

# 连续指令驱动市场的信息交易概率： 一种新的方法<sup>①</sup>

李广川<sup>1,2</sup>，刘善存<sup>1</sup>，邱菀华<sup>1</sup>

(1 北京航空航天大学经济管理学院, 北京 100193 2 深圳证券交易所, 深圳 518010)

**摘要:** 通过构造“成交积极性”变量, 提出了一种在连续竞价指令驱动市场中估计信息交易概率的新方法, 运用有序 probit 方法和马尔科夫转换技术对模型进行了估计. 对该方法与基于 PIN 框架的经典信息交易概率模型进行比较分析, 同时研究了信息交易概率和资产收益波动、报价价差、成交持续期之关系的日内效应. 研究结果发现, 模型较之基于 PN 框架的经典信息交易概率模型对价差具有更好的解释力; 信息交易概率和资产收益波动、报价价差、成交持续期等股票交易特征之关系具有显著的日内效应. 该结论经过面板回归模型检验具有稳健性.

**关键词:** 指令驱动连续竞价市场; 信息交易概率; 交易特征; 日内效应

**中图分类号:** F830.9 **文献标识码:** A **文章编号:** 1007-9807(2010)10-0008-13

## 0 引言

信息交易概率是一种衡量证券市场信息不对称程度的直接测度, 指所有成交的交易中知情者参加交易所占的比率<sup>②</sup>. 一直以来, 度量信息不对称程度的大小成为金融市场微观结构领域的重要命题<sup>③</sup>. 早期的研究均以寻找代理变量的方法对信息不对称进行间接测度. 例如, Jaffe 和 Winkel<sup>[2]</sup> 等认为买卖价差可以作为信息不对称程度的简单代理等等. 国内学者的研究包括, 汪毅慧等<sup>[3]</sup> 以报价价差与交易价差之间的差异作为信息不对称成本的代理变量. 王春峰和王燕<sup>[4]</sup> 以交易规模和与信息事件有关的交易时期作为对交易中信息的度量等. 代理变量作为一种间接测度的缺点主要在于“噪声信息”的存在, 即代理变量含有除真实价值信息以外的噪声信号. 例如, 不少学者将价差进行分解, 只用其中的逆向选择成分对

信息不对称进行测度, 如 Huang 和 Stoll<sup>[5]</sup>, 吴文峰等<sup>[6]</sup> 等.

鉴于间接测度的缺陷, 学者们开始尝试寻找信息不对称的直接测度. Easley et al<sup>[7]</sup> 提出的信息交易概率模型 (即 PN 模型) 是其中很有影响力的工作. 该模型利用做市商每日收到的买卖委托指令流个数, 为每只股票每季度计算出一个信息交易概率值. Qin Lei 和 Wu<sup>[8]</sup> 考虑了不同交易者市场到达强度的时变特性将该模型扩展到具有时变特性的信息交易概率测度. 很多后续工作<sup>[9-12]</sup> 借用该模型框架测度了指令驱动市场上的信息交易概率. 然而, 由于该类模型是基于报价驱动市场的, 在应用于指令驱动市场时, 会遇到委托指令流不能成交的风险问题. 该类模型的另一个局限是要求有足够的交易数据和时间期限, 但当时间间隔变小时, 特别是考虑每一笔交易时, 信息交易概率较难被算出, 因此在时间序列方面的

① 收稿日期: 2008-01-17 修订日期: 2009-12-25

基金项目: 国家自然科学基金资助项目 (70871002 70671006); 全国优秀博士学位论文作者专项基金资助项目 (200466).

作者简介: 李广川 (1982-), 男, 河北廊坊人, 博士, 深圳证券交易所博士后. Email: bfen\_lg@gnail.com

② 严格意义上, 所有提交订单中信息交易者所提交的订单比率更能反应市场信息不对称的程度. 但是, 由于数据的不能获取和指令修改、撤销等复杂因素影响等原因, 本文所研究的信息交易概率聚焦于所有“交易”中为信息性交易的比例. 事实上, 这也是目前绝大部分信息交易概率文献所共同关注的角度.

③ 关于市场微观结构理论, 参见文献 [1] 给出的综述.

应用并不广泛。

鉴于 PIN 模型的上述缺陷, Nyholm<sup>[13]</sup>提出了一个新的模型, 该模型可以估算出每一笔交易属于信息性交易的概率, 同时也规避了委托指令流不能成交的风险问题。该模型的主要思想是, 做市商会在判断上一笔成交的买卖方向和交易者类型的基础上, 对报价中点进行不同的调整, 以促成下一笔成交。Ma et al<sup>[14]</sup>运用一个类似的模型框架研究了台湾股票市场的信息交易概率, 其模型架构思想为, 每笔成交价会因成交是否属于信息性交易而进行不同的调整, 而成交量可以作为传递信息量的变量。国内学者攀登、施东晖<sup>[15]</sup>则基于类似的框架度量了中国股市的信息交易概率。

在考虑像中国股票市场这样的连续竞价指令驱动市场时, 直接应用上述框架可能会导致错误的判断。首先, 市场交易机制不同。Nyholm<sup>[13]</sup>提出的模型主要针对报价驱动市场, 在应用于指令驱动市场时同样会遇到委托指令流不能成交的风险问题。Ma et al<sup>[15]</sup>提出的模型则适用于像台湾股票市场那样的连续集合竞价指令驱动市场<sup>④</sup>。更重要的是, 两模型的估计方法不适合连续竞价指令驱动市场。两种模型分别以报价价差中位数的调整和成交价的调整作为市场对信息量进行反馈的研究变量, 由于最小报价单位的存在, 二者事实上都是离散变量, 而两种模型都为它们赋予连续分布加以估计, 这在数据离散化严重的连续竞价指令驱动市场尤为不适用。

针对该情形, 提出一个在连续竞价指令驱动市场中估计信息交易概率的模型。首先, 构造一个名为“成交积极性”的有序离散变量, 并认为其反映了市场上信息量的变化。然后, 建立一个成交积极性与成交量的线性模型, 并通过一个离散指示变量区分信息性交易和非信息性交易条件下两种不同的线性关系。最后, 利用中国股票市场的分笔成交数据, 结合有序 probit 方法和马尔可夫状态转移技术, 对模型进行了估计。运用该模型得出的

信息交易概率与被学者们广泛运用的 PN 模型及其扩展时变 TPN 模型进行了比较分析。通过对不同价差的解释能力进行分析发现, 基于模型估计的信息交易概率相对于 PN 和 TPN 具有更好的有效性和解释力。

在此基础上, 进一步考查了信息交易概率与资产收益波动、报价价差和成交持续期之关系的日内效应。已有研究表明, 股票交易特征往往具有明显的“日内效应”, 即一天当中某段时间内的值明显地异于其他时间内的值<sup>[16-19]</sup>。许多学者认为信息性交易是引发“日内效应”的一个重要因素。例如, McInish 和 Wood<sup>[13]</sup>的研究认为报价价差的日内效应来自于私有信息; Admati 和 Pfleiderer<sup>[20]</sup>认为资产价格波动的日内效应与信息交易者的交易策略有关; Easley 和 O'Hara<sup>[21]</sup>等指出, 成交持续期的长短与信息的有无及好坏等有关, 间接表明成交持续期的日内效应可能与日内信息分布的不均有关。国内学者何诚颖等<sup>[22]</sup>研究发现逆向选择成本的日内模式呈“L”形, 且刚开盘时交易活跃的主要原因是信息性交易。然而, 这些研究并没有涉及信息性交易与股票交易特征之关系是否具有“日内时变性”。本文运用定量方法研究了各个交易特征与信息交易概率在日内不同时间区间的相关性, 并用图形描绘了这种相关性的日内变化特征。结果确实发现, 信息交易概率与股票交易特征不仅具有相关性, 而且具有明显的日内效应。

