

学校编码: 10384

分类号_____密级_____

学号: 15620060153076

UDC _____

厦 门 大 学

博 士 学 位 论 文

股票信息风险测度研究

A Study on Measuring Information Risk of Stocks

杨 伟

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: 2009年 4月

论文答辩时间: 2009年 月

学位授予日期: 2009年 月

答辩委员会主席: _____

评 阅 人: _____

2009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

内 容 摘 要

对股票信息风险进行准确的测度无论对资产定价、风险管理还是市场绩效的衡量都有着重要意义。如果一国股票市场的信息风险比较高，或者信息不对称的程度比较严重，无疑会极大地损害市场的公平与效率。各国股票市场监管者都力图能降低股票市场的信息风险或信息不对称的程度，以维护市场的公平和效率。然而，各国股市仍然存在着不同程度的信息风险或信息不对称现象。

那么股票信息风险是否是一种系统性风险？如果信息风险是一种系统性风险，那么投资者持有信息风险高的股票就应该要求获得一个比较高的风险溢价。而要对信息风险是否是一种系统性风险进行实证检验，首先需要对信息风险进行准确的测度。Easley, Kiefer, O'Hara and Paperman (1996)^[1]最早提出了直接测度信息风险的 PIN 模型，此后该模型成为测度信息风险的炙手可热的模型。然而，PIN 模型隐含的买卖指令之间的负相关关系与实际数据中买卖指令之间的正相关关系并不相符。此外，PIN 模型中隐含的买卖指令的方差与实际数据中买卖指令相对较大的方差也不能很好地匹配。

本文在 Easley, Kiefer, O' Hara and Paperman (1996)^[1]提出的经典的 PIN 模型基础上，通过增加交易动机，提出了修正的 PIN 模型。本文基于中国股票的逐笔交易数据，利用修正的 PIN 模型对我国股票具有的信息风险进行的实证研究表明，修正的 PIN 模型隐含的买卖指令之间的相关性和买卖指令的方差能够更好地与实际数据相匹配。

本文分别利用经典的 PIN 模型和修正的 PIN 模型对交易活跃程度不同的股票具有的信息风险进行了实证研究，发现股票具有的信息风险大小同交易活跃程度之间呈负相关关系。经典的 PIN 模型由于忽视了市场指令流冲击事件发生时引起的交易动机，倾向于高估股票所具有的信息风险。

本文采用滚动 (rolling) 的方法构造出了股票信息风险的时间序列，并对股票信息风险和股票收益率之间的关系进行了实证研究，发现股票信息风险对股票收益率并不存在显著的影响，而与市场流动性相关的市场指令流冲击事件发生时引起的交易概率对股票收益率有着持续显著的影响。

关键词：信息风险；测度；实证研究

Abstract

Measuring information risk of stocks accurately has important significance on asset pricing, risk management and measuring the market performance. If information risk of stocks is higher, or the extent of information asymmetry is more serious in one country, then the justice and efficiency of market in the country will be damaged. Every country's supervisors of stock market try to reduce information risk or the extent of information asymmetry in stock market in order to maintain the justice and efficiency of market. However, there is still different extent of information risk or information asymmetry in every county's stock market.

Then, is information risk of stocks a systematic risk? If it is, investors who hold stocks with higher information risk should require obtaining higher risk premium. When we want to test whether information risk is a systematic risk empirically, we firstly need to measure information risk accurately. Easley, Kiefer, O'Hara and Paperman (1996) first put forward the PIN model used to directly measure information risk of stocks, hereafter this model became one of the most popular models used to measure information risk. However, the negative correlation between buy and sell orders implied in the PIN model is not accordant with the positive correlation between buy and sell orders in real data. Furthermore, the variance of buy and sell orders implied in the PIN model can't match the larger variance of buy and sell orders in real data either.

This paper put forward a corrected PIN model by adding trading motive based on the classical PIN model put forward by Easley, Kiefer, O'Hara and Paperman (1996). Based on the trading data tick by tick of Chinese stocks, this paper applied the corrected PIN model to measure information risk of stocks empirically, and found that the correlation between buy and sell orders and the variance of buy and sell orders implied in the corrected PIN model can better match the real data.

This paper empirically measured information risk of stocks with different trading activity using the classical PIN model and the corrected PIN model respectively, and found that information risk of stocks is negative with trading activity. The classical PIN model tended to overestimate information risk of stocks due to ignoring the trading

motive aroused by market order flow shock.

This paper constructed the time series of information risk of stocks using the method of rolling, and did an empirical study on the relation between information risk of stocks and stock returns, found that information risk had not significant impact on stock returns, while the probability of market order flow shock which is related to liquidity had persistent and significant impact on stock returns.

Key Words: Information Risk; Measurement; Empirical Study

目 录

导 论.....	1
第 1 章 信息风险测度的文献综述.....	5
§1.1 信息风险的定义.....	5
§1.2 信息风险测度的国外文献综述.....	6
§1.3 信息风险测度的国内文献综述.....	12
第 2 章 信息风险测度的模型.....	15
§2.1 经典的PIN模型.....	15
§2.2 修正的PIN模型.....	22
第 3 章 中国股票市场信息风险测度实证研究.....	30
§3.1 中国股票市场交易机制概述.....	30
§3.2 经典的PIN模型的估计结果.....	32
§3.2 修正的PIN模型的估计结果.....	48
第 4 章 中国股票市场信息风险性质研究.....	63
§ 4.1 中国股票市场信息风险时间序列特征研究.....	63
§ 4.2 中国股票市场信息风险与股票收益率关系研究.....	73
第 5 章 结论及未来进一步的研究方向.....	84
§5.1 本文主要结论.....	84
§5.2 未来进一步的研究方向.....	85
附 录.....	87
参考文献.....	100
后 记.....	106

