

上海证券交易所

# 上海股市的知情交易概率 及其影响因素

---

上海证券交易所研究中心

2006年4月



上海證券交易所  
SHANGHAI STOCK EXCHANGE

# 上海股市的知情交易概率及其影响因素

研究中心 施东晖

## 摘要

- 上海股市知情交易的无条件概率接近 40%，远大于 Easley, Kiefer, O'Hara & Paperman (1996) 和 Nyholm (2002, 2003) 对纽约证券交易所实证分析的结果。
- 知情交易对价格的影响程度远远大于不知情交易，是知情交易导致了股票价格根本性的大幅度的变化，而非知情交易只对股票价格形成短暂而微小的冲击。
- 知情交易概率的日内形态具有如些特点：开盘后的知情交易概率水平最高，随着交易的进行，信息逐渐传播，知情交易概率也随之下降，但在收盘前又有所增加。
- 公司的基本面因素会影响股票的知情交易概率，其中，流通市值与知情交易概率为负相关关系；而财务杠杆比率、主营业务增长率以及股权集中度与知情交易概率为正相关关系。
- 高流动性和低波动性会吸引知情交易者进场交易，反过来知情交易者的进场交易将导致流动性降低和波动性增大。

## 一、如何衡量市场的信息不对称性

信息不对称性和内幕交易行为一直都是各国股市在迈向有效市场路途上备受关注的课题，但如何衡量信息的不对称性却颇具争议。Hasbrouck(1991)认为，信息不对称应该是下列三项变量的函数：一是潜在知情交易者的比率；二是发现私有信息的概率；三是私有信息的准确性。然而，上述观点在实证研究上却面临相当的困难，因为这三个变量无法直接根据市场数据获得，而必须通过相关的非

外生变量进行间接衡量。

基于上述思路，早期的研究文献大都以买卖价差、非系统性风险、内部人持股、公司成长性等指标来间接衡量市场的信息不对称。但事实上，这些度量指标要么是信息不对称的结果(如价差和非系统性风险)，要么是信息不对称的原因(内部人持股、公司成长性)，因此并不能够对信息不对称程度形成清晰而准确的量化描述，并且很难在不同证券或不同市场间进行比较和分析。直到最近几年，Easley，Kiefer，O'Hara 和 Paperman(1996)，Handa，Schwartz 和 Tiwari(1999)，Nyholm(2000)以及 Ma，Hsieh 和 Chen(2000)的研究提出了新的理论分析思路，用知情交易概率(Probability of Informed Trading)来描述市场上到底有多少的交易(或投资者)属于知情交易(知情交易者)，从而来直接衡量市场的信息非对称程度。

作为转轨经济过程中发展起来的新兴市场，中国证券市场的发育尚不成熟，上市公司的信息披露不够规范，加上监管手段缺乏及执法力度不够，使得投资者之间存在较严重的信息不对称性。因此，对中国股市知情交易概率的分析具有重要的现实意义：一方面可以使监管层对证券市场中的信息不对称水平有一个总体的认识，然后通过完善交易机制和强化信息披露来降低信息不对称水平；另一方面有助于我们更加深刻地理解信息在价格形成机制中的角色和作用，动态了解知情交易的概率水平，甚至可以基于此监控和追踪内幕交易等违法行为。本文的目的就是通过建构竞价交易制度下知情交易概率的测度模型，对上海股市的知情交易概率进行定量分析，并分析相关的影响因素。

本文结构如下：除本节引言外，第二节为研究文献综述；第三节基于上海股市的竞价交易制度，提出一个测度知情交易概率的新模型；第四节介绍研究样本与数据；第五节运用上海股市的高频交易数据，对上海股市的知情交易概率及其影响因素进行了实证分析；第六节为本文的结论部份。

## 二、对既有研究文献的综述

Bagehot(1971)指出，根据投资者所掌握信息的不同，可以将其分为知情交易者(Informed Trader)和非知情交易者(Uninformed Trader)。其中，知情交易者拥有关于证券真实价值的私有信息；而非知情交易者主要根据公共信息或谣言进行交

易，成为噪音交易者(Noise Trader)，或是仅仅出于流动性需求而进行交易，成为流动性交易者(Liquidity Trader)。沿袭这一思路，近年来金融学术界提出了一系列计量模型，来测度和评估市场的知情交易概率和信息不对称水平。

Easley、Kiefer、O'Hara 和 Paperman(1996) ,以及 Easley、Kiefer 和 O'Hara(1997) 的论文最早对知情交易概率这一问题进行了研究。他们的研究以做市商市场为背景，建立了一个做市商在每一交易日面临的委托到达过程的模型，在此基础上以决策树的方法计算出知情交易者和非知情交易者在面临好消息、坏消息和没有消息时的委托单到达数，然后以知情交易者的期望委托单到达数为分子，所有交易者的期望委托单到达数为分母，求出知情交易概率。由于做市商市场下投资者的委托数与成交笔数等同，因此 Easley 等人定义的知情交易概率实质上是市场总成交笔数中知情交易者的成交笔数所占的比例。Easley 等人的研究在学术界引起了较大反响，随后，Heidle & Huang(2000)、Ma, Hsieh & Chen(2000)等在模型参数、估计方法等方面对这一理论架构进行了修正和拓展