## 1 模型及信息交易概率

### 1.1 成交积极性

在连续竞价指令驱动市场上, 假定交易者根据拥有资产价格信息的情况提出报价、报量与当前市场最优报价、报量的间距, 则这种间距反映了交易者交易意愿的强烈程度<sup>⑤</sup>。例如, 如果报价低于当前市场的最优卖价, 或者高于当前市场的最

④ 台湾股票市场的正常交易期间, 每若干分钟进行一次集合竞价的撮合, 以成交量最大原则给出成交价。

⑤ 需要说明两点。第一, 这里的假定仍是基于理性框架的, 即投资者要么条件于已有信息进行投资决策, 要么是噪声交易者, 这种设定与 PIN 框架是一致的。不可否认的是, 依据行为金融理论, 投资者的指令提交行为有可能受到心理因素的影响而显示出非理性行为。然而, 本文的重点并非探讨理性与非理性框架孰对孰错(这本身仍是目前传统金融学与行为金融学争论的焦点问题), 这也超出了本文的研究范畴。第二, 投资者决策所依据的信息可能有二, 其一为有关公司未来现金流的客观信息, 其二为一些专业投资者进行专业分析之后得到的主观信息。即使后者不一定与公司未来现金流有关, 也不影响这些投资者与拥有客观信息的投资者具有相似的指令提交行为。从这种意义上说, 本文所度量的信息交易概率涵盖了这种基于主观信息的交易行为。

优买价,即交易者提交的指令更具竞争性,说明交易者希望达成交易的意愿很强烈,具有很高的交易积极性.反之则说明交易者不急于成交.

在完整指令流数据难于获取的中国股票市场,可以通过分笔成交数据来间接观测交易者提交报价、报量的策略.当  $t$  时刻发生一笔交易时,通过观测成交价、成交量和  $t-1$  时刻最优报价、报量的间距,可以近似地判断交易者交易意愿的强烈程度.为此,构造了名为“成交积极性”的有序离散变量  $C_t$  (见表 1),取值从 1 到 5 数值越大,代表成交积极性越高,即交易者具有越强烈的交易意愿.

表 1 成交积极性  
Table 1 Trading aggressiveness

成交积极性	赋值依据
1	$t$ 时刻成交价介于 $t-1$ 时刻最优买价和最优卖价之间.
2	$t$ 时刻成交价等于 $t-1$ 时刻最优买价同时成交量小于 $t-1$ 时刻买一量;或者 $t$ 时刻成交价等于 $t-1$ 时刻最优卖价同时成交量小于 $t-1$ 时刻卖一量.
3	$t$ 时刻成交价等于 $t-1$ 时刻最优买价同时成交量等于 $t-1$ 时刻买一量;或者 $t$ 时刻成交价等于 $t-1$ 时刻最优卖价同时成交量等于 $t-1$ 时刻卖一量.
4	$t$ 时刻成交价等于 $t-1$ 时刻最优买价同时成交量大于 $t-1$ 时刻买一量;或者 $t$ 时刻成交价等于 $t-1$ 时刻最优卖价同时成交量大于 $t-1$ 时刻卖一量.
5	$t$ 时刻成交价小于 $t-1$ 时刻最优买价或者大于 $t-1$ 时刻最优卖价.

由于交易意愿的强烈程度取决于交易者拥有资产价格信息的多少,因此成交积极性反映了信息量的变化.

在研究分笔信息交易概率的类似框架中, Nyholm<sup>[13]</sup>, Ma et al<sup>[14]</sup> 分别使用报价价差的变化、成交价变化等变量来反映信息量的变化.这在他们所研究的特定交易机制下是适用的,却并不适用于连续竞价指令驱动市场.首先,从机制角度考虑,报价价差变化所反映的信息量变化是基于做市商视角的,这在我们的框架中显然是不合适的.其次,从实证角度考虑,在纽约和台湾这些交易频率相对较低的市场上,由于相邻两笔交易的

间隔时间足够长,无论是报价价差还是成交价,其变化一般都是显著的,也就能有效反应信息量变化.而在中国这种连续竞价指令驱动市场上,相邻两笔交易的时间间隔往往低于 10 second 造成绝大多数相邻两笔交易的成交价和报价价差均相同,即二者无变化,因此它们不能有效反映信息量的变化.从成交积极性的定义可以看出,它本质上是将价格变化进行了细分,使得即使相邻两笔交易的成交价相同,也完全有可能根据交易量的不同而被划分为不同的成交积极性水平,这更为有效地反映了信息量的变化.这种方法的一个不足是使得价格产生巨大变化的成交被统一划分为同一积极性水平.但是,由于交易过于频繁,这种类型的成交所占的比例远远低于价格产生微小变化情形的比例,以至于对本文整体结论的影响可以忽略不计.正因为这些原因,本文认为成交积极性更为科学地区分了不同类型交易者的微观行为,能够有效反映连续竞价指令驱动市场中信息量的变化.

成交积极性是一个有序离散变量  $C_t$ . 在本文中,其理论分布与一个不可观测的连续随机变量  $C_t^*$  的理论分布相关联.具体来说,二者关系满足

$$C_t = \begin{cases} 1 & \text{若 } C_t^* \in (-\infty, \delta_1) \\ 2 & \text{若 } C_t^* \in (\delta_1, \delta_2) \\ 3 & \text{若 } C_t^* \in (\delta_2, \delta_3) \\ 4 & \text{若 } C_t^* \in (\delta_3, \delta_4) \\ 5 & \text{若 } C_t^* \in (\delta_4, +\infty) \end{cases} \quad (1)$$

其中,  $\delta_i, i = 1, 2, \dots, 5$  为  $C_t$  发生突变的阈值,作为模型参数被估计得出.如果  $C_t^*$  的条件累积分布函数为  $Y(C_t^* | F_t)$ ,  $F_t$  为  $t$ 时刻之前可得信息,那么  $C_t$  的条件密度函数如式 (2) 所示,该密度函数将作为构造似然函数的基础.

$$P(C_t | F_t) = \begin{cases} Y(\delta_1 | F_t) & C_t = 1 \\ Y(\delta_i | F_t) - Y(\delta_{i-1} | F_t) & C_t = 2, 3, 4 \\ 1 - Y(\delta_4 | F_t) & C_t = 5 \end{cases} \quad (2)$$

### 1.2 模型

假定市场上仅存在信息交易者和非信息交易者两种交易者,前者基于私有信息和公开信息进行交易,后者仅基于公开信息交易.在  $t-1$ 时刻

至  $t$  时刻的时间期间, 令市场上释放的公开信息集合为  $\Omega_{t-}^u$ , 私有信息集合为  $\Omega_{t-}^i$ . 成交积极性反映信息量的变化表明,  $t$  时刻的成交积极性  $C_t$  将是  $\Omega_{t-}^u$  和  $\Omega_{t-}^i$  的函数

$$C_t = f_t(\Omega_{t-}^u, \Omega_{t-}^i) \quad (3)$$

由于  $\Omega_{t-}^u$  和  $\Omega_{t-}^i$  不可观测, 间接的信息代理变量就变得很重要. 大量研究混合分布假说的文献<sup>[23]</sup>都将成交量看作信息流的代理变量, 遵循这一思路, 本文认为基于公开信息的成交量  $Q_t^u$  可以作为  $\Omega_{t-}^u$  的代理变量, 基于私有信息的成交量  $Q_t^i$  可以作为  $\Omega_{t-}^i$  的代理变量, 则有

$$C_t = f_t(Q_t^u, Q_t^i) \quad (4)$$

下面考虑具体的函数形式. 本文假设成交积极性与成交量呈线性关系. 这是因为, 根据本文的定义, 成交积极性是对价、量同时变化程度的一种度量. 其中, 量的变化与成交量成线性关系是合理的, 此外, Kyle<sup>[24]</sup>认为价格的变化与成交量之间也呈线性关系.

