟合导致的泛化能力下降也愈发凸显, 而采取集成算法则是提升模型泛化能力的有效方式之一.

**表 1　多种模型识别准确率对比**

| Panel A | 样本内准确率 | | | | | | |
|---|---|---|---|---|---|---|---|
| 训练集/验证集样本区间 | 2013–2014 | 2013–2015 | 2013–2016 | 2013–2017 | 2013–2018 | 2013–2019 | Avg |
| Logistic Regression | 97.917% | 93.611% | 93.574% | 92.769% | 92.698% | 92.202% | 93.759% |
| Random Forest | 92.484% | 90.248% | 87.564% | 87.815% | 86.357% | 85.179% | 88.274% |
| XGBoost | 100.000% | 99.982% | 99.964% | 99.907% | 99.866% | 99.788% | 99.918% |
| SVM | 98.802% | 96.937% | 96.651% | 95.917% | 95.152% | 94.715% | 96.362% |
| 1 Layer Neural Network | 96.637% | 96.783% | 98.248% | 97.903% | 97.871% | 98.265% | 97.618% |
| 3 Layers Neural Network | 98.883% | 98.782% | 99.370% | 99.524% | 99.584% | 99.593% | 99.289% |
| 5 Layers Neural Network | 99.632% | 99.700% | 99.188% | 99.849% | 99.833% | 99.858% | 99.677% |
| Ensemble Neural Network | 99.483% | 99.509% | 99.413% | 99.744% | 99.789% | 99.781% | 99.620% |
| Panel B | 样本外准确率 | | | | | | |
| 测试集样本区间 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | Avg |
| Logistic Regression | 81.600% | 65.288% | 77.148% | 78.543% | 80.683% | 84.274% | 77.923% |
| Random Forest | 88.000% | 72.212% | 76.174% | 79.013% | 75.391% | 81.224% | 78.669% |
| XGBoost | 94.255% | 80.769% | 89.283% | 85.356% | 85.535% | 86.935% | 87.022% |
| SVM | 85.236% | 73.365% | 78.211% | 81.754% | 84.627% | 85.875% | 81.512% |
| 1 Layer Neural Network | 80.727% | 80.000% | 86.182% | 86.766% | 88.979% | 89.098% | 85.292% |
| 3 Layers Neural Network | 86.545% | 85.962% | 91.231% | 91.073% | 93.112% | 92.970% | 90.149% |
| 5 Layers Neural Network | 86.909% | 87.308% | 90.345% | 91.543% | 93.081% | 93.900% | 90.514% |
| Ensemble Neural Network | 86.109% | 87.404% | 92.028% | 91.229% | 93.331% | 93.705% | 90.635% |

### 3.3.2 变量重要性分析与知情交易行为特征

进一步明晰集成神经网络判别知情交易行为的背后机理并进一步刻画知情交易者的交易行为特征, 本文采用置换检验 (permutation test) 的方法对基于高频数据生成的 81 个特征进行了重要性排序, 表 2 展示了重要性程度较高的部分变量.

结果可见开盘行为、订单买卖报价差与日内成交价格偏度三类指标高居榜首, 表明知情交易者主要以开盘前集合竞价、开盘时段快速拉升或打压股价和短期集中布单扩大报价差并形成价格骤变为主要的交易手段. 同时尾盘操纵和订单簿深度相关指标也表现出了较强的判别作用, 与已有文献结论 (Aitken et al. (2018), 孙广宇等 (2021)) 和金融实务领域的相关经验保持一致, 表明知情交易者常利用投资者的有限注意力特性, 在开盘时段普通投资者反应不及和尾盘时段投资者注意力相对分散等时间段进行交易, 尽量避免自身计划受到市场波动的干扰. 此外本文所使用的 81 个特征均对知情交易识别表现出了正向的判别效果, 进一步证实了基于高频交易数据监测知情交易行为的有效性.

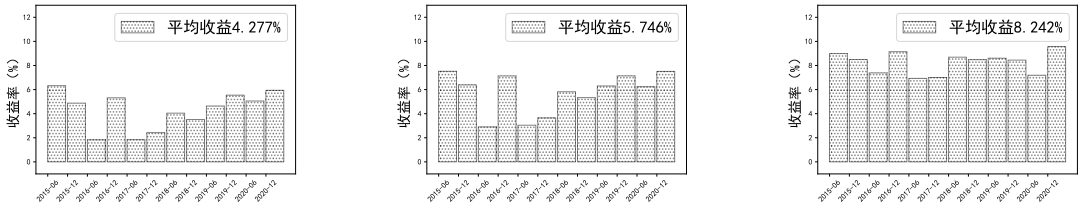
**表 2 特征重要性排序**

| 特征变量 | 重要性程度 | 特征变量 | 重要性程度 |
|---|---|---|---|
| 开盘涨跌幅 | 6.890% | 订单簿深度 2 STD | 2.430% |
| 相对买卖报价差 Mean | 6.840% | 订单簿深度 1 STD | 2.386% |
| 相对有效买卖报价差 Max-Min | 5.943% | 订单簿深度 1 Mean | 2.159% |
| 分笔价格 Skew | 5.215% | 短期价格波动比 Mean | 2.087% |
| 相对买卖报价差 STD | 5.198% | 短期流动指标 Min | 1.850% |
| 相对有效买卖报价差 STD | 5.048% | 订单簿深度 1 Max | 1.797% |
| 相对有效买卖报价差 Max | 4.539% | 短期价格波动比 Min | 1.598% |
| 开盘涨跌幅 15 min | 4.222% | 订单簿深度 1 Max-Min | 1.532% |
| 相对买卖报价差 Max | 4.201% | 分笔交易量 Kurt | 1.522% |
| 相对有效买卖报价差 Mean | 3.753% | 短期流动指标 Skew | 1.465% |
| 开盘交易量 15 min | 3.746% | 订单簿深度 2 Mean | 1.458% |
| 相对买卖报价差 Max-Min | 3.434% | 相对买卖报价差 Skew | 1.444% |
| 开盘涨跌幅 10 min | 2.934% | 分笔平均交易量 STD | 1.312% |
| 分笔交易量 Skew | 2.908% | 尾盘每笔交易量 15 min | 1.309% |
| 分笔交易量 Mean | 2.738% | 尾盘交易量 5 min | 1.192% |
| 相对买卖报价差 Min | 2.689% | 相对有效买卖报价差 Skew | 1.178% |
| 分笔平均交易量 Min | 2.530% | 分笔平均交易量 Kurt | 1.139% |
| 订单簿深度 2 Max-Min | 2.511% | 短期价格波动比 Kurt | 1.107% |
| 尾盘涨跌幅 5 min | 2.491% | 分笔平均交易量 Skew | 1.060% |
| 订单簿深度 2 Max | 2.487% | 尾盘涨跌幅 10 min | 1.043% |

### 3.3.3 知情交易事件的拓展识别与知情交易当日收益统计

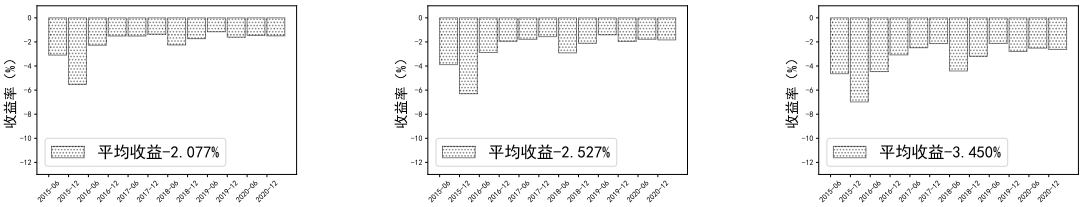
前述的训练集样本仅局限于与公告信息提前泄露相关的知情交易, 因而会对某些并不以公告形式发布, 但对股票价格具有显著影响的信息所引发的知情交易事件存在遗漏. 本文的基本假设在于针对不同特质的内幕信息, 知情交易者所采用的交易手法不会产生显著变化, 因而可以将本文的知情交易行为识别框架进行逐股、逐日的识别范围拓展.