Contents

Introduction	1
1. Literature Review of Measuring Information Risk	5
1.1 The Definition of Information Risk	5
1.2 International Literature Review of Measuring Information Risk.....	6
1.3 Domestic Literature Review of Measuring Information Risk.....	12
2. The Model of Measuring Information Risk	15
2.1 The Classical PIN Model.....	15
2.2 The Corrected PIN Model.....	22
3. Empirical Study on Measuring Information Risk in Chinese Stock	
Market	30
3.1 The Review of Trading Mechanisms in Chinese Stock Market.....	30
3.2 The Estimation Results of the Classical PIN Model.....	32
3.3 The Estimation Results of the Corrected PIN Model.....	48
4. Study on Character of Information Risk in Chinese Stock Market	63
4.1 Study on Time Series Characteristic of Information Risk in Chinese Stock	
Market	63
4.2 Study on Relation between Information Risk and Stock Returns in Chinese	
Stock Market	73
5. Conclusions and Further Research in Future	84
5.1 The Main Conclusions of This Paper	84
5.2 Further Research in Future	85
Appendix	87
References	100
Acknowledgement	106

图表目录

图 2.1	经典的 PIN 模型中股票在一个交易日内的交易过程	16
图 2.2	修正的 PIN 模型中股票在一个交易日内的交易过程	25
图 3.1	对各组股票换手率的 Kruskal-Wallis 非参数检验	34
图 3.2	经典的 PIN 模型估计得到的参数 Kruskal-Wallis 非参数检验	41
图 3.3	经典的 PIN 模型估计得到的各组股票信息风险的 Kruskal-Wallis 非参数检验	46
图 3.4	修正的 PIN 模型估计得到的参数 Kruskal-Wallis 非参数检验	56
图 3.5	修正的 PIN 模型估计得到的各组股票信息风险和市场指令流冲击事件发生时引起的交易概率的 Kruskal-Wallis 非参数检验	60
图 4.1	经典的 PIN 模型参数估计结果的时间序列	64
图 4.2	经典的 PIN 模型估计出的信息风险时间序列	65
图 4.3	修正的 PIN 模型参数估计结果时间序列	67
图 4.4	修正的 PIN 模型估计出的信息风险时间序列	68
图 4.5	修正的 PIN 模型估计出的市场指令流冲击事件发生时引起的交易概率时间序列	69
表 3.1	各组股票在样本期内的日均换手率描述统计	33
表 3.2	各组股票在样本期内的日均买卖单数量的描述性统计	35
表 3.3	各组股票的买单和卖单数量、方差及其相关性的分位数统计	35
表 3.4	经典的 PIN 模型中各组股票的参数估计值及信息风险的分位数统计	39
表 3.5	经典的 PIN 模型估计得到的参数分布的 Kruskal-Wallis 非参数检验	40
表 3.6	经典的 PIN 模型估计得到的参数分布的 Mann-Whitney 非参数检验	42
表 3.7	经典的 PIN 模型估计得到的各组股票信息风险的描述性统计	43
表 3.8	经典的 PIN 模型估计得到的信息风险 Kruskal-Wallis 非参数检验	45
表 3.9	经典的 PIN 模型估计得到的信息风险的 Mann-Whitney 非参数检验	46
表 3.10	经典的 PIN 模型隐含的买卖单、方差及买卖单之间的相关性在不同分位数下的统计	47

表 3.11	修正的 PIN 模型与经典的 PIN 模型的似然比检验	50
表 3.12	修正的 PIN 模型中各组股票的参数估计值及信息风险的分位数统计	51
表 3.13	修正的 PIN 模型估计得到的参数分布的 Kruskal-Wallis 非参数检验 ..	53
表 3.14	修正的 PIN 模型估计得到的参数分布的 Mann-Whitney 非参数检验 ..	56
表 3.15	经典的 PIN 模型和修正的 PIN 模型估计得到的各组股票的信息风险比较	58
表 3.16	修正的 PIN 模型估计得到的信息风险 Kruskal-Wallis 非参数检验	59
表 3.17	修正的 PIN 模型估计得到的信息风险的 Mann-Whitney 非参数检验 ..	60
表 3.18	修正的 PIN 模型中隐含的各组股票的买卖单均值、方差及它们之间的相关性	61
表 4.1	钢铁行业股票由经典的 PIN 模型估计得到的参数描述性统计	66
表 4.2	钢铁行业股票由修正的 PIN 模型估计出来的参数描述性统计	69
表 4.3	经典的 PIN 模型估计得到的股票信息风险时间序列的序列相关检验 ..	71
表 4.4	修正的 PIN 模型估计得到的股票信息风险时间序列的序列相关检验 ..	72
表 4.5	修正的 PIN 模型估计得到的市场指令流冲击事件发生时引起的交易概率时间序列的序列相关检验	73
表 4.6	指数和股票开盘收盘收益率描述性统计	75
表 4.7	信息风险和市场指令流冲击事件发生时引起的交易概率描述性统计 ..	75
表 4.8	PIN 对股票收益率的影响回归结果	76
表 4.9	NPIN 对股票收益率的影响回归结果	77
表 4.10	PMOS 对股票收益率的影响回归结果	78
表 4.11	PIN 和 PMOS 对股票收益率的影响回归结果	79
表 4.12	PIN、NPIN、PMOS 和 NSO 之间的相关性	80
表 4.13	NSO 对股票收益率的影响回归结果	81
表 4.14	PIN 和 NSO 对股票收益率的影响回归结果	81
表 4.15	PIN、NPIN、PMOS 和换手率之间的相关性	83

导 论

一、研究背景

信息风险或信息不对称问题一直都是各国股票市场微观结构中备受关注的问題之一。当市场上存在信息风险或信息不对称的情形时——即一部分交易者相对于另外的交易者拥有与资产真实价值相关的私有信息时，这部分知情交易者利用其掌握的私有信息进行知情交易便可以从中获利^①。如果一国股票市场的信息风险比较高，或者信息不对称的程度比较严重，无疑会极大地损害市场的公平与效率。各国股票市场监管者都力图能降低股票市场的信息风险或信息不对称的程度，以维护市场的公平和效率。然而，各国股市仍然存在着不同程度的信息风险或信息不对称现象。

