

# 价格跳跃前大中小单的行为特征和信息含量<sup>①</sup>

万 谍<sup>1</sup>, 杨晓光<sup>2</sup>

(1. 浙江工商大学金融学院, 杭州 310018; 2. 中国科学院数学与系统科学研究院, 北京 100190)

**摘要:** 利用中证800指数成分股3年的分笔交易数据, 将股票成交订单分成大单、中单和小单, 然后分析跳跃前的交易指令与跳跃的关联性及其对跳跃的预测能力. 对跳跃样本和匹配有消息发布的跳跃样本进行事件分析和回归分析, 发现跳跃前有明显的、逐渐加强的、与跳跃方向相对应的异常交易, 而且异常交易指令与跳跃的发生概率、方向、大小和匹配的信息有不同程度的显著关联性; 进一步的预测分析也发现, 在控制流动性指标后, 异常交易指令可以提升模型对跳跃的预测效果. 这些结果显示市场在跳跃前存在着信息泄露, 且信息泄露随时间扩散. 此外, 事件分析、回归模型和预测效果均显示: 跳跃前中单的信息含量一般高于大单和小单, 说明中国股市在极端价格变化前存在隐秘交易

**关键词:** 价格跳跃; 交易指令; 知情交易; 隐秘交易; 跳跃预测

**中图分类号:** F832.5   **文献标识码:** A   **文章编号:** 1007-9807(2019)10-0037-18

## 0 引 言

跳跃指的是价格过程的不连续点. 跳跃代表着高强度的信息对市场价格形成了显著的冲击. 但跳跃时刻并非是信息首次冲击市场的时间, 有大量研究表明跳跃前存在信息泄露. 美国市场中, Jiang等<sup>[1]</sup>发现跳跃前5min的流动性指标, 如买卖价差、市场深度, 对跳跃有显著预测作用. Boudt和Petitjean<sup>[2]</sup>发现跳跃前1h有效价差和成交笔数有异常的变化. 针对中国市场, Wan等<sup>[3]</sup>发现跳跃前0.5h的成交笔数、每笔成交量和买卖深度有异常变化, 且变化的方向与跳跃的方向一致, 显示着跳跃前的信息泄露. 但是, 单纯地从流动性指标无法细致描述与跳跃相关的信息泄露过程. 注意到信息需要通过投资者的下单行为才能传递到市场上, 而下单的大小不仅在统计意义上对应着不同的投资者类型, 而且订单金额的显著改变也昭示着投资者下单策略的改变. 在跳跃发生前, 信息是如何通过投资者的下单行为泄露到市场中

的, 泄露信息的指令会否与跳跃的发生有关系? 本文基于分笔交易数据, 按照交易金额将成交订单分为大、中、小单, 分析各类指令的使用者在跳跃前的行为特征及指令间的信息含量差异, 从而更清晰地展现跳跃前的信息泄露过程.

研究跳跃前信息泄露过程, 至少有两方面的意义: 其一, 研究信息含量随指令大小的分布情况, 可以了解信息优势者的下单策略. 有大量研究发现, 信息交易者偏好采用大单迅速完成交易, 使得大单造成的市场冲击更明显<sup>[4,5]</sup>; 但也有研究指出, 信息交易者可能通过拆单隐藏交易目的, 使得中小单的市场影响超过大单<sup>[6,7]</sup>. 因此, 并不能断定跳跃前大单包含着主导信息. 本文将通过比较跳跃前各类指令的信息含量来分析信息优势者的下单策略, 并讨论其可能的原因. 其二, 研究跳跃前投资者的行为特征与跳跃的关联性, 可以更好地了解跳跃的可预测性. 出乎意料的消息发布会引发跳跃<sup>[8-10]</sup>, 但信息的发布要事后才能确定, 不能用

① 收稿日期: 2017-06-11; 修订日期: 2019-03-08

基金项目: 国家自然科学基金资助项目(71501170; 71431008; 71532013)

作者简介: 万 谍(1986—), 四川眉山人, 博士, 副教授, Email: wandie@amss.ac.cn

于预测跳跃,所以应该从事前的信息泄露过程中寻找跳跃的预测因子.之前分析跳跃可预测性的文章,大都只证实跳跃前流动性指标对跳跃发生概率的预测能力<sup>[1,2]</sup>,本文将探讨跳跃前的各类交易指令对跳跃的预测作用.

本文的具体研究方案如下.使用由中国 A 股市场 777 支个股的 3 年分笔交易数据构成的大型数据集,并按金额大小的分位点将成交订单分为大单、中单和小单,然后将日内价格跳跃作为事件,为跳跃股票配对一支股价和市值都接近的无跳跃股票,考察跳跃前大、中、小单在跳跃样本和配对样本之间的差异性.如果某类指令在两类样本间差异明显且该差异都表现为正跳跃前买入、负跳跃前卖出,则说明该指令的使用者拥有更多的信息.之后进一步通过回归分析检测各类指令与跳跃的发生概率、方向和大小的关联性,进而比较大、中、小单的信息含量,推测信息优势者的下单策略及其与跳跃的关系.

由于并非所有的跳跃都有公开的消息发布匹配<sup>[8-10]</sup>,所以对于统计方法检测出的跳跃而言,既可能是公开发布的消息造成的,也可能源于隐秘的私有消息.由于隐秘消息不会有公开的记载,难以通过资料进行辨识.只有那些与公开发布的消息有显著关联的跳跃样本有显性的材料,可以作为检验本文研究问题的稳健性分析样本.因此,除了采用单纯的跳跃样本外,还将跳跃与业绩快报、业绩预测报告和分红送转报告等消息发布匹配,测试有信息匹配的跳跃前的异常订单及其与跳跃和发布信息的关联性,从而更稳健地比较各类指令的信息含量.

最后,为了测试跳跃前异常指令对跳跃的预测能力,将通过随机抽样将跳跃及其匹配样本分成训练样本和预测样本,检测控制了流动性指标后,大单、中单和小单对跳跃发生概率、方向和大小预测能力差异.如果信息交易者提前获知引发跳跃的信息并据此交易,则前面发现的与跳跃显著相关的异常指令应该能够在流动性指标的基础上提升模型的预测能力.此外,如果某类指令相比其他指令能更多地提升模型的预测能力,其信息含量也将更高,因此预测能力的比较还可以进一步加强先前步骤中对各类指令的信息含量差异检测的稳健性.

## 1 文献回顾及假设

理论和实证研究都指出信息交易者会偏好使用某类指令来完成交易.理论研究方面,Kyle<sup>[11]</sup>指出信息交易者试图通过分散成交来掩饰其目的.Easley 和 O'Hara<sup>[12]</sup>的模型指出,做市商会通过以往提交订单的大小来推断新消息发布的可能性,使得订单大小和提交顺序会影响价格冲击,因此信息交易者会权衡指令的冲击成本和信息优势带来的获利,从而选择最优的指令大小和提交顺序.Holden 和 Subrahmanyam<sup>[13]</sup>发现,在多人掌握不同质信息的情况下,信息交易者会互相竞争而用大单交易,而 Foster 和 Viswanathan<sup>[14]</sup>则强调交易指令的大小由其掌握信息的相关度决定.总结现有的理论模型结果,如果信息相关性较强,或者流动性足够好,信息优势者会偏好大单交易;而当意外消息足够突然,以至于只有少数人掌握信息时,投资者可能在竞争市场份额的压力不大或者流动性受限的情况下,采用中小单减小价格冲击同时隐藏交易目的,以获得更大的收益.

针对信息交易者的订单大小选择策略的实证研究,分为两类:一类针对某个样本期内的所有交易指令,并不针对具体某个事件.比如,Barclay 和 Warner<sup>[6]</sup>针对纽交所股票 3 年交易数据的研究发现,与其数量占比比较,中单造成的价格冲击更大,从而证实信息优势者主要采取中单隐蔽成交.Chakravarty<sup>[7]</sup>进一步证实,价格冲击更大的中单主要是由机构提交的.Alexander 和 Peterson<sup>[15]</sup>发现投资者倾向于使用 500 股、1 000 股和 1 500 股这样的整数单位下单,且隐秘交易者倾向于用中等规模的整单来隐藏其交易.Chan 和 Fong<sup>[4]</sup>发现对于纽交所的股票,大单的不平衡对收益率的影响比小单的不平衡更大.Cai 等<sup>[16]</sup>基于上证 A 股 2001 年 9 月的分笔交易数据的研究发现,虽然大单和中单都造成了与其订单规模不成比例的价格冲击,但大单的累计价格冲击最显著.朱燕建等<sup>[5]</sup>也证实中国市场大单的价格影响力最高.另一类的研究,主要针对某个具体事件,考察事件发生前信息交易者的下单策略.比如,Tong 等<sup>[17]</sup>发现在大小非股票解禁公告前大单的量会大幅增加.Barclay 和 Warner<sup>[6]</sup>发现要约收购事件前中单

的价格冲击比大单和小单更强,从而证明美国市场中隐秘交易在具体事件和一般情形下都成立。

相比美国市场无论是一般情形还是具体事件前都有隐秘交易的证据<sup>[6,7,15]</sup>,中国市场中缺乏针对隐秘交易的实证证据,更缺乏对隐秘交易形式的实证刻画。朱燕建等<sup>[5]</sup>认为相对于发达市场中更常用的拆单隐秘成交策略,中国市场中内幕交易惩处力度偏低使得信息投资者偏好采用大单迅速完成交易。但事实上,Cai 等<sup>[16]</sup>和朱燕建等<sup>[5]</sup>都只针对样本期内的所有指令,并没有聚焦于某个事件。Tong 等<sup>[17]</sup>针对的大小非解禁样本,其发生的突然性和保密性并不强,所以并没有发现隐秘交易的存在。根据上面的理论模型结果<sup>[12-14]</sup>,当意外消息足够突然时,投资者有更大的可能选择拆单隐藏交易目的。而价格跳跃代表着意外信息的冲击突然爆发,中国市场的信息交易者会否在跳跃前权衡大单的冲击成本和收益后拆单呢? 本文将通过分析各类订单与跳跃及跳跃匹配信息的关联性来测试如下的隐秘交易假设:

**H1** 跳跃前中单信息含量比大、小单更高。

H1 成立要求事件分析前有异常的中单数量和规模,且其方向与跳跃一致(即正跳跃前买入,负跳跃前卖出),同时中单数量和规模与跳跃及跳跃匹配的信息的相关性比大、小单更强。