$t$  时刻的成交量  $Q_t$  为  $Q_t^u$  与  $Q_t^i$  之和. 因此函数形式设定如下

$$C_t = \lambda(Q_t^u + Q_t^i) + \varepsilon_t = \lambda Q_t + \varepsilon_t \quad (5)$$

其中,  $\lambda$  为常系数,  $\varepsilon_t$  为随机误差项, 彼此独立同分布, 满足  $\varepsilon_t \sim N(0, \sigma^2)$ ,  $\sigma^2$  为方差, 同时  $E(\varepsilon_t | Q_t) = 0$

为研究私有信息对成交积极性的影响并进一步计算信息交易概率, 添加一个离散指示变量  $I_t$ . 当  $t$  时刻发生的交易由信息交易者发起时,  $I_t = 1$ ; 当  $t$  时刻发生的交易由非信息交易者发起时,  $I_t = 0$ . 同时, 将式 (5) 修改为

$$C_t = (\lambda_u + \lambda_i \cdot I_t) Q_t + \varepsilon_t \quad (6)$$

$\lambda_u$  和  $\lambda_i$  均为常系数. 这样一来, 由于  $I_t$  的存在, 式 (6) 实际演变为非线性模型. 当信息交易者发起交易时, 成交量对成交积极性影响的系数为  $\lambda_u + \lambda_i$ , 当非信息交易者发起交易时, 成交量对成交积极性影响的系数仅为  $\lambda_u$ .

对式 (6) 进行估计时会遇到两个难题. 第一, 作为有序离散变量的  $C_t$  尚未与一个连续随机变量相关联, 因此无法获得其条件密度函数; 第二,  $I_t$  不能被观测. 首先, 运用有序 probit 回归模型解决第一个难题. 考虑一个不可观测的连续随机变量  $C_t^*$ , 代替式 (6) 中的  $C_t$ . 可知  $C_t^*$  的条件累积分布函数为  $\Phi((C_t^* - (\lambda_u + \lambda_i \cdot I_t) Q_t) / \sigma)$ ,  $\Phi(\cdot)$

为标准正态分布的累积分布函数. 由式 (2) 易得  $C_t$  的条件密度函数  $P(C_t | F_t)$ . 其次, 根据已有的相关研究, 马尔可夫状态转换方法可以很好地解决  $I_t$  不能被观测这一问题<sup>[13-14]</sup>. 因此, 本文同样应用这一方法解决该问题.

值得讨论的一个问题是: 如果  $\lambda_i > 0$  那么是否意味着信息交易者一定会提交更积极的指令呢? 答案是不一定. 因为在有序 Probit 模型框架下,  $\lambda_i > 0$  仅仅意味着信息交易者提交的指令处于高积极性水平的“概率”较非信息交易者高. 这个结果是重要的, 因为已有研究表明, 信息交易者有时会模仿非信息交易者的交易策略从而达到隐藏信息的目的. 因此, 本文的框架并不排除这种可能性. 但是, 严格区分信息交易者的隐藏交易和非信息交易者的交易已经超出了本文的范畴.

两个难题都得以解决后, 需要估计的对数似然函数为

$$L(C_t | F_t) = \sum_{t=1}^T \log P(C_t | F_t) \quad (7)$$

附录给出了应用上述两方法计算对数似然函数的技术细节.

进一步将交易者分为买方和卖方, 并研究双方的信息交易行为是否显著不同. 首先采用 Lee 和 Ready<sup>[25]</sup> 提出的标记检验 (tick test) 方法区分每笔交易的发起方向. 然后, 当由买方发起促成成交时, 令  $\lambda_t = \lambda_b^t$ ; 当由卖方发起促成成交时, 令  $\lambda_t = \lambda_s^t$ . 因此, 需要估计的系数向量为  $\omega = (\lambda_u, \lambda_b^t, \lambda_s^t, p_{00}, p_{11}, \delta_1, \delta_2, \delta_3, \delta_4, \sigma)$ , 可以通过极大化似然函数  $L(C_t | F_t)$  得到.

在估计模型时, 为保证得到准确的估计结果, 对参数进行了一定的变换. 例如, 通过 logit 变换以保证  $0 \leq p_{00}, p_{11} \leq 1$ ; 通过对数变换以保证  $\sigma \geq 0$ . 参数的标准差则通过 delta 方法来计算.

### 1.3 信息交易概率

尽管  $P(I_t = m | C_t, F_t)$  也可以直接用来计算信息交易概率, 但它只利用了  $t$  时刻之前的可得信息, 而不是全部时期样本的信息. 如果从全部时期的角度看,  $P(I_t = m | C_t, F_t)$  就只能是  $t$  时刻状态的“事前判断”, 因为已获得了全部时期的样本信息, 就可以进行“事后判断”, 即可以计算出  $P(I_t = m | C_t, F_t)$ ,  $m = 0, 1$ . 显然利用该值计算

信息交易概率, 结果能够反应全部时期样本的信息. 因此, 首先利用  $\text{Kin}^{[26]}$  给出的迭代算法求出该值, 见附录式 (F-6).

基于该结果, 任一时间段内的信息交易概率  $PI_{t_1 \leq t \leq t_2}$  可由下式求出

$$PI_{t_1 \leq t \leq t_2} = \frac{\sum_{t=t_1}^{t_2} P(I_t = 1 | C_T, F_T) \cdot Q_t}{\sum_{t=t_1}^{t_2} Q_t} \quad (8)$$

类似于 Ma et al<sup>[14]</sup> 的研究, 考虑了成交量的加权.

由式 (8) 可以看出, 当某一时间段缩小到  $t$  时刻发生的一笔成交时, 该笔成交的信息交易概率  $PI_t$  为  $I_t$  等于 1 的平滑概率. 即

$$PI_t = P(I_t = 1 | C_T, F_T) \quad (9)$$

## 2 样本与数据

本文使用的数据来源为北京色诺芬信息服务有限公司开发的中国股票市场分笔成交数据库, 选取上证 50 指标股在 2003-07-01 至 2003-09-30 的高频交易数据为样本. 对于任一只股票, 每

一笔成交数据记录了股票代码、日期、成交价、成交量、成交时间、成交时刻买(卖)一至买(卖)三价、成交时刻买(卖)一至买(卖)三量等信息. 需要特别指出的是, 由于交易频率过于频繁和数据记录方法的原因, 每一笔成交可能是由若干笔更小的成交组合而成的. 我们认为, 第一, 本文模型并不像 PN 模型那样要利用买卖交易笔数来估计信息交易概率, 如果组成单笔成交的若干小成交中信息交易者发起的成交占多数, 那么它被判断为信息性交易的概率与单独一笔更可能由信息交易者发起的成交被判断为信息性交易的概率并无不同, 这对计算股票一段时间信息交易概率的影响更小. 第二, 由于相邻两笔成交的时间间隔足够短(数据记录中的成交持续期经常为 3 至 5 second), 将其视为单笔成交并不会产生太大的误差.

为提高估计准确性, 对样本数据进行如下预处理. 首先, 为避免错误数据的影响, 把样本期内存在错误数据记录的 5 只股票剔除, 剩余 45 只股票作为分析样本. 其次, 为避免集合竞价和连续竞价两种机制的相互影响, 剔除了在连续竞价期外的所有成交数据. 表 2 给出了样本数据的部分横截面描述统计量.

