

学校编码: 10384  
学号: 15620131152104

分类号\_\_密级\_\_  
UDC\_\_

廈門大學

碩 士 学 位 论 文

基于聚类估计方法下的知情交易 DY-PIN 模型及相关  
研究

Empirical Research of Informed Trading DY-PIN Model Based on  
Hierarchical Clustering Method

朱振宇

指导教师姓名: 郑振龙教授  
专 业 名 称: 金融工程  
论文提交日期: 2016 年 4 月  
论文答辩时间: 2016 年 4 月  
学位授予日期: 2016 年 6 月

答辩委员会主席: \_\_  
评阅人: \_\_

2016 年 4 月

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

( ) 1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

( ) 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

## 摘要

知情交易的研究是金融市场微观理论的一个重要组成部分。在证券交易当中，知情交易体现为一种信息不对称行为，从风险管理角度上看，度量知情交易被看作是对信息风险的度量。在以往的文献中，学者通过构建符合市场交易特征的微观理论模型来从交易数据中挖掘和度量知情交易。信息风险度量的意义在于提供了观测金融市场风险的新信息，这些新信息有利于提前发现市场的整体风险状况和维护金融市场的合理运行，信息风险的度量同时为风险监管带来了新的要求，促进金融市场制度设计的完善。对于投资者来说，度量信息风险能够促进其合理规划自身的投资组合和分散信息风险、并做出理性的投资决策。

本文在文献综述中对以往文献中的知情交易模型进行了回顾和分析比较，结合中国的数据特征，选取了 DY-PIN 参数模型作为本文的实证模型。在正文部分，本文针对 DY-PIN 参数模型的传统似然估计方法进行了创新，提出了聚类方法来估计 DY-PIN 模型（称为 DY-HAC-PIN 方法）。为了比较两种方法，本文借鉴了以往文献中的模拟实验法，从模型估计准确性和收敛速度两个角度来比较 DY-HAC-PIN 方法和似然估计方法，结论显示 DY-HAC-PIN 方法的估计结果在两个方面均显著优于似然估计方法。为了突破聚类统计方法固有的局限性，本文进一步提出把 DY-HAC-PIN 方法同似然估计相结合，提出了因子 DY-HAC-PIN 方法，这种方法与 DY-HAC-PIN 方法的实验比较中显示因子 DY-HAC-PIN 方法是更好的估计方法。

本文的实证分析分为两个部分，第一部分本文把因子 DY-HAC-PIN 方法应用于中国股票数据中，对沪深 300 指数成分股进行了特征分类，实证结果表明换手率越低、市值越低的股票的知情交易概率越大。第二个部分检验了 PIN 模型和 DY-PIN 模型是否拟合了对实际数据的内在统计属性，包括实际数据的买卖相关性、累积分布等统计属性，实证结果显示 DY-PIN 模型对实际数据的拟合效果优于 PIN 模型，但两者拟合效果都受限于理论模型的分布假设。

**关键字：**知情交易概率；DY-PIN 模型；因子 DY-HAC-PIN 方法

## Abstract

Informed trading is an important part of Micro Market ' s Theory .In security trading ' s activity ,informed trading commonly could be view as information asymmetry ,and in the risk management way ,measuring informed trading could be seen as measuring information risk .In the past related literature ,plenty of scholars measured informed trading probability from real data by constructing Micro ' s theory model .The meaning of measuring information risk and informed trading lies in providing a new angel to observe the financial market .Measuring information risk also be beneficial to early discover the risk level of financial market and maintain reasonable operation of financial market .For common investors ,measuring their portfolio ' s information risk would be beneficial to diversify their portfolio ' s information risk .

This paper reviews the past ' s informed trading model ,and makes comparisons among these models .Based on the character of China ' s financial data ,this paper adopts the DY-PIN parametric model as this paper ' s empirical model .This paper carries an innovation of DY-PIN model ' s estimated method and take the hierarchical agglomerative clustering (HAC) method to estimate the DY-PIN model (DY-HAC-PIN Method).In the designed experiment ,DY-HAC-PIN method behaves better than traditional likelihood estimated method .In consideration of the HAC method ' s defect ,this paper combine the likelihood method and HAC method as an integrated estimated method ,called as Factor DY-HAC-PIN Method .Factor DY-HAC-PIN method behaves best among mentioned methods in the designed experiment .