本文采用滚动监测的方法保证结论的样本外性质, 由 2013 年至 $T-1$ 年的训练集样本训练所得的模型, 对 $T$ 年所有股票, 基于其高频交易数据逐日计算其当天出现 “买入型” 知情交易、“卖出型” 知情交易和未发生知情交易的概率指数, 三者之和为 1. 图 3 和图 4 分别展示了拓展识别出的 “买入型” 知情交易事件和 “卖出型” 知情交易事件的当日收益均值, 且根据其属于对应类别的概率的分位数进行了进一步的分组统计. 结果可见 “买入型” 知情交易的当日收益均值为 4.3% 左右, “卖出型” 知情交易的当日收益约为 $-2.1\%$, 二者的非对称性可能与中国股市的卖空限制相关. 进一步可见伴随着分位数逐渐上升, 知情交易当日收益率呈现出显著的单调趋势, 侧面证实了本文识别结果的准确性.




(a) 全部样本　　(b) 前 60%　　(c) 前 20%

**图 3　集成神经网络识别的“买入型”知情交易发生当日收益率统计**




(a) 全部样本　　(b) 前 60%　　(c) 前 20%

**图 4　集成神经网络识别的“卖出型”知情交易发生当日收益率统计**

## 4　实证结果

### 4.1　双变量组合分析

本文首先采用双重分组的方法, 在控制多种不同因素的前提下检验知情交易股票价格异象的存在并探究其在截面维度上的异质性.

#### 4.1.1　基本面特征与知情交易股票价格异象

首先本文将股票按照其上月知情交易程度分为 5 组, 其中 No 组为上月无知情交易事件的股票, 其余股票按照“买入型”知情交易概率和“卖出型”知情交易概率之和的月度均值等分为 4 组, 由低 (Low) 到高 (High) 依次排序, 表明股票的知情交易程度由低到高, 并构建等权重投资组合. 表 3 结果可见组合收益表现出了稳定的单调递减特征, 基于本文指标构建的多空组合可以获得显著的月度收益. 具体而言在做多无知情交易股票或低知情交易程度股票, 同时做空高知情交易程度股票的情况下可以分别获得约 1.38% (No-High) 和 1.14% (Low-High) 的月均收益率. 表 3 进一步汇报了控制基本面因素后 (规模 Size, 账面市值比 B/M, 市盈率 E/P) 的双重分组检验结果, 结果可见基于本文指标构建的多空组合仍可获得显著的正向收益. 具体而言, 多空组合在高市值、低 B/M、低 E/P 的股票中表现更好, 表明知情交易所造成的风险溢价在股价被高估的大市值企业中更加显著. 这一发现与陈国进等 (2019) 指出的知情交易行为对沪深 300 成分股影响更加明显的结论相互支撑.

**表 3 控制基本面特征**

| | 知情交易指数 | | | | | No-High | | Low-High | |
|---|---|---|---|---|---|---|---|---|---|
| | No | Low | 2 | 3 | High | mean | $t$ statistic | mean | $t$ statistic |
| ALL | 1.601 | 1.367 | 1.366 | 0.893 | 0.226 | 1.376*** | 2.79 | 1.141*** | 2.93 |
| Panel A: By size | | | | | | | | | |
| Low | 2.019 | 1.659 | 1.668 | 1.229 | 1.175 | 0.767 | 1.11 | 0.488 | 0.91 |
| 2 | 1.423 | 1.466 | 0.871 | 0.181 | 0.079 | 1.114** | 2.06 | 1.387*** | 2.90 |
| High | 1.306 | 0.614 | 1.260 | 1.516 | 0.248 | 1.187** | 2.37 | 0.487 | 0.87 |
| Panel B: By book-to-market ratio | | | | | | | | | |
| Low | 1.531 | 1.380 | 1.227 | 0.342 | −0.222 | 1.705*** | 3.15 | 1.635*** | 3.56 |
| 2 | 1.756 | 1.636 | 1.555 | 0.618 | 0.208 | 1.492*** | 2.67 | 1.427** | 2.40 |
| High | 1.653 | 1.152 | 1.322 | 1.570 | 0.551 | 1.029* | 1.73 | 0.689 | 1.22 |
| Panel C: By E/P ratio | | | | | | | | | |
| Low | 1.664 | 1.574 | 1.220 | 0.852 | 0.163 | 1.451*** | 2.70 | 1.450*** | 3.03 |
| 2 | 1.795 | 1.501 | 1.339 | 0.709 | 0.157 | 1.547** | 2.60 | 1.307*** | 2.76 |
| High | 1.456 | 1.035 | 0.979 | 0.788 | 0.629 | 0.866 | 1.56 | 0.411 | 0.80 |

注: *、**、*** 分别表示 10%、5% 和 1% 的显著性水平.

#### 4.1.2 量价特征与知情交易股票价格异象

由于本文的知情交易指标是基于市场的高频交易数据计算所得, 则有必要检验本文投资组合在传统的量价指标之下是否仍能获得显著的正向收益, 证实基于高频数据识别的知情交易指标提供了相对传统指标的额外信息. 本文分别从价格反转 (上月收益 past month return, 上月收盘价均值 price)、交易行为 (换手率 turnover, 成交量 trade volume) 和个股波动性 (volatility) 三个角度进行了双重分组检验, 结果如表 4 所示. 结果表明在多种传统量价指标之下本文所构建的投资组合收益显著性保持不变, 其中在高换手率和高成交量的股票中价格异象更加明显, 潜在原因是知情交易者倾向于操纵高流动性股票以避免在其交易过程中可能出现的流动性风险.

#### 4.1.3 公司内部人持股与知情交易股票价格异象

进一步地, 已有文献常从投资者的身份出发研究知情交易现象, 尤其关注内部人和机构投资者的相关交易行为. 因此本文分别采用机构投资者持股比例 (institution holds) 和前 10 大股东持股比例 (Top10 Holds) 衡量机构和大股东对于公司的影响力, 探究内部人和机构投资者持股与知情交易现象之间是否存在必然的联系. 结果如表 5 所示, 与以往文献保持一致, 本文证实存在高机构投资者持股和高股权集中度的公司更易出现信息泄露与知情交易现象, 组合收益在相应公司中更加显著, 表明机构投资者与大股东更加倾向于通过知情交易行为在股票市场中输送或谋取利益.