表 2 样本数据横截面描述统计量

Table 2 Cross-Sectional descriptive statistics of sample data

变量	均值	标准差	分位数				
平均深度	670.546	1253.528	最小值	25	50	75	100
				110.003	218.006	421.666	492.100
平均持续期	43.436	23.353	18.368	24.301	37.033	54.491	126.398
平均价差	0.016	0.012	0.010	0.011	0.012	0.015	0.086
收益标准差	0.0069	0.0112	0.0008	0.0011	0.0014	0.0056	0.0497
平均成交量	6036.742	3046.719	2781.344	4246.569	5721.005	6728.367	18939.960

注 1: 相关变量的计算如下: 收益 = 当期成交价的对数 - 上期成交价的对数; 深度 = 买一量 + 卖一量; 持续期 = 当期成交时间 - 上期成交时间; 价差 = 卖一价 - 买一价.

注 2: 表中“变量”一栏中的平均值变量指各支股票样本期内的平均值, 收益标准差指各支股票每期收益在样本期内的标准差; “均值”一栏指全部股票样本对应“变量”一栏内各变量的均值. 其余栏变量的计算类似于“均值”栏.

## 3 模型估计结果与分析

运用 Gauss 软件对所有股票样本进行估计,

结果列于表 3 限于篇幅, 没有列出所有股票的全部系数估计值, 只列出了部分系数统计量和比较统计量<sup>⑥</sup>, 这些统计量已经足够得出清晰和有效的结论.

⑥  $t$  检验表明所有系数都统计显著, 故没有给出系数标准差统计量.

表 3 估计系数统计表  
Table 3 Statistics of estimators

系数统计量							
系数	均值	标准差	分位数				
			最小值	25	50	75	100
$\lambda_u$	0.128	0.119	0.033	0.073	0.096	0.158	0.828
$\lambda_i^b$	1.232	1.327	0.357	0.578	0.879	1.254	8.170
$\lambda_i^s$	1.982	5.572	0.294	0.610	0.919	1.339	37.798
$p_{00}$	0.868	0.093	0.625	0.815	0.914	0.941	0.989
$p_{11}$	0.843	0.102	0.561	0.783	0.886	0.920	0.956
$\delta_1$	-14.2618	7.6208	-39.8946	-15.6174	-9.7491	-9.5480	-8.9676
$\delta_2$	7.2385	4.4919	3.6775	4.5908	5.0448	8.3326	21.7840
$\delta_3$	8.9327	4.9556	4.5603	6.0105	6.5897	9.2913	24.6410
$\delta_4$	14.7770	7.8301	9.3415	9.9233	10.3962	15.7863	40.4268
$\sigma$	8.1987	3.7195	5.0007	6.0007	6.6719	8.8681	19.7642
$PI$	0.380	0.090	0.056	0.335	0.368	0.455	0.522
比较统计量							
比较关系	百分比			Kruskal-Wallis 检验			
$\lambda_i > 0$	100.00%			—			
$\lambda_i^b > 0$	100.00%			—			
$\lambda_i^s > 0$	100.00%			—			
$\lambda_i^s > \lambda_i^b$	60.00%			0.9968			
$p_{00} > p_{11}$	71.11%			0.1591			

注 1: 表中“系数统计量”部分有关信息交易概率的统计量的计算步骤为, 首先, 计算基于每只股票全部样本的唯一信息交易概率值; 然后, 对所有 45 个值进行相应统计量的计算。

注 2: 表中“比较统计量”部分“百分比”栏的值, 为所有比较关系成立的股票样本占全部样本的比例数。

注 3: 表中“比较统计量”部分“Kruskal-Wallis 检验”栏的值, 只对具有比较关系的两组系数值是否来源于同一分布进行了检验, 没有对系数值和常数的比较进行检验。

注 4:  $PI$  值的计算基于全部样本期。

首先, 对于所有股票,  $\lambda_u$ ,  $\lambda_i^b$ ,  $\lambda_i^s$  均大于 0 这一结果说明, 信息性交易会提高指令属于高积极性水平的概率。正如前文所述, 尽管这一结果并不排除信息交易者模仿非信息交易者之策略以隐藏信息这一行为的可能, 但在一定程度上仍表明, 在中国股票市场上, 信息交易者往往提供更具竞争性的指令 (如提交更高的买价、更低的卖价, 或更大的报量) 以尽快地促成交易。

其次, 尽管  $\lambda_i^s$  的均值和绝大部分分位数 (最小值除外) 都大于  $\lambda_i^b$ , 同时 60% 的股票样本的  $\lambda_i^s$  大于  $\lambda_i^b$ , 但这不表明由卖方发起的交易比由买方发起的交易具有更高的信息交易概率。Kruskal-Wallis 检验的结果表明, 不能拒绝  $\lambda_i^s$  与  $\lambda_i^b$  来自同一总体的假设, 即信息在买方交易者和

卖方交易者中的分布没有显著的差别。

最后, 对  $p_{00}$  和  $p_{11}$  关系进行比较的结果类似于对  $\lambda_i^s$  和  $\lambda_i^b$  的比较。尽管从均值、分位数和百分比的角度看,  $p_{00}$  都大于  $p_{11}$ , 但 Kruskal-Wallis 检验的结果却不能拒绝  $p_{00}$  和  $p_{11}$  来自同一总体的假设, 即信息性交易和非信息性交易具有基本相同的聚集性。已有研究均表明非信息性交易具有更大的聚集性<sup>[13-14]</sup>, 本文的研究发现了中国市场上与之相反的证据。

#### 4 与经典信息交易概率模型的分析

为检验运用本文方法计算的信息交易概率的

有效性和解释能力,这一部分对本文模型与经典的信息交易概率模型进行比较分析.鉴于PN模型被广泛接受和使用,选取该模型及扩展该模型的时变TPN模型<sup>[8]</sup>,分别用于检验本文模型在横截面特性上和时间序列特性上的有效性和解释能力.分别运用SAS和Gauss软件对两个模型在所有样本上进行了估计.限于篇幅,没有给出两模型的似然函数方程和信息交易概率计算方法,请读者参考原文.

遵循并扩展Qin Le和Wu<sup>[8]</sup>比较TPN模型和PN模型的思路,分别建立了一个横截面回归模型和一个面板回归模型,用于检验本文模型、PN模型和TPN模型对不同价差的解释能力,以比较各信息交易概率模型的相对有效性及解释能力.式(10)和式(11)分别给出了横截面回归模型和面板回归模型<sup>⑦</sup>.

$$Spread_i = \alpha_0 + \alpha_1 VPI_i + \alpha_2 VPN_i + \alpha_3 VTPN_i + \alpha_4 VOL_i + \varepsilon_i \quad i = 1 \dots, 45 \quad (10)$$

$$Spread_{i,t} = \beta_1 VPI_{i,t} + \beta_2 VTPN_{i,t} + \beta_3 VOL_{i,t} + \sum_{i=1}^{45} \gamma_i \delta(i) + \eta_{i,t} \quad i = 1 \dots, 45 \quad (11)$$

表5 分析变量横截面描述统计量

Table 5 Cross-sectional descriptive statistics of variables

变量	均值	标准差	分位数				
			最小值	25	50	75	100
PI	0.378	0.094	0.056	0.334	0.359	0.464	0.522
PIN	0.133	0.049	0.083	0.099	0.124	0.148	0.337
TPN	0.174	0.061	0.076	0.129	0.160	0.216	0.295
Spread(open)	0.022	0.014	0.007	0.014	0.017	0.025	0.075
Spread(close)	0.014	0.006	0.010	0.011	0.012	0.016	0.041
VOL	16.983	0.798	15.428	16.481	17.026	17.526	18.779
PRICE	8.920	3.469	3.647	6.295	8.299	11.309	16.723