At last ,this paper ' s empirical analysis consists of two parts .First part of empirical analysis consists of estimating China ' s financial data .Empirical results show that the probability of informed trading exists a negative correlation with the turnover ratio of the equity security .The other part of empirical analysis examines whether estimated EHO-PIN model and DY-PIN model can reflect the statistical properties of real data .The result shows the fitting effect of DY-PIN model is better than EHO-PIN model .

**Keywords:** Probability of informed trading; DY-PIN; Factor DY-HAC-PIN method.

# 目录

<b>第一章 导论</b> .....	<b>1</b>
1.1 选题背景.....	1
1.2 研究内容.....	2
1.3 本文的主要创新和贡献 .....	3
1.4 本文结构.....	4
<b>第二章 文献综述</b> .....	<b>6</b>
2.1 知情交易模型的估计方法改进 .....	6
2.2 知情交易模型的演进 .....	7
2.3 知情交易模型的实际应用.....	9
<b>第三章 模型与理论基础</b> .....	<b>13</b>
3.1 PIN 模型及其估计方法.....	13
3.2 DY-PIN 模型及其估计方法 .....	20
3.3 因子 DY-HAC-PIN 估计方法及实验比较.....	28
<b>第四章 实证研究</b> .....	<b>35</b>
4.1 数据来源及统计特征 .....	35
4.2 实证检验及估计结果 .....	36
<b>第五章 拟合检验：设计和实证</b> .....	<b>43</b>
5.1 数据统计特征的模型拟合检验 .....	45
5.2 买卖相关性的模型拟合检验 .....	51
<b>第六章 结论与展望</b> .....	<b>55</b>
6.1 主要研究结论 .....	55
6.2 后续的研究展望.....	56
<b>参考文献</b> .....	<b>57</b>
<b>致谢</b> .....	<b>59</b>

## Contents

<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Background.....	1
1.2 Focus of the Research.....	2
1.3 Contributions.....	3
1.4 Framework of the Research.....	4
<b>Chapter 2 Literature Review</b> .....	<b>6</b>
2.1 Estimated methods of informed trading model.....	6
2.2 Evolution of informed trading model.....	7
2.3 Application of informed trading model.....	9
<b>Chapter 3 Models and theory hypotheses</b> .....	<b>13</b>
3.1 PIN model and estimation.....	13
3.2 DY-PIN model and estimation.....	20
3.3 factor DY-HAC-PIN estimation and experimental comparison..	28
<b>Chapter 4 Empirical Research</b> .....	<b>35</b>
4.1 Data and Description.....	35
4.2 Empirical result.....	36
<b>Chapter 5 Fitting test</b> .....	<b>43</b>
5.1 Fitting test of statistics.....	45
5.2 Fitting test of correlation.....	51
<b>Chapter 6 Summary and Conclusion</b> .....	<b>55</b>
6.1 Research Conclusion.....	55
6.2 Future Research.....	56
<b>References</b> .....	<b>57</b>

# 第一章 导论

## 1.1 选题背景

金融学研究在过去的 60 年内不断地演进和发展, 研究主题经历了资产组合理论 (CAPM)、资产定价理论 (APT)、有效市场理论 (EMH)、衍生品及期权定价 (BSM)、行为金融学、市场微观理论等各个历史热门领域。现在学界的关注热点集中在市场微观理论, 市场微观理论的主要研究内容包括市场结构和设计、价格信息和价格发现、交易成本和时间成本、信息与披露、做市商行为和投资者行为等。市场微观理论的成熟和发展, 有利于促进金融市场制度和设计的完善和进步, 维护好金融市场微观层面的生态平衡, 从而充分发挥金融市场对于资源的优化配置作用, 充分发挥金融市场对于实体经济促进的功能。对于参与金融市场的投资者来说, 市场微观理论的发展有利于投资者能够更理智地看待金融市场, 正确地认识金融市场发展的规律, 对于自己的投资行为能够更加科学地规划。