我国股票市场由于长期以来的政治和经济因素使得信息不对称现象比其他国家更为严重，股票市场具有的信息风险也更高。根据深圳证券交易所《2007年证券市场主体违法违规情况报告》，我国证券市场“2007年被查处的内幕交易与市场操纵有9宗，其中上市公司高管内幕交易5宗，证券公司内幕交易2宗，其他主体市场操纵2宗。”同时，该报告还指出，由于法律规则和证据制度的制约，我国证券市场上还存在大量未被查处或虽被查处但是以其他违法类型定性，诸如“信息披露违规”定性的内幕交易和市场操纵行为。我国证券市场上内幕交易的盛行，极大地损害了我国证券市场的公平与效率。

而要能较好地对内幕交易进行查处，首先需要能对股票具有的信息风险或信息不对称的程度进行测度。由于知情交易不能从市场上直接观测到，早期的文献主要从间接的角度、采用一些知情交易的替代变量来衡量市场中的信息风险或信息不对称程度，如 Bagehot (1971)^[2]和 Jaffe and Winkler (1976)^[6]等将买卖价差作为信息不对称的简单测度。但是这些方法都不能明确地测度交易者在市场中面临的信息不对称状况，而且这些结果都没有标准化，让人无法对不同市场上的信息风

^① 从 Bagehot (1971)^[2]开始，Copeland and Galai (1983)^[3]、Glosten and Milgrom (1985)^[4]以及 Easley and O'Hara (1987)^[5]等提出证券市场的参与者可以区分为知情交易者 (informed traders) 和非知情交易者 (uninformed traders) 两类。所谓知情交易者指的是拥有与资产真实价值相关的私有信息的交易者；与此相对应，市场上除了知情交易者之外的交易者便是非知情交易者，这一类的交易者在交易时不是没有依据与资产真实价值相关的私有信息就是所依据的信息为公开信息或是谣言。

险或信息不对称的严重程度进行直接比较。Easley, Kiefer, O'Hara and Paperman (1996)^[1]最早提出了直接测度信息风险的模型,此后, Easley, Kiefer and O'Hara (1997a)^[7]、Easley, Kiefer and O'Hara (1997b)^[8]、Easley, O'Hara and Paperman (1998)^[9]、Brockman and Chung (2000)^[10]、Easley, O'Hara and Saar (2001)^[11]、Easley, Hvidkjaer and O'Hara (2002)^[12]、Easley and O'Hara (2004)^[13]、Vega (2006)^[14]、Boehmer, Grammig and Theissen (2007)^[15]等大量学者以该模型为基础,对各个证券市场的信息风险进行了估计和应用。

虽然Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的PIN模型目前已成为对证券具有的信息风险进行测度的炙手可热的模型,但是也有不同学者对其提出了一些质疑。Venter and De Jongh (2006)^[16]从统计的角度发现PIN模型隐含的买卖指令之间的负相关关系与实际数据中买卖指令之间的正相关关系并不相符。Duarte and Young (2008)^[17]的实证研究也证实PIN模型不能满足实际数据中买卖指令之间的正相关关系,此外, PIN模型也不能匹配实际数据中买卖指令相对较大的方差。

因此,我们有必要对 Easley, Kiefer, O'Hara and Paperman (1996)^[1]的 PIN 模型进行修正,以便使得修正后的模型能够更加准确地对信息风险进行测度。

二、研究意义

对股票信息风险进行准确的测度无论对资产定价、风险管理还是市场绩效的衡量都有着重要意义。

1、有助于进行资产定价

信息风险是否是一种系统性风险目前在学术界还是一个备受争议的问题。如果信息风险是一种系统性风险,那么投资者持有信息风险高的股票就应该要求获得一个比较高的风险溢价。而要对信息风险是否是一种系统性风险进行实证检验,首先需要对信息风险进行准确的测度。因此,对于信息风险测度的研究,将有助于进行资产定价。

2、有助于投资者进行风险管理

虽然信息风险或信息不对称是投资者面临的一种普遍现象,但是对不同股票具有的信息风险或信息不对称程度进行准确测度,可以使投资者更好地了解不同股票具有的信息风险或信息不对称程度,帮助投资者回避信息风险较高或信息不

对称程度严重的股票，从而保障投资者的权益。因此，对与信息风险测度的研究，将有助于投资者进行风险管理。

3、有助于衡量市场绩效

市场绩效包括公平性、效率性和透明度等方面的含义。如果一个国家股票市场上具有的信息风险较高或信息不对称程度较为严重，那么意味着这个国家股票市场较为缺乏公平性和效率性，透明度也比较低。而要对市场绩效进行衡量，也要求能够对市场上具有的信息风险或信息不对称程度进行准确的测度，从而为监管层采取市场治理措施提供依据。因此，对于信息风险测度的研究，将有助于衡量市场绩效。

三、主要创新

本文对股票信息风险测度的相关问题进行了研究，主要创新之处在于：

1、对 Easley, Kiefer, O' Hara and Paperman (1996)^[1]提出的测度信息风险的经典 PIN 模型进行了修正。经典的 PIN 模型隐含的买单和卖单数量之间的负相关关系与实际数据中买单和卖单数量之间的正相关关系并不相符，并且经典的 PIN 模型隐含的买单和卖单数量的方差与实际数据中买卖指令相对较大的方差也不是很匹配。本文通过在交易动机中增加由于市场指令流冲击引起的交易，提出了修正的 PIN 模型。本文基于中国股票的逐笔交易数据，利用修正的 PIN 模型对我国股票具有的信息风险进行的实证研究表明，修正的 PIN 模型隐含的买单和卖单数量之间的相关性为正，这能够较好地匹配实际数据中买单和卖单数量之间的正相关关系。此外，修正的 PIN 模型也能够较好地匹配实际数据中买单和卖单数量较大的方差。