注1 “变量”栏中,PI为基于本文模型估计的信息交易概率;PIN为基于PN模型估计的信息交易概率;TPN为基于TPN模型估计的信息交易概率;Spread(open)为开盘后10min平均价差;Spread(close)为收盘前10min平均价差;VOL为对数日交易额;PRICE为日收盘价.全天空平均价差描述统计量见表2

从表5中可以看出,TPN取值整体上要高于PIN,这与Qin Le和Wu<sup>[8]</sup>研究的结果是一致的.同时,基于本文模型估计的PI取值整体上要高于

其中,在横截面回归方程中,Spread为价差变量,在实证分析中将分别采用开盘后10min平均价差,收盘前10min平均价差和全天平均价差.VPI为股票价格和本文信息交易概率的乘积,VPN为股票价格和PN模型信息交易概率的乘积,VTPN为股票价格和TPN模型信息交易概率的乘积,所有信息交易概率均为根据各个模型计算出的唯一值(TPN模型为日信息交易概率在样本期内的均值),股票价格为日收盘价在样本期内的平均值.VOL为日均交易额的对数值.其余所有变量的意义同横截面回归方程,但均为日变量时间序列.PN模型由于不能计算日信息交易概率值而未包含在面板回归方程中.

根据已有研究,信息交易概率对价差有正向影响,交易量对价差有负向影响.因此,本文期望VPI、VPN和VTPN变量有正的系数值,VOL变量则有负的系数值.如果本文模型优于PN模型和TPN模型,那么期望VPI在各个回归模型中的系数估计值均为正且显著,同时VPN和VTPN变量系数估计值的显著性较低或不显著.表5给出了用于回归分析的各变量的横截面描述统计量.

PN和TPN.如果实证分析表明,PI相对于PN和TPN具有更好的有效性和解释力,那么这在一定程度上表明,基于PN框架估计的信息交易概率

⑦ 在面板回归方程中,采用横截面固定效应面板回归,即每支股票有不同的截距估计值,但有相同的回归变量系数估计值.δ(i)为对应各支股票的哑变量,这种描述方式的目的是清晰地表示不同截距估计量,在实际估计过程中,Eviews软件会自动添加一个公共截距,所估计出的个股截距为与此公共截距的偏差.

有可能低估了信息交易的比例. 表 6 给出了式 (10) 和式 (11) 的估计结果. 限于篇幅, 表中仅列出了重要系数的估计结果.

表 6 回归方程估计结果表

Table 6 Estimation results of regression equations

回归变量	开盘后 10 min 平均价差		收盘前 10 min 平均价差		全天平均价差	
面板 A: 横截面回归分析						
<i>VPI</i>	0.0036	(3.6760)	0.0006	(2.4977)	0.0011	(3.3804)
<i>VTPN</i>	-0.0061	(-2.6118)	0.0003	(0.3764)	0.0028	(3.6573)
<i>VPN</i>	0.0127	(4.6959)	0.0024	(3.3580)	0.0001	(0.1136)
<i>VOL</i>	-0.0079	(-4.6743)	-0.0012	(-3.0405)	-0.0014	(-2.5399)
面板 B: 面板回归分析						
<i>VPI</i>	0.0005	(2.0405)	0.0006	(2.7112)	0.0009	(7.8314)
<i>VTPN</i>	0.0064	(4.3231)	0.0003	(0.3978)	0.0027	(7.9254)
<i>VOL</i>	-0.0046	(-6.6303)	-0.0008	(-2.3790)	-0.0003	(-2.0855)

注 1: 不同价差栏分别代表因变量为相应价差时的估计结果. 任一价差栏包括两列, 第 1 列为与行变量相对应的系数估计值, 第 2 列括号内的数值为相应系数的 *t* 统计量. 一般情形下, *t* 统计量绝对值大于 2 意味着系数显著.

注 2: 面板 A 给出了对应式 (10) 的估计结果, 面板 B 给出了对应式 (11) 的估计结果.

从表 6 的结果可以看出, 在横截面回归分析中, *VTPN* 变量的系数仅在对全天平均价差的解释中为正且显著, 在对开盘后 10 min 平均价差的解释中符号为负, 在对收盘前 10 min 平均价差的解释中不显著; *VPN* 系数则在对全天平均价差的解释中不显著. 与之相对的是, *VPI* 系数在对全部三种价差的解释中均为正且显著. 在面板回归分析中, *VTPN* 系数在对收盘前 10 min 平均价差的解释中不显著, 与之相对的是, *VPI* 系数在对全部三种价差的解释中均为正且显著. 以上结果表明, 基于本文模型估计的信息交易概率 *PI* 相对于 *PN* 和 *TPN* 具有更好的有效性和解释力. 这一结果同时表明, 基于 *PN* 框架估计的信息交易概率可能低估了信息交易的比例.

## 5 信息交易概率与股票交易特征之关系的日内效应

### 5.1 信息交易概率与股票交易特征的日内效应

首先给出信息交易概率与股票交易特征的日内效应. 图 1 描绘了信息交易概率和资产收益波动、报价价差、成交持续期等股票交易特征的日内效应.

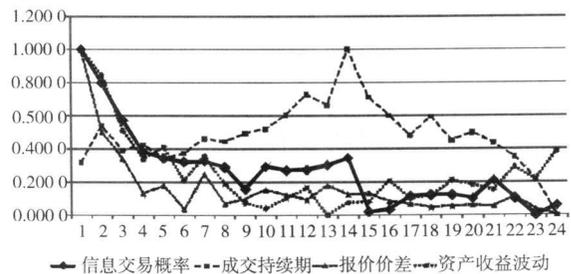


图 1 信息交易概率与股票交易特征的日内效应

Fig. 1 Intraday effect of probability of informed trading and stock's trading characteristics

注 1: 数据的计算步骤为: 第 1 步, 将每天的交易时间分成相等的 24 个区间, 每个区间 10 min, 然后分别求出每个区间内各变量的唯一值. 首先针对每一支股票在每一天内计算, 信息交易概率根据式 (13) 求出; 资产收益波动为分笔成交收益序列在 10 min 内的标准差; 成交持续期和报价价差都是取 10 min 内的平均值. 然后, 各变量的值在全部样本期内取平均. 第 2 步, 在每个 10 min 区间内, 取各变量的值在全部样本股票内的中位数, 形成各变量的日内数值序列. 第 3 步, 为便于将各变量放在同一个图中, 对各变量进行一定的变换. 以信息交易概率为例, 变换方法为

$$PI_i^* = (PI_i - MIN_{PI}) / (MAX_{PI} - MIN_{PI})$$

其中,  $PI_i$  表示第 *i* 个 10 min 区间内的信息交易概率在变换之前的值,  $PI_i^*$  表示变换之后的值.  $MIN_{PI}$  表示所有区间的信息交易概率值的最小值,  $MAX_{PI}$  表示所有区间的信息交易概率值的最大值.

注 2: 横坐标表示区间的序号, 纵坐标表示变换后各变量值的大小.

从图 1 可以得出如下结论

第一, 信息交易概率在开盘后取值较高, 然后缓慢下降, 40 m in 后达到一个稳定的水平. 中午开盘后, 信息交易概率在略有上升后显著下降, 然后处于一个基本稳定的水平. 从全天来看, 信息交易概率具有与已有研究结果类似的“L”型, 且上午的信息交易概率大于下午的信息交易概率. 本文认为, 信息交易概率在一天的两次开盘后都具有较高的取值, 是由于停盘时期的累计信息在开盘后需要释放的结果.