**表 4　控制量价特征**

| | 知情交易指数 | | | | | No-High | | Low-High | |
|---|---|---|---|---|---|---|---|---|---|
| | No | Low | 2 | 3 | High | mean | $t$ statistic | mean | $t$ statistic |
| Panel A: By Past month return | | | | | | | | | |
| Low | 1.806 | 1.882 | 1.929 | 0.915 | 0.445 | 1.339** | 2.17 | 1.502* | 1.91 |
| 2 | 1.599 | 1.196 | 1.228 | 2.380 | 0.623 | 0.954* | 1.72 | 0.652 | 1.30 |
| High | 1.252 | 1.082 | 0.896 | 0.463 | −0.483 | 1.717*** | 3.11 | 1.561*** | 3.27 |
| Panel B: By price | | | | | | | | | |
| Low | 1.943 | 1.407 | 1.640 | 1.756 | 0.551 | 1.221** | 2.19 | 0.919* | 1.94 |
| 2 | 1.686 | 1.661 | 1.198 | 0.570 | −0.516 | 2.186*** | 4.12 | 2.135*** | 3.93 |
| High | 1.774 | 1.394 | 0.716 | 0.035 | −0.673 | 2.363*** | 2.84 | 2.144*** | 2.68 |
| Panel C: By turnover | | | | | | | | | |
| Low | 1.320 | 1.170 | 1.574 | 1.075 | 0.422 | 0.802* | 1.73 | 0.671 | 1.20 |
| 2 | 1.813 | 1.352 | 1.731 | 1.021 | 0.641 | 1.099** | 2.00 | 0.745 | 1.37 |
| High | 2.107 | 1.746 | 0.033 | −0.328 | −0.249 | 2.042*** | 2.92 | 1.888*** | 2.90 |
| Panel D: By trade volume | | | | | | | | | |
| Low | 1.769 | 1.615 | 1.779 | −0.217 | 2.549 | −0.938 | −0.66 | −1.144 | −0.76 |
| 2 | 1.787 | 1.459 | 1.471 | 1.200 | 1.644 | −0.178 | −0.22 | −0.188 | −0.26 |
| High | 1.731 | 0.841 | 0.899 | 0.423 | −0.478 | 1.606** | 2.21 | 1.194* | 1.71 |
| Panel E: By volatility | | | | | | | | | |
| Low | 1.271 | 0.442 | 0.542 | 0.239 | −0.045 | 1.062 | 1.41 | 0.522 | 0.71 |
| 2 | 1.972 | 1.722 | 1.904 | 1.184 | 0.685 | 1.223** | 2.00 | 1.032** | 2.03 |
| High | 1.894 | 1.418 | 0.883 | 0.233 | 0.021 | 1.751*** | 3.12 | 1.342** | 2.64 |

注: *、**、*** 分别表示 10%、5% 和 1% 的显著性水平.

**表 5　控制内部人与机构投资者持股特征**

| | 知情交易指数 | | | | | No-High | | Low-High | |
|---|---|---|---|---|---|---|---|---|---|
| | No | Low | 2 | 3 | High | mean | $t$ statistic | mean | $t$ statistic |
| Panel A: By institution holds | | | | | | | | | |
| Low | 1.909 | 1.451 | 0.979 | 0.808 | 0.978 | 0.918 | 1.43 | 0.473 | 0.76 |
| 2 | 1.825 | 0.905 | 2.068 | 1.156 | 0.915 | 0.858 | 1.41 | −0.061 | −0.10 |
| High | 2.396 | 1.212 | 1.331 | 0.942 | −0.030 | 1.919** | 2.49 | 0.924 | 1.16 |
| Panel B: By Top10 holds | | | | | | | | | |
| Low | 1.725 | 0.913 | 1.983 | 0.872 | 1.272 | 0.213 | 0.31 | −0.317 | −0.44 |
| 2 | 1.794 | 1.197 | 1.572 | 0.919 | 0.820 | 0.942 | 1.31 | 0.255 | 0.35 |
| High | 2.328 | 1.119 | 1.177 | 1.177 | −0.037 | 2.116*** | 3.59 | 1.259** | 2.16 |

注: *、**、*** 分别表示 10%、5% 和 1% 的显著性水平.

### 4.2 因子回归与 Fama-Macbeth 回归检验

在此本文采用因子回归的方法进一步检验基于本文指标构建的投资组合在当前主流定价因子模型下是否仍具有稳定的超额收益, 结果如表 6 所示. Panel A 为做多无知情交易股票 (No) 并做空高知情交易程度股票 (High) 时的等权多空组合收益回归检验结果, 可见截

**表 6 因子回归检验**

| | CAPM | FF3 | Carhart4 | FF5 | CH3 | CH4 |
|---|---|---|---|---|---|---|
| Panel A: 做多无知情交易股, 做空高知情交易股 | | | | | | |
| intercept | 0.0129*** | 0.0124** | 0.0118** | 0.0152*** | 0.0147*** | 0.0160*** |
| | (2.67) | (2.57) | (2.51) | (2.87) | (2.84) | (3.18) |
| MKT | 0.1389* | 0.1059 | 0.0468 | 0.0867 | 0.0704 | −0.0018 |
| | (1.93) | (1.47) | (0.62) | (1.06) | (0.86) | (−0.02) |
| SMB | | 0.1725 | 0.3772** | −0.0801 | 0.1143 | 0.3008* |
| | | (1.26) | (2.31) | (−0.33) | (0.74) | (1.77) |
| HML | | 0.0092 | 0.0694 | −0.0869 | | |
| | | (0.07) | (0.53) | (−0.47) | | |
| UMD | | | 0.2269** | | | |
| | | | (2.16) | | | |
| RMW | | | | −0.2364 | | |
| | | | | (−0.94) | | |
| CMA | | | | 0.1374 | | |
| | | | | (0.58) | | |
| VMG | | | | | −0.1678 | −0.0115 |
| | | | | | (−0.94) | (−0.06) |
| PMO | | | | | | −0.3606** |
| | | | | | | (−2.52) |
| $N$ | 72 | 72 | 72 | 72 | 72 | 72 |
| $R^2$ | 0.051 | 0.089 | 0.148 | 0.113 | 0.091 | 0.172 |
| Panel B: 做多低知情交易股, 做空高知情交易股 | | | | | | |
| intercept | 0.0109*** | 0.0108*** | 0.0104*** | 0.0131*** | 0.0125*** | 0.0136*** |
| | (2.87) | (2.82) | (2.76) | (3.14) | (3.04) | (3.45) |
| Panel C: 做多无知情交易股 | | | | | | |
| intercept | 0.0134 | 0.0127 | 0.0123 | 0.0197 | 0.0190 | 0.0193 |
| | (1.20) | (1.12) | (1.07) | (1.59) | (1.57) | (1.58) |
| Panel D: 做多低知情交易股 | | | | | | |
| intercept | 0.0113 | 0.0111 | 0.0109 | 0.0176 | 0.0168 | 0.0169 |
| | (0.98) | (0.94) | (0.92) | (1.37) | (1.35) | (1.34) |
| Panel E: 做空高知情交易股 | | | | | | |
| intercept | 0.0004 | 0.0003 | 0.0005 | 0.0045 | 0.0043 | 0.0033 |
| | (0.04) | (0.02) | (0.04) | (0.33) | (0.32) | (0.25) |

注: *、**、*** 分别表示 10%、5% 和 1% 的显著性水平.

距项 $t$ 值均大于 2.5, 表明本文组合收益无法被现有的因子模型所解释, 具有显著的超额收益. Panel B 则汇报了做多低知情交易程度股票 (Low) 并做空高知情交易程度股票 (High) 时的组合超额收益检验结果. 结果可见多空组合仍然存在显著的超额收益, 截距项 $t$ 值均大于 2.8. Panel C, Panel D, Panel E 分别为多头、空头组合的回归检验结果.

表 7 为 Fama-Macbeth 回归检验结果, 本文以下月股票收益为被解释变量, Informed Trading Score 为本文模型计算的 “买入” 与 “卖出” 型知情交易指数之和的月度均值, Informed Trading Dummy 为表示该股票当月是否出现知情交易的虚拟变量. 控制变量包括 Beta, 市值 (size), 账面市值比 (BM), 市盈率倒数 (EP), 个股月波动率 (volatility), 月换手率 (turnover) 和交易量 (trade volume). 回归 (2) 和 (3) 分别控制基本面和量价特征相关变量, 旨在证实本文的知情交易指标所涵盖的信息量无法被传统指标所覆盖. 结果可见在分别控制了基本面以及市场因素后本文指标均在 5% 以上的置信水平下显著为负, 其中 $t$ 值均经过了 Newey-West (1987) 3 阶滞后调整, 进一步证实了知情交易股票价格异象的存在.