Nyholm(2002)提出了完全不同的知情交易概率测度方法，该方法主要构建在做市商的报价行为和价差关联性的基础上。做市商根据所有可观察到的信息来推断一笔交易是来自知情还是非知情交易者，进而可算出一段时间内的知情交易概率。相对于以前的研究文献，Nyholm(2002)的优点在于运用了成交信息，并且可以估计每一笔交易或者每一段交易时间内的知情交易概率，这样就可以研究知情交易概率的变化规律，也符合不同时间知情交易概率不同的常识。Nyholm(2003)对模型又进行了相应的拓展，增加考虑了交易量这一因素。

Handa, Schwartz & Tiwari(2003)首先研究了委托驱动市场中的知情交易概率，他们也采用了决策树方法，但是面对决策树的不再是做市商，而是非知情交易者。因此，该文中的知情交易概率是交易者中知情交易者的比例，并可根根据买卖价差、委托失衡等信息，通过一定的结构化模型倒推出来。但是，Handa et.al 的模型有着极其严格但又不符合实际的假定，比如假定知情交易者只以市价委托的形式进行交易，而不能采用限价委托。

总体上来说，上述这些研究利用理论模型和计量经济学方法，在一定的框架内可以估计市场中知情交易发生的概率，即利用交易和委托数据度量市场的信息非对称程度，这些模型得到的估计量有更好的经济学含义和现实意义，与信息非

对称程度的直觉意义也更加接近。但是，由于各国(地区)股市的交易机制具有相当大的差异，例如，我国沪深证券市场属于纯粹的订单驱动市场，只有限价订单没有市价订单，不能够直接应用 Nyholm (2002, 2003) 的模型，因此我们借鉴其建立模型的方法，从完全不同的假设推导出能够适用在竞价市场的知情交易概率测度模型，并且在参数估计上采用的数学方法也不相同。

### 三、竞价市场中测量知情交易概率的模型

证券市场上的交易者可以区分为两类：一类是知情交易者(Informed Trader)，是指获得了关于证券真实价值的私有信息，同时利用这些私有信息进行交易的投资者；另一类是非知情交易者(Uninformed Trader)，这类交易者在交易时可能处于流动性需求而进行交易，成为流动性交易者(Liquidity Trader)；或者是根据公共信息或谣言进行交易，成为噪音交易者(Noise Trader)。

定义某股票在时点  $t$  时的交易量为  $Q_t$ ，该交易量由知情交易者的交易量  $Q^i$  和非知情交易者的交易量  $Q^u$  所组成。根据 Kyle (1985) 在订单驱动连续竞价市场模式下定义的递归线性均衡 (Recursive Linear Equilibrium)，交易价格与成交量存在线性关系，即

$$\Delta P_t = \lambda(Q^i + Q^u) + \varepsilon_t \quad (1)$$

式中的  $\lambda$  反映了交易量对市场价格的影响，为市场深度的倒数。市场深度定义为市场价格变动一个单位所需要的交易量，如果市场需要很大的交易量才能够使市场价格波动一个单位，则说明市场深度大。 $\Delta P_t$  代表第  $t$  笔交易引起的价格变化，为第  $t$  笔交易价格  $P_t$  与第  $t-1$  笔交易价格  $P_{t-1}$  之差。 $\varepsilon_t$  为市场中其他因素引起的价格变化，比如公开信息的到达使得市场各方同步调整价格，假设  $\varepsilon_t$  服从白噪声过程。

为了使价格变化与交易量具有对应关系，我们定义一个虚拟变量  $D_t$ ，当交

易是由买方主动引起时， $D_t$  等于 1；而当交易由卖方主动引起时， $D_t$  等于 -1。

将(1)式改写为：

$$\Delta P_t = \lambda Q_t D_t + \varepsilon_t \quad (2)$$

为了衡量知情交易对股票价格及交易量的影响，我们引入一个指标变量  $I_t$ ，当交易量由知情交易者引起时， $I_t = 1$ ；当交易量由不知情交易者引起时， $I_t = 0$ 。

这样，可以将(2)式表示为：

$$\Delta P_t = (\alpha_0 + \alpha_1 I_t) Q_t D_t + \varepsilon_t \quad (3)$$

上式中， $\alpha_0$  为非知情交易对价格变化的影响， $\alpha_1$  为知情交易相对于不知情交易对价格变化的额外影响

$I_T$  又被称作体制 (Regime) 变量，存在  $\{0,1\}$  两个状态，取值为 1 的无条件概率记作  $p$ ，则取值为 0 的无条件概率为  $1-p$ 。因为是否为知情交易由交易者拥有的信息的自然状态决定，所以我们假定其分布为跨期的 i.i.d.。则可得  $\Delta P_T$  的无条件密度

$$f(\Delta P_T) = \frac{(1-p)}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(\Delta P_T - \alpha_0 Q_T)^2}{2\sigma^2}\right\} + \frac{p}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{[\Delta P_T - (\alpha_0 + \alpha_1) Q_T]^2}{2\sigma^2}\right\} \quad (4)$$