2、本文选取换手率作为交易活跃程度的衡量指标，分别利用经典的 PIN 模型和修正的 PIN 模型对交易活跃程度不同的股票具有的信息风险进行了实证研究，发现股票具有的信息风险大小同交易活跃程度之间呈负相关关系，换手率越高的股票具有的信息风险越低。而经典的 PIN 模型由于忽视了市场指令流冲击事件发生时引起的交易动机，倾向于高估股票所具有的信息风险。

3、本文采用滚动 (rolling) 的方法构造出了股票信息风险的时间序列，发现股票具有的信息风险是一个时变的过程，并且存在很强的序列相关性。本文对股票

信息风险和股票收益率之间的关系进行了实证研究，发现股票信息风险对股票收益率并不存在显著的影响，而与市场流动性相关的市场指令流冲击事件发生时引起的交易概率则对股票收益率有着持续显著的影响。

四、论文结构

本文的剩余部分安排如下：第一章在对信息风险进行定义的基础上，对信息风险测度的国内外研究文献进行了较为全面的回顾和评述；第二章在详细介绍 Easley, Kiefer, O' Hara and Paperman (1996)^[1]提出的测度信息风险的经典 PIN 模型的基础上，对该模型进行了修正，提出了修正的 PIN 模型；第三章基于中国股票的逐笔交易数据，分别利用经典的 PIN 模型和修正的 PIN 模型对交易活跃程度不同的股票具有的信息风险进行了测度；第四章在采用滚动（rolling）的方法构造出中国股票信息风险的时间序列的基础上，对中国股票的信息风险同股票收益率之间的关系进行了实证研究；第五章是结论和未来进一步的研究方向。

第1章 信息风险测度的文献综述

§ 1.1 信息风险的定义

一、信息的定义

信息一词具有不同的含义，当我们把信息理解为积累下来的关于世界的资料和论据时，信息是一个存量的概念；在更多的时候，当信息以消息的形式出现时，信息是一个存量的增量的概念。阿罗曾指出，所谓信息就是指根据条件概率原则能有效地改变后验概率的任何可以观察的结果。信息集是状态空间的一个分划（Partition），有了关于某件事的信息，可以此为条件修正对该事件发生可能性的后验信念。如果概率是客观的，则信息揭示了客观事物的某一侧面，可以使人们对客观事物的认识又前进一步。如果认为概率是主观的信念，信息的到来则可以促使人们修正自己的信念。由于人们对客观信息的理解和鉴别常常因人而异，信息的内容往往就有了主观性的一面，这有助于理解人们把噪声当作信息使用的行为和某些人散布谣言蛊惑他人的动机。

二、信息风险的定义

信息对证券市场价格的决定至关重要。证券价格的发现过程实际上就是证券价格反映与证券相关的信息的过程，在这一过程中投资者拥有的信息是否对称将对证券市场价格的形成产生至关重要的影响。

从 Bagehot (1971)^[2]开始，Copeland and Galai (1983)^[3]、Glosten and Milgrom (1985)^[4]以及 Easley and O'Hara (1987)^[5]等提出证券市场的参与者可以区分为知情交易者（informed traders）和非知情交易者（uninformed traders）两类。知情交易者是指拥有信息优势的那些交易者，非知情交易者则是指那些不具备信息优势的交易者。需要指出的是，有的文献把知情交易者拥有的信息定义为私有信息，知情交易者等同于内幕交易者，事实上这些信息也可以是公开信息，只要证券价格没有充分反映与其相关的所有信息，知情交易者就可以根据这些信息进行交易，其交易行为也属于知情交易。

Easley and O'Hara (2004)^[13]论证了具有更多信息不对称的股票拥有更高的预期收益率。他们构造了一个具有非对称信息的理性预期资产定价模型并发现在其他条件不变情况下，非知情投资者要求获得溢价以持有具有更高的信息不对称的公司的股票。在该模型中，信息不对称的效应是不可分散的，因为非知情投资者预期会应信息不对称而损失，因此要求对该预期损失进行补偿。

因此，在本文中，信息风险即是指投资者由于在某种资产上的信息不对称而遭受损失的可能性，它是对投资者面临的信息不对称程度的衡量，其大小通常用知情交易概率来衡量^①。因为在交易过程中，知情交易者的行为往往被非知情交易者所掩盖，我们无法确切知道谁是知情交易者，也无法确切知道某一笔交易是否为知情交易，只能从概率的角度来进行判断。某种资产的知情交易概率越大，意味着该资产上信息不对称的程度越严重，那么该资产具有的信息风险也就越高。

§ 1.2 信息风险测度的国外文献综述

一、对信息风险（信息不对称）测度的相关文献

（一）对信息风险（信息不对称）的间接测度

由于信息风险（信息不对称）不能从市场上直接观测到，因此国外学者早期的文献主要从间接的角度采用一些替代变量来对信息风险（信息不对称）进行测度。

Bagehot (1971)^[2]首先将买卖价差作为信息不对称的简单测度，用来解释做市商面临的信息不对称问题。他认为如果市场存在过多的知情交易者，那么做市商将被迫设定足够大的买卖价差来弥补与知情交易者交易带来的损失，因此，买卖价差越大，知情交易存在的概率也越大。在这之后，Jaffe and Winkler (1976)^[6]、Copeland and Galai (1983)^[3]、Morse and Ushman (1983)^[18]、Glosten and Milgrom (1985)^[4]、Glosten and Harris (1988)^[19]、Foster and Viswanathan (1990)^[20]、McInish and Wood (1992)^[21]、Foster and Viswanathan (1993)^[22]和Coller and Yohn (1997)^[23]等都采用买卖价差作为对信息不对称的简单测度。

尽管在早期的文献中，买卖价差是最广泛被用来作为信息不对称程度的间接

^① Easley, Hvidkjaer and O'Hara (2002)^[12]最早用知情交易概率来衡量股票具有的信息风险。

测度，其他学者也提出了对信息不对称的许多不同的测度方法。

Benston and Hagerman (1974)^[24]将非系统性风险作为信息不对称或知情交易的替代变量。Stoll (1978)^[25]和 Stoll (1989)^[26]将换手率作为信息不对称或知情交易的替代变量。Hasbrouck (1991)^[27]采用方差分解的办法将价格变动的方差分解为与交易有关的部分和与交易无关的部分，然后将与交易有关的价格变动的方差作为信息不对称的绝对衡量指标。Chiang and Venkatesh (1988)^[28]将公司内部人持股比例（公司内部人持股占公司流通在外股数的比例）作为信息不对称的测度。Sarin, Shastri and Shastri (2000)^[29]将公司的股权结构（股权集中度）作为信息不对称的测度。