### 4.3 机制检验

基于 Easley et al. (2011) 的交易流毒性 (flow toxicity) 理论, 由于普通投资者处于信息劣势的位置多数情况下其与知情交易者的交易行为将导致其蒙受损失, 因而知情交易者占比的上升将降低普通投资者参与市场的积极性. 在此情况下, 普通投资者要求市场提供更低的买价和更高的卖价进行风险补偿并由此导致了买卖价差的扩大, 进而引发了股票的流动性风险. 根据上述理论, 如果知情交易现象是通过买卖价差引发市场的流动性风险从而参与市场定价, 则首先在知情交易发生的当期应当观察到股票买卖价差的明显扩张. 而在下一期, 由于前期知情交易者已经使用了其所持有的内幕信息并进行了交易, 信息不对称程度得到释放, 因此股票的知情交易指数应对下一期的买卖价差具有显著的负向预测作用.

本文采用前述的高频交易数据计算了股票的月度买卖价差度量指标, 并采用经过 Newey-West (1987) 调整后的 Fama-Macbeth 回归验证机制假设, 结果如表 8 所示. Panel A 为知情交易指数对于当期股票买卖价差的回归结果, 结果可见知情交易现象的出现将显著增大股票的买卖价差, 导致流动性风险的产生. 与此同时 Panel B 证实当期的知情交易指标对下一期的买卖价差具有显著的负向预测能力, 表明信息不对称程度的降低导致了流动性风险的释放, 与前文证实的知情交易指数对下期股票收益具有负向预测作用的结论完全吻合.

## 5 稳健性检验

### 5.1 竞争指标检验

VPIN 指标是 Easley et al. (2012) 在 Easley et al. (2002) 提出了 PIN 指标后结合分时高频数据改进计算方法所提出的全新指标, 思路上采用买卖订单的不均衡性度量市场中的知情交易程度, 并假设在信息完全对称的市场中买卖订单应当趋向均衡. VPIN 一方面通过使用高频数据提高了对交易行为的刻画精度, 同时避免了 PIN 值估计方法中对待估参数分布进行主观假设可能导致的潜在偏差, 因而近年来被学者广泛采用 (陈国进等 (2019)).

本文依据 Easley et al. (2012) 的 BVC 算法计算个股的月度 VPIN 指数, 并采用双重分组、多空组合收益相互回归和 Fama-Macbeth 回归三种方式验证本文指标的稳健性. 表 9 的

**表 7 Fama-Macbeth 回归检验**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| informed trading score | $-0.0403^{***}$ | $-0.0382^{***}$ | $-0.0276^{**}$ | $-0.0262^{**}$ |
| | $(-3.73)$ | $(-3.43)$ | $(-2.44)$ | $(-2.29)$ |
| beta | | 0.0002 | | 0.0003 |
| | | (0.18) | | (0.29) |
| Ln(size) | | $-0.0029^{*}$ | | $-0.0040^{**}$ |
| | | $(-1.74)$ | | $(-2.35)$ |
| Ln(B/M) | | $0.0014^{*}$ | | 0.0010 |
| | | (1.94) | | (1.17) |
| Ln(E/P) | | 0.0147 | | 0.0169 |
| | | (1.14) | | (1.31) |
| volatility | | | $0.3930^{*}$ | 0.3750 |
| | | | (1.80) | (1.64) |
| turnover | | | $-0.0002^{***}$ | $-0.0002^{***}$ |
| | | | $(-5.21)$ | $(-5.29)$ |
| trade volume | | | $-1.98\times10^{-12}$ | $-6.74\times10^{-13}$ |
| | | | $(-0.84)$ | $(-0.31)$ |
| intercept | $0.0145^{**}$ | $0.0811^{**}$ | 0.0135 | $0.101^{**}$ |
| | (2.42) | (2.03) | (1.52) | (2.62) |
| $N$ | 164937 | 156439 | 164854 | 156363 |
| $R^2$ | 0.009 | 0.021 | 0.033 | 0.042 |
| | (1) | (2) | (3) | (4) |
| informed trading dummy | $-0.0119^{***}$ | $-0.0116^{***}$ | $-0.0069^{**}$ | $-0.0064^{**}$ |
| | $(-3.34)$ | $(-3.07)$ | $(-2.39)$ | $(-2.28)$ |
| beta | | $-0.0002$ | | 0.0003 |
| | | $(-0.24)$ | | (0.23) |
| Ln(size) | | $-0.0038^{**}$ | | $-0.0041^{**}$ |
| | | $(-2.22)$ | | $(-2.43)$ |
| Ln(B/M) | | $0.0013^{*}$ | | 0.0009 |
| | | (1.89) | | (1.11) |
| Ln(E/P) | | 0.0176 | | 0.0167 |
| | | (1.41) | | (1.30) |
| volatility | | | 0.3250 | 0.2940 |
| | | | (1.47) | (1.25) |
| turnover | | | $-0.0002^{***}$ | $-0.0002^{***}$ |
| | | | $(-5.49)$ | $(-5.23)$ |
| trade volume | | | $-5.11\times10^{-11**}$ | $-3.57\times10^{-11*}$ |
| | | | $(-2.27)$ | $(-1.86)$ |
| intercept | $0.0143^{**}$ | $0.1020^{**}$ | $0.0161^{*}$ | $0.1070^{***}$ |
| | (2.48) | (2.46) | (1.77) | (2.76) |
| $N$ | 164937 | 156439 | 164854 | 156363 |
| $R^2$ | 0.007 | 0.018 | 0.030 | 0.038 |

注: $^{*}$、$^{**}$、$^{***}$ 分别表示 10%、5% 和 1% 的显著性水平, 括号内为 Newey-West 调整后的 $t$ 值.

**表 8　知情交易与流动性风险**

| Panel A: 知情交易与买卖价差 ($T$) | | | | |
|---|---|---|---|---|
| | 买卖报价差 | 买卖报价差 | 有效买卖报价差 | 有效买卖报价差 |
| informed trading score | $0.448^{***}$ | | $0.601^{***}$ | |
| | (9.55) | | (6.85) | |
| informed trading dummy | | $0.136^{***}$ | | $0.167^{***}$ |
| | | (11.27) | | (8.11) |
| controls | Yes | Yes | Yes | Yes |
| $N$ | 154896 | 154896 | 154896 | 154896 |
| $R^2$ | 0.071 | 0.053 | 0.097 | 0.077 |
| Panel B: 知情交易与买卖价差 ($T+1$) | | | | |
| | 买卖报价差 | 买卖报价差 | 有效买卖报价差 | 有效买卖报价差 |
| informed trading score | $-0.0455^{**}$ | | $-0.1540^{***}$ | |
| | $(-2.33)$ | | $(-5.01)$ | |
| informed trading dummy | | $-0.0097^{**}$ | | $-0.0627^{***}$ |
| | | $(-2.12)$ | | $(-8.75)$ |
| beta | $0.0176^{***}$ | $0.0177^{***}$ | $0.0261^{***}$ | $0.0264^{***}$ |
| | (4.43) | (4.38) | (3.78) | (3.84) |
| Ln(size) | $-0.0073^{*}$ | $-0.0078^{*}$ | $0.0368^{***}$ | $0.0361^{***}$ |
| | $(-1.85)$ | $(-1.92)$ | (3.18) | (3.09) |
| Ln(B/M) | $0.0017^{*}$ | 0.0015 | $-0.0046^{***}$ | $-0.0044^{**}$ |
| | (1.93) | (1.57) | $(-2.69)$ | $(-2.62)$ |
| Ln(E/P) | $-0.0189$ | $-0.0174$ | $-0.3270^{***}$ | $-0.3290^{***}$ |
| | $(-0.62)$ | $(-0.56)$ | $(-4.48)$ | $(-4.52)$ |
| volatility | $2.9220^{***}$ | $2.7790^{***}$ | $7.9150^{***}$ | $7.8340^{***}$ |
| | (7.56) | (7.45) | (4.60) | (4.73) |
| turnover | $0.0004^{***}$ | $0.0003^{***}$ | $0.0008^{***}$ | $0.0007^{***}$ |
| | (5.79) | (5.33) | (8.86) | (8.67) |
| trade volume | $-1.61\times10^{-11***}$ | $-2.03\times10^{-11***}$ | $-7.43\times10^{-11***}$ | $-8.33\times10^{-11***}$ |
| | $(-3.87)$ | $(-4.64)$ | $(-6.33)$ | $(-7.03)$ |
| intercept | 0.114 | 0.129 | $-0.880^{***}$ | $-0.849^{***}$ |
| | (1.30) | (1.45) | $(-3.00)$ | $(-2.89)$ |
| $N$ | 154920 | 154920 | 154920 | 154920 |
| $R^2$ | 0.027 | 0.027 | 0.060 | 0.059 |

注: $^{*}$、$^{**}$、$^{***}$ 分别表示 10%、5% 和 1% 的显著性水平, 括号内为 Newey-West 调整后的 $t$ 值.