本文的研究重心集中于市场微观理论中的价格信息和价格发现领域, 价格信息和价格发现主题的主要研究重点不仅在于私有信息是如何传导给金融市场、是否存在知情交易以及知情交易是如何影响金融资产价格和交易量, 还包括资产价格波动率、交易量变化的日内变化模式等。在证券交易中, 知情交易是一种信息不对称的模式, 学者希望通过构建知情交易模型来从实际数据中度量知情交易的存在概率, 这种知情交易的度量也被看作是对信息风险的度量, 信息风险的度量对于金融市场的监管者来说, 有利其能够发现市场的整体风险状况和使制度设计能够合理地监管信息风险, 对于投资者来说能够正确构造自身的投资组合和分散信息风险。

本文的研究目标在于改进知情交易的度量模型和估计方法, 回答何种模型和何种估计方法才能更准确的度量知情交易概率。现有的研究中主要存在两方面的局限性, 一方面是虽然已经有很多文献对知情交易模型的估计方法进行了研究, 但却始终在最大似然估计的框架下改进, 很少有文献突破这个框架来对估计方法进行创新; 另一方面是知情交易模型目前最主要的有 EHO-PIN 模型和 DY-PIN 模型, 虽然学者还在不断地补充和优化, 但却很少有文献研究这些模型的估计结果



是否反映了真实数据中的内在属性,比如真实数据的分布特征和相关性是否在估计结果中得到了反映。

因此,本文的研究目的就在于补充现有知情交易研究中存在的局限和不足,在现有的研究基础上,系统性描述现有的模型是如何对知情交易建模,并且创新地通过机器学习领域中的算法来改进估计方法,并经过模拟实验和实证研究来验证新的估计方法的优良之处,希望能够回答(1)什么样的知情交易模型能够更好的刻画知情交易和度量知情交易概率?(2)用什么样的估计方法来估计该模型是偏差最小、估计效果最好的?(3)传统的似然估计方法的局限性有哪些?是否有新的估计方法能够克服这些局限性,得到更稳定的估计结果?(4)现有的知情交易模型是否能够拟合市场真实买卖数据的内在统计特征和相关性?(5)现有的度量知情交易的解决办法能否匹配于中国证券市场的数据特征?

## 1.2 研究内容

本文所有的研究内容是基于 PIN (probability of informed trading) 模型的理论改进和估计方法改进。针对本文的研究目的,本文在知情交易概率度量的模型上选择了 DY-PIN 模型 (Jefferson Duarte and Lance Young, 2007), 与最先提出来的 PIN 模型 (Easley, Kiefer, O'Hara and Paperman, 1996) 相比, DY-PIN 模型考虑到了股票真实交易数据中买卖正相关的现象,这种现象很大程度是由于投资者分歧引起的,所以 DY-PIN 模型在原有的 PIN 模型的框架下,加入了系统流动性冲击因子来更好地刻画了交易数据的内在现象。在估计方法的选择上,现有的文献对于 DY-PIN 模型局限于最大似然估计方法,然而这种方法存在着两方面的缺陷,包括对于数据数值比较大的估计容易出现浮点数溢出的问题以及初始值选择不当引起的局部解问题,两方面的缺陷都将导致 DY-PIN 模型的估计准确性下降,所以本文借鉴了 QUAN, David and Wang (2015) 应用于 PIN 模型估计的 HAC (Hierarchical agglomerative clustering) 方法,创新性地引入 HAC 方法来对 DY-PIN 模型进行估计,并且针对 DY-PIN 模型的特点对 HAC 方法进行的重新设计和改进,本文把这种估计方法称为 DY-HAC-PIN 方法。为了检验 DY-HAC-PIN 方法的有效性,本文进行了模拟实验的测试,与现有的似然估计方法相比, DY-HAC-PIN 方法的估计结果更加稳定,估计效率也有很大提高,但 DY-HAC-PIN

方法也存在自身的局限性，所以本文最终把 DY-HAC-PIN 估计方法同似然估计法相结合，提出了因子 DY-HAC-PIN 估计方法，因子 DY-HAC-PIN 方法在模拟实验中表现更加出色，估计的效果超越 DY-HAC-PIN 方法和似然估计方法。