关于跳跃的可预测性,之前的研究试图分析和证实跳跃与信息的出乎意料程度的关联性<sup>[8-10,18-20]</sup>。但信息的出乎意料程度事后才可知,并不能用于跳跃的预测。Fuss 等<sup>[21]</sup>发现跳跃前过量的小道消息(noisy ticker news)在控制跳跃前日内交易和消息的异常量后仍旧对跳跃有显著解释力。Jiang 等<sup>[1]</sup>指出跳跃之前的流动性变化(深度、买卖价差等)在控制了消息的出乎意料部分后对跳跃的发生概率有预测能力。Boudt 和 Petitjean<sup>[2]</sup>发现跳跃前的流动性冲击对跳跃形成有明显的推动作用。过量的小道消息和流动性变化都代表着信息可能提前泄露,其对跳跃的预测能力其实意味着泄露的信息与引发跳跃的信息高度相关。泄露的信息总是要通过投资者的下单和

交易作用于市场,如果泄露的信息确实是引发跳跃的信息,那么异常的交易指令就应该与跳跃相关从而对跳跃有预测能力。因此,有必要进一步考察跳跃前投资者的交易指令对跳跃的预测作用。如此,本文假设:

**H2** 跳跃前投资者的异常交易指令对跳跃有预测能力。

H2 成立意味着跳跃前有代表信息泄露的异常指令,且这些指令的使用者确实知晓引发跳跃的信息,使得其交易结果对跳跃有预测能力。由于跳跃包含 3 个维度:跳跃发生概率、跳跃方向和跳跃大小,本文将通过随机抽样把样本分类为训练样本和测试样本,同时在原跳跃样本和有信息匹配的跳跃样本中测试各类指令对跳跃的 3 个维度的预测能力。

## 2 数 据

本文的分笔数据来自万得数据库。该分笔数据提供单笔交易的成交价和成交量,同时提供最近的 10 档报价。样本包含 2013 年至 2015 年中,中证 800 成分股(以 2015 年底为准)所有成交数据。中证 800 包含上海和深圳交易所中流动性最佳的 800 支股票,且该样本期内中国股市恰好经历了低位震荡、牛市和熊市这 3 类市场状态,包含了各类可能的市场环境,从而可以增进结果的稳健性。由于有一些成分股在 2013 年之后才上市,将这些数据长度不足的股票剔除后,样本中还剩下 777 支股票。

分笔交易数据可以将成交指令按规模进行分类。不同于 Zhang 和 Chang<sup>[22]</sup>按照资金量和成交手数的绝对分类方法,本文的分类方法与 Tong 等<sup>[17]</sup>类似:将所有的股票按照 2014 年 12 月 31 日的市值和股价的中位数分为了 4 类,计算出每类股票在该日所有成交的分位点,按照交易金额的分位点可以将所有样本期内的成交分为 3 类(小于 25%,25%至 95%,95%以上)<sup>②</sup>。最终的分值计算结果如表 1 所示。表 1 中报告了各类

② 选取 2014 年 12 月 31 日的交易的原因有 2 个:首先是计算机的能力只够计算 1 个交易日的某类股票的所有成交序列的分位点;而 2014 年底恰好上证指数升至 3200 点左右,位于牛市和整个样本期的大致中点,故其单笔成交较有代表性。选取其他的交易日进行分类并不会导致结果有很大的不同,因为这些分位点的计算涉及的序列长度超过 200 万,反映了多支股票的参与者的交易行为,而这些行为导致的指令分布在整体上应该是较稳健的。

股票在2014年12月31日的各笔成交金额的5个分位点和均值;所有的成交按照25%和95%分位点划分为小单、中单和大单。低于25%的单笔成交额都在5500元以下,称之为小单;超过95%分位点的交易额都在70000元以上,称之为大单;其余的成交指令,称之为中单<sup>③</sup>。选取25%和95%作为大、小单的阈值,一则是为了获取充分多

的大单和小单样本以保证数据建模的可靠性,二则是希望大单和小单能分别对机构和散户的交易有一定的代表性,从而便于对大中小单信息含量差异的解读<sup>④</sup>。表1中的粗体即为各类股票的大小单阈值,即对于某一支股票,其大小单的划分方法即为其对应的分类中每笔成交金额的25%和95%分位点。

表1 单个交易日全样本股票交易的描述性统计

Table 1 Summary statistics of stock trading sample in one trading day

分位点	5%	25%	50%	均值	75%	95%
面板 A: 全样本股票(777支)						
市值/亿元	31.88	70.80	112.23	313.65	222.60	916.15
收盘价/元	5.17	8.89	13.71	17.17	21.12	39.57
成交金额/元	1 676.27	<b>4 277.24</b>	9 532.06	25 671.78	21 991.10	<b>93 924.27</b>
交易量/股	99.96	322.02	721.55	1 960.91	1 682.25	7 130.81
面板 B: 收盘价超过13.71元且市值大于112亿元的股票(221支)						
市值/亿元	116.20	151.24	225.77	451.63	398.16	1 718.31
收盘价/元	14.41	16.60	21.92	26.19	30.30	47.81
成交金额/元	2 517.25	<b>5 313.60</b>	11 453.86	32 980.97	27 000.26	<b>121 453.82</b>
交易量/股	98.92	230.64	502.67	1 439.73	1 175.17	5 319.72
面板 C: 收盘价低于13.71元且市值大于112亿元的股票(175支)						
市值/亿元	117.86	152.30	215.41	690.85	426.78	2104.14
收盘价/元	4.03	6.45	8.72	8.58	10.79	13.09
成交金额/元	859.56	<b>3 610.76</b>	7 987.50	23 245.79	19 096.96	<b>82 231.88</b>
交易量/股	101.87	451.55	1 006.12	3 009.58	2 465.46	10 572.12
面板 D: 收盘价超过13.71元且市值小于112亿元的股票(164支)						
市值/亿元	20.67	45.07	67.08	66.30	87.11	105.70
收盘价/元	14.38	16.54	20.28	24.12	28.07	43.71
成交金额/元	2 328.33	<b>4 511.37</b>	10 494.62	26 287.41	23 527.04	<b>96 904.83</b>
交易量/股	97.72	211.09	490.30	1 216.63	1 084.60	4 481.04
面板 E: 收盘价低于13.71元且市值小于112亿元的股票(217支)						
市值/亿元	31.98	56.80	71.37	71.65	87.36	106.68
收盘价/元	4.78	7.59	9.52	9.34	11.30	13.31
成交金额/元	935.65	<b>3 547.56</b>	8 003.62	19 517.81	17 874.08	<b>72 281.52</b>
交易量/股	101.31	401.72	905.54	2 254.47	2 058.03	8 366.11

注: 面板 A 报告了所有股票在2014年12月31日收盘时的市值、收盘价、当日单笔成交金额、当日单笔成交量的5%、25%、50%、均值、75%、95%分位点; 面板 B 至面板 E 是将所有股票按照全样本当日收盘价和市值的中位数分类后, 分别报告了这些分位点数值。数据都来自 WIND 数据库。由于平均每支股票每日交易笔数超过 10 万笔, 计算机内存不足以计算全部股票所有成交金额和交易量的分位点(无法存储一个长度超过 200 万的向量), 所以成交额和交易量的分位点实际上是各个股票相应的分位点的平均值。

③ 本文认为小单应该是以散户的交易为主, 大单应该是以机构交易为主。对于足够大金额的单自然可以认为是机构所为, 而足够小的单可以认为散户为主的原因有三: 其一, 机构虽然可以采用算法交易(如 VWAP、TWAP)拆分大单, 但拆单增加了等待的机会成本<sup>[23]</sup>, 若要及时完成交易则不可能拆分过细。Barclay 和 Warner<sup>[6]</sup>也指出小单的盈利能力对于机构而言太小, 信息交易者倾向于使用中等规模以上的单交易; 其二, 国内的算法交易刚刚发展起来, 使用率偏低; 其三, 国内散户交易量远超机构(根据上交所 2015 年统计年鉴数据, 2014 年上交所的自然投资者交易账户数量占比达 99% 以上, 其交易量占 85% 以上)。至于中单, 本文认为对机构和散户来说都有可能。

④ 需要注意的是, 划分大小单的目的主要是为了使得不同规模的订单之间有足够的信息量差异。Tong 等<sup>[17]</sup>是按照中位数的 8 倍和 1/4 划分大小单。相比他们的样本时期(2005 年), 本文的样本已经是 8 年后, 通胀累积的情况下, 8 和 1/4 倍的划分已经不足以代表机构和散户的差异。故本文优先选择 25% 和 95% 分位点。没有选择 5% 分位点是因为, 25% 分位点已经低于 5 500 元, 按照单笔最低 5 元的手续费, 5 500 元已经属于超过平均券商手续费率(约万分之四)的亏损交易, 一般应该是散户采用。

本文使用 5 min 区间来考察跳跃前各类指令的变化情况<sup>⑤</sup>。为了保证跳跃的突然性和代表性, 采用了组合方法以稳健地检测跳跃: 只有同时被 Barndorff-Nielsen 和 Shephard<sup>[25]</sup> 方法检测出当天有跳跃且 Lee 和 Mykland<sup>[8]</sup> 方法也在同一天找到了日内跳跃才可以认定跳跃的存在<sup>⑥</sup>。在整个样本(3 年, 777 支股票)中, 共检测到 13 802 个日内跳跃和 12 722 个含有跳跃的交易日。平均每年每支股票跳跃 5 次。所有这些跳跃中, 超过 92% 是单跳跃(即一天只有一个跳跃), 即 11 715 个单跳跃, 其中 9 852 个单跳跃是正跳跃, 1 863 个单跳跃是负跳跃。所以在中国市场中, 发生极端价格暴

涨的可能性超过了极端下跌的可能性。各类指令的数量和平均大小在无跳跃期、跳跃前和跳跃时点的日度、5 min 描述性统计分析的结果在表 2 中显示。表 2 同时还报告了跳跃前和无跳跃时的收益率(*RE*)、累计成交量(*V*), 10 档买价和 10 档卖价的平均等待成交量(*BLQ*、*SLQ*)。总体而言, 跳跃时比无跳跃时有更活跃、更激烈的成交和更充分的流动性供给, 而跳跃前 1 天的交易量和流动性都较小(相比无跳跃期), 而跳跃前 5 min 的交易已经开始活跃和激进起来, 显示信息有临近发布时提前泄露的可能。下面的事件分析将给出更具体的结果。

表 2 各类指令的成交笔数和每笔成交量的描述性统计

Table 2 Summary statistics of number of trades and average trade size for all kinds of orders