第二, 资产收益波动在开盘后的取值最高, 然后下降, 在收盘前取值再次上升, 呈“U”型; 报价价差同样在开盘后取值最高, 然后下降, 40 m in 后达到一个稳定水平并持续至收盘, 呈“L”型; 成交持续期在开盘后和收盘前的取值较低, 在中盘时期特别是中午开盘后的一段时间内取值较高, 呈“倒 U”型. 以上股票交易特征的日内效应, 除报价价差略有不同外, 均与已有的针对国外股票市场的研究结果类似.

### 5 2 信息交易概率与股票交易特征之关系的日内效应

直观上, 从图 1 可以看出, 信息交易概率与各个股票交易特征的日内效应具有一定的相关性. 如在开盘后的一段时间内, 信息交易概率与资产收益波动、报价价差都具有较高的取值, 然后逐渐下降. 那么, 这种相关性在日内是“一成不变”的吗? 本文将通过一种简单的定量方法研究该相关性的日内变化特征, 即日内效应.

首先, 构造一系列虚拟变量  $K_i, i = 1, 2, \dots, 24$  如下:

$$K_i = \begin{cases} 1 & \text{若时间处在第 } i \text{ 个区间} \\ 0 & \text{若时间未在第 } i \text{ 个区间} \end{cases} \quad (12)$$

然后, 构造一系列变量  $PI_{it}, i = 1, 2, \dots, 24$  如下:

$$PI_{it} = PI_t \cdot K_i \quad (13)$$

其中,  $PI_t$  为由式 (9) 求出的信息交易概率.

令  $V_t, S_t, D_t$  分别表示资产收益波动、报价价差和成交持续期. 通过计算  $PI_{it}$  与  $V_t, S_t, D_t$  的相关系数, 得到如下 3 个序列 ( $i = 1, 2, \dots, 24$ )

$$Cor_{it}^v = Corr(PI_{it}, V_t) \quad (14)$$

$$Cor_{it}^s = Corr(PI_{it}, S_t) \quad (15)$$

$$Cor_{it}^d = Corr(PI_{it}, D_t) \quad (16)$$

这 3 个序列表示了信息交易概率与资产收益

波动、报价价差、成交持续期的相关性的日内变化特征, 即日内效应, 见图 2

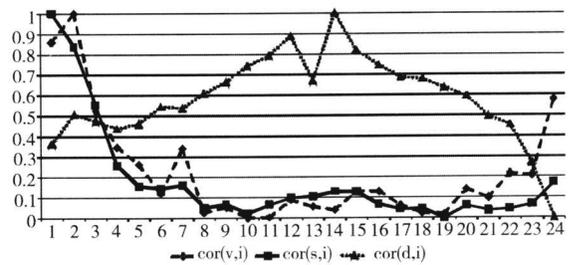


图 2 信息交易概率与股票交易特征相关性的日内效应

Fig 2 Intraday effect of correlation between probability of informed trading and stock's trading characteristics

注 1: 各变量的值均为全部股票样本取值序列的中位数. 所有变量都运用图 1(注 1 第 3 步)的方法进行了变换.

注 2: 横坐标表示区间的序号, 纵坐标表示  $PI_{it}$  与股票交易特征的相关系数.

注 3:  $Cor(v, i)$  表示  $Cor_{it}^v, Cor(s, i)$  表示  $Cor_{it}^s, Cor(d, i)$  表示  $Cor_{it}^d$ .

从图 2 可以得出如下结论

第一, 信息交易概率与股票交易特征的相关性具有显著的日内效应. 信息交易概率与资产收益波动、报价价差的相关性在日内基本呈现“U”型(或尾部上翘的“L”型), 观察原始数据发现, 这种相关性在开盘后和收盘前为正, 开盘后正相关性最强, 然后缓慢下降, 40 m in 后基本处于不相关水平. 信息交易概率与成交持续期的相关性在日内基本呈现“倒 U”型, 观察原始数据后发现, 这种相关性在开盘后和收盘前为负, 在中午停盘前后为正, 收盘前负相关性最强. 以上结论证明信息交易概率与股票交易特征不仅具有相关性, 而且具有明显的日内时变性.

第二, 信息交易概率与股票交易特征的相关性, 同股票交易特征本身具有类似的日内效应. 对比图 1 和图 2 可以发现这一现象. 该结论进一步佐证了信息性交易是促成股票交易特征日内变化的重要因素.

### 5 3 稳健性检验

尽管上述定量方法简单并直观地给出了信息交易概率与日内交易特征之关系的日内效应, 但不得不承认的是, 这种仅使用样本中位数来绘制日内效应图形的做法可能会带来“是否疏漏了其他可能信息”的质疑. 本文构造了两类面板回归模型来验证上述结论的稳健性.

第 1 类, 仍使用 10 m in 区间数据构造如下三个面板回归模型:

$$\begin{cases}
 V_{ij} = \sum_{j=1}^{24} b_j^v \varphi(j) PI_{ij} + \sum_{j=1}^{24} \hat{c}_j^v \varphi(j) VOL_{ij} + \sum_{i=1}^{45} \gamma_i \delta(i) + \xi_{ij}^v \\
 S_{ij} = \sum_{j=1}^{24} b_j^s \varphi(j) PI_{ij} + \sum_{j=1}^{24} \hat{c}_j^s \varphi(j) VOL_{ij} + \sum_{i=1}^{45} \gamma_i \delta(i) + \xi_{ij}^s \\
 D_{ij} = \sum_{j=1}^{24} b_j^d \varphi(j) PI_{ij} + \sum_{j=1}^{24} \hat{c}_j^d \varphi(j) VOL_{ij} + \sum_{i=1}^{45} \gamma_i \delta(i) + \xi_{ij}^d
 \end{cases}$$

$i = 1, \dots, 45 \quad j = 1, \dots, 24$  (17)

其中,  $V_{ij}$ ,  $S_{ij}$ ,  $D_{ij}$ ,  $PI_{ij}$ ,  $VOL_{ij}$  分别代表第  $i$  支股票第  $j$  个区间 (10 m in) 的资产收益波动、平均报价价差、平均成交持续期、信息交易概率和平均交易量,  $\varphi(j)$  为对应各个区间的哑变量,  $\delta(i)$  为对应各支股票的哑变量, 因此这些模型均为横截面固定效应面板回归模型, 不仅不同股票的截距项不同, 解释变量的系数在不同区间上也不同. 模型运用二阶段广义最小二乘方法进行估计, 交易量作为工具变量, 同时使用了横截面加权方法以消除横截面异方差的影响.

第 2 类, 使用分笔高频数据构造如下两组面板回归模型

$$\begin{cases}
 S_{i,t,j} = d_j^s PI_{i,t,j} + \hat{c}_j^s VOL_{i,t,j} + \sum_{i=1}^{45} \gamma_{ij} \delta(i,j) + \xi_{i,t,j}^s \\
 D_{i,t,j} = d_j^d PI_{i,t,j} + \hat{c}_j^d VOL_{i,t,j} + \sum_{i=1}^{45} \gamma_{ij} \delta(i,j) + \xi_{i,t,j}^d
 \end{cases}$$

$i = 1, \dots, 45 \quad t = 1, \dots, T_{i,j} \quad j = 1, \dots, 24$  (18)

其中, 每组模型 (方程) 均包含  $j = 1, \dots, 24$  个面板回归模型, 每个 10 m in 区间对应一个模型.  $S_{i,t,j}$ ,  $D_{i,t,j}$ ,  $PI_{i,t,j}$ ,  $VOL_{i,t,j}$  分别代表第  $j$  个模型中, 第  $i$  支股票第  $t$  笔成交发生时的报价价差、成交持续期、信息交易概率和交易量.  $\delta(i,j)$ ,  $T_{i,j}$  分别为第  $j$  个模型中, 第  $i$  支股票对应的哑变量以及第  $i$

支股票所在区间发生的成交笔数. 这些模型也为横截面固定效应面板回归模型, 但不同股票仅仅在截距项上不同, 解释变量的系数在同一区间中的横截面上是相同的. 同时, 由于不同股票在不同区间的成交笔数不同, 这些面板回归模型是非平衡的. 该模型提供了信息交易概率与日内交易特征之关系微观层面上更为深入的考查, 因为其变量取值均以“分笔”为单位. 由于资产收益波动在“分笔”的层面上无法算出, 因此未包含在上述分析中.