基于因子 DY-HAC-PIN 的估计方法，本文利用沪深 300 指数成分股中市值规模处于中位数的公司股票数据进行了知情交易的实证检验，检验和比较 DY-PIN 模型和 EHO-PIN 模型对于真实数据统计特征的拟合效果，包括真实数据内在的相关性、分布特征(累积函数和密度函数)等。最后，本文运用 2014 年全年的交易数据，对于沪深 300 股票组合中的股票进行市值、换手率、高管持股比例以及融资融券余额的四种分类，运用因子 DY-HAC-PIN 方法估计不同分类特征的股票组合的知情交易概率，并比较不同分类特征下知情交易概率的特征和区别。

本文研究结论为从参数估计结果的稳定性、偏差和效率上看，因子 DY-HAC-PIN 估计方法是目前对于 DY-PIN 模型的最优估计方法。在实证研究中，本文发现对于换手率低、市值较小、大股东持股比例较高、融资融券余额比例较高的股票组合的知情交易概率较高，这与之前知情交易实证文献的结果相似。在对实际数据的内在属性拟合检验中，本文发现 DY-PIN 模型是比 PIN 模型更好的拟合模型，更好地拟合了实际数据的统计特征，但两者都受限于模型理论设定，不能完全地拟合好实际数据的统计特征。

### 1.3 本文的主要创新和贡献

本文的主要创新和贡献如下：

首先，本文的最大的贡献在于创新性地提出了对 DY-PIN 模型估计的聚类估计法，又称为 DY-HAC-PIN 方法，这种估计方法是对原有似然估计法的创新，能够有效地减少了原有似然估计法存在的局部解和浮点数溢出的问题，最后把 DY-HAC-PIN 方法同似然估计法结合，提出了估计效果更好的因子 DY-HAC-PIN 方法。本文从理论和模拟实验两个角度上说明了聚类估计法（包括 DY-HAC-PIN 方法和因子 DY-HAC-PIN 方法）的优点和局限性，并且从本文的实证结果上看，因子 DY-HAC-PIN 估计方法非常适合应用于我国换手率高、数据数值高的股票市场环境。本文采用创新的聚类估计方法的思路在于，知情交易 PIN 模型和 DY-PIN 模型 DY-HAC-PIN 估计方法都是把交易日的状态进行划分，如存在好消息的交易

日、坏消息的交易日和没有私有信息的交易日，DY-PIN 模型会在各种私有消息的条件之下进一步分为存在流动性冲击和非流动性冲击两种状态，出现 6 种交易日状态。所以从宏观的角度看，知情交易 PIN 模型和 DY-PIN 模型都是一种分类的方法，在统计学上最常用的分类方法就是聚类方法，所以本文创新地把聚类方法引入 DY-PIN 模型的参数估计中是符合模型理念，从实际效果上看，聚类方法对 DY-PIN 模型的参数估计无论中估计偏差和估计稳定性上都优于似然估计法。

其次，本文通过对沪深 300 指数成分股进行多项因子的分组，包括换手率、市值、高管持股比例和融资融券余额，揭示了知情交易概率在不同分组特征下的规律，验证知情交易概率与这些因子的相关性。

最后，很少有文献研究知情交易模型（EHO-PIN 模型、DY-PIN 模型）是否能够拟合实际数据的统计属性，本文结合了因子 DY-HAC-PIN 估计方法填补了这方面的研究空白。本文用中国股票市场的实际数据验证了 PIN 模型和 DY-PIN 模型是否是对实际数据中的相关性、以及各自分布特征能够有效拟合，结论显示 DY-PIN 模型对于实际数据的拟合的确是比 PIN 模型更好，但固有理论假设使得难以完全拟合实际数据的内在属性。

## 1.4 本文结构

本文的行文结构安排如下：

第一章是导论，介绍了本文的写作背景、动机、主要的研究内容和主要创新和贡献、第二章是文献综述，该部分分为三个部分来介绍，第一部分介绍了现有的文献对知情交易概率模型的创新和发展；第二部分介绍现有的文献对知情交易概率模型度量 and 估计方法的改进和创新；第三部分介绍了现有文献对知情交易概率模型的实际应用，主要包括知情交易预测等方向；