变量	无跳跃参考期		跳跃前 2 d		跳跃日		无跳跃 5 min 区间		跳跃前 5 min 区间		跳跃时刻	
	均值	中位数	均值	中位数	均值	中位数	均值	中位数	均值	中位数	均值	中位数
<i>RE</i>	-0.001	0.121	-0.010	0.210	1.612	2.088	0.004	0	0.064 2	0	1.773	2.105
<i>V</i>	27.71	11.27	26.27	11.29	38.48	17.09	0.573	0.190	0.834	0.296	3.265	1.335
<i>BLQ</i>	4.208	1.229	3.762	1.234	3.256	1.202	8.494	2.273	6.388	2.293	7.006	2.481
<i>SLQ</i>	3.405	1.095	3.709	1.174	3.839	1.292	6.871	1.987	8.260	2.464	7.132	2.501
<i>Sbtrades</i> /百万笔	1 862	903	1 740	902	2 534	1 385	38.79	15	54.85	23	231.9	109
<i>Mbtrades</i> /百万笔	4 814	2 236	4 554	2 240	6 875	3 636	100.3	37	153	61	714	348
<i>Lbtrades</i> /百万笔	340.4	127	321.7	128	500.5	228	7.09	2	11.39	4	53.66	24
<i>Sstrades</i> /百万笔	2 077	1 050	1 904	1 021	2 401	1 347	43.27	18	48.44	23	112.8	54
<i>Mstrades</i> /百万笔	5 183	2 510	4 804	2 477	6 188	3 343	108	44	123.2	57	304	151
<i>Lstrades</i> /百万笔	353	142	327.7	141	440.1	205	7.35	2	8.86	3	21.63	9
<i>Sbmsize</i> /手	186.8	175.6	186.3	175	187.4	176.2	187.2	166.7	188.5	169.1	195.9	180.9
<i>Mbmsize</i> /手	1 559	1 408	1 559	1 415	1 551	1 416	1 548	1 358	1 580	1 400	1 569	1 429
<i>Lbmsize</i> /手	16 813	14 504	16 689	14 454	17 200	14 856	15 731	12 390	16 598	13 076	18 616	15 339
<i>Ssmsize</i> /手	184.1	171.4	183.1	170.6	183.6	169.5	183.8	163.2	183.7	164.2	189.8	172.0
<i>Msmsize</i> /手	1 554	1 405	1 555	1 413	1 548	1 408	1 552	1 369	1 570	1 396	1 536	1 380
<i>Lsmsize</i> /手	16 927	14 606	16 742	14 489	17 090	14 878	16 038	12 650	16 367	13 000	17 710	14 247

注: 表中报告了在无跳跃参考期(跳跃前 15 d 至前 5 d)、跳跃前 2 d、跳跃日以及无跳跃 5 min 区间、跳跃前 10 min、跳跃时刻(即发生跳跃的 5 min 区间)这 6 个样本期内的收益率(百分比)、累计交易量、买方 10 档平均深度(万股)、卖方 10 档平均深度(万股)、各类指令的笔数和每笔成交量的均值和中位数。这 6 个样本期都是针对事件分析中的日内跳跃所选定的观测期, 故观测数依次是 69 120, 13 824, 6 192, 331 776, 13 824, 6 192。前 3 个样本期内的指标都是日度数据, 而后 3 个样本期内的指标都是 5 min 数据。其中 *msize* 表示每笔成交量, *trades* 表示成交笔数, 字母 *S*、*M*、*L* 表示小单、中单和大单, 而 *b*、*s* 表示买卖方向。

⑤ 根据 Andersen 等<sup>[24]</sup> 的建议, 5 min 数据可以较好的平衡样本数量和微观结构噪声。

⑥ 具体的检测算法和程序可以联系作者提供。

### 3 事件分析

#### 3.1 方法

本文的事件分析关注跳跃前日度和日内的投资者行为动态. 日度数据常用于研究消息发布的影响或发布前的市场结构性变动, 比如, Tong 等<sup>[17]</sup>研究了股权分置改革中, 解禁消息发布前大单和小单的日度变化情况. 而近期的研究更多地关注日内的交易区间以便于在高频时段对事件的发生和影响作精确的测试. Boudt 和 Petitjean<sup>[2]</sup>将日内跳跃作为事件, 分析了其前后 1 h 的流动性变化. 本文的事件分析法是两者的组合, 具体的方法分为如下 4 个步骤.

**步骤 1** 明确事件和事件窗口. 事件是检测出的日内单个跳跃. 事件窗口是跳跃前 15 d 和日内跳跃前 10 min<sup>⑦</sup>. 把跳跃前 15 d 至跳跃前 6 d 作为参考期, 然后分析跳跃前 2 d 和日内跳跃前 10 min 的交易情况与参考期的差异. 首先, 为了避免可能的日内跳跃互相影响<sup>[2]</sup>, 只用单跳跃作为事件进行研究. 其次, 由于参考期在跳跃前 15 d, 只有之前 15 d 没有跳跃的单跳跃可以作为事件样本. 此外, 跳跃前的参考期和观察期应该是较连续的交易日, 而不是有长期的停牌, 要求这 15 d 的时间跨度在 20 个工作日内, 并且该跳跃有合适的配对股票(见步骤 2). 又因需要跳跃前有 10 min 窗口供进行日内分析, 所以开盘 10 min 内的单跳跃应该剔除, 最后剩余 6 912 个日内跳跃样本, 其中 5 769 个正跳跃, 1 143 个负跳跃. 最终的事件分析样本中, 正负跳跃的比例都与全样本

相近, 故选出的跳跃仍具有代表性.

**步骤 2** 配对股票. 由于行业或市场的系统性因素变化会导致跳跃前的时段许多股票发生类似的交易行为和股价变动, 因此仅仅考察参考期和跳跃前的差异不足以反映跳跃相关的变动. 为了将跳跃相关联的变化与市场或行业的时间趋势变动区别开来, 用 Tong 等<sup>[17]</sup>的方法来为发生跳跃事件的股票进行配对. 配对股票与跳跃股票应该在表 1 中同属一类, 且在该类中两者的收盘价最接近<sup>⑧</sup>. 同时, 配对股票之前 15 d 和之后 5 d 没有发生跳跃.

**步骤 3** 标准化指标. 事件分析关注的是跳跃前各类交易指令的变化情况, 且每类交易指令按买卖方向进行了分类. 每类成交指令的数量和平均成交量分别度量了该类投资者的活跃程度和激进程度. 因此, 主要分析的指标构建如下

$$XBR_t = XBtrades_t / trades_t, \quad (1)$$

$$XZ_t = XBvolume_t / XBtrades_t,$$

$$XSR_t = XStrades_t / trades_t, \quad (2)$$

$$XSZ_t = XSvolume_t / XStrades_t,$$

式中  $t$  表示某个交易日或某个 5 min 交易区间. 'B' 和 'S' 分别表示主动买和主动卖, 而 'R' 表示所占比率, 'Z' 表示平均成交量. 'X' 可以是 L、M、S, 分别表示大单、中单和小单, 即  $LBR_t$  表示在  $t$  时间主动买的大单数量占全部交易笔数的比率<sup>⑨</sup>,  $MSZ_t$  表示在  $t$  时间主动卖的中单指令的平均成交量<sup>⑩</sup>, 其余符号含义以此类推.

为了使得这些指标在日度和日内都可以在不同的股票和不同的交易日之间进行比较, 按照

⑦ Kyle<sup>[11]</sup>的多期模型指出, 信息交易者会选择交易时段和交易资产以最小化交易造成的信息冲击, 因此信息交易者可能并不会等到最后一刻, 而是会较早的开始交易. 因此将关注期从 Boudt 和 Petitjean<sup>[2]</sup>的日内延长到跳跃前 2 d, 以观察早期的信息泄露是否存在.

⑧ 未报告的结果中, 本文采用了随机配对的方法进行稳健性检验, 发现配对方法对结果影响不大, 相关结果可以联系作者提供.

⑨ Jones 等<sup>[26]</sup>发现成交笔数是交易量的主要决定因素, 而 Barclay 和 Warner<sup>[6]</sup>发现相比交易量占比, 成交笔数的占比更能够反映该类指令的累计价格影响比例. 成交笔数代表了这类指令的交易活跃程度, 且成交笔数占比率可以更好地比较不同时段和不同股票的交易指令的分布情况. 故本文主要构建成交笔数的占比率来进行事件研究.

⑩ Chan 和 Fong<sup>[4]</sup>发现不是成交笔数, 而是指令的规模对于价格波动的影响更剧烈. 故本文采用某类指令的平均成交量来度量激进程度, 讨论其激进的交易是否反映其掌握着信息优势. Barclay 和 Warner<sup>[6]</sup>、Tong 等<sup>[17]</sup>都采用了某类指令导致的累计价格变化占总的价格变动的比例来度量激进程度, 但正如 Barclay 和 Warner<sup>[6]</sup>所言, 该度量无法分辨指令的短期冲击和长期影响从而存在偏误, 且本文的分析会在较高频的 5 min 区间, 这类偏误更加明显, 短期冲击导致估计的价格比例指标方差极大, 稳定性和可靠性不足. 故为了度量的准确性, 本文只报告了平均成交量的结果.

Boudt 和 Petitjean<sup>[2]</sup>的方法来对其进行标准化. 标准化的主要思路是计算该指标与同样交易时段的正常情形的百分比差异. 正常情形就是参考期(跳跃前 15 d 至跳跃前 6 d)内该指标在该交易时段的中位数. 即定义标准化的指标为

$$sdR_t = R_t - Med\{R_t | t = -6, \dots, -15\} \quad (3)$$

$$sdZ_t = \frac{Z_t}{Med\{Z_t | t = -6, \dots, -15\}} - 1 \quad (4)$$

式中  $R_t$  表示某个比率指标,  $Z_t$  表示某个实际值度量的指标. 比率指标已经是百分比数据, 故标准化就是减去参考期内该指标的中位数, 而实际值指标需要除以参考期内中位数再减 1. 类似于 Boudt 和 Petitjean<sup>[2]</sup>, 日内指标标准化后, 就基本消除了指标的日内周期性变化对结果的影响.

**步骤 4 检验配对差异.** 事件分析的原假设: 在没有信息泄露的情况下, 跳跃样本和配对样本的指令分布不应有明显差异. 具体操作是针对指标配对差异进行检测. 配对差异类似于双重差分方法, 指的是标准化指标在跳跃股票和配对股票之间的差异. 测试配对差异是否显著区别于 0, 就可以考察相对于参考期的变化中, 跳跃股票是否显著区别于配对股票. 若证明某类指标的配对差异显著大于 0, 则说明存在信息泄露, 且信息交易者采用了这类异常指令进行交易. 若指标配对差异显著小于 0, 则说明信息交易者减少了这类指令的使用. 由于大多数指标都不是正态分布的, 本文用 Kruskal-Wallis 检验来测试各类指标的中位数在跳跃样本和配对样本中是否有显著区别. 检验将分别针对样本指标的正跳跃序列和负跳跃序列. 对于每个指标, 分析样本有 6 912 个, 正跳跃序列长度 5 769, 负跳跃序列长度 1 143. 事件分析报告了各指标配对差异的中位数和  $p$  值.