然而，需要指出的是，这些替代变量要么反映的是信息不对称的结果（买卖价差、非系统性风险、换手率和与交易有关的价格变动方差），要么是信息不对称的原因（内部人持股比例和股权结构），因此都不能明确地测度交易者在市场中面临的信息不对称状况，而且这些结果都没有标准化，让人无法对不同市场上的信息不对称的严重程度进行直接比较。

（二）对信息风险（信息不对称）的直接测度

对信息风险（信息不对称）的直接测度从 20 世纪 90 年代中后期才发展起来。

Easley, Kiefer, O'Hara and Paperman (1996)^[1]最早提出直接测度信息风险（信息不对称）的模型，该模型建立在 Easley and O'Hara (1987)^[5]和 Easley and O'Hara (1992)^[30]的序贯交易模型基础之上，以逐笔交易的混合分布的方法实现了对信息风险（信息不对称）的直接测度。他们以决策树的方法计算出知情交易者和非知情交易者在好消息、坏消息和没有消息下的订单到达数的期望值，然后以知情交易者订单的期望到达数为分子，以所有交易者订单的期望到达数为分母，求出知情交易概率——PIN（Probability of Informed Trading），用来衡量信息风险（信息不对称）的大小。他们实际上是以异常的指令流非平衡（order flow imbalance）来度量信息不对称的严重程度，同时他们也证明了不存在从价格到指令流的明显的反馈。他们利用纽约股票交易所的交易数据，估计了每只股票的信息风险大小，发现交易活跃的股票的信息风险较低，交易不活跃的股票的信息风险较高^①。

Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出 PIN 模型后，该模型立即成为

^① Easley, Kiefer, O'Hara and Paperman (1996)^[1]用交易量高低来反映交易的活跃程度。

测度信息风险的经典模型，随后大多数涉及信息风险的文献都直接采用该模型，该模型被广泛应用于解决证券市场微观结构中的各种问题。

Easley, Kiefer and O'Hara (1997a)^[7]进一步考虑了交易量对信息风险大小的影响，发现交易量的影响并不显著。Easley, O'Hara and Paperman (1998)^[9]研究了金融分析师的数量对股票信息风险的影响，发现金融分析师数量高的股票具有较低的信息风险。Brockman and Chung (2000)^[10]利用香港证券交易所的股票作为研究对象，发现较高的信息风险会导致较高的买卖价差，同时他们发现成交量越大的股票，其信息风险有越小的现象。Easley, O'Hara and Saar (2001)^[11]研究了股票分割如何影响信息风险，他们发现股票拆分会同时吸引知情交易者和非知情交易者，其中知情交易的增加不会引起交易信息内容的可察觉的变化，因此股票拆分并不会导致信息风险的提高。Straser (2002)^[31]利用PIN模型比较了美国证券交易委员会在2000年10月23日通过上市公司公平揭露法案 (Regulation Fair Disclosure) 前后上市公司股票信息风险的变化，结果发现在实行该法案后，股票信息风险反而有了显著上升。Barclay and Hendershott (2003)^[32]利用PIN模型比较了在正常交易时间之外进行交易的股票具有的信息风险之间的差异，发现股票在开盘前进行交易比股票在收盘后进行交易具有更高的信息风险^①。Brown, Hillegeist and Lo (2004)^[33]利用PIN模型研究了公司的电话会议 (Conference Calls) 与股票信息风险之间的关系，发现公司的电话会议活动频率与股票的信息风险之间呈显著的负相关关系。Chung, Li and McInish (2005)^[34]的研究发现即使在控制了股票自身属性特征之后，交易对价格的影响和交易方向上的序列相关性仍然同股票信息风险的大小之间具有显著的正相关关系。Vega (2006)^[14]利用PIN模型研究了公开信息和私有信息如何影响公司盈利公布后的价格漂移，他发现具有高的信息风险的股票在公司盈利公布后的价格漂移较低。Aktas, de Bodt, Declerck and Van Oppens (2007)^[35]利用PIN模型研究了公司发生兼并收购事件前后信息风险的变化，发现信息风险在公司发布兼并收购事件前趋于下降，而在公司发布兼并收购事件后趋于上升。Chen, Goldstein and Jiang (2007)^[36]利用信息风险 (PIN) 作为公司股票价格的信息含量指标，发现在控制了公司管理层掌握的信息等因素后，信息风险 (PIN) 同公司投资对股票价格的

^① 大多数美国证券交易所的交易时间为美国东部时间的上午 9:30 到下午 4:00，信息技术的发展使得投资者在交易所交易时间之外也可以进行交易，如美国投资者可以通过像 Instinet、Island 和 Archipelago 等其他电子通讯网 (Electronic Communications Networks) 在交易所交易时间之外进行交易。

敏感性之间存在正相关关系。Aslan, Easley, Hvidkjaer and O'Hara (2007)^[37]研究了公司特征与股票信息风险之间的关系,发现规模较小的公司、较年轻的公司、具有更多内部人持股或机构持股的公司、具有更少的金融分析师数量的公司、托宾Q较低的公司以及换手率较高的公司的股票更有可能具有较高的信息风险。Albuquerque, De Francisco and Marques (2008)^[38]在PIN模型的基础之上研究了知情交易者基于公司特有的私人信息和市场范围内的私人信息进行交易的情形,他们的实证研究证实了市场范围内的私人信息的存在,并发现市场范围内的私人信息不仅可以预测行业股票收益率还可以预测汇率收益率。