若已知  $\Delta P_T$ ，则可以形成  $I_T = 1$  的条件概率判断，即交易为知情交易的概率判断。由条件概率的定义，可以得出

$$\Pr(I_T = 1 | \Delta P_T) = \frac{\Pr(\Delta P_T, I_T = 1)}{f(\Delta P_T)} = \frac{p \times f(\Delta P_T | I_T = 1)}{f(\Delta P_T)} \quad (5)$$

根据 (7) 式就可以完成知情交易概率的测度问题。但是在进行测度之前，还必须解决未知参数  $\alpha_0, \alpha_1, \sigma^2, p$  的估计问题。令未知参数组成向量

$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \sigma^2, p)'$  , 假设有  $N$  个时间间隔的观察数据  $\Delta P_T$  和  $Q_T$  , 则对数似然函数为

$$L(\boldsymbol{\theta}) = \sum_{T=1}^N \log \{f(\Delta P_T; \boldsymbol{\theta})\} \quad (6)$$

该似然函数无法求出解析解, 只能采用数值算法, 我们这里提供一种高效率的算法。取对数似然函数  $L(\boldsymbol{\theta})$  的导数并令其等于零, 以求解  $\hat{\boldsymbol{\theta}}$  使  $L(\boldsymbol{\theta})$  达到最大值,  $\hat{\boldsymbol{\theta}}$  便是  $\boldsymbol{\theta}$  的最优估计。对数似然的导数为

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{T=1}^N \left\{ \frac{1}{f(\Delta P_T)} \times \frac{\partial f(\Delta P_T)}{\partial \boldsymbol{\theta}} \right\} \quad (7)$$

根据 (6) 式我们可以求出  $\frac{\partial f(\Delta P_T)}{\partial \boldsymbol{\theta}}$  , 得

$$\begin{aligned} \frac{\partial f(\Delta P_T)}{\partial \alpha_0} &= \frac{(\Delta P_T - \alpha_0 Q_T) Q_T}{\sigma^2} \Pr(\Delta P_T, I_T = 0) \\ &+ \frac{[\Delta P_T - (\alpha_0 + \alpha_1) Q_T] Q_T}{\sigma^2} \Pr(\Delta P_T, I_T = 1) \end{aligned} \quad (8)$$

$$\frac{\partial f(\Delta P_T)}{\partial \alpha_1} = \frac{[\Delta P_T - (\alpha_0 + \alpha_1) Q_T] Q_T}{\sigma^2} \Pr(\Delta P_T, I_T = 1) \quad (9)$$

$$\begin{aligned} \frac{\partial f(\Delta P_T)}{\partial \sigma^2} &= \left[ \frac{(\Delta P_T - \alpha_0 Q_T)^2}{2\sigma^4} - \frac{1}{2\sigma^2} \right] \Pr(\Delta P_T, I_T = 0) \\ &+ \left[ \frac{[\Delta P_T - (\alpha_0 + \alpha_1) Q_T]^2}{2\sigma^4} - \frac{1}{2\sigma^2} \right] \Pr(\Delta P_T, I_T = 1) \end{aligned} \quad (10)$$

$$\frac{\partial f(\Delta P_T)}{\partial p} = \Pr(\Delta P_T | I_T = 1) - \Pr(\Delta P_T | I_T = 0) \quad (11)$$

将 (10)(11)(12)(13) 式分别代入 (9) 式, 并令其等于零, 可得

$$\hat{\alpha}_0 = \frac{\sum_{T=1}^N \{ \Delta P_T Q_T - \hat{\alpha}_1 Q_T^2 \Pr(I_T = 1 | \Delta P_T) \}}{\sum_{T=1}^N Q_T^2} \quad (12)$$

$$\hat{\alpha}_1 = \frac{\sum_{T=1}^N \{ (\Delta P_T Q_T - \hat{\alpha}_0 Q_T^2) \Pr(I_T = 1 | \Delta P_T) \}}{\sum_{T=1}^N \{ Q_T^2 \Pr(I_T = 1 | \Delta P_T) \}} \quad (13)$$

$$\hat{\sigma}^2 = \frac{\sum_{T=1}^N \{ (\Delta P_T - \hat{\alpha}_0 Q_T)^2 + (-2\Delta P_T \hat{\alpha}_1 Q_T + 2\hat{\alpha}_0 \hat{\alpha}_1 Q_T^2 + \hat{\alpha}_1^2 Q_T^2) \Pr(I_T = 1 | \Delta P_T) \}}{N} \quad (14)$$

$$\hat{p} = \frac{\sum_{T=1}^N \{ \Pr(I_T = 1 | \Delta P_T) \}}{N} \quad (15)$$

由  $\theta$  的记作  $\theta^{(0)}$  的任意初始猜测开始，根据 (7) 式计算  $\Pr(I_T = 1 | \Delta P_T)$ ，然后计算式 (14)(15)(16)(17) 的右侧项，则左侧得出一个新的估计  $\theta^{(1)}$ 。用该估计  $\theta^{(1)}$  重新计算  $\Pr(I_T = 1 | \Delta P_T)$  和式 (14)(15)(16)(17) 的右侧项，可得  $\theta^{(2)}$ 。持续这一迭代方式直至  $\theta^{(m+1)}$  与  $\theta^{(m)}$  的差比指定的收敛标准小为止。该迭代算法为 Dempster, Laird & Rubin (1977) 建立的 EM 原理的特例，可以证明该算法的每次迭代都增加了似然函数的数值。