### 3.2 事件分析结果

表 3 报告了大、中、小单的主动买和主动卖的

成交笔数占比的事件分析结果. 从日度配对差异来看, 跳跃前早期的信息泄露应该是非常有限的, 正跳跃前 1 d 的大买单和大卖单都有显著的上升, 负跳跃前的大买单有显著的减小, 其余都不显著, 由此可以看出大单交易者是具备信息优势的, 在跳跃前会相应地增加或减少激进成交的数量. 正跳跃前显著上升的大卖单比例可能是源于信息不对称和异质信念: 由激进成交增加所暴露的信息并不足以推断未来的股价变化, 可能引发其余不知情机构因避险而卖出股票. 日内的配对差异给出的结果更显著. 正跳跃前 10 min 内, 买方相对比重显著增加, 其中以中单增加最多, 其次小单, 再次大单<sup>①</sup>; 卖方相对比重显著减少, 其中以中单减少最多, 其次小单、再次大单; 且该趋势随时间增大. 负跳跃前买方相对比重显著减少, 其中以中单减少最多, 其次小单, 再次大单; 卖方相对比重显著增加, 其中以中单增加最多, 其次小单、再次大单; 该趋势也随时间增大. 总结表 3 的结果, 异常的买卖指令与跳跃方向基本一致, 且趋势逐步增大, 说明信息泄露是接近跳跃之前发生的, 而且是逐步扩散的. 投资者倾向于使用中小单作为操作工具, 这与隐秘交易方面的研究是一致的<sup>[6,7,15]</sup>.

表 4 报告的是大、中、小单, 主买和主卖的每笔成交量的事件分析结果. 从日度配对结果来看, 不同于表 3 中成交笔数占比的情况, 跳跃前的较早时期就已经有着激烈的成交预示着信息提前泄露: 正跳跃前 2 d 的小单和中单, 无论买卖方向, 都有显著高于配对股票的每笔成交量. 该结果说明信息交易者在早期为防止信息提前泄露, 倾向于使用较大规模的中小单进行交易, 且考虑到小单成本过高, 中单的规模相对平时增幅更大. 而负跳跃前的日度配对差异中, 只有 1 d 的中单规模有显著增加, 且主卖中单的增幅最大, 这也说明信息含量较大的是中单.

① 注意, 大、中、小单的比率配对差异都显著为正并不是矛盾的. 某类指令的显著性仅说明这一序列指标的中位数显著大于同时段的配对股票的相应序列值的中位数, 并不是说任意时刻 3 类指标的占比都上升, 且因为对于某一个样本来说, 这 3 个占比之和为 1, 所以这个检验仅说明在部分跳跃之前, 该类指标的数量有较大的增加.

表3 跳跃前成交笔数占比的事件分析结果

Table 3 Event study results of the ratios of number of trades before jumps

面板 A: 正跳跃前的各类订单成交笔数情况								
变量	日度配对差异				日内配对差异			
	[-2, -1]		[-1, 0]		[-10, -5]		[-5, 0]	
	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>
<i>SBR</i>	0.052	0.390	0.041	0.204	0.694***	0	1.013***	0
<i>MBR</i>	-0.078	0.891	0.108	0.525	4.079***	0	6.453***	0
<i>LBR</i>	0.021	0.223	0.022**	0.017	0.260***	0	0.653***	0
<i>SSR</i>	0.106	0.786	-0.076	0.293	-1.311***	0	-2.226***	0
<i>MSR</i>	-0.123	0.394	-0.113	0.350	-3.391***	0	-4.866***	0
<i>LSR</i>	0.006	0.867	0.031**	0.038	0.000	0.578	-0.247***	0

  

面板 B: 负跳跃前的各类订单成交笔数情况								
变量	日度配对差异				日内配对差异			
	[-2, -1]		[-1, 0]		[-10, -5]		[-5, 0]	
	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>
<i>SBR</i>	0.130	0.598	0.045	0.779	-0.414**	0.031	-0.568***	0.005
<i>MBR</i>	-0.036	0.377	-0.431	0.204	-1.395***	0.001	-1.998***	0
<i>LBR</i>	-0.037	0.344	-0.037*	0.094	0.000	0.584	-0.169***	0.001
<i>SSR</i>	0.227	0.605	0.165	0.309	0.308**	0.038	1.086***	0
<i>MSR</i>	0.052	0.587	0.360	0.452	1.501***	0.001	2.010***	0
<i>LSR</i>	0.029	0.406	0.000	0.451	0.062**	0.021	0.131**	0.017

注: 报告了跳跃前 2 d 和跳跃前 10 min 各类指令在此时间段的占比的事件分析结果, 其中 *LBR*、*MBR*、*SBR* 分别表示大单、中单、小单中的买单占比, *LSR*、*MSR*、*SSR* 分别表示大单、中单、小单中的卖单占比. 所有的比率都已经按照式(3)进行了标准化; 面板 A 中观测数为 5 769 个, 面板 B 中观测数为 1 143 个; ‘\*’, ‘\*\*’ 和 ‘\*\*\*’ 分别表示在 90%、95% 和 99% 置信水平下显著.

表4 跳跃前每笔成交量的事件分析结果

Table 4 Event study results of average trade size before jumps

面板 A: 正跳跃前的各类订单每笔成交量情况								
变量	日度配对差异				日内配对差异			
	[-2, -1]		[-1, 0]		[-10, -5]		[-5, 0]	
	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>
<i>SBZ</i>	0.285***	0.001	0.273**	0.013	0.742***	0.004	0.984***	0
<i>MBZ</i>	0.594**	0.010	0.633***	0	2.451***	0	3.902***	0
<i>LBZ</i>	0.274	0.292	0.577	0.184	5.236***	0	7.905***	0
<i>SSZ</i>	0.098*	0.086	0.018	0.727	0.191	0.178	0.148	0.734
<i>MSZ</i>	0.662***	0.002	0.814***	0	0.921**	0.028	2.080***	0
<i>LSZ</i>	0.139	0.343	0.119	0.625	1.068	0.533	1.968**	0.050

  

面板 B: 负跳跃前的各类订单每笔成交量情况								
变量	日度配对差异				日内配对差异			
	[-2, -1]		[-1, 0]		[-10, -5]		[-5, 0]	
	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>	中位数	<i>p</i>
<i>SBZ</i>	0.425	0.698	0.283	0.274	0.612**	0.037	0.524*	0.088
<i>MBZ</i>	0.286**	0.045	0.329	0.267	1.297	0.248	-0.215	0.716
<i>LBZ</i>	0.418	0.959	-0.093	0.678	0.989	0.844	0.907	0.740
<i>SSZ</i>	0.355	0.455	-0.119	0.485	3.157***	0	2***	0.002
<i>MSZ</i>	0.788**	0.040	0.613	0.343	-0.269	0.264	0.476	0.217
<i>LSZ</i>	0.721	0.882	-0.790	0.142	4.195	0.103	-0.151	0.293

注: 报告了跳跃前 2 d 和跳跃前 10 min 各类指令在此时间段的每笔成交量的事件分析结果, 其中 *LBZ*、*MBZ*、*SBZ* 分别表示主动买的大单、中单和小单的每笔成交量, *LSZ*、*MSZ*、*SSZ* 分别表示主动卖的大单、中单和小单的每笔成交量. 所有的每笔成交量指标都已经按照式(4)进行了标准化; 检验的观测数与表 3 一致. ‘\*’, ‘\*\*’ 和 ‘\*\*\*’ 分别表示在 90%、95% 和 99% 置信水平下显著.

从每笔成交量的日内配对差异来看,正跳跃前 10 min 内,买方下单规模显著增加,其中以大单规模增加最多,其次中单,再次小单;卖方下单规模也是显著增加,其中以中单增加最多,大单要在临近跳跃时才增加,小单不显著;且上述的趋势是逐渐增大的。正跳跃的结果说明,接近跳跃之前发生信息泄露,且逐步扩散;买方倾向于加大每单的规模;不知情的卖方可能被买方挂出的大单吸引而大单卖出。而负跳跃前 10 min 内,只有小单规模显著增加,而且卖方小单增加更多。可能的原因是因负跳跃前价格的持续下跌引发散户的关注度下降<sup>⑫</sup>,并没有足够大的流动性支撑信息交易者采用大额中单隐秘交易或用大单竞争交易,信息交易者只能采用较大规模的小卖单偷偷出货,从而导致了小卖单的每笔成交量显著高于配对股票。

总的来说,事件分析证明了跳跃前存在信息泄露,且信息泄露随时间逐步扩散。从显著异常的日内指标来看,除了正跳跃前的大买单和负跳跃前的小单的规模显著大于中单,其他情形下都是显著异常的中单数量或规模最大,即信息泄露更多时候是通过中单实现的,从而证实中单的信息含量更高,即事件分析结果对 H1 提供了初步支持。

## 4 交易指令与跳跃的关联性

### 4.1 回归方程设定

事件分析结果显示跳跃前有异常的交易行为,反映有信息提前泄露。需要进一步证明这些异常交易行为中所携带的信息就是导致跳跃的信息,而且这些异常交易行为对跳跃有预测能力。本节将测试各类指令与跳跃的关联性,进一步比较大中小单的信息含量。事后检测出的跳跃包含的信息有 3 个维度:跳跃的发生时间、跳跃的方向和

跳跃的大小。如果跳跃前有信息泄露且信息优势者采用了某类指令完成交易,则该指令应该与跳跃的发生概率、方向和大小有显著的关联性。下文将从 3 个维度建立不同的回归方程来检验各类指令与跳跃的关联性并测试 H1。

如果要分析 6 类指令(大中小单、买卖方向)与跳跃的关联性,需要引入的变量数量过多,且会有很严重的多重共线性。为了减少变量个数并避免可能的多重共线性,根据 Chan 和 Fong<sup>[4]</sup>的做法只专注于某类指令的买卖不平衡性,并构建回归指标如下

$$\begin{aligned} R_t^{(X)} &= \ln(XBtrades_t/XStrades_t) \\ Z_t^{(X)} &= \ln(XBZ_t/XSZ_t) \end{aligned} \quad (5)$$