第三章是介绍了本文所使用的知情交易概率模型和度量方法，重点介绍了本文进行创新的估计方法，包括加入聚类估计方法的 DY-HAC-PIN 估计方法以及结合了聚类估计和似然估计两种方法的因子 DY-HAC-PIN 估计方法。之后本文对两种创新估计方法进行估计有效性的模拟实验，并比较了两种估计方法；

第四章是本文的第一部分实证内容，本文通过对沪深 300 指数成分股的不同特征分类，包括换手率、高管持股比例、市值大小、融资融券月，来检验不同分

组条件下的 PIN 模型和 DY-PIN 模型估计的知情交易概率的变化特征。

第五章是本文第二部分实证,这部分检验和比较了现有的 PIN 模型和 DY-PIN 模型对实际数据中的内在分布特征的拟合效果以及用经验 Copula 形式来比较两种模型对实际数据中相依关系的拟合。

第六章是本文的结论和展望部分,总结了现有的结论和对后续研究的展望。

厦门大学博硕士论文摘要库

## 第二章 文献综述

市场微观结构理论作为金融理论研究的前沿，这套理论主要研究资产价格在特定规则下的变化过程，分为市场结构和设计、价格信息和价格发现、交易和时间成本、做市商和投资者行为等主题。其中，价格信息和价格发现这个主题的主要研究内容在于金融信息是如何传导给金融市场，是否存在知情交易以及如何影响金融资产价格和交易量。关于金融市场是否存在知情交易，最早是由 Kyle(1985) 研究显示知情交易者存在着私有信息，并且知情交易者采用一种最佳的订单策略来进行知情交易来获利。关于知情交易的文献主要集中在是否存在知情交易、知情交易的发生方式和知情交易模型的度量及应用（PIN 模型及其优化模型）三个方面。与本文相关的文献可以分为三个方向：知情交易模型的估计方法改进、知情交易模型的演进、知情交易模型的应用。

### 2.1 知情交易模型的估计方法改进

首先，在 PIN 的度量估计方面，由于 PIN 的估计是通过对实际数据求最大似然函数来校准未知参数，而且经过 Easley 和 O' Hara 的实证研究，需要 60 天的 buy 和 sell 数据（其中，buy 为以买方发起的成交量，sell 为以卖方发起的成交量）才能得到性质良好的估计结果。Easley, Hvidkjaer 和 O' Hara (2010) 研究发现如果某只股票的 buy 和 sell 数值过大时，PIN 估计的似然函数值会超过最大浮点数范围，即浮点数溢出的问题，从而使得 PIN 的估计出现很大偏差，所以 Easley, Hvidkjaer 和 O' Hara (2010) 对似然函数进行了对数因子化的改进（后面简称 EHO 估计），提高了估计的有效性。随后，Lin 和 Ke (2011) 研究指出，EHO 的对数因子化方式会导致另一个问题的出现，就是最后 PIN 的估计结果会出现明显的低估，大约 44% 的股票估计 PIN 值会出现下偏风险，而且成交量活跃的股票比不活跃的股票更易出现下偏风险，所以 Lin 和 Ke (2011) 也提出了新的似然函数对数因子化的方式（后面简称 LK-PIN 估计），包含了两种改进原则，第一种是计算  $e^{x+y}$  而不是  $e^x e^y$ ，这种计算方法更能避免浮点数溢出；第二