#### 四、研究样本与数据

本文实证研究的数据取自上海证券市场的逐笔申报和成交数据库。申报数据库包含 11 个字段的信息，分别是：申报编号、申报日期、申报时间、证券代码、股东代码、买卖方向、申报价格、申报数量、申报余额、成交标识、申报席位。成交数据库包含 13 个字段的信息，分别是：成交编号、成交日期、申报时间、成交时间、证券代码、股东代码、买卖方向、成交价格、成交数量、本次余额、



成交金额、申报编号、席位代码。

样本区间是 2004 年 4 月至 2004 年 9 月共 117 个交易日。由于样本时间比较长,为防止数据量过于庞大,我们以上证 180 指数成份股票作为研究对象。基于其科学客观的选择方法,通常认为上证 180 指数的成份股票具有很好的代表性,能够代表上海证券市场的概貌和运行状况。另外,需要指出的是,由于集合竞价的交易撮合方式完全不同于连续竞价并且交易稀少,所以本文的研究剔除了集合竞价部分,只选取了连续竞价时段的数据,即上午 9:30 到 11:30 和下午 13 点到 15 点之间的数据。

## 五、实证结果及分析

### 1. 上海股市知情交易概率的描述性分析

为使估计出来的  $\alpha_0$  和  $\alpha_1$  值不至于太小,价格变化  $\Delta P_T$  以分为单位,交易量  $Q_t$  以万元为单位。分别取 1 分钟和 10 分钟作为计算  $\Delta P_T$  和  $Q_t$  的时间间隔,令  $\theta^{(0)} = [0.1, 1, 1, 0.5]$  应用 (7)(13)(14)(15) 式进行迭代,终止标准为  $\theta^{(m+1)}$  与  $\theta^{(m)}$  相比每一项的差异都不超过 0.1%, 我们计算出 180 只股票的参数向量,其均值和标准差列于表 2。由于我们在前面的迭代算法中没有限制参数的取值范围,即  $\alpha_0 \geq 0$ 、 $\alpha_1 \geq 0$ 、 $\sigma^2 > 0$  和  $0 \leq p \leq 1$ , 在个别计算中会出现  $p > 1$  的奇异点,解决方法是不理会该奇异点,取一个不同的初始点进行迭代,便可以找到新的合理的结果。

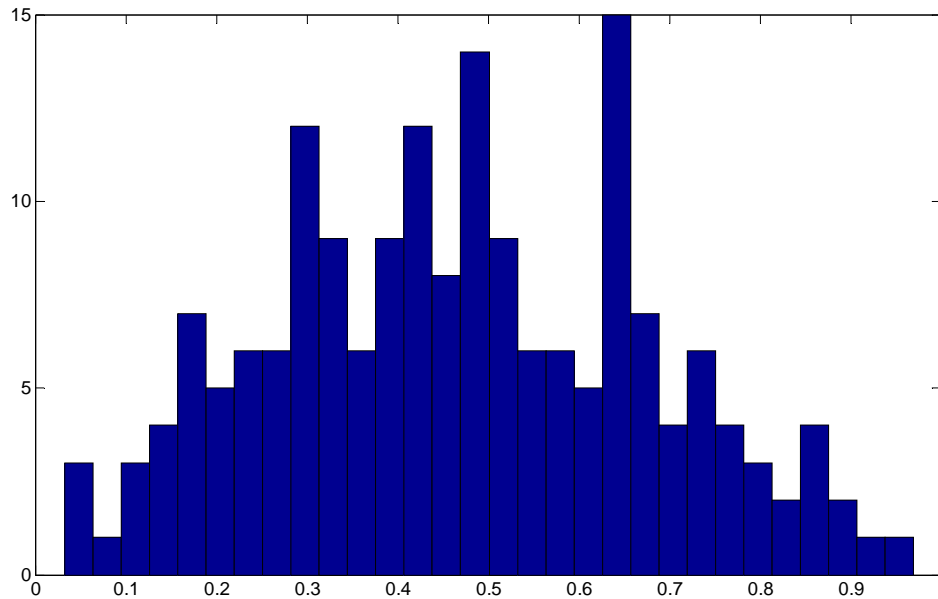
表 1 参数估计结果

		$\alpha_0$	$\alpha_1$	$\sigma^2$	$p$
取 1 分钟为 时间间隔	均值	0.102178	1.298788	1.277962	0.429714
	标准差	0.084928	0.567020	0.614398	0.174131
取 10 分钟为 时间间隔	均值	0.044756	0.174810	7.516469	0.377360
	标准差	0.071867	0.060993	4.860570	0.206596