式中  $R_t^{(X)}$  表示  $t$  时间某类指令的主动买笔数和主动卖笔数的百分比差异<sup>⑬</sup>,而  $Z_t^{(X)}$  表示  $t$  时间某类指令中主动买和主动卖平均成交量的百分比差异。其中  $X$  可以是  $L$ 、 $M$ 、 $S$  (分别表示大、中、小单)。此外,由于要同时分析跳跃前 1 d 与跳跃前 5 min 的交易,用  $DR_t^L, DR_t^M, DR_t^S$  和  $DZ_t^L, DZ_t^M, DZ_t^S$  表示相应的跳跃前的日度交易。

先按照 Jiang 等<sup>[1]</sup>的方法来检测各类指令与跳跃发生概率的关联性。模型的构建如下

$$\begin{aligned} \ln\left[\frac{p(\text{jump} | \text{match})}{1 - p(\text{jump} | \text{match})}\right] &= c + \\ &\sum_{k=S,M,L} (\beta_1^{(k)} DR^{(k)} + \beta_2^{(k)} DZ^{(k)}) + \\ &\sum_{k=S,M,L} (\beta_3^{(k)} R^{(k)} + \beta_4^{(k)} Z^{(k)}) + \\ &\text{control} + \varepsilon \end{aligned} \quad (6)$$

即本文主要关注的是跳跃前 1 d 和跳跃前 5 min 各类指令是否含有预测跳跃发生的准确信息<sup>⑭</sup>。控制变量中包含跳跃前 1 d 和前 5 min 的收益率、交易量、卖方深度和买方深度(10 档买价或卖价中平均等待成交量)。控制住这些变量,事实上是

<sup>⑫</sup> Wan 等<sup>[3]</sup>发现中国市场个股正跳跃前流动性较充足,但负跳跃前流动性较缺乏。

<sup>⑬</sup> 此处不再使用占总成交的比例首先是为了避免完全共线性,且根据 Chan 和 Fong<sup>[4]</sup>的结果,某类指令中的买卖交易不平衡对收益率和波动率有显著的影响。

<sup>⑭</sup> 加入更多的滞后项可以分析更长的交易时段,但对于分析哪种指令含有更多信息却帮助不大,却可能导致更多的多重共线性。本文尝试了滞后 2 期的模型,需要更多的联合检验以判断某类指令与跳跃的显著关联性。为了简洁,只报告滞后 1 期的结果。

排除流动性指标所含信息的影响后,专注于分析各类指标所含信息与跳跃发生的关联性. 方程(6)的回归样本是事件分析中的日内跳跃和相应的配对样本.

事件分析显示正跳跃和负跳跃前的异常指令有显著的不同,显示有人可能知道了消息的好坏. 为了准确推断哪类指令含有准确预测跳跃方向的信息,设定模型如下

$$\ln\left[\frac{p(\text{positive} | \text{jump})}{1 - p(\text{positive} | \text{jump})}\right] = c + \sum_{k=S,M,L} (\beta_1^{(k)} DR^{(k)} + \beta_2^{(k)} DZ^{(k)}) + \sum_{k=S,M,L} (\beta_3^{(k)} R^{(k)} + \beta_4^{(k)} Z^{(k)}) + \text{control} + \varepsilon \quad (7)$$

判断跳跃的方向是以已知将会有跳跃为条件的,即分析的样本是只有日内跳跃而不包括配对样本.

事件分析也发现跳跃前有投资者会激烈地交易,但这些激烈的交易并不足以证实信息交易者知晓消息发布的确切内容并能够分析其对市场造成的影响. 为了检验是否某类指令含有的信息足够推断未来消息发布造成的市场影响,构建模型如下

$$E(\text{size} | \text{jump}) = c + \sum_{k=S,M,L} (\beta_1^{(k)} DR^{(k)} + \beta_2^{(k)} DZ^{(k)}) + \sum_{k=S,M,L} (\beta_3^{(k)} R^{(k)} + \beta_4^{(k)} Z^{(k)}) + \text{control} + \varepsilon \quad (8)$$

式中  $E(\text{size} | \text{jump})$  表示以跳跃为条件,跳跃的大小,即对应的 5 min 收益率的绝对值. 本文分正负跳跃两类样本来讨论指令的信息含量是否与跳跃规模显著相关. 方程(7)和方程(8)都同样控制了跳跃前 1 d 和前 5 min 的收益率、交易量、卖方深度和买方深度.

#### 4.2 回归分析结果

表 5 报告了方程(6)~方程(8)的估计结果,其中前 2 个回归是针对正跳跃发生概率和负跳跃发生概率的回归,使用的样本是事件分析中的日内跳跃样本及其对应的配对样本. 故此回归结果既可以推断与跳跃发生概率显著相关的指令类型,也可以对事件分析结果测试稳健性. 根据表

5,跳跃前 1 d 的小单笔数的买卖差异、小单和中单规模的买卖差异与正跳跃发生概率显著相关,而其余的日度变量都不显著. 该发现为事件分析中的推测提供了支持: 信息交易者采取了隐秘交易策略,使得中小单的信息含量增大. 注意,因为信息泄露越早就会越早将价格往不利的方向冲击,信息交易者的总利润就越低,所以跳跃前 1 d 的小单显著可能只是更隐蔽的拆单. 跳跃前 5 min 的中单和大单的买卖成交笔数差异以及大单的买卖规模差异都显著为正. 中单与正跳跃发生概率显著相关与 Barclay 和 Warner<sup>[6]</sup> 以及 Chakravarty<sup>[7]</sup> 发现一致,而显著的大单则与朱燕建等<sup>[5]</sup> 及 Cai 等<sup>[16]</sup> 的发现一致. 正如 Chakravarty<sup>[7]</sup> 所强调的,价格下跌时段由于存在卖空限制,以及下跌股票的流动性限制,信息交易的难度较高. 负跳跃前 1 d 的所有变量都不显著,而跳跃前 5 min 内所有的成交笔数差异变量都显著为负,即卖单越多负跳跃概率越大,而中单的买卖笔数差异的  $p$  值最小,也说明准确的信息聚集在中等规模的指令.

表 5 的第 3 个回归针对的是方程(7),使用的是事件分析中所有的日内跳跃样本. 回归结果显示,跳跃前 1 d 并没有哪类指令含有准确推断跳跃方向的信息. 跳跃前 5 min 的变量中,中单和大单的买卖成交笔数差异、中单和大单规模的买卖差异都与正跳跃发生概率显著正相关,这说明信息交易者会大量使用中单掩盖交易目的,但在正跳跃时也会趁流动性较好而使用较大规模的大单竞争市场份额.

表 5 的最后 2 个回归对方程(8),使用的样本分别是日内正跳跃样本和日内负跳跃样本. 结果显示,跳跃前 1 d 的中单买卖笔数差异与正跳跃大小显著正相关,而小单的买卖笔数差异与正跳跃大小显著负相关. 小买单越少正跳跃越大,说明散户有可能被更高价的买单吸引而为信息交易者提供流动性. 跳跃前 5 min 的小买单比例增加、中等买单减少和中等买单规模增加都有助于正跳跃幅度增加. 正跳跃前若吸引了足量的散户提供流动性,且信息交易者可以拆分出更大规模的单以隐蔽其交易目的时,正跳跃更容易被推高. 然而,日内负跳跃的大小因负面信息对散户的吸

引力较小,流动性供给不足之下<sup>[3]</sup>,其大小难以反应在某类指令上.跳跃前 1 d 的小卖单和大卖单的规模增加都显示负跳跃幅度会显著增加,这体现信息交易者既可能因流动性不足而用小单缓慢出逃,也可能被迫用大单仓促的成交.而跳跃前 5 min 的交易与跳跃大小无显著相关性,说明负跳跃发生时,流动性不足会导致指令含有的信息量不足,即泄露的信息不能充分的反映到市场中.

总的来说,表 5 的结果与事件分析是一致的:中单与跳跃发生概率、方向和大小有更高的关联性,说明主要结论支持 H1 成立,即中国市场存在隐秘交易.表 5 的结果还对 H2 提供初步的支持:事件分析中的异常指令,包括中单数量和规模,正跳跃时的大买单和负跳跃时的小卖单都与跳跃有不同程度的关联性,说明确实信息交易者知晓跳跃有关的信息并以此交易,其交易指令可能对跳跃有预测作用,后面会作进一步测试.

表 5 跳跃前指令对跳跃概率、方向和大小的影响分析

Table 5 Impact of pre-jump orders on the probability, direction and size of jumps

变量	正跳跃概率		负跳跃概率		跳跃方向		正跳跃大小		负跳跃大小	
	系数	p 值	系数	p 值	系数	p 值	系数	p 值	系数	p 值
DSR	0.242**	0.041	-0.004	0.988	0.277	0.250	-0.208*	0.089	-0.035	0.873
DMR	-0.028	0.649	-0.026	0.854	0.104	0.398	0.213***	0.001	0.028	0.798
DLR	-0.076	0.557	-0.035	0.902	0.303	0.238	-0.023	0.864	-0.041	0.852
DSZ	0.637*	0.053	0.637*	0.053	0.039	0.953	-0.452	0.186	-1.578***	0.006
DMZ	0.898*	0.051	0.898*	0.051	-0.528	0.587	-0.644	0.192	-0.074	0.934
DLZ	0.283	0.116	0.283	0.116	0.244	0.486	-0.152	0.409	-0.696**	0.019
SR	-0.019	0.738	-0.251*	0.073	0.034	0.777	0.148**	0.020	0.076	0.520
MR	0.178***	0.000	-0.127**	0.024	0.470***	0.000	-0.071***	0.005	-0.029	0.523
LR	0.271***	0.000	-0.178*	0.072	0.548***	0.000	0.014	0.764	0.092	0.248
SZ	-0.038	0.724	-0.038	0.724	0.116	0.635	-0.008	0.941	0.438	0.110
MZ	-0.087	0.305	-0.087	0.305	0.360**	0.031	0.170*	0.067	-0.101	0.488
LZ	0.121**	0.024	0.121**	0.024	0.222**	0.034	-0.024	0.671	-0.113	0.248
N	11 538		2 286		6 912		5 769		1 143	
R <sup>2</sup>	0.426		0.385		0.552		0.136		0.359	

注: DSR, DMR, DLR 表示跳跃前 1 d 的该类别(从小到大)中主动买与主动卖笔数的百分比差异,SR,MR, LR 表示跳跃前 5 min 的成交笔数百分比差异,DSZ,DMZ,DLZ 以及 SZ,MZ, LZ 分别表示跳跃前 1 d 和 5 min 的该类指标的买卖平均成交量的百分比差异;表中的前 2 个回归是方程(6)的回归结果,分别对正跳跃及其配对样本和负跳跃及其配对样本进行回归分析.第 3 个回归是针对方程(7)的,所用样本是跳跃样本.后面 2 个回归是方程(8)的结果,所用样本分别是正跳跃样本和负跳跃样本.所有回归方程中都控制了跳跃前 1 d 和跳跃前 5 min 的收益率、交易量、卖方深度和卖方深度,为了节省空间,未报告这些控制变量的估计结果.‘\*’, ‘\*\*’和 ‘\*\*\*’ 分别表示在 90%、95% 和 99% 置信水平下显著.