虽然Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的PIN模型目前已被广泛应用于证券市场微观结构领域中,但也有不同学者对其提出了一些质疑。Venter and De Jongh (2006)^[16]从统计的角度发现PIN模型隐含的买卖指令之间的负相关关系与实际数据中买卖指令之间的正相关关系并不相符。Boehmer, Grammig and Theissen (2007)^[15]考察了将交易错误分类^①对信息风险及其估计的影响,他们的研究发现错误的交易分类会导致下偏的信息风险估计值。Duarte and Young (2008)^[17]的实证研究也证实PIN模型隐含的买卖指令之间的负相关关系与实际数据中买卖指令之间的正相关关系并不相符,此外, PIN模型隐含的买卖指令的方差与实际数据中买卖指令相对较大的方差也不能很好地匹配。

为此, Nyholm (2002)^[40]提出了完全不同的信息风险测度方法,他以纽约证券市场为研究对象,模型主要是根据专家 (specialist) 参与市场交易的行为进行构建。专家根据所有可观察到的信息来推断一笔交易是来自知情交易者还是非知情交易者,进而算出一段时间内的信息风险。Nyholm (2002)^[40]的方法相对于以前研究的优势在于可以估计每一笔交易或者每一小段交易时间内的信息风险,这样就可以研究信息风险的变化规律,也符合不同时间信息风险不同的常识,而不是像Easley, Kiefer, O'Hara and Paperman (1996)^[1]等的模型那样,一个股票只有一个知情交易概率密度。Nyholm (2003)^[41]对Nyholm (2002)^[40]模型又进行了相应的拓展,增加了对交易量的考虑。

^① PIN模型的估计需要判断一笔交易是由买方发起还是由卖方发起,而大部分交易数据都不提供交易的买卖方向,对此国外学者通常采用 Lee and Ready (1991)^[39]的买卖判断准则,即将一笔交易的成交价与买卖报价的中点相比较,高于中点的这笔交易为卖,低于中点的交易为买,如果成交价与买卖报价的中点相同,则与上一笔成交价相比较,以此来判断出一笔交易的买卖方向。

然而，上述这些学者对信息风险测度的研究都是以报价驱动市场交易机制为研究对象，而很少有学者对指令驱动市场交易机制^①下的信息风险测度进行研究。

Handa, Schwartz and Tiwari (2003)^[42]在研究买卖价差的决定因素中涉及到信息风险的测度问题，他们从非知情交易者的角度出发，采用同 Easley, Kiefer, O' Hara and Paperman (1996)^[1]相类似的方法，研究了指令驱动市场交易机制下信息风险的测度问题。然而他们的模型有着非常严格却又不符合实际的假定，如假定知情交易者只能进行市价委托不能进行限价委托。Ma, Hsieh and Chen (2007)^[43]对 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型进行了拓展，并对台湾股票市场的信息风险进行了实证研究。台湾的股票市场虽然属于指令驱动市场，然而其实行的是短时间内多次集合竞价的交易机制，与我国股票市场实行开盘集合竞价、交易时间内实行连续竞价的交易机制并不相同，因此他们得出的结论并不具有一般性。

二、信息风险（信息不对称）与资产收益率之间的关系

（一）信息风险与资产收益率关系的理论研究

国外学者首先从理论上对信息风险是否是资产收益的显著决定因素进行了探讨。

Wang (1993)^[44]最早对信息不对称或知情交易对资产收益的影响进行了研究，他建立了一个跨期的资产定价模型，模型中假设有一种无风险资产和一种风险资产可供投资，他的研究认为拥有信息优势的知情交易者是导致逆向选择的源头，而非知情交易者据此要求承受风险不对称的风险溢价。Easley and O'Hara (2004)^[13]建立了一个理性预期的均衡资产定价模型，发现当各资产的公开和私有信息存在差异，市场均衡时非知情交易者会对私有信息较多的股票要求较多的风险溢价，使股票的收益率产生横截面的差异。

另一方面，Hughes (2005)^[45]和 Lambert, Leuz and Verrecchia (2007)^[46]证明了在大经济中非对称信息关于预期收益率的效应是可分散的。他们论证了非对称信息在 Easley and O'Hara (2004)^[13]中被定价是因为在他们的模型中资产的数量是有限的，因此非对称信息的风险不能被分散掉。

^① 在报价驱动交易机制下，每一只证券都有专门的做市商，也就是做市商为争取提供买卖报价，并在自己所报的价格水平上接受交易者的买卖；在指令驱动交易机制下，投资者以电子竞价交易的方式进行交易。

(二) 信息风险与资产收益率关系的实证研究

信息风险是否是资产收益率的显著决定因素不仅是一个理论问题，也是一个实证问题。

Easley, Hvidkjaer and O'Hara (2002)^[12]对信息风险是否为资产收益的决定因素进行了实证研究，他们利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型估计出信息风险，结果发现除了 Fama and French (1992)^[47]提出的三因子之外，信息风险确实对资产收益有显著为正的影晌。他们的研究发现两只股票的信息风险之间十个百分点的差异导致了它们年预期收益率的 250 个基点的差异。

Hughson and Kang (2007)^[48]利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型研究了知情交易对股票预期收益率的影响，发现信息风险是股票预期收益率的显著决定因素，而且信息风险与股票预期收益率之间的这种关系呈现出非线性的特征。

Lu and Wong (2008)^[49]利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型估计出了台湾股票的信息风险，并对信息风险是否是台湾股票收益率的显著定价因子进行了实证检验，发现信息风险是台湾股票横截面收益率的显著定价因子。他们的研究发现在台湾股票市场上信息风险增加十个百分点平均要求股票提供额外四到七个百分点的收益率。

Li, Wang, Wu and He (2008)^[50]就流动性风险和信思风险对国债的预期收益率的影响进行了实证研究，他们用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型估计出信息风险，发现在控制了其他的系统性风险因子和债券特征之后，预期国债收益率和流动性风险以及信息风险之间有很强的正相关关系。