$\alpha_0$  衡量的是不知情交易对价格的影响幅度，而  $\alpha_1$  衡量的是知情交易对价格的影响幅度。从表 1 我们可以发现， $\alpha_1$  远远大于  $\alpha_0$ ，说明知情交易对价格的影响程度远远大于不知情交易，是知情交易导致了股票价格根本性的大幅度的变化，而不知情交易只对股票价格形成短暂而小的冲击。以 10 分钟为时间间隔算得的  $\alpha_0$  和  $\alpha_1$ ，比以 1 分钟为时间间隔算得的  $\alpha_0$  和  $\alpha_1$  要小，说明在较长的时间段内用无论不知情交易还是知情交易都需要更大的交易金额才能够推动股票价格变化。 $P$  为对知情交易无条件概率的测度结果，取时间间隔为 1 分钟和 10 分钟计算得出的结果基本一致，都接近 40%，但前者大于后者，说明在更长的时间间隔内知情交易容易被掩盖，难于被发现。就其数值来说，我国证券市场的知情交易概率远大于 Easley, Kiefer, O'Hara & Paperman(1996) 和 Nyholm(2002, 2003) 对纽约证券交易所实证分析的结果。这种差异也许可以用我国证券市场的成熟、信息传递不畅等原因来解释，但是由于测度模型和市场微观结构有很大差别，因此不能武断地下此结论。

在 4 个参数中，我们更关注的是知情交易无条件概率  $P$  值。把以 10 分钟为时间间隔算得的 180 只股票  $P$  值的分布示于图 1，会发现其分布接近于正态。事实上，以 1 分钟为时间间隔算得的  $p$  值的分布与图 1 十分相似，为了缩减篇幅，不再列出。应用检验分布是否为正态分布的 Jarque-Bera 检验。两者都在 5% 的置信水平上通过了检验。图 1 中知情交易无条件概率  $p$  值最大的一些股票值得特别关注，因此本研究方法在某种程度上也有助于证券交易监管。

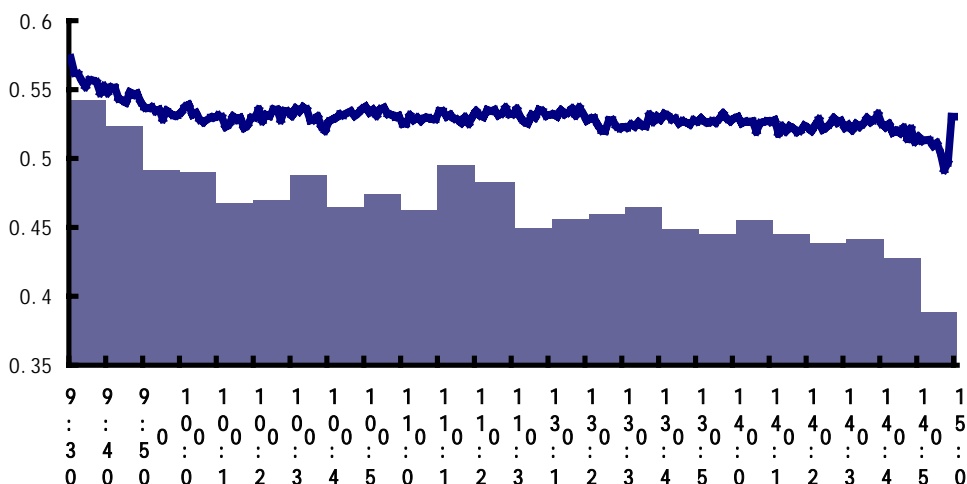
图 1 上证 180 指数成份股知情交易无条件概率  $P$  值分布



## 2. 知情交易概率的时间影响因素

测度出每一个时间间隔的知情交易条件概率  $\Pr(I_T = 1 | \Delta P_T)$ ，取 117 个交易日和 180 只股票的平均值，即得知情交易概率的日内变化型态，如图 2 所示。线状图为以 1 分钟为时间间隔算得的知情交易条件概率；柱状图为 10 分钟的。任何时刻，1 分钟时间间隔的知情交易条件概率都要高于 10 分钟的，说明在长时间间隔下知情交易被掩盖，难于被发现。从图 2 的知情交易概率日内型态我们可以发现，经过一夜的信息累积，投资者之间的信息不均衡拉大，开盘后的知情交易概率水平最高，随着交易的进行，信息逐渐传播，知情交易概率也随之下降。开盘后 40 分钟到收盘前 20 分钟的知情交易概率水平基本稳定，只是中午收盘前的 20 分钟的知情交易概率略有升高。收盘前的 20 分钟知情交易概率水平呈下降趋势。收盘前 2 分钟知情交易概率水平快速升高，原因可能有两方面，一方面收盘临近，知情交易者急于完成交易；另一方面收盘价格在全天成交价格中影响力最大，庄家有操纵的动机。除了收盘前 2 分钟，图 2 绘出的知情交易概率的日内型态与 Nyholm (2002) 根据纽约证券交易所数据得出的结果基本一致，这也说明我国沪深证券市场的收盘价格确定机制的确有变革的必要。

图 2 上证 180 指数成份股知情交易无条件概率  $P$  值日内变化形态



接下来，我们以一般线性模型来检验知情交易概率的日内型态，为防止参数过多，取以 10 分钟为时间间隔算得的  $\Pr(I_T = 1 | \Delta P_T)$  进行回归，模型如下

$$\Pr(I_T = 1 | \Delta P_T) = \alpha_0 + \sum_{i=1}^{24} \alpha_i I_{i,T} + \varepsilon_T \quad (16)$$