## 5 交易指令在有消息发布的跳跃前表现

虽然上述事件分析中的跳跃样本是按照 Lee 和 Mykland<sup>[8]</sup>的方法检测出日内跳跃并经过 Barndorff-Nielsen 和 Shephard<sup>[25]</sup>的方法筛选,但并不肯定这些跳跃足够突然,也无法确定跳跃幅度能够从多大程度上准确代表信息的重要程度.本节将 Lee 和 Mykland<sup>[8]</sup>方法检测出的日内跳跃与

企业业绩快报、企业业绩预测报告和分红送转报告的公告日匹配,然后分析经信息匹配过的跳跃前的大中小单与跳跃和相关信息的关联性,可以进一步稳健测试 H1.如果代表隐秘交易的中单在有信息匹配的跳跃前仍旧表现出异常,且异常的中单与跳跃匹配的信息关联显著,就说明 H1 成立,即隐秘交易确实存在,且隐秘交易者知晓引发跳跃的意外信息,是知情交易者.而选取这 3 类信息发布的原因在于,它们的发布有一定的突然性且发布日期数据可得,并有研究证实中国市场中

这些信息能够引发价格突变<sup>[27-29]</sup>。根据 Wind 金融数据库,2013 年至 2015 年中证 800 成分股总共发布了 966 份业绩快报(包括季报、中报和年报),共发布 3 676 份业绩预测报告,分红公告 1 527 份,送股公告 455 份。因为公告有可能发生在周末和节假日,公告信息也可能泄露而提前引发跳跃<sup>[30]</sup>,所以在信息发布日前后 3 d 之内的跳跃都认为与该信息匹配。按照该匹配算法,最后匹配了 1 073 个跳跃。若有跳跃与超过 1 种信息匹配,则选择与其日期最接近的信息匹配。按照日内分析的思路,要求前 10 d 没有跳跃且跳跃发生在开盘 10 min 以后才能作为分析样本,最终选出 628 个信息相关的跳跃,其中正跳跃 457 个,负跳跃 171 个。

表 6 是将信息匹配的跳跃按照 3.1 节介绍的

方法进行事件分析,考察与消息发布相关联的跳跃前大中小单交易的情况。由于跳跃前 1 d~2 d 的交易都与配对股票没有显著差异,故表 6 中并没有报告日度对比结果。表 6 的面板 A 显示的是跳跃前 10 min 和 5 min 的各类订单占比情况。有信息匹配的正跳跃前,买单显著增加而卖单显著减少,且增减幅度最大的是中单,即中单比大单和小单含有更多的信息量,说明 H1 成立。有信息匹配的负跳跃前,订单比例并无明显差异,可能是用坏消息牟利的难度更大。面板 B 报告跳跃前 10 min 和 5 min 的各类订单平均成交量的情况,其中只有正跳跃前中买单和负跳跃前中卖单的每笔成交量显著大于配对股票。这进一步证实跳跃前信息交易者知晓信息的方向且采用了中单隐蔽的成交。

表 6 信息匹配的跳跃前日内分析

Table 6 Intraday analyses in the day before information-matched jumps

面板 A: 有信息匹配的跳跃前日内订单成交笔数情况								
变量	正跳跃配对差异				负跳跃配对差异			
	[-10, -5]		[-5, 0]		[-10, -5]		[-5, 0]	
	中位数	p	中位数	p	中位数	p	中位数	p
SBR	0.672	0.658	1.678***	0.003	-1.606	0.300	-0.714	0.515
MBR	1.632	0.216	9.996***	0.000	-0.365	0.960	0.508	0.941
LBR	0.174	0.148	1.271***	0.000	-0.020	0.550	0.000	0.752
SSR	0.264	0.855	-3.467***	0.000	0.921	0.231	0.197	0.453
MSR	-1.425	0.153	-7.333***	0.000	0.368	0.518	0.378	0.510
LSR	0.053	0.658	-0.305***	0.007	-0.226	0.906	0.549	0.273
面板 B: 有信息匹配的跳跃前日内每笔成交量情况								
变量	正跳跃配对差异				负跳跃配对差异			
	[-10, -5]		[-, 0]		[-10, -5]		[-5, 0]	
	中位数	p	中位数	p	中位数	p	中位数	p
SBZ	5.647	0.874	-2.454	0.241	21.275	0.420	-6.354	0.341
MBZ	-0.607	0.631	23.290**	0.025	-20.161	0.938	-21.940	0.204
LBZ	-9.539	0.596	-21.834	0.732	-51.972	0.158	6.073	0.709
SSZ	-13.533	0.215	-10.555	0.755	-14.953	0.094	4.626	0.599
MSZ	-4.839	0.438	-1.601	0.912	5.990	0.456	39.815*	0.051
LSZ	21.617	0.975	-6.600	0.425	-60.626	0.875	13.690	0.209

注:表中报告了用业绩报告、业绩预告报告和分红送转报告发布日匹配的跳跃前 10 min 各类指令在此时间段的占比和每笔成交量的事件分析结果;LBR、MBR、SBR 分别表示大单、中单、小单中的买单占比,LSZ、MSZ、SSZ 分别表示主动卖的大单、中单和小单的每笔成交量,\*、\*\*和\*\*\*分别表示在 90%、95%和 99%置信水平下显著。

表 7 中将有关信息匹配的跳跃按表 5 的方法进行回归分析,测试这些跳跃的概率、大小和方向与哪些类型的指令最相关。表 7 的结果有大量支持 H1 成立

的证据:跳跃前 5 min,中买单相对越多则正跳跃概率和正跳跃幅度都越大;跳跃前 5 min,中卖单相比中买单越多则负跳跃概率越大;跳跃前 1 d 的中卖单

越多则负跳跃越大. 总之, 表 7 与表 5 的结果都证实, 不论是通过其他跳跃检测法筛选, 还是用信息匹配筛选, 选出的跳跃都能与以中单为主的异常指令显著相关, 从而都支持 H1 成立.

表 7 跳跃前订单对信息匹配跳跃的影响分析

Table 7 Impact of pre-jump orders on information-matched jumps

变量	正跳跃概率		负跳跃概率		跳跃方向		正跳跃大小		负跳跃大小	
	系数	<i>p</i>	系数	<i>p</i>	系数	<i>p</i>	系数	<i>p</i>	系数	<i>p</i>
<i>DSR</i>	0.163	0.937	-3.827	0.420	-0.860	0.816	-0.362	0.804	-0.223	0.369
<i>DMR</i>	-1.323	0.423	3.794	0.231	2.240	0.369	2.049*	0.075	1.116	0.168
<i>DLR</i>	25.496	0.110	-1.725	0.954	-23.850	0.343	12.637	0.223	-41.412	0.127
<i>DSZ</i>	0.034	0.951	0.174	0.871	-0.189	0.862	-0.393	0.319	-0.153	0.865
<i>DMZ</i>	-0.056	0.509	-0.113	0.580	0.167	0.284	-0.048	0.456	-0.154*	0.059
<i>DLZ</i>	0.001	0.836	-0.005	0.370	$9.56 \times 10^{-3}$	0.077	-0.003	0.219	0.002	0.658
<i>SR</i>	-0.235	0.829	0.444	0.847	2.541	0.182	1.501*	0.070	0.371	0.812
<i>MR</i>	1.881***	0.003	-3.426***	0.009	4.543***	0.000	0.760*	0.099	-0.194	0.831
<i>LR</i>	7.154*	0.087	1.380	0.847	12.540*	0.068	2.061	0.482	5.372	0.365
<i>SZ</i>	0.036	0.886	0.168	0.768	-0.227	0.685	0.250	0.267	0.524	0.291
<i>MZ</i>	0.004	0.860	-0.038	0.437	0.048	0.200	-0.024	0.131	-0.005	0.878
<i>LZ</i>	-0.000 112	0.896	0.000 107	0.948	-0.000 390	0.816	-0.000 297	0.688	-0.003 ***	0.008
<i>N</i>	914		342		628		457		171	
<i>R</i> <sup>2</sup>	0.476		0.411		0.571		0.342		0.359	

注: 表中报告了用业绩报告、业绩预测报告和分红送转报告发布日匹配的跳跃前 1 d 和前 5 min 的大中小单与跳跃发生概率、跳跃大小和跳跃方向的关系, 即分别用这些样本估计了式(6)~式(8)的回归. 表 7 的配对采用的是随机配对方法; *DSR*, *DMR*, *DLR* 表示跳跃前 1 d 的该类别(从小到大)中主动买与主动卖笔数的百分比差异, *SR*, *MR*, *LR* 表示跳跃前 5 min 的成交笔数百分比差异, *DSZ*, *DMZ*, *DLZ* 以及 *SZ*, *MZ*, *LZ* 分别表示跳跃前 1 d 和 5 min 的该类指标的买卖平均成交量的百分比差异. “\*”、“\*\*”和“\*\*\*”分别表示在 90%、95% 和 99% 置信水平下显著.

表 8 考察大中小单与跳跃相关信息的关联性. 对于业绩预测报告和业绩快报, 本文选择净利润增长率作为衡量信息重要程度的指标, 而对分红, 选择的指标则是每股派息额(税后); 对于送转股, 由于送股比例和转增比例数量级差异大, 无法相互比较, 且单独的送股或转增样本太少, 不足以稳健估计回归模型, 所以送转股的样本没有用于与指令的关联性分析. 由于净利润增长率和派息额的影响程度并不一定是按指标值线性增长的, 在建模时按照 30% 和 70% 分位点将指标变换成序数变量, 然后构建有序 Probit 模型来分析各类订单与指标的关联性. 比如针对业绩快报和预报, 净利润增长率(或预测净利润增长率)共有 374 个观测, 大于 70% 分位点(约为利润增长 70%) 为 1, 小于 30% 分位点(约为利润增长 1.8%) 则为 -1, 其余为 0, 由此

定义了有序 Probit 模型的因变量, 其自变量选取类似于方程(8). 从对利润增长率的有序 Probit 回归结果看, 跳跃前 1 d 和跳跃前 5 min 的中买单相比中卖单越多, 净利润增长越多, 而使用大单和小单的交易者都有不同程度的方向判断错误, 从而证实掌握最精确信息的交易者是中单交易者. 每股派息的回归结果与此类似, 跳跃前 1 d 的中买单越多, 跳跃前 5 min 的中买单越大, 则派息额就越大, 也提供了对 H1 的支持. 除此之外, 跳跃前 5 min 的大买单比大卖单越多, 则信息越乐观, 也说明信息交易者会根据信息重要程度选择是否采用更激进的交易策略. 此外, 表 8 的结果还对 H2 提供了支持: 与信息发布显著相关的大单和中单表明这些异常指令的使用者知晓引发跳跃的信息, 从而这些指令对跳跃有一定的预测能力.