种改进原则是用  $\log(e^{(x+y)-m} + e^{z-m}) + m$ ,  $m = \max(x+y, x)$ , 替代  $\log(e^{(x+y)} + e^z)$ 。这种新的对数因子化方式进一步提高 PIN 模型参数估计的有效性。其次, 利用似然函数估计 PIN 模型的另一个问题则在于如何选取初始值的问题, 如果选择的初始值不合适, 则会使得最终的估计时间变长以及估计结果有效性的降低, 容易陷入边界解的陷阱。Yan 和 Zhang (2012) 的实证研究发现, 在估计 PIN 模型的过程中, 出现边界估计结果的频率很高, 他们提出了用穷举法搜索的办法择优选取概率参数  $\{\alpha, \delta, \gamma\}$  初始值, 每个概率参数的初始值为  $\{0.1, 0.3, 0.5, 0.7, 0.9\}$  (总共 125 个初始值选择), 并且由于 PIN 模型隐含着 buy 和 sell 的均值为:  $E(B) = \alpha(1 - \delta)\mu + \varepsilon_b$ ,  $E(S) = \alpha\delta\mu + \varepsilon_s$ , 则密度参数  $\{\mu, \varepsilon_s, \varepsilon_b\}$  可以用矩估计的方法来设定初始值, 通过对 125 个初始值进行似然估计, 最后则选取似然值最高的参数作为估计参数, 这种方法能够有效地规避局部解的问题。虽然对数因子化似然函数的办法 (EHO-PIN 估计和 LK-PIN 估计) 和穷举搜索 (LK-PIN 估计) 选取初始值相结合确实改善了估计结果的准确性, 但是由于增加的 125 个初始值使得计算步骤的估计时间变长, 效率仍然不高。综合这些考虑, Quan Gan、Wang Chun Wei 和 David Johnstone (2015) 研究提出了聚类方法来估计 PIN 值, 与现有的估计方法相比, 聚类估计方法的准确性不会下降, 同时估计效率得到了很大的提高, 这种基于聚类估计方法脱离了原有似然估计方法的框架和局限, 直接对实际 buy 和 sell 数据划分为三类 (好消息、坏消息、无消息), 并且在 QWD 设计的模拟实验中表现跟 YZ-LK-PIN 估计的效果几乎相同, 但是 QWD 方法的估计效率有了很大的提高。

## 2.2 知情交易模型的演进

其次, 在 PIN 的理论假设方面, Duarte 和 Young (2007) 研究发现美国 NYSE 交易所中 95% 的股票的每日 buy 和 sell 数据在统计上是呈现正相关, 而 PIN 模型中隐含的 buy 和 sell 数据的关系却是负相关, 基于这个实际现象, 他们拓展了 PIN 模型的理论假设, 使新模型适用于金融市场上 buy 和 sell 统计上正相关的情况, 这个新模型称为 DY-PIN 模型。DY-PIN 模型提出一种系统流动性冲击事件, 当系统流动性冲击事件发生时, 市场中的 buy 和 sell 数量可以同时增加, 这种系统流动性冲击事件出现的原因之一在于某种公开事件的发生会导致投资

者出现分歧,这种分歧会导致 buy 和 sell 数量同时大量增长,事实上这种分歧也是股票交易量的重要来源。DY-PIN 模型不仅改进了 PIN 模型的理论假设,提出改进后的知情交易概率的计算方法(称为 Adj-PIN),同时 DY-PIN 模型也能够计算出市场中系统流动性冲击事件发生的概率 PSOS (probability of a symmetric order flow shock)。在 Duarte 和 Young (2007) 的实证部分进一步显示具有高的系统流动性冲击事件发生概率的公司通常在大多数交易日内具有比较低的交易量,并且在某些交易日会出现交易日的暴增,所以 Duarte 和 Young 也把 PSOS 看作是非流动性的代理变量,实证研究也表明 PSOS 同其他流动性测度高度相关,证明了 PSOS 可以当作非流动性的代理变量。在实证中, PIN 模型的知情交易概率和 Adj-PIN 高度相关(约为 0.67), PIN 模型知情交易概率和 PSOS 值的相关性也很高(约为 0.71),而 Adj-PIN 值与 PSOS 值相关性却相对低(约为 0.22),所以 PIN 模型知情交易概率包含着一部分非流动性因子,该文认为以前 PIN 在资产定价中效应有一部分可以用非流动性因子来解释,而 Adj-PIN 更能反应资产定价中的信息不对称的风险。郑振龙和杨伟(2010)的实证研究发现,中国的股票交易数据表明大部分的股票的每日 buy 和 sell 数量呈正相关,所以使用 DY-PIN 模型估计知情交易概率更适合,在使用 DY-PIN 模型对中国股票信息风险的测量时,这能够较好地匹配实际数据中的正相关关系。该文章选取换手率作为划分股票组合的依据,发现 DY-PIN 值的大小同换手率呈负相关关系,即说明了换手率低的股票更容易出现知情交易,而换手率高的股票则不容易出现知情交易,这种结果与传统的知情交易的实证研究是吻合的。Easley et al (2010a) 提出基于估计高频数据知情交易概率的方法,称之为 VPIN 模型 (Volume Synchronized Probability of informed trading)。这种模型以交易量为计算刻度,代替了 PIN 模型中的交易日时间刻度,在交易量的单位中需找知情交易。VPIN 方法首先计算样本期中所有交易日的平均交易量,设定一个主观的每日交易块数量  $n$ ,从而得到每个交易块的交易量,并用变动价格正态分位数的方法从每个交易块中估计以买方发起交易量大小和以卖方发起交易量大小,并计算两者之差的绝对值作为知情交易的交易量,最终得到知情交易概率。VPIN 方法的优点在于其是非参方法,不存在模型参数估计误差的问题,从理论上可以证明 VPIN 方法是 PIN 模型的新的估计形式。Anthony Tay (2009) 提出了同样基于高频数据的