式中  $I_i$  为虚拟变量，若处在交易日内的第  $i$  个时段则取值为 1，若处在其余时段则取值为 0。

180 只股票的算术平均  $\alpha_i$  值列于表 3，括号里的数字为  $\alpha_i$  通过 5% 显著性检验的股票个数，从中我们可以发现除了开盘后 20 分钟这一时段大部分股票的知情交易概率特别高之外，其余时段的知情交易概率均不受时间显著影响。

表 2 知情交易概率的日内型态的线性回归结果

$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
0.388311	0.153856	0.135156	0.103211	0.101761	0.0796
(156)	(132)	(63)	(62)	(43)	(46)
$\alpha_7$	$\alpha_8$	$\alpha_9$	$\alpha_{10}$	$\alpha_{11}$	$\alpha_{12}$
0.08105	0.099406	0.076161	0.085744	0.07445	0.1071
(36)	(39)	(40)	(42)	(56)	(73)
$\alpha_{13}$	$\alpha_{14}$	$\alpha_{15}$	$\alpha_{16}$	$\alpha_{17}$	$\alpha_{18}$

0.094506 (57)	0.06145 (29)	0.067672 (36)	0.071422 (34)	0.076389 (45)	0.060289 (41)
$\alpha_{19}$	$\alpha_{20}$	$\alpha_{21}$	$\alpha_{22}$	$\alpha_{23}$	$\alpha_{24}$
0.056906 (33)	0.066717 (30)	0.05675 (36)	0.050416 (28)	0.052882 (22)	0.039289 (21)

需要补充说明的是，除了知情交易概率的日内形态之外，我们还研究了日间形态，比如周效应和月初月末效应等，但是没有发现显著并有意义的特征。

### 3. 知情交易概率的公司影响因素分析

上市公司的特征也会影响股票的知情交易概率。我们以流通市值  $Size$ 、财务杠杆比率  $Leverage$ 、无形资产占总资产比率  $Tangible$ 、主营业务增长率  $Growth$ 、流通股占总股本的比例  $Proportion$ 、前十大流通股东持有市值比例  $Holdings$  这 6 个指标与知情交易无条件概率  $p$  值作多元线性回归，模型如下

$$p_i = \alpha_0 + \alpha_1 Size_i + \alpha_2 Leverage_i + \alpha_3 Tangible_i + \alpha_4 Growth_i + \alpha_5 Proportion_i + \alpha_6 Holdings_i + \varepsilon_i \quad (17)$$

在选取 (19) 式右侧的自变量时，我们遵循了这样一个原则，即自变量不能是信息本身。例如，如果加入每股收益增长率这一自变量，那么其他自变量对于知情交易无条件概率  $p$  来说都无足轻重了，但是这样的回归也不再有意义，因为每股收益增长率其实就是信息本身。在计算自变量的时候也遵循了上述原则，所有的指标都取的是三年平均值，而不是一年的数值。

回归结果列于表 3，除了流通股占总股本的比例这一因素不显著外，其余因素均对知情交易概率有显著影响。

流通市值的大小与知情交易概率负相关，并且在 1% 的水平上显著，与 Easley, Hvidkjaer & O'Hara (2000) 和 Nyholm (2002) 的结果相一致。原因在于大盘股经营比较稳健，信息披露规范，适合长期投资，而小盘股经营波动性比较大，容易受到庄家的操纵。或者也可以用 Chiang & Venkatesh (1988) 的解释，公司规模越小，则内部人越少，内部信息成为公众信息的速度就越慢，私有信息的保留度越高，自然证券市场的知情交易概率就会维持在一个比较高的水平。

财务杠杆比率高的上市公司更加充分地利用了融资工具，但财务成本也更

高，导致其经营业绩有更大幅度的波动，因此也更容易受到知情交易者的光顾，统计上在 5% 的置信水平上显著。

无形资产比有形资产转化为利润的途径更加间接，因此也会造成上市公司的业绩波动较大，引起知情交易概率的增加，但是此项统计只在 10% 的置信水平上显著。

事实上，我们认为财务杠杆比率和无形资产占总资产比率这两个指标与知情交易概率正相关的原因不仅仅在于会引起经营业绩的波动，还有一个原因可能在于利润操纵。我国上市公司进行利润操纵的主要手法便是调整财务费用和无形资产估值，而利润操纵势必引起知情交易概率的大幅度提高。

主营业务增长率是衡量一个公司成长性的最常用也是最主要的指标，它与知情交易概率正相关的结果与 Hedge & McDermott (2000) 的结论相一致。成长型的公司在价值评估上本来就存在很大的不确定性，并且所处竞争环境也是多变的，知情交易者所掌握的信息就相对比较准确，也更有价值，因此知情交易概率水平会比较高。

股权结构也是可能应该知情交易概率的因素，我们用两个指标来分析。一个指标是流通股占总股本的比例，结果是并不显著，这样的结果也好理解，在 161 个样本股票中绝大多数的控股方均为国有非流通股股东，至于流通股比例高一点或低一点并没有什么影响；另一个指标是前十大流通股股东持有市值比例，结果是在 5% 置信水平上正相关，说明股权集中度高会导致高知情交易概率。

表 3 知情交易无条件概率与公司特征影响因素的横截面分析

	$\alpha$	$t$ 值	显著性水平
流通市值 $\times 10^{-8}$	-0.1379	-3.12	<1%
财务杠杆比率	0.0393	2.07	<5%
无形资产占总资产比率	0.0132	1.75	<10%
主营业务增长率	0.1279	2.97	<5%
流通股占总股本的比例	-0.0129	-0.53	不显著
前十大流通股股东持有市值比例	0.1261	2.37	<5%