表 8 跳跃前订单对有跳匹配的信息的影响分析

Table 8 Impact of pre-jump orders on the jump-matched information

变量	业绩增幅		每股派息	
	系数	<i>p</i>	系数	<i>p</i>
<i>DSR</i>	-2.402***	0.000	-1.750*	0.084
<i>DMR</i>	1.512**	0.038	2.573***	0.000
<i>DLR</i>	-6.089***	0.000	27.351***	0.000
<i>DSZ</i>	-0.538*	0.094	0.601	0.242
<i>DMZ</i>	0.069	0.174	-0.025	0.428
<i>DLZ</i>	-0.001	0.483	-0.000469	0.447
<i>SR</i>	-0.248	0.390	0.407	0.398
<i>MR</i>	0.577*	0.099	0.256	0.340
<i>LR</i>	6.312***	0.000	4.628***	0.000
<i>SZ</i>	0.380*	0.079	0.414	0.170
<i>MZ</i>	0.000159	0.237	0.000822**	0.025
<i>LZ</i>	0.000677	0.208	-0.002	0.115
<i>N</i>	374		243	
伪 <i>R</i> <sup>2</sup>	0.340		0.355	

注: 报告略去了收益、交易量和深度等控制变量的系数。\*, \*\*, 和 \*\*\* 分别表示在 90%、95% 和 99% 置信水平下显著。

总结表 6 ~ 表 8 的结果, 换用消息发布公告匹配的跳跃, 做与前文类似的事件分析和回归分析, 包括用大中小单对信息本身回归, 其结果都支持中国市场中的知情交易者采用了隐秘交易策略, 因此 H1 成立是较为稳健的。

## 6 各类指令对跳跃的预测能力测试

本节考察各类指令对跳跃的预测效果, 并测试 H2。从表 5、表 7 和表 8 的结果已经可以看出, 跳跃之前有交易指令与跳跃的发生概率、方向、大小以及跳跃相匹配的信息显著相关, 显示有信息泄露, 且信息交易者知晓引发跳跃的信息并据此进行交易。因此, 猜测这些与跳跃显著相关的指令应该对跳跃有预测能力。Jiang 等<sup>[1]</sup>只考察了跳跃前流动性变化对跳跃发生概率的预测作用, 本节将在控制流动性指标的基础上, 进一步考察跳跃前大中小单对跳跃发生概率、方向和大小的预测效果。本文将只含有流动性指标的模型作为基准模型, 设为模型 1

$$Y = c + DRE + DV + DBLQ + DSLQ + RE + V + BLQ + SLQ + \varepsilon \quad (9)$$

其中 *Y* 是因变量, 按照跳跃概率、方向和大小依次与式(6) ~ 式(8) 中的因变量相同, 而 *DRE*、*DV*、*DBLQ*、*DSLQ*、*RE*、*V*、*BLQ*、*SLQ* 都是式(6) ~ 式(8) 中的控制变量, 分别表示跳跃前 1 d 和跳跃前 5 min 的收益率、波动率、买方平均深度和卖方平均深度。因此, 模型 1 是只用跳跃前收益、波动和流动性来预测跳跃的概率、大小和方向。模型 2 是在式(9) 中加入按式(5) 定义的大单数量和规模的买卖不平衡; 模型 3 是在式(9) 中加入中单数量和规模的买卖不平衡; 模型 4 是在式(9) 中加入小单数量和规模的买卖不平衡; 而模型 5 就是式(6) ~ 式(8), 即包含所有指令。用模型 1 ~ 模型 5 来测试交易指令是否能够在流动性指标的基础上提升预测跳跃的能力, 从而测试 H2; 此外, 模型 2 ~ 模型 4 的结果比较, 还可以进一步稳健检测各类指令的信息含量差异, 从而也能够对 H1 进行测试。

表 9 中报告了模型 1 ~ 模型 5 用于跳跃样本(面板 A) 和用于有信息匹配的跳跃样本(面板 B) 的预测效果。预测的方法是, 对原样本用不放回随机抽样, 抽取一半的样本作为训练样本估计模型, 再用剩余的样本进行预测。为了比较预测效

果,对于分类模型(跳跃概率和方向预测),计算 *KS*、*AUROC* 和 *AR* 值<sup>⑮</sup>,对于线性模型(跳跃大小预测),计算平均绝对误差 *MAE* 即预测值与真实值差异的绝对值求平均)。随机抽样并预测的步骤进行 1 000 次,相应得到的 *KS*、*AUROC*、*AR* 和 *MAE* 值是 1 000 次预测效果的平均值。

从表 9 的结果看,无论是用跳跃样本,还是用有信息匹配的样本,4 个预测效果指标都一致显示,模型 5 的预测效果比模型 1 好。这说明,交易指令能够在流动性指标的基础上提升对跳跃的预测效果,结合表 5 和表 7 中存在与跳跃概率、方向和大小显著相关的指令,证明了跳跃前的异常指

令确实对跳跃有预测能力,从而证明 H2 成立。表 9 中,还用粗体数字标出了相对于模型 1,对预测效果提升最明显的指令类型(模型 2、模型 3 和模型 4 分别对应大、中、小单)。面板 A 和面板 B 的结果都显示,在控制流动性指标的基础上,中单对跳跃的预测能力最佳,其次是大单,再次是小单,即信息交易者更偏好采用中单完成交易,使得中单含有更多与跳跃有关的信息,该结果再次说明 H1 是稳健成立。此外,大单主要在预测正跳跃概率和跳跃方向时预测能力最佳,而小单主要在负跳跃概率预测时效果最佳,也与表 3、表 4 和表 5 得到的结果一致。

表 9 各类指令对跳跃的预测能力比较

Table 9 Comparisons on the forecasting ability of all kinds of orders on jumps

面板 A: 使用跳跃样本的预测效果比较						
变量	预测项目	模型				
		1	2	3	4	5
<i>KS</i>	正跳概率预测	0.207	0.225	<b>0.252</b>	0.219	0.267
	负跳概率预测	0.121	0.126	0.128	<b>0.149</b>	0.143
	跳跃方向预测	0.439	0.544	<b>0.584</b>	0.524	0.602
<i>AUROC</i>	正跳概率预测	0.510	<b>0.573</b>	0.542	0.522	0.584
	负跳概率预测	0.514	0.531	<b>0.539</b>	0.535	0.543
	跳跃方向预测	0.869	<b>0.889</b>	<b>0.889</b>	0.876	0.899
<i>AR</i>	正跳概率预测	0.129	0.144	<b>0.159</b>	0.141	0.159
	负跳概率预测	0.067	0.073	<b>0.089</b>	0.082	0.084
	跳跃方向预测	0.101	0.116	<b>0.124</b>	0.113	0.127
<i>MAE</i>	正跳大小预测	0.943	0.940	<b>0.894</b>	0.900	0.897
	负跳大小预测	5.378	0.957	0.916	<b>0.912</b>	0.961
面板 B: 使用有信息匹配的跳跃样本的预测效果						
变量	预测项目	模型				
		1	2	3	4	5
<i>KS</i>	正跳跃概率预测	0.280	0.285	<b>0.319</b>	0.281	0.320
	负跳跃概率预测	0.139	<b>0.221</b>	0.173	0.178	0.223
	跳跃方向预测	0.514	0.579	<b>0.583</b>	0.521	0.623
<i>AUROC</i>	正跳跃概率预测	0.587	<b>0.645</b>	0.598	0.589	0.647
	负跳跃概率预测	0.477	0.514	<b>0.521</b>	0.499	0.519
	跳跃方向预测	0.810	0.819	<b>0.829</b>	0.806	0.841
<i>AR</i>	正跳跃概率预测	0.117	0.194	<b>0.223</b>	0.193	0.208
	负跳跃概率预测	0.045	0.075	<b>0.111</b>	0.103	0.145
	跳跃方向预测	0.186	<b>0.219</b>	0.194	0.182	0.226
<i>MAE</i>	正跳跃大小预测	1.003	0.998	<b>0.877</b>	0.903	0.887
	负跳跃大小预测	0.924	0.889	0.813	<b>0.791</b>	<b>0.806</b>

注: *KS*、*AUROC* 和 *AR* 值是跳跃发生概率和方向的预测报告, *MAE* 是跳跃大小平均绝对误差的预测报告; 所有的预测都是样本外预测: 即采用不放回的随机抽样方法, 从原样本中抽取一半的观测作为训练样本估计模型, 而用剩余的样本拟合模型, 比较真实值和预测值的差异。因此, 面板 A 中样本量是表 5 中对应的一半, 而面板 B 中的样本量是表 7 中对应的一半; 表中粗体数字为 3 类指令中预测效果最好的指令。

⑮ 根据 Giudici<sup>[31]</sup>, *KS* 值是 *KS* 曲线的最大差异值, *AUROC* 是 *ROC* 曲线中预测模型与随机模型组成的面积占比, *AR* 是 *CAP* 曲线中预测模型与随机模型组成的面积占比。这 3 个值都是正向指标, 即值越大, 预测效果越好。

## 7 结 束 语

本文利用中证 800 指数成分股 3 年的分笔交易数据,将交易指令拆分为大单、中单和小单,并分买卖方向讨论了各类指令与跳跃发生的关联性以及其对跳跃的预测能力.事件分析表明,跳跃前有明显的信息泄露,且信息泄露随时间扩散;跳跃前有正跳跃前买入、负跳跃前卖出的异常指令,其异常程度随跳跃临近而增加.进一步的回归和预测分析发现,泄露的信息确实是引发跳跃的信息,且信息交易者在据此交易牟利;投资者的交易指令与跳跃的发生概率、方向、大小以及匹配的信息有不同程度的显著关联性,且在控制流动性指标的模型中,加入投资者的交易指令可以提升模型的预测效果.通过事件分析、回归模型和预测效果比较都一致地得出结论:跳跃前中单的信息含量一般高于大单和小单,说明中国股市中存在隐秘交易.此外,结合以往研究的结论<sup>[3]</sup>,正跳跃前流动性供给充足,而负跳跃前流动性供给缺乏;正跳跃前大单交易显著增加,信息含量较大,而负跳跃前小单交易增加显著,信息含量较大,说明信息优势者的交易策略与流动性有关;以中单隐秘交易为主,在流动性充足的情况下会采用大单竞争份额,在流动性不足的情况下会用小单隐藏交易目的并缓慢成交.