虽然上述学者对信息风险是否是资产收益率的显著决定因素进行的实证研究都发现信息风险是股票收益率显著为正的决左因素，但是也有学者的研究得出了不同的结论。

Mohanram and Rajgopal (2006)^[51]利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型估计出的信息风险进行了资产定价检验，发现在加入 Fama and French (1992)^[47]提出的市场因子、规模因子和账面市值因子这三个因子和 Carhart (1997)^[52]提出的惯性因子之后，信息风险对股票收益率仍然有显著为正的影晌。但是，当他们进一步采用 Daniel and Titman (1997)^[53]提出的因子荷载 (Factor Loading)

的办法就信息风险的因子荷载对股票收益率的影响进行检验时，发现在控制信息风险的特征因素后，信息风险的因子荷载对股票收益率的影响并不显著。这意味着信息风险可能并不是资产收益率的显著决定因素。

Kubota and Takehara (2008)^[54]以在东京股票交易所交易的股票为研究对象，利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型估计出了日本股票的信息风险，并对信息风险是否是日本股票收益率的显著定价因子进行了实证检验，发现虽然信息风险对日本股票横截面收益率有正的影响，但是这种影响并不显著。

§ 1.3 信息风险测度的国内文献综述

一、国内学者对信息风险的研究现状

国内学者对信息风险（信息不对称）的测度以及信息风险是否是资产收益的显著决定因素的相关研究还处于起步阶段，而且现有的研究多是直接将 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型应用到我国的股票市场上。

国内的学者最初利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的 PIN 模型对我国股票市场上的信息风险与我国股票市场的买卖价差、流动性、波动性之间的关系进行了研究。

杨之曙和姚松瑶（2004）^[55]对我国上海股市的信息风险进行了实证研究，发现成交量大的股票具有较低的信息风险，且信息风险是我国股市买卖价差的显著决定因素之一。王春峰、董向征和房振明（2005）^[56]对我国上证 180 指数中股票的信息风险与市场流动性、波动性之间的关系进行了实证研究，发现沪市信息风险与流动性呈正相关，与波动性呈负相关。李朋和刘善存（2006）^[57]对我国上证 50 指数中股票的信息风险与买卖价差之间的关系进行了实证研究，发现信息风险对买卖价差具有显著的正向影响。

近两年来，国内学者利用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的模型对信息风险（信息不对称）是否是我国股票市场收益率的显著决定因素进行了实证研究，然而得出的结论并不一致。

杨栋和郭玉清（2007）^[58]对我国沪深股票市场的信息风险与股票收益率之间的关系进行了实证研究，发现信息风险是我国股市收益率的显著为正的決定因素。

王展翔（2007）^[59]对我国沪深股市的信息风险进行了实证研究，发现信息风险与股票收益率呈显著的正相关关系。

然而，韩立岩、郑君彦和李东辉（2008）^[60]对我国上海股市进行的实证研究却发现我国股市的信息风险与股票收益率呈显著的负相关关系。

二、对信息风险测度研究现状的总结

从上面对信息风险测度研究的相关文献回顾中，我们可以得到以下结论：

1、在信息风险测度的相关文献中，主要以Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的PIN模型和Nyholm (2002)^[40]、Nyholm (2003)^[41]这两个模型为主。对比这两个模型，我们可以发现虽然两者都以做市商或者专家为基础，但是Nyholm (2002)^[40]和Nyholm (2003)^[41]的模型主要通过推测专家（specialist）对交易对手的估计来确定市场中的信息风险，因此该模型无法应用于指令驱动市场；而Easley, Kiefer, O'Hara and Paperman (1996)^[1]的PIN模型中的知情交易概率本质上是通过市场订单的不平衡性来推测信息风险，因此这种思想同样适用于指令驱动市场下的信息风险的度量。

2、目前国内外学术界对信息风险的测度基本上都采用 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的 PIN 模型。对信息风险进行准确测度对资产定价以及市场绩效的衡量有着重要影响。然而，Venter and De Jongh (2006)^[16]和 Duarte and Young (2008)^[17]的实证研究对 PIN 模型度量信息风险的准确性提出了质疑，因此，我们有必要对 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的 PIN 模型进行修正，以便使得修正后的模型能够更加准确地对信息风险进行度量。

3、在信息风险与资产收益率之间关系的相关文献中，国外学者无论在理论方面还是在实证方面得出的结论都不一致。在理论方面，有的学者认为信息风险是一种系统性风险应该被定价（如Wang (1993)^[44]和Easley and O'Hara (2004)^[13]）；而有的学者却认为信息风险并不是一种系统性风险而是一种可分散的风险（如Hughes (2005)^[45]和Lambert, Leuz and Verrecchia (2007)^[46]）。在实证方面，有的学者通过实证检验发现信息风险是资产收益率的显著为正的決定因素（如Easley, Hvidkjaer and O'Hara (2002)^[12]、Hughson and Kang (2007)^[48]、Lu and Wong (2008)^[49]和Li, Wang, Wu and He (2008)^[50]）；但也有学者通过实证检验发现信息风险对资产

收益率的影响并不显著（如Mohanram and Rajgopal (2006)^[51]和Kubota and Takehara (2008)^[54]）。然而，需要指出的是，上述学者在对信息风险是否是资产收益率的显著决定因素进行检验的过程中都采用Easley, Kiefer, O' Hara and Paperman (1996)^[1]的PIN模型来测度信息风险，而正如前文所述，信息风险的测度准确与否会对检验结果造成重要的影响。因此，我们有必要在对Easley, Kiefer, O' Hara and Paperman (1996)^[1]的PIN模型进行修正以便能够更加准确地测度信息风险的基础上，重新就信息风险与资产收益率之间的关系进行实证检验。