#### 4. 知情交易概率的市场影响因素分析

知情交易作为发生在证券市场中的交易行为,势必要受到市场中各类因素的影响,反过来知情交易也会影响市场的运行。流动性、波动性和透明度是证券市场的三大主要特征,它们都与知情交易概率有密切关系。知情交易者通常选择流动性好的时刻进入市场,这样他们的行为更容易被掩盖,付出的交易成本也低;反过来,市场上知情交易概率较高时,不知情交易者通常选择不进行交易,以避免损失,从而降低了流动性。因此流动性与知情交易概率可能存在双向因果关系,即流动性影响知情交易概率,知情交易概率同时影响流动性。在波动性高的时刻,知情投资者不愿意进行交易,因为交易价格不确定,并且交易成本较高;反过来,较高的知情交易概率会导致较高的波动性。所以波动性也与知情交易概率可能存在双向因果关系。透明度应该与知情交易概率只存在单向的因果关系,即高透明度导致低知情交易概率。然而,上述关系只是我们的推测,是否真的存在还需要实证分析。

遗憾的是,我们很难直接实证分析知情交易概率与市场流动性、波动性和透明度的关系。这是因为,一方面,流动性、波动性和透明度这三者本身就是一个指标体系,彼此互相影响,比如通常认为高流动性同时伴随着低波动性;另一方面,流动性、波动性和透明度这三者各自也并不存在一个公认的简单可测的指标来衡量。因此,如果我们尝试直接分析知情交易概率与流动性、波动性和透明度这三者之间的关系,不但会篇幅大大增加,超出本文允许的限度,而且会陷入这些指标是否真实准确地衡量了流动性、波动性和透明度的争论,而这些争论并不属于本文的范畴。为了避免这样的问题,我们直接分析知情交易概率与价格变化、交易金额和买卖价差三个指标之间的关系。这三个指标不但简单易测,并且分析它们与知情交易概率之间的关系还可以大致了解流动性、波动性和透明度与知情交易概率的关系。

为明确知情交易概率与价格变化、交易金额和买卖价差之间的关系,特别是因果关系,我们采用时间序列的向量自回归(VAR)方法。但是,这里不能直接采用 $p$ 值做回归,因为它只是在整个样本时间段内知情交易的无条件概率,并非

时间序列，我们采用条件概率  $\Pr(I_T = 1 | \Delta P_T)$  作 VAR 回归。也不能直接采用价格变化  $\Delta P_T$  作回归，因为  $\Delta P_T$  是有正负号的，而  $\Pr(I_T = 1 | \Delta P_T)$  没有正负号，知情交易可能是买入也可能是卖出，所以我们采用价格变化  $\Delta P_T$  的绝对值  $|\Delta P_T|$  作回归。在应用 VAR 方法之前，还需要作知情交易条件概率  $\Pr(I_T = 1 | \Delta P_T)$ 、价格变化绝对值  $|\Delta P_T|$ 、交易金额  $Q_T$  和买卖价差  $Spread_T$  的平稳性检验，否则不能应用 VAR 方法。经过我们对 180 只股票在样本区间的 ADF (Augmented Dickey-Fuller) 平稳性检验，有 172 只股票所有 4 个指标都通过 5% 置信水平的平稳性检验，则取这 172 只股票作为我们后面回归的样本。然后还需要确定 VAR 模型的滞后阶数，为避免滞后阶数过多，我们采用 10 分钟时间间隔，这样我们取一阶滞后就可以了，因为我们很难认为 20 分钟前的交易信息还对当前的决策有很大的影响。由于知情交易条件概率  $\Pr(I_T = 1 | \Delta P_T)$  与价格变化绝对值  $|\Delta P_T|$ 、交易金额  $Q_T$  和买卖价差  $Spread_T$  的 VAR 模型完全一样，我们只列出与交易金额  $Q_T$  的 VAR 模型如下

$$\begin{aligned} \Pr(I_T = 1 | \Delta P_T) &= \beta_0 + \beta_1 \Pr(I_{T-1} = 1 | \Delta P_{T-1}) + \beta_2 Q_T + \beta_3 Q_{T-1} + \varepsilon_T \\ Q_T &= \gamma_0 + \gamma_1 Q_{T-1} + \gamma_2 \Pr(I_T = 1 | \Delta P_T) + \gamma_3 \Pr(I_{T-1} = 1 | \Delta P_{T-1}) + \mu_T \end{aligned} \quad (18)$$

其中  $\beta_0$ 、 $\gamma_0$  为常数项； $\beta_1$ 、 $\gamma_1$  为一阶自回归系数； $\beta_2$ 、 $\gamma_2$ 、 $\beta_3$ 、 $\gamma_3$  为滞后不同阶数的回归系数； $\varepsilon_T$ 、 $\mu_T$  为两式的误差项。