PIN-AACD 模型, 这种方法在原有的 PIN 模型中加入了计量统计中的非对称自回归间隔模型(缩写为 AACD), 主要依据是知情交易发生后, 股票的成交速度会加快, 交易间隔会变短, 把这些交易间隔数据加入到 PIN 模型中, 使得日内的数据就足够计算出知情交易概率。PIN-AACD 模型假设时间间隔为指数分布, 最后同样使用最大似然的估计方法求得知情交易概率。Sanders (2014) 提出一种动态的日内知情交易概率测度方法, 称为 DPIN 方法。这种估计方法避免了 PIN 模型中对数据长度的要求(至少 60 个交易日)和似然估计误差问题, 它把每个交易日划分为多个 15 分钟的间隔, 通过对 15 分钟时间段的非预期收益的正负情况判断知情交易的发生方向, 正的非预期收益出现时, 以买方发起的成交量被识别为知情交易, 同理负的非预期收益出现时, 以卖方发起的成交量被识别为知情交易。通过实证研究, DPIN 值高的股票具有小市值、低流动性、低成交量的特征, 从日内的表现看, DPIN 值高的时间段往往集中在开盘后和收盘前的一段时间内, 这些结论是符合知情交易的特征的。James Petchey (2015) 在实证研究中对比 PIN-AACD 模型和 PIN 模型在兼并重组事件前的表现, 结果发现 PIN 值在兼并重组事件前出现下降, 而 PIN-AACD 值则在兼并重组事件前出现上升, 说明 PIN-AACD 模型对知情交易的识别效果更好。

## 2.3 知情交易模型的实际应用

在对于 PIN 模型及其衍生模型的实际应用中, 学者的研究首先关注 PIN 值是否能够预测股票的未来收益以及是否能够在资产定价中作为一个信息风险因子, 这种定价效应是否有效在学界中出现了分歧, 有些学者的实证研究支持 PIN 值对股票未来收益有预测作用, 有些学者的实证研究则表明 PIN 值在一些明显具有知情交易事件之前并没有变化和提示预测, 并讨论了 PIN 模型失效的原因, 另外一些学者则对比 PIN 模型和其他衍生模型(PIN-AACD、VPIN 等)识别知情交易的能力。Easley et al (2002) 研究发现 PIN 值对股票收益预测具有显著的决定作用, 有高 PIN 值的股票具有更高的回报率, 该文发现 PIN 值相差 10 个百分点的两只股票, 则年收益相差 2.5%, 同时发现小市值股票的 PIN 值相对于大市值的股票高。随后的一些实证研究却得出了相反的结论, Aktas, et, al. (2007) 通过划分三个事件窗口期 $[-180, -66]$ ,  $[-65, -6]$ ,  $[+3, +63]$ , 第一个窗口作为基准,



Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.