Panel A 为双重分组结果, 结果可见基于本文指标构建的多空组合仍然具有正向的显著收益且组合收益的单调性保持不变. Panel B 中本文将股票按照本文指标和 VPIN 指标分别分成 5 组, 计算各自多空组合的收益序列并相互回归. 结果可见, 以本文指标构建的多空组合收益作为被解释变量时, 截距项在 5% 的置信水平下显著异于 0, 表明基于 VPIN 指标构建的组合其收益无法解释本文组合收益; 相反以 VPIN 指标构建的多空组合收益为被解释变量时, 其截距项失去显著性. 此外本文将 VPIN 作为控制变量加入 Fama-Macbeth 回归检验当中, 表 10 结果可见本文指标的显著性并未受到任何影响, 证实本文指标对于股价的预测能力强于 VPIN 指标, 并非 VPIN 的替代变量.

**表 9 VPIN 指标双变量分组和组合收益回归**

| Panel A: By VPIN | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 知情交易指数 | | | | | No-High | | Low-High | |
| | No | Low | 2 | 3 | High | mean | $t$ statistic | mean | $t$ statistic |
| Low | 1.284 | 0.887 | 0.961 | 0.764 | 0.123 | $1.201^{**}$ | 2.04 | $0.879^{**}$ | 2.04 |
| 2 | 1.697 | 1.766 | 1.512 | 0.617 | 0.554 | $1.033^{*}$ | 1.85 | $1.169^{**}$ | 2.27 |
| High | 1.663 | 1.575 | 0.981 | 0.721 | 0.063 | $1.554^{***}$ | 2.94 | $1.513^{***}$ | 3.12 |

| Panel B: 组合收益相互回归 | | | | | |
|---|---|---|---|---|---|
| | 基于知情交易指数的多空组合收益 | | | | |
| | $\alpha$ | $t_\alpha$ | $\beta$ | $t_\beta$ | $R^2$ |
| 基于 VPIN 指数的多空组合收益 | $0.0122^{**}$ | 2.32 | $0.2324^{*}$ | 1.92 | 0.05 |
| | 基于 VPIN 指数的多空组合收益 | | | | |
| | $\alpha$ | $t_\alpha$ | $\beta$ | $t_\beta$ | $R^2$ |
| 基于知情交易指数的多空组合收益 | 0.0017 | 0.32 | $0.2152^{*}$ | 1.92 | 0.05 |

注: $^{*}$、$^{**}$、$^{***}$ 分别表示 10%、5% 和 1% 的显著性水平.

**表 10 加入 VPIN 指标的 Fama-Macbeth 回归检验**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| informed trading score | $-0.0410^{***}$ | | $-0.0258^{**}$ | |
| | $(-3.95)$ | | $(-2.33)$ | |
| informed trading dummy | | $-0.0121^{***}$ | | $-0.00656^{**}$ |
| | | $(-3.60)$ | | $(-2.41)$ |
| VPIN | $-0.0862^{**}$ | $-0.0722^{*}$ | $-0.0773^{*}$ | $-0.0805^{*}$ |
| | $(-2.13)$ | $(-1.84)$ | $(-1.88)$ | $(-1.96)$ |
| controls | No | No | Yes | Yes |
| $N$ | 164852 | 164852 | 156305 | 156305 |
| $R^2$ | 0.013 | 0.010 | 0.045 | 0.042 |

注: $^{*}$、$^{**}$、$^{***}$ 分别表示 10%、5% 和 1% 的显著性水平, 括号内为 Newey-West 调整后的 $t$ 值.

### 5.2 市值加权投资组合回归结果

前文的组合收益均采用等权重计算方法, 为进一步夯实本文结论, 本文进一步计算了多空组合的市值加权收益并进行因子回归检验. 结果见附录表 A3, 可见 Panel A 截距项依然显著异于 0 且 $t$ 值均在 2.3 以上, 与正文结论完全一致.

## 6 结论与启示

本文基于中国 A 股的日内高频交易数据细致刻画了投资者的行为特征并使用基于 Stacking 算法的集成神经网络实现了针对知情交易行为的样本外精准识别, 由此构建了股票知情交易程度的全新度量指标. 基于全新指标, 本文进一步验证了知情交易股票价格异象的存在性, 并对其形成机制进行了进一步的探讨. 具体而言本文发现由于信息不对称所导致的流动性风险, 使得高知情交易倾向股需提供额外的风险补偿吸引普通投资者的进入, 基于本文计算的知情交易指数所构建的月度多空组合可以获得稳定的超额收益, 知情交易因素显著参与到了中国股票市场的定价过程之中. 同时本文证实知情交易股票价格异象在截面维度上存在差异, 在市值较大、流动性较高、大股东和机构投资者持股比例更高的股票中更加明显.

本文结论对于维护市场环境和优化投资策略均具有重要的指导意义. 从市场监管的视角出发, 本文基于高频交易数据和深度学习算法所构建的模型框架实现了针对知情交易行为的实时监控, 提升监管效率的同时为进一步维护市场公平、促进资本市场资源配置功能提供了参考. 具体而言本文证实知情交易者以集中布单扩大价差并形成价格骤变为主的交易手段使得知情交易呈现出了更强的隐蔽性与复杂性, 监管者应当充分利用当前丰富的数据要素资源, 借助大数据、云计算以及人工智能等新兴技术手段在信号捕捉与信息挖掘上的优势构建更加完善的自动监管平台, 提高知情交易的行为成本并从提升信息透明度的角度促进资本市场的可持续发展. 此外, 从资产定价理论与量化投资的角度本文拓展了中国 A 股市场价格异象的相关研究, 探讨了知情交易因素与股票横截面收益之间的内在联系并有助于提升中国资本市场的定价效率. 同时本文为基于知情交易因素的量化投资策略提供了参考, 实现了金融理论与业界实务的紧密结合.

当然本文研究仍然存在较大的拓展空间, 进一步的研究方向可能包括将知情交易行为的识别从股票层面向账户层面进行延伸, 为相关机构提供更加直接的监管策略与工具. 对于市场投资者而言股票是否发生知情交易事件是投资决策的重要参考, 而市场监管者则需在此之上进一步筛选出潜在的知情交易账户, 从而进行相应的监管与处罚. 受限于数据的可获得性已有文献 (Easley et al. (2008), Chen and Zhao (2012), 陈国进等 (2019), 李志辉和孙广宇 (2020)) 并未深入探讨账户层面的知情交易者识别问题, 但这一问题对于投资者行为研究和资本市场监管均具有重大的理论与实践意义, 从而留下了广阔的研究空间. 因此对于未来研究而言, 如何构建合理的深度学习算法框架并依托账户层面的逐笔交易数据实现知情交易者的精准判别应当成为相关领域的重点问题.

## 参　考　文　献

蔡宁, (2012). 信息优势、择时行为与大股东内幕交易 [J]. 金融研究, (5): 179–192.

Cai N, (2012). Information Advantage, Timing and Insider Trading[J]. Journal of Financial Research, (5): 179–192.

陈国进, 张润泽, 谢沛霖, 赵向琴, (2019). 知情交易、信息不确定性与股票风险溢价 [J]. 管理科学学报, 22(4): 53–74.