对每只通过平稳性检验的股票估计 VAR 模型的参数。152 只股票各自估计出的参数的算术平均值列于表 4，其中括号里面的数字为通过 5% 显著性水平的股票数目。绝大多数股票通过了知情交易条件概率  $\Pr(I_T = 1 | \Delta P_T)$  与价格变化绝对值  $|\Delta P_T|$ 、交易金额  $Q_T$  和买卖价差  $Spread_T$  的双向 Granger 因果关系检验，显著性水平为 5%，说明知情交易概率是动态变化的，不但受市场其他因素的影响，反过来也影响市场的其他因素。 $T$  期知情交易概率  $\Pr(I_T = 1 | \Delta P_T)$  与  $T-1$  期知情交易概率  $\Pr(I_{T-1} = 1 | \Delta P_{T-1})$  正相关说明知情交易概率有聚集现象，通常当期较高的知情交易概率也预示着下一期的高知情交易概率。 $T$  期知情交易概率



$\Pr(I_T = 1|\Delta P_T)$  与  $T-1$  期的价格变化  $|\Delta P_{T-1}|$ 、交易金额  $Q_{T-1}$  和买卖价差  $Spread_{T-1}$  负相关,这说明知情交易者通常选择价格变化小、交易金额低和价差窄的时机入场,以实现成交价格确定、成交不引起市场注意和交易成本低的目的。而  $T$  期知情交易概率  $\Pr(I_T = 1|\Delta P_T)$  与  $T$  期的价格变化绝对值  $|\Delta P_T|$ 、交易金额  $Q_T$  和买卖价差  $Spread_T$  正相关,这又说明知情交易者的入场导致了价格变化增大、交易金额增加和价差变宽。 $T-1$  期知情交易概率  $\Pr(I_{T-1} = 1|\Delta P_{T-1})$  与  $T$  期的价格变化绝对值  $|\Delta P_T|$  和交易金额  $Q_T$  负相关,而与买卖价差  $Spread_T$  正相关,说明不知情交易者意识到知情交易的存在后,为防止信息上的劣势导致损失从而降低了交易意愿,使得价格变化减小、交易金额减少和价差变宽。进一步引申下去,我们可以得出这样的结论,即高流动性和低波动性吸引知情交易者进场交易,反过来知情交易者的进场交易导致流动性降低和波动性增高。

表 4 VAR 模型的实证结果

应变量	$\Pr(I_T = 1 \Delta P_T)$	$\Pr(I_{T-1} = 1 \Delta P_{T-1})$	$ \Delta P_T $	$ \Delta P_{T-1} $
$\Pr(I_T = 1 \Delta P_T)$	/	0.172692 (147)	0.012298 (139)	-0.019754 (128)
$ \Delta P_T $	0.45951 (132)	-0.032153 (127)	/	0.329783 (151)
159 只股票通过了 $\Pr(I_T = 1 \Delta P_T)$ 与 $ \Delta P_T $ 双向 Granger 因果关系检验				
应变量	$\Pr(I_T = 1 \Delta P_T)$	$\Pr(I_{T-1} = 1 \Delta P_{T-1})$	$Q_T$	$Q_{T-1}$
$\Pr(I_T = 1 \Delta P_T)$	/	0.157632 (147)	0.037258 (132)	-0.019754 (127)
$Q_T$	1.47653 (145)	-0.143753 (127)	/	0.396831 (151)
148 只股票通过了 $\Pr(I_T = 1 \Delta P_T)$ 与 $Q_T$ 双向 Granger 因果关系检验				
应变量	$\Pr(I_T = 1 \Delta P_T)$	$\Pr(I_{T-1} = 1 \Delta P_{T-1})$	$Spread_T$	$Spread_{T-1}$
$\Pr(I_T = 1 \Delta P_T)$	/	0.187213 (147)	0.019274 (137)	-0.010654 (118)
$Spread_T$	0.85643 (135)	0.095851 (119)	/	0.135731 (143)

## 六、结论

本文建立了知情交易概率的测度模型，并将之运用于上海股市的实证研究，得到了如下一些研究结论：

(1) 知情交易对价格的影响程度远远大于不知情交易，是知情交易导致了股票价格根本性的大幅度的变化，而不知情交易只对股票价格形成短暂而小的冲击。

(2) 上海股市知情交易的无条件概率接近 40%，远大于 Easley, Kiefer, O'Hara & Paperman (1996) 和 Nyholm (2002, 2003) 对纽约证券交易所实证分析的结果。

(3) 知情交易概率的日内型态具有一些特点：开盘后的知情交易概率水平最高，随着交易的进行，信息逐渐传播，知情交易概率也随之下降，但在收盘前又有所增加。

(4) 公司的基本面因素会影响股票的知情交易概率，其中，流通市值与知情交易概率负相关；而财务杠杆比率、主营业务增长率以及股权集中度与知情交易概率为正相关关系。

(5) 高流动性和低波动性吸引知情交易者进场交易，反过来知情交易者的进场交易导致流动性降低和波动性增高。

上述结论在理论和实践两个层面都具有重要价值。从理论层面来看，本文提出的知情交易测度模型可以检验市场微观结构理论模型的有效性，以及验证行为金融中的相关模型和结论。从实践层面来看，对知情交易的研究将有利于充分理解订单驱动市场的信息传递机制，可以作为跨市场跨机制比较知情交易风险高低的基础指标，有利于证券交易所及市场监管部门为提高市场质量做出合理的政策选择。