本文证实了中国市场在极端价格变化前存在隐秘交易.美国市场中无论是一般情形下还是突发事件前都存在隐秘交易<sup>[6]</sup>,而以往研究认为中国市场的内幕交易监管不严,故内幕交易者更偏好使用大单交易<sup>[5,16]</sup>.但本文研究表明,在突发事件前中国股市的内幕交易者仍然会选择拆单以隐藏交易目的.隐秘交易的存在意味着信息含量并不是随订单金额单调递增的,突发事件前隐秘交

易的增加会使得中单也具有较大的信息含量.

本文进一步深入分析了跳跃的可预测性.通过分笔交易数据,深入挖掘跳跃前的信息泄露过程,发现跳跃前有代表信息泄露的异常指令且这些指令与跳跃的发生概率、方向和大小以及匹配的信息都有显著关联性,且能够在控制流动性指标的基础上提升模型对跳跃的预测效果.之前的研究主要是从信息的出乎意料程度、流动性变化等方面来研究跳跃的可预测性,而本文是从微观的交易层面发现投资者的交易指令能够作为跳跃预测指标,深化了前人对跳跃可预测性的研究.

进一步引申本文的结果,可以推测机构和大户比散户更具信息优势,且隐秘交易主要是机构投资者所为.在以散户交易量为主的中国市场中,最具信息量的交易更可能是机构和大户做出的.较大规模的中单或小单以及流动性充足时的大单与跳跃的发生概率、方向和大小以及跳跃匹配的信息都有不同程度的关联性,而小单的数量大多数情况都是不显著的,这说明以数量占优的散户并不能通过更多的小单交易来从跳跃过程中牟利,反而是数量较少但资金占优的机构和大户具备着信息优势,他们拆分出的较大规模的中小单与跳跃以及引发跳跃的意外信息显著相关.

本文的发现给监管者提供了参考.对于以跳跃为代表的极端风险,监管部门可以从交易指令(比如中单和大单的数量、中小单的规模)出发来构建相应的跳跃预警因子进行监控.对于内幕交易的监管,除了在一般情形下监测代表大单交易的主力资金流量、大单净量(DDE)等指标外,监管者还可以通过计算异常的中小单数量和规模来检测隐蔽内幕交易的存在性,如此可以提高对内幕交易的打击精准度,进而构建更公平的市场环境,促进资源的有效配置.

### 参 考 文 献:

- [1] Jiang G J, Lo I, Verdellhan A. Information shocks, liquidity shocks, jumps, and price discovery: Evidence from the US treasury market [J]. *Journal of Financial and Quantitative Analysis*, 2011, 46(2): 527 - 551.
- [2] Boudt K, Petitjean M. Intraday liquidity dynamics and news releases around price jumps: Evidence from the DJIA stocks

- [J]. *Journal of Financial Markets*, 2014, 17(1): 121–149.
- [3] Wan D, Wei X, Yang X. Liquidity dynamics around intraday price jumps in Chinese stock market [J]. *Journal of Systems Science and Complexity*, 2017, 30(2): 434–463.
- [4] Chan K, Fong W M. Trade size, order imbalance, and the volatility-volume relation [J]. *Journal of Financial Economics*, 2000, 57(2): 247–273.
- [5] 朱燕建, 蒋岳祥, 李 奥. 信息拥有者的交易单选择策略与监管的执法强度 [J]. *金融研究*, 2012, (6): 154–166.  
Zhu Yanjian, Jiang Yuexiang, Li Ao. The order selection strategies of informed traders and the legal enforcement of supervision [J]. *Journal of Financial Research*, 2012, (6): 154–166. (in Chinese)
- [6] Barclay M J, Warner J B. Stealth trading and volatility: Which trades move prices? [J]. *Journal of Financial Economics*, 1993, 34(3): 281–305.
- [7] Chakravarty S. Stealth-trading: Which traders' trades move stock prices? [J]. *Journal of Financial Economics*, 2001, 61(2): 289–307.
- [8] Lee S S, Mykland P A. Jumps in financial markets: A new nonparametric test and jump dynamics [J]. *Review of Financial Studies*, 2008, 21(6): 2535–2563.
- [9] Evans K P. Intraday jumps and US macroeconomic news announcements [J]. *Journal of Banking & Finance*, 2011, 35(10): 2511–2527.
- [10] Lahaye J, Laurent S, Neely C J. Jumps, cojumps and macro announcements [J]. *Journal of Applied Econometrics*, 2011, 26(6): 893–921.
- [11] Kyle A S. Continuous auctions and insider trading [J]. *Econometrica*, 1985, 53(6): 1315–1335.
- [12] Easley D, O'hara M. Price, trade size, and information in securities markets [J]. *Journal of Financial Economics*, 1987, 19(1): 69–90.
- [13] Holden C W, Subrahmanyam A. Long-lived private information and imperfect competition [J]. *The Journal of Finance*, 1992, 47(1): 247–270.
- [14] Foster F D, Viswanathan S. Strategic trading when agents forecast the forecasts of others [J]. *The Journal of Finance*, 1996, 51(4): 1437–1478.
- [15] Alexander G J, Peterson M A. An analysis of trade-size clustering and its relation to stealth trading [J]. *Journal of Financial Economics*, 2007, 84(2): 435–471.
- [16] Cai B M, Cai C X, Keasey K. Which trades move prices in emerging markets? : Evidence from China's stock market [J]. *Pacific-Basin Finance Journal*, 2006, 14(5): 453–466.
- [17] Tong W H S, Zhang S, Zhu Y. Trading on inside information: Evidence from the share-structure reform in China [J]. *Journal of Banking & Finance*, 2013, 37(5): 1422–1436.
- [18] Andersen T G, Bollerslev T, Diebold F X, et al. Micro effects of macro announcements: Real-time price discovery in foreign exchange [J]. *American Economic Review*, 2003, 93(1): 38–62.
- [19] Andersen T G, Bollerslev T, Diebold F X, et al. Real-time price discovery in global stock, bond and foreign exchange markets [J]. *Journal of International Economics*, 2007, 73(2): 251–277.
- [20] Lee S S. Jumps and information flow in financial markets [J]. *Review of Financial Studies*, 2012, 25(2): 439–479.
- [21] Füss R, Grabellus M, Mager F, et al. Something in the air: Information density, news surprises, and price jumps [J]. *Journal of International Financial Markets, Institutions and Money*, 2018, 53(3): 50–75.
- [22] Zhang Z, Chang T S. Trade size and the cross-section of stock returns: Informed versus noise traders in the Chinese growth enterprise market [J]. *International Journal of Business*, 2014, 19(4): 322–335.
- [23] 王宇超, 李心丹, 刘海飞. 算法交易的市场影响研究 [J]. *管理科学学报*, 2014, 17(1): 57–71.  
Wang Yuchao, Li Xindan, Liu Haifei. Market impact of algorithmic trading [J]. *Journal of Management Sciences in China*, 2014, 17(1): 57–71. (in Chinese)

- [24] Andersen T G, Bollerslev T, Diebold F X, et al. Modeling and forecasting realized volatility [J]. *Econometrica*, 2003, 71 (2): 579 – 625.
- [25] Barndorff-Nielsen O E, Shephard N. Power and bipower variation with stochastic volatility and jumps [J]. *Journal of Financial Econometrics*, 2004, 2(1): 1 – 37.
- [26] Jones C M, Kaul G, Lipson M L. Information, trading, and volatility [J]. *Journal of Financial Economics*, 1994, 36(1): 127 – 154.
- [27] 李馨予, 罗 婷. 业绩预测历史准确度的声誉效应 [J]. *金融研究*, 2014, (1): 152 – 166.  
Li Xinyu, Luo Ting. The reputation effect of prior management forecast accuracy [J]. *Journal of Financial Research*, 2014, (1): 152 – 166. (in Chinese)
- [28] 宋逢明, 姜 琪, 高 峰. 现金分红对股票收益率波动和基本面信息相关性的影响 [J]. *金融研究*, 2010, (10): 103 – 116.  
Song Fengming, Jiang Qi, Gao Feng. The impact of cash dividend on the relevance between stock return volatility and fundamental information [J]. *Journal of Financial Research*, 2010, (10): 103 – 116. (in Chinese)
- [29] 李心丹, 俞红海, 陆 蓉, 等. 中国股票市场“高送转”现象研究 [J]. *管理世界*, 2014, (11): 133 – 145.  
Li Xindan, Yu Honghai, Lu Rong, et al. A study on the ‘high stock dividends’ in Chinese stock market [J]. *Management World*, 2014, (11): 133 – 145. (in Chinese)
- [30] Zhou H, Zhu J Q. An empirical examination of jump risk in asset pricing and volatility forecasting in China’s equity and bond markets [J]. *Pacific-Basin Finance Journal*, 2012, 20(5): 857 – 880.
- [31] Giudici P. *Applied Data Mining: Statistical Methods for Business and Industry* [M]. Hoboken: John Wiley & Sons, 2005.

## Behavioral characteristics and informativeness of large, medium and small orders before price jumps

WAN Die<sup>1</sup>, YANG Xiao-guang<sup>2</sup>

1. School of Finance, Zhejiang Gongshang University, Hangzhou 310018, China;

2. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

**Abstract:** Based on tick data of CSI800 index components during three years, this paper classifies the transaction orders into large, medium and small ones, and analyzes the relationship between pre-jump orders and jumps and the predictive power of pre-jump orders on jumps. Event study and regressions based on both the jump sample and information matched with the jump sample show that there are obvious and gradually strengthening abnormal trades before the jumps consistent with the jump directions, and that these abnormal trades have significant correlations with the probability, direction, size and matched information of the jumps. Further forecasting analyses show that abnormal trades can also promote the model’s performance on forecasting jumps after controlling liquidity measures. These results indicate that there are information leakages before jumps in the market, and that the leakages spread over time. Besides, event study, regressions and forecasting power comparisons all show that the medium orders before jumps are more informative than large and small orders, so stealth trading exists before extreme price changes in Chinese stock markets.

**Key words:** price jumps; trading order; informed trading; stealth trading; jump prediction