Chen G J, Zhang R Z, Xie P L, Zhao X Q, (2019). Informed Trading, Information Uncertainty and Stock Risk Premium[J]. Journal of Management Sciences in China, 22(4): 53–74.

韩立岩, 郑君彦, 李东辉, (2008). 沪市知情交易概率 (PIN) 特征与风险定价能力 [J]. 中国管理科学, 16(1): 16–24.

Han L Y, Zheng J Y, Li D H, (2008). The Feature of Probability of Informed Trading and Risk Pricing in Shanghai Stock Market[J]. Chinese Journal of Management Science, 16(1): 16–24.

何诚颖, 陈锐, 薛冰, 何牧原, (2021). 投资者情绪、有限套利与股价异象 [J]. 经济研究, 56(1): 58–73.

He C Y, Chen R, Xue B, He M Y, (2021). Investor Sentiment, Limited Arbitrage and Stock Price Anomalies[J]. Economic Research Journal, 56(1): 58–73.

李志辉, 孙广宇, (2020). 中国股票市场内幕交易对信息效率的影响 —— 基于内幕交易行为的识别与监测[J]. 南开学报 (哲学社会科学版), (5): 136–145.

Li Z H, Sun G Y, (2020). The Impact of Insider Trading on Information Efficiency in China's Stock Market[J]. Nankai Journal (Philosophy, Literature and Social Science Edition), (5): 136–145.

李志辉, 王近, 李梦雨, (2018). 中国股票市场操纵对市场流动性的影响研究 —— 基于收盘价操纵行为的识别与监测 [J]. 金融研究, (2): 135–152.

Li Z H, Wang J, Li M Y, (2018). A Study on China's Stock Market Manipulation's Effects on Market Liquidity: Based on Closing Price Manipulation Behavior's Identification and Monitoring[J]. Journal of Financial Research, (2): 135–152.

邵新建, 贾中正, 赵映雪, 江萍, 薛熠, (2014). 借壳上市、内幕交易与股价异动 —— 基于 ST 类公司的研究[J]. 金融研究, (5): 126–142.

Shao X J, Jia Z Z, Zhao Y X, Jiang P, Xue Y, (2014). Reverse Merger, Insider Trading and Abnormal Market Reaction: Evidence from ST Listed Companies in China[J]. Journal of Financial Research, (5): 126–142.

沈冰, 冉光和, 钟剑, (2012). 我国股票市场知情交易的形成及策略分析 [J]. 管理世界, (1): 170–171.

Shen B, Ran G H, Zhong J, (2012). An Analysis of the Forming of Insider Transactions in China's Stock Market and an Analysis of the Strategy[J]. Journal of Management World, (1): 170–171.

孙广宇, 李志辉, 杜阳, 王近, (2021). 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, (9): 151–169.

Sun G Y, Li Z H, Du Y, Wang J, (2021). Does Market Manipulation Reduce the Information Efficiency of China's Stock Market? Empirical Evidence from Shanghai A-share Market's High Frequency Trading Data[J]. Journal of Financial Research, (9): 151–169.

吴育辉, 吴世农, (2010). 股票减持过程中的大股东掏空行为研究 [J]. 中国工业经济, (5): 121–130.

Wu Y H, Wu S N, (2010). Tunneling Behaviors during Large Shareholder's Stock Selling Periods[J]. China Industrial Economics, (5): 121–130.

徐龙炳, 李琛, 陈倩雯, (2021). 信息型市场操纵与财富转移效应研究 —— 基于上市公司内部人减持的视角[J]. 财经研究, 47(5): 4–18.

Xu L B, Li C, Chen Q W, (2021). Information Based Market Manipulation and Wealth Transfer Effect: Evidence from Insiders' Shares Selling[J]. Journal of Finance and Economics, 47(5): 4–18.

许泳吴, 徐鑫, 朱菲菲, (2022). 中国 A 股市场的" 大单异象" 研究 [J]. 管理世界, 38(7): 120–136.

Xu Y H, Xu X, Zhu F F, (2022). The "Large-Volume Trading Anomaly" in China's A-Share Market[J]. Journal of Management World, 38(7): 120–136.

杨之曙, 姚松瑶, (2004). 沪市买卖价差和信息性交易实证研究 [J]. 金融研究, (4): 45–56.

Yang Z S, Yao S Y, (2004). An Empirical Study of Bid-ask Spread and Information-based Trading in

Shanghai A-share Market[J]. Journal of Financial Research, (4): 45–56.

曾庆生, (2008). 公司内部人具有交易时机的选择能力吗? —— 来自中国上市公司内部人卖出股票的证据[J]. 金融研究, (10): 117–135.

Zeng Q S, (2008). Can Insiders "Time" the Market When they Sell Their Corporate Stock? Evidence from China's Stock Market[J]. Journal of Financial Research, (10): 117–135.

周志华, (2016). 机器学习 [M]. 北京: 清华大学出版社.

Zhou Z H, (2016). Machine Learning[M]. Beijing: Tsinghua University Press.

朱红兵, 张兵, (2020). 价值性投资还是博彩性投机? —— 中国 A 股市场的 MAX 异象研究 [J]. 金融研究, (2): 167–187.

Zhu H B, Zhang B, (2020). Investment or Gambling? The MAX Anomaly in China's A-Share Stock Market[J]. Journal of Financial Research, (2): 167–187.

Aitken M J, Aspris A, Foley S, de B Harris F H, (2018). Market Fairness: The Poor Country Cousin of Market Efficiency[J]. Journal of Business Ethics, 147(1): 5–23.

Ali U, Hirshleifer D, (2017). Opportunism as a Firm and Managerial Trait: Predicting Insider Trading Profits and Misconduct[J]. Journal of Financial Economics, 126(3): 490–515.

Andersen T G, Bondarenko O, (2014). VPIN and the Flash Crash[J]. Journal of Financial Markets, 17(1): 1–46.

Benabou R, Laroque G, (1992). Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility[J]. The Quarterly Journal of Economics, 107(3): 921–958.

Breiman L, (1996). Stacked Regressions[J]. Machine Learning, 24(1): 49–64.

Carhart M M, (1997). On Persistence in Mutual Fund Performance[J]. The Journal of Finance, 52(1): 57–82.

Chen Y, Zhao H, (2012). Informed Trading, Information Uncertainty, and Price Momentum[J]. Journal of Banking and Finance, 36(7): 2095–2109.

Cumming D, Ji S, Peter R, Tarsalewska M, (2020). Market Manipulation and Innovation[J]. Journal of Banking and Finance, 120(11): 105957.

Czech R, Huang S, Lou D, Wang T, (2021). Informed Trading in Government Bond Markets[J]. Journal of Financial Economics, 142(3): 1253–1274.

Dang C, Foerster S, Li Z, Tang Z, (2021). Analyst Talent, Information, and Insider Trading[J]. Journal of Corporate Finance, 67(2): 101803.

De Bondt W F, Thaler R, (1985). Does the Stock Market Overreact?[J]. The Journal of Finance, 40(3): 793–805.

Dittmar A, Field L C, (2015). Can Managers Time the Market? Evidence Using Repurchase Price Data[J]. Journal of Financial Economics, 115(2): 261–282.

Easley D, de Prado M M L, O'Hara M, (2011). The Microstructure of the Flash Crash: Flow Toxicity, Liquidity Crashes and the Probability of Informed Trading[J]. Journal of Portfolio Management, 37(2): 118–128.

Easley D, de Prado M M L, O'Hara M, (2012). Flow Toxicity and Liquidity in a High-frequency World[J]. The Review of Financial Studies, 25(5): 1457–1493.

Easley D, Engle R F, O'Hara M, Wu L, (2008). Time-Varying Arrival Rates of Informed and Uninformed Trades[J]. Journal of Financial Econometrics, 6(2): 171–207.