通过估计上述两类模型, 分别得到了  $b_j^v$ ,  $b_j^s$ ,  $b_j^d$ ,  $d_j^s$ ,  $d_j^d$ ,  $j = 1, \dots, 24$  系数序列, 它们在绝大多数情形下是显著的, 限于篇幅这里没有列出结果. 这些系数序列反映了信息交易概率在不同区间下对日内交易特征影响模式的不同, 即反映了二者关系的日内形态. 把这些系数序列按照图 1 (注 1 第 3 步) 的方法进行了变换以方便将它们放在同一张图中. 图 3 给出了变换后的  $b_j^v$ ,  $b_j^s$ ,  $b_j^d$ ,  $d_j^s$ ,  $d_j^d$ ,  $j = 1, \dots, 24$  序列图.

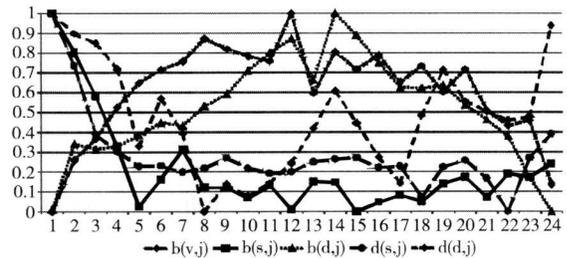


图 3  $b_j^v$ ,  $b_j^s$ ,  $b_j^d$ ,  $d_j^s$ ,  $d_j^d$  系数序列图

Fig 3  $b_j^v$ ,  $b_j^s$ ,  $b_j^d$ ,  $d_j^s$ ,  $d_j^d$  Estimator series

注 1 横坐标表示区间的序号, 纵坐标表示  $b_j^v$ ,  $b_j^s$ ,  $b_j^d$ ,  $d_j^s$ ,  $d_j^d$  值.

注 2  $b(v,j)$  表示  $b_j^v$ ,  $b(s,j)$  表示  $b_j^s$ ,  $b(d,j)$  表示  $b_j^d$ ,  $d(s,j)$  表示  $d_j^s$ ,  $d(d,j)$  表示  $d_j^d$ .

从图 3 可以看出,  $b_j^s$ ,  $d_j^s$  序列基本呈现“U”型 (或尾部上翘的“L”型),  $b_j^d$ ,  $d_j^d$  序列基本呈现“倒 U”型, 尽管  $b_j^v$  序列在中部有一定的起伏, 但整体趋势仍呈现“U”型. 对比图 2 以上序列的日内形态分别与  $Cor_i^s$ ,  $Cor_i^d$ ,  $Cor_i^v$  序列的日内形态整体上具有一致性. 这一结果证明, 本文所采用的简单定量方法所得的结论是具有稳健性的, 较好地捕捉了信息交易概率与日内交易特征之关系的日内效应.

### 6 结束语

本文提出了一个在连续竞价指令驱动市场中

估计信息交易概率的模型. 该模型有效解决了数据离散化问题, 能够有效计算连续竞价指令驱动市场的任一笔交易的信息交易概率. 本文对该模型和基于 PIN 框架的经典信息交易概率模型进行了比较分析, 发现前者对价差具有更好的解释力. 进一步地, 本文发现基于本文模型的信息交易概率与股票交易特征之关系具有显著的日内效应.

尽管本文研究了信息交易概率与交易特征之关系的“日内效应”, 但关于二者关系的符号并没有作深入讨论, 仅根据原始计算数据进行了简要说明. 这是因为, 正如本文所指出的, 由于在日内不同时间区间(如开盘、中盘和收盘), 信息交易

概率与交易特征的相关性可以为正或负, 因此在高频数据下对二者关系的符号下定论是武断的. 鉴于该问题需要更为深入的讨论, 本文建议将此议题作为未来研究的方向.

本文首先为连续竞价指令驱动市场的信息性交易研究提供了新的方法. 由于克服了数据离散化问题, 我们得以运用该方法更准确地和直接地度量信息性交易的强弱程度, 同时能任意调整度量的时间范围. 其次, 本文确认了信息交易概率与股票交易特征之关系的日内效应的存在, 这为更深入地研究股票交易特征的形成原因, 以及信息交易对投资者行为和交易特征的影响模式等, 提供了进一步的指引.

## 参考文献:

- [1] 李平, 曾勇, 唐小我. 市场微观结构理论综述 [J]. 管理科学学报, 2003, 6(5): 87-98  
Li Ping, Zeng Yong, Tang Xiaowo. Survey of market microstructure theory [J]. Journal of management sciences in China, 2003, 6(5): 87-98 (in Chinese)
- [2] Jaffe J F, Winkler R L. Optimal speculation against an efficient market [J]. The Journal of Finance, 1976, 31(1): 49-61.
- [3] 汪毅慧, 廖理, 邓小铁. 不对称信息, 交易成本和投资者保护: 内地(中国大陆)和香港的比较研究 [J]. 金融研究, 2003, 10: 27-36  
Wang Yihui, Liao Li, Deng Xiaotie. A symmetric information, trading cost and investor protection: Comparison between mainland and HongKong [J]. Journal of Financial Research, 2003, 10: 27-36 (in Chinese)
- [4] 王春峰, 王燕. 测量股票交易中的信息含量 [J]. 管理工程学报, 2004, 18(3): 42-46  
Wang Chunfeng, Wang Yan. Measuring the information content of stock trades [J]. Journal of Industrial Engineering and Engineering Management, 2004, 18(3): 42-46 (in Chinese)
- [5] Huang R, Stoll H. The components of the bid-ask spread: A general approach [J]. Review of Financial Studies, 1997, 10(4): 995-1034
- [6] 吴文锋, 朱云, 吴冲锋, 等. B股向境内居民开放对市场信息不对称的影响——买卖价差分解方法 [J]. 管理科学学报, 2007, 10(6): 57-64  
Wu Wenfeng, Zhu Yun, Wu Chongfeng, et al. Effect of B shares opening to domestic investors on information asymmetry - Using bid-ask spread decomposition method [J]. Journal of Management Sciences in China, 2007, 10(6): 57-64 (in Chinese)
- [7] Easley D, et al. Liquidity, information and infrequently traded stocks [J]. The Journal of Finance, 1996, 51(4): 1405-1436
- [8] Qian Leijun, and G. Wu. Time-varying informed and uninformed trading activities [J]. Journal of Financial Markets, 2005, 8: 153-181
- [9] MA T, HSEH M, Chen J. The probability of informed trading and the performance of stock in an order-driven market [C]. 9th Conference on the Theories and Practices of Securities and Financial Markets, 2000
- [10] Brockman P, Chung D Y. Informed and uninformed trading in an electronic order-driven environment [J]. The Financial Review, 2000, 35(2): 125-146
- [11] 刘善存, 李朋. 信息性交易概率和信息风险溢价研究 [J]. 中国金融学, 2006, 1: 116-129  
Liu Shancun, Li Peng. The probability of informed trading and information risk premium [J]. China Journal of Finance,