Easley D, Hvidkjaer S, O'Hara M, (2002). Is Information Risk a Determinant of Asset Returns[J]. The Journal of Finance, 57(5): 2185–2221.

Easley D, Kiefer N M, O'Hara M, Paperman J B, (1996). Liquidity, Information and Infrequently Traded Stocks[J]. The Journal of Finance, 51(4): 1405–1436.

Fama E F, French K R, (1993). Common Risk Factors in the Returns on Stocks and Bonds[J]. Journal of Financial Economics, 33(1): 3–56.

Fama E F, French K R, (2015). A Five-factor Asset Pricing Model[J]. Journal of Financial Economics, 116(1): 1–22.

George T J, Hwang C, (2010). A Resolution of the Distress Risk and Leverage Puzzles in the Cross Section of Stock Returns[J]. Journal of Financial Economics, 96(1): 56–79.

Jegadeesh N, Titman S, (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency[J]. The Journal of Finance, 48(1): 65–91.

Liu J, Stambaugh R F, Yuan Y, (2019). Size and Value in China[J]. Journal of Financial Economics, 134(1): 48–69.

Newey W K, West K D, (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix[J]. Econometrica, 55(3): 703–708.

Novy-Marx R, (2013). The Other Side of Value: The Gross Profitability Premium[J]. Journal of Financial Economics, 108(1): 1–28.

O'Hara M, (1997). Market Microstructure Theory[M]. London: Blackwell Publishers.

Wolpert D H, (1992). Stacked Gerneralization[J]. Neural Networks, 5(2): 241–260.

# 附录

附录 A

**表 A1　训练集样本生成方法**

| 步骤名称 | 操作说明 |
|---|---|
| 步骤 1: 筛选 “清洁窗口公告” | 基于 A 股上市公司的公告发布日期数据, 若上市公司在 $t-6$ 至 $t-1$ 日没有公告发布, 则 $t$ 日发布的公告则为清洁窗口公告. |
| 步骤 2: 筛选 “价格敏感性公告” | 在清洁窗口公告集合中, 以 $t-210$ 至 $t-10$ 之间的 200 个交易日为参数估计区间, 使用股票日收益率和市场日收益率数据拟合 CAPM 模型获得回归系数, 并计算公司股票在公告发布前后 $t-6$ 至 $t+6$ 的异常收益率 (abnormal return, AR). $$R_{i,t}=\hat{\alpha}+\hat{\beta}_i R_t^M+\varepsilon_{i,t},$$ $$AR_{i,t}=R_{i,t}-\left(\hat{\alpha}+\hat{\beta}_i R_t^M\right).$$ 若 $t-6$ 至 $t+6$ 之间的异常收益率在 10% 的置信水平下显著异于 0, 则判定此公告为价格敏感性公告. |
| 步骤 3: 筛选 “信息泄露公告” | 在价格敏感性公告集合中, 若 $t-6$ 至 $t-1$ 期间的异常收益率在 10% 的置信水平下显著异于 0, 表明在公告发布前公告信息已经对股价产生了显著影响, 则判定此公告内容发生了信息泄露. |
| 步骤 4: 筛选 “知情交易发生日” | 在被判定为发生信息泄露的公告中, 使用日度成交量数据, 若在公告发布前 $t-6$ 至 $t-1$ 期间内, 某个交易日的成交量大于 $t-260$ 至 $t-10$ 期间日成交量均值的 3 个标准差, 则判定此交易日为知情交易行为的发生日期. 具体如下: $$\Delta_{i,t-j}>\frac{1}{250}\sum_{k=t-260}^{k=t-11}\Delta_{i,k-j}+3\sigma_{i,t},\qquad \forall j=1,\cdots,6.$$ |
| 步骤 5: 设置训练集样本标签 | “买入型” 知情交易样本: 知情交易发生日当天收益为正, 则为划分为类别 1, 命名为 “买入型” 知情交易日. “卖出型” 知情交易样本: 知情交易发生日当天收益为负, 则划分为类别 2, 命名为 “卖出型” 知情交易日. 未发生知情交易样本: 在 “价格敏感性公告” 样本集合中, 选取公告前时段内超额收益不显著、公告后超额收益显著的事件样本, 表明该公告未发生信息泄露. 所得事件样本中公告前时段内未出现异常交易现象的交易日, 则划分为类别 0, 构成未发生知情交易的样本集合. |

**表 A2　特征变量说明**

| 开盘操纵指标 | |
|---|---|
| 开盘涨跌幅 | 开盘价相对前日收盘价涨跌幅 |
| 开盘交易量 5 min | 开盘后 5 分钟累计交易量 |
| 开盘交易量 10 min | 开盘后 10 分钟累计交易量 |
| 开盘交易量 15 min | 开盘后 15 分钟累计交易量 |
| 开盘每笔交易量 5 min | 开盘后 5 分钟平均每笔交易量 |
| 开盘每笔交易量 10 min | 开盘后 10 分钟平均每笔交易量 |
| 开盘每笔交易量 15 min | 开盘后 15 分钟平均每笔交易量 |

**表 A2 (续)**

| | |
|---|---|
| 开盘操纵指标 | |
| 开盘涨跌幅 5 min | 开盘后 5 分钟价格相对当日开盘价涨跌幅 |
| 开盘涨跌幅 10 min | 开盘后 10 分钟价格相对当日开盘价涨跌幅 |
| 开盘涨跌幅 15 min | 开盘后 15 分钟价格相对当日开盘价涨跌幅 |
| 尾盘操纵指标 | |
| 尾盘交易量 5 min | 收盘前 5 分钟至收盘时刻间累计交易量 |
| 尾盘交易量 10 min | 收盘前 10 分钟至收盘时刻间累计交易量 |
| 尾盘交易量 15 min | 收盘前 15 分钟至收盘时刻间累计交易量 |
| 尾盘每笔交易量 5 min | 收盘前 5 分钟至收盘时刻间平均每笔交易量 |
| 尾盘每笔交易量 10 min | 收盘前 10 分钟至收盘时刻间平均每笔交易量 |
| 尾盘每笔交易量 15 min | 收盘前 15 分钟至收盘时刻间平均每笔交易量 |
| 尾盘涨跌幅 5 min | 当日收盘价相对收盘前 5 分钟价格涨跌幅 |
| 尾盘涨跌幅 10 min | 当日收盘价相对收盘前 10 分钟价格涨跌幅 |
| 尾盘涨跌幅 15 min | 当日收盘价相对收盘前 15 分钟价格涨跌幅 |
| 订单簿特征 | |
| 订单簿深度 1 | $\text{买价}_1 \times \text{买入数量}_1 + \text{卖价}_1 \times \text{卖出数量}_1$ |
| 订单簿深度 2 | $\sum_{i=1}^{5}(\text{买价}_i \times \text{买入数量}_i + \text{卖价}_i \times \text{卖出数量}_i)$ |
| 相对买卖报价差 | $\frac{\text{买价}_1-\text{卖价}_1}{0.5\times(\text{买价}_1+\text{卖价}_1)}$ |
| 相对有效买卖报价差 | $2\times\frac{\lvert\text{成交价}-0.5\times(\text{买价}_1+\text{卖价}_1)\rvert}{0.5\times(\text{买价}_1+\ \text{卖价}_1)}$ |
| 高频量价特征 | |
| 分笔价格 | 日内高频成交价 |
| 分笔交易量 | 日内高频成交量 |
| 分笔平均交易量 | 每 3 s 间隔累计成交量除以期间成交笔数 |
| 短期价格波动比 | 每 5 min 间隔内成交价极差除以成交价最低值 |
| 短期流动性指标 | 每 5 min 间隔期间成交价除以期间交易额 |

注: Mean、STD、Max、Min、Max-Min、Skew、Kurt 分别代表均值、标准差、最大值、最小值、极差、偏度和峰度, 用以度量相应特征日内变化情况.