- 2006, 1: 116–129. (in Chinese)
- [12] 杨之曙, 姚松瑶. 沪市买卖价差和信息性交易实证研究[J]. 金融研究, 2004, 4: 45–56  
Yang Zhi-shu, Yao Song-yao. Bid-ask spread and informed trading: Evidence from Shanghai stock exchange[J]. Journal of Financial Research, 2004, 4: 45–56 (in Chinese)
- [13] Nyholm K. Estimating the probability of informed trading[J]. Journal of Financial Research, 2002, 25(4): 485–505
- [14] Ma T, Yang J, Kaohsiung T. Measuring the Probability of Informed Trading in an Order-Driven Auction Market and a Comprehensive Analysis on the Determinants of Informed Trading[Z]. Working Paper, 2004. (in Chinese)
- [15] 攀登, 施东辉. 知情交易的测度模型及其影响因素分析[J]. 管理世界, 2006, 6: 18–26  
Pan Deng, Shi Dong-hui. The measure model of informed trading and its determinants[J]. Management World, 2006, 6: 18–26 (in Chinese)
- [16] Andersen T G, Bollerslev T. Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements and longer run dependencies[J]. The Journal of Finance, 1998, 53(1): 219–265
- [17] Mclnish T H, Wood R A. An analysis of intraday patterns in bid/ask spreads for NYSE stocks[J]. The Journal of Finance, 1992, 47(2): 753–764
- [18] Chan K, Chung Y P, Johnson H. The intraday behavior of bid-ask spreads for NYSE stocks and CBOE options[J]. The Journal of Financial and Quantitative Analysis, 1995, 30(3): 329–346
- [19] Engle R F, Russell J R. Autoregressive conditional duration: A new model for irregularly spaced transaction data[J]. Econometrica, 1998, 66(5): 1127–1162
- [20] Admati A R, Pfleiderer P. A theory of intraday patterns: Volume and price variability[J]. Review of Financial Studies, 1988, 1(1): 3–40
- [21] Easley D, O'Hara M. Time and the process of security price adjustment[J]. The Journal of Finance, 1992, 47(2): 577–605
- [22] 何诚颖, 卢宗辉, 何兴强, 等. 中国股票市场逆向选择成本研究[J]. 经济研究, 2009, 2: 68–80  
He Cheng-ying, Lu Zong-hui, HE Xing-qiang, et al. A study on the adverse selection cost of Chinese stock exchange[J]. Economic Research Journal, 2009, 2: 68–80 (in Chinese)
- [23] Tauchen G E, Pitts M. The price variability-volume relationship on speculative markets[J]. Econometrica, 1983, 51(2): 485–505
- [24] Kyle A S. Continuous auctions and insider trading[J]. Econometrica, 1985, 53(6): 1315–1335
- [25] Lee C M C, Ready M J. Inferring trade direction from intraday data[J]. The Journal of Finance, 1991, 46(2): 733–746
- [26] Kim C J. Dynamic linear models with Markov-switching[J]. Journal of Econometrics, 1994, 60(1): 1–22
- [27] Hamilton J. Time Series Analysis[M]. Princeton: Princeton University Press, 1994.

## The probability of informed trading in continuous order driven market: A new method

LI Guang-chuan<sup>1, 2</sup>, LIU Shan-cun<sup>1</sup>, QIU Wan-hua<sup>1</sup>

1. School of Economics and Management, Beihang University, Beijing 100193, China

2. Shenzhen Stock Exchange, Shenzhen 518010, China

**Abstract** Based on the variable “Transaction Aggressiveness” put forward in this paper, a new method to estimate probability of informed trading in continuous auction order driven market is proposed. We then do a comparative analysis between our method and classic models based on framework. We also research on intraday effects of the relations between probability of informed trading and trading characteristics such as asset re-

turn volatility, quoted spread and trading duration. We conclude that firstly, our model has more explanatory power on spread in comparison with classic models based on framework. Secondly, intraday effects of the relations between probability of informed trading and trading characteristics do exist which are robust based on panel regression analysis.

**Keywords** continuous auction order driven market probability of informed trading trading characteristic intraday effect

附录

1) 似然函数  $L(C_t | F_t)$  的构造

首先, 不考虑  $I_t$  不可观测这一限制, 即给定  $I_t$  值,  $C_t$  的条件概率  $P(C_t | I_t = m, F_t)$ , 可由式 (F-1) 给出. 其中  $C_t = 1, 2, \dots, 5, m = 0, 1, F_t = (C_{t-1}, C_{t-2}, \dots, C_1, Q_t, Q_{t-1}, \dots, Q_1)$  为至  $t$  时刻的信息集.

$$P(C_t | I_t = m, F_t) = \begin{cases} \Phi\left(\frac{\delta_1 - (\lambda_u + \lambda_i \cdot m)Q_t}{\sigma}\right) & C_t = 1 \\ \Phi\left(\frac{\delta_{c_t} - (\lambda_u + \lambda_i \cdot m)Q_t}{\sigma}\right) - \Phi\left(\frac{\delta_{c_{t-1}} - (\lambda_u + \lambda_i \cdot m)Q_t}{\sigma}\right) & C_t = 2, 3, 4 \\ 1 - \Phi\left(\frac{\delta_4 - (\lambda_u + \lambda_i \cdot m)Q_t}{\sigma}\right) & C_t = 5 \end{cases} \quad (F-1)$$

则  $C_t$  与  $I_t$  无关的条件密度函数可由下式求出

$$P(C_t | F_t) = \sum_{m=0}^1 P(I_t = m | F_t) \cdot P(C_t | I_t = m, F_t) \quad (F-2)$$

现在考虑  $I_t$  不能被观测的问题, 该问题导致  $P(I_t = m | F_t)$  未知, 则式 (F-2) 给出的条件概率不能被求出. 这里应用马尔可夫状态转换方法解决该问题.

令  $I_t$  服从一阶马尔可夫过程,  $P = \{p_{mn}\}, m, n = 0, 1, 0 \leq p_{mn} \leq 1$  为静态状态转移矩阵, 其中

$$p_{mn} = p(I_t = n | I_{t-1} = m) \quad (F-3)$$

令  $P(I_t = m | F_t), m = 0, 1$  为  $t$  时刻  $I_t = m$  的事前推断概率. 以此为基础, 给定初始值  $P(I_1 = m | F_1)$  及相应参数的初始值,  $P(I_t = m | F_t)$  可以通过以下两步迭代过程获得

$$P(I_{t-1} = m | C_{t-1}, F_{t-1}) = \frac{P(I_{t-1} = m | F_{t-1}) \cdot P(C_{t-1} | I_{t-1} = m, F_{t-1})}{P(C_{t-1} | F_{t-1})} \quad (F-4)$$

$$P(I_t = m | F_t) = \sum_{n=0}^1 p_{mn} \cdot p(I_{t-1} = n | C_{t-1}, F_{t-1}) \quad (F-5)$$

以上过程可参见 Hamilton<sup>[27]</sup> 对此的证明. 至此,  $P(C_t | F_t)$  已可完全求出, 因此式 (7) 给出的似然函数方程可被构造得出.

2) Kim 平滑概率迭代算法

$$P(I_t = m | C_T, F_T) = \begin{cases} \frac{P(I_T = m | F_T) \cdot P(C_T | I_T = m, F_T)}{P(C_T | F_T)} & t = T \\ P(I_t = m | C_t, F_t) \cdot \sum_{n=1}^2 p_{mn} \cdot \frac{P(I_{t+1} = m | C_t, F_t)}{P(I_{t+1} = m | F_{t+1})} & 1 \leq t \leq T \end{cases} \quad (F-6)$$