**表 A3 因子回归检验 (市值加权组合)**

| | CAPM | FF3 | Carhart4 | FF5 | CH3 | CH4 |
|---|---|---|---|---|---|---|
| Panel A: 做多无知情交易股, 做空高知情交易股 | | | | | | |
| intercept | 0.0126** | 0.0127** | 0.0124** | 0.0181*** | 0.0133** | 0.0139** |
| | (2.38) | (2.38) | (2.32) | (3.20) | (2.31) | (2.40) |
| MKT | 0.1168 | 0.1148 | 0.0871 | 0.0682 | 0.0779 | 0.0445 |
| | (1.48) | (1.45) | (1.02) | (0.78) | (0.86) | (0.46) |
| SMB | | 0.1527 | 0.2484 | −0.3487 | 0.0951 | 0.1858 |
| | | (1.01) | (1.34) | (−1.34) | (0.56) | (0.95) |
| HML | | 0.1721 | 0.2003 | 0.0209 | | |
| | | (1.20) | (1.36) | (0.11) | | |
| UMD | | | 0.1062 | | | |
| | | | (0.89) | | | |
| RMW | | | | −0.5016* | | |
| | | | | (−1.87) | | |

**表 A3 (续)**

| | CAPM | FF3 | Carhart4 | FF5 | CH3 | CH4 |
|---|---|---|---|---|---|---|
| CMA | | | | 0.2006 | | |
| | | | | (0.79) | | |
| VMG | | | | | −0.0684 | 0.0113 |
| | | | | | (−0.35) | (0.05) |
| PMO | | | | | | −0.1689 |
| | | | | | | (−1.02) |
| $N$ | 72 | 72 | 72 | 72 | 72 | 72 |
| $R^2$ | 0.030 | 0.062 | 0.073 | 0.138 | 0.043 | 0.058 |
| Panel B: 做多低知情交易股, 做空高知情交易股 | | | | | | |
| intercept | 0.0075 | 0.0076 | 0.0071 | 0.0126** | 0.0091* | 0.0093* |
| | (1.58) | (1.58) | (1.50) | (2.48) | (1.76) | (1.77) |

注: *、**、*** 分别表示 10%、5% 和 1% 的显著性水平.

**表 A4　因子回归检验 (去除壳价值因素)**

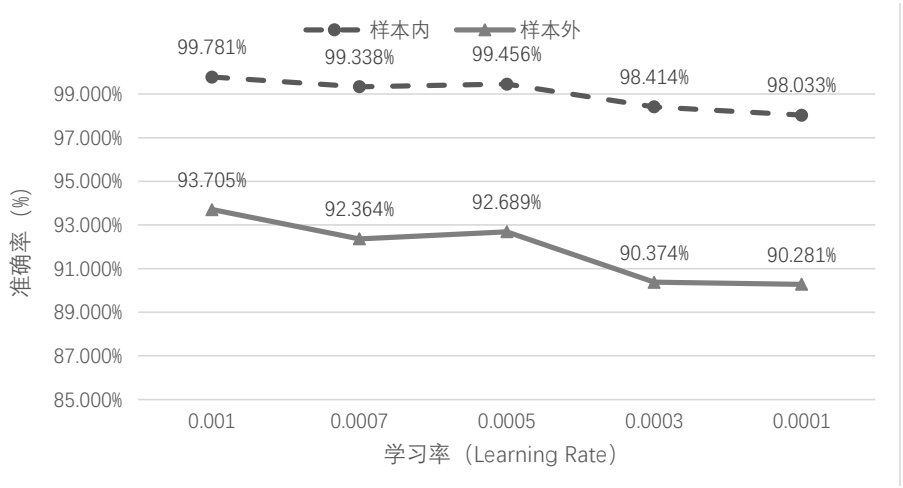
| | CAPM | FF3 | Carhart4 | FF5 | CH3 | CH4 |
|---|---|---|---|---|---|---|
| Panel A: 做多无知情交易股, 做空高知情交易股 | | | | | | |
| intercept | 0.0127** | 0.0122** | 0.0117** | 0.0142** | 0.0141** | 0.0151*** |
| | (2.50) | (2.38) | (2.31) | (2.51) | (2.54) | (2.79) |
| MKT | 0.0943 | 0.0844 | 0.0263 | 0.0854 | 0.0494 | −0.0155 |
| | (1.25) | (1.10) | (0.33) | (0.98) | (0.57) | (−0.17) |
| SMB | | 0.0640 | 0.2651 | −0.1040 | 0.0548 | 0.2270 |
| | | (0.44) | (1.51) | (−0.40) | (0.33) | (1.24) |
| HML | | −0.0364 | 0.0228 | −0.1639 | | |
| | | (−0.26) | (0.16) | (−0.82) | | |
| UMD | | | 0.2231* | | | |
| | | | (1.98) | | | |
| RMW | | | | −0.1047 | | |
| | | | | (−0.39) | | |
| CMA | | | | 0.2075 | | |
| | | | | (0.82) | | |
| VMG | | | | | −0.1279 | 0.0110 |
| | | | | | (−0.67) | (0.05) |
| PMO | | | | | | −0.3194** |
| | | | | | | (−2.07) |
| $N$ | 72 | 72 | 72 | 72 | 72 | 72 |
| $R^2$ | 0.022 | 0.039 | 0.092 | 0.055 | 0.038 | 0.097 |
| Panel B: 做多低知情交易股, 做空高知情交易股 | | | | | | |
| intercept | 0.0109** | 0.0109** | 0.0104** | 0.0128*** | 0.0122** | 0.0135*** |
| | (2.50) | (2.47) | (2.41) | (2.66) | (2.55) | (3.00) |

注: *、**、*** 分别表示 10%、5% 和 1% 的显著性水平. 为检验本文结论是否受到 A 股市场壳价值因素的干扰, 本文参照 Liu et al. (2019) 对于 A 股市场壳价值因素的相关研究在此汇报了剔除市值占比处于后 30% 的小市值股票后的组合超额收益检验结果. 结果可见剔除壳价值因素后组合超额收益仍然存在, 本文结论保持稳健.

**表 A5　不同市场平稳性下集成神经网络模型的预测准确率**

| | 集成神经网络模型的样本外识别准确率 | | |
|---|---|---|---|
| | 低市场波动区间 | 高市场波动区间 | 高波动区间-低波东区间 |
| 2015 | 87.994% | 97.688% | 9.694% |
| 2016 | 88.112% | 88.216% | 0.104% |
| 2017 | 97.143% | 91.740% | −5.403% |
| 2018 | 91.749% | 89.866% | −1.883% |
| 2019 | 91.908% | 93.333% | 1.425% |
| 2020 | 91.645% | 93.557% | 1.912% |
| Avg | 91.425% | 92.400% | 0.975% |

注：本文依据上证指数波动率的中位数将资本市场环境逐年划分为高波动区间和低波动区间，并用集成神经网络模型对处于高波动区间和低波动区间的测试集数据样本分别进行预测判别，以此比较在不同市场平稳程度下本文模型的预测结果是否存在显著变化．结果可见总体上本文模型的预测准确率不受市场平稳程度的影响，平均而言高波动区间的预测准确率较低波动区间上升 0.975%，不足 1%，证实了本文模型在不同市场平稳环境下的稳健性．




注：本图旨在证实集成神经网络模型的判别能力在不同超参数设定下的稳定性．具体本文以 2013–2019 年的数据样本划分为训练集与验证集进行模型训练，以 2020 年数据样本作为测试集进行模型的样本外预测能力检验．结果可见相较最优参数，在其他学习率参数设定下模型的识别准确率确实存在一定程度的降低，但总体而言模型的样本内外识别准确率保持稳定并均维持在较高水平，证明了结论的稳健性．

**A1　不同学习率参数设定下集成神经网络模型的识别准确率**