第 2 章 信息风险测度的模型

§ 2.1 经典的 PIN 模型

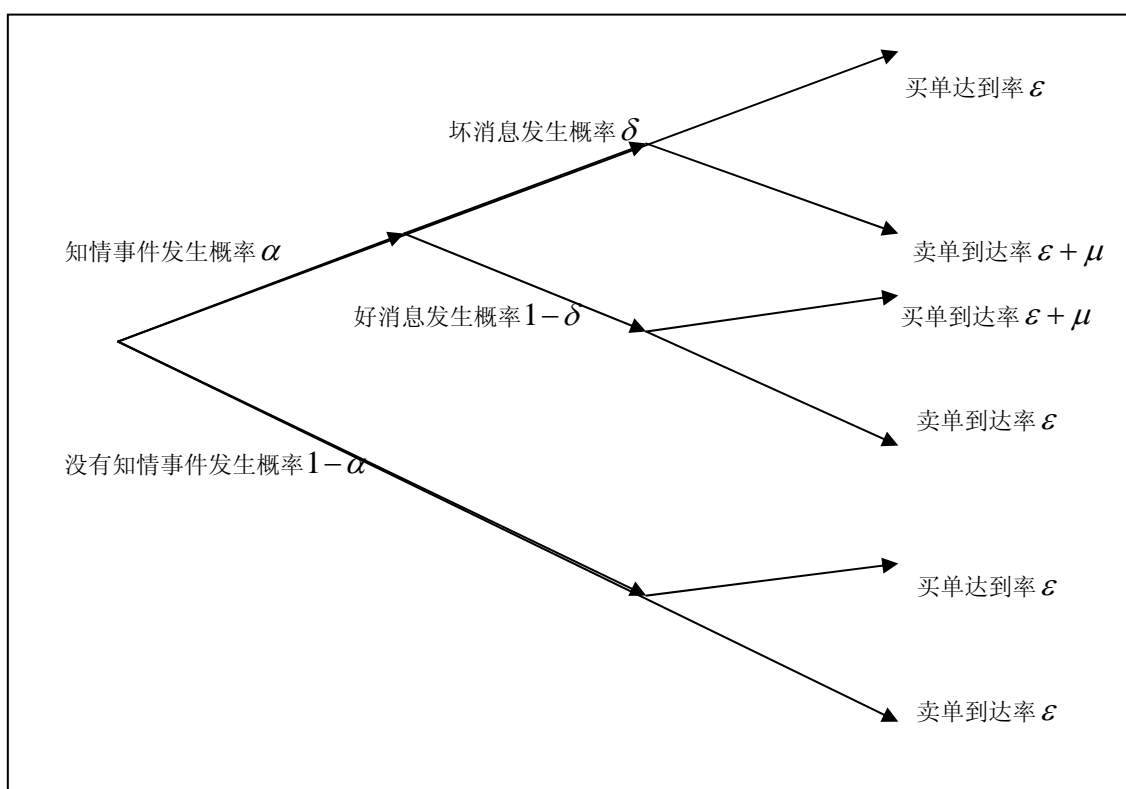
正如前一章中的文献回顾所述，对信息风险（信息不对称）的直接测度从 20 世纪 90 年代中后期才发展起来。

Easley, Kiefer, O'Hara and Paperman (1996)^[1]最早提出直接测度信息风险（信息不对称）的模型，该模型建立在序贯交易模型基础之上，以逐笔交易的混合分布的方法实现了对信息风险（信息不对称）的直接测度。他们以决策树的方法计算出知情交易者和非知情交易者在好消息、坏消息和没有消息的情况下订单到达数的期望值，然后以知情交易者订单的期望到达数为分子，以所有交易者订单的期望到达数为分母，求出知情交易概率——PIN（Probability of Informed Trading），以衡量信息风险（信息不对称）的大小。他们实际上是以异常的指令流非平衡（order flow imbalance）来度量信息不对称的严重程度，同时也证明了不存在从价格到指令流的明显的反馈。他们利用纽约股票交易所的交易数据，估计出了股票具有的信息风险，发现交易活跃的股票具有较低的信息风险，交易不活跃的股票具有较高的信息风险。

下面本文首先对 Easley, Kiefer, O'Hara and Paperman (1996)^[1]的 PIN 模型进行较为详细的介绍。

一、经典的 PIN 模型的基本假设

Easley, Kiefer, O'Hara and Paperman (1996)^[1]假设市场上只存在知情交易者和非知情交易者，交易者在 $i = 1, 2, \dots, I$ 个交易日内与做市商进行风险资产与现金的交易，在每个交易日内时间是连续的， $t \in [0, T]$ ，做市商在任何时候都根据他所决定的买卖价格与交易者进行交易。假定做市商为风险中性者，且期望利润为零，以下推导出的买卖价格都是做市商根据过去的所有交易信息求出的期望价格。图 2.1 显示了经典的 PIN 模型中股票在一个交易日内的整个交易过程。



注： α 为某个交易日发生知情事件的概率； δ 为坏消息发生的概率； μ 为知情交易到达率； ε 为非知情交易到达率。

图 2.1 经典的 PIN 模型中股票在一个交易日内的交易过程

在每个交易日前，首先由自然决定与资产真实价值有关的知情事件发生的概率 α ，且每个交易日知情事件发生的概率 α 彼此独立。这些知情事件的发生又区分为坏消息和好消息，其对应概率分别为 δ 和 $1-\delta$ 。在每个交易日，信息会完全反映到资产价值上，不会影响到下一个交易日的资产价值。假设证券在第 i 个交易日结束后的价值为一随机数 V_i ， $i=1,2,\dots,I$ 。假设在第 i 个交易日有好消息发生时，证券在交易结束后的价值为 V_i^g ；在第 i 个交易日有坏消息发生时，证券在交易结束后的价值为 V_i^b ；在第 i 个交易日没有消息发生时的价值为 V_i^n ，且 $V_i^b < V_i^n < V_i^g$ 。市场上的交易由知情交易者和非知情交易者进行，在一个交易日内，知情交易者的指令到达率和非知情交易者的指令到达率服从泊松分布且假设每个交易日之间的指令到达率彼此独立。非知情交易者的买单和卖单到达率为 ε ，所有指令到达率

定义为每个交易日内指令的平均达到率。模型假定所有的知情交易者都为风险中性者，一旦其洞察到有好消息时利润最大化的交易策略则是买进股票，反之，当其洞察到有坏消息时则卖出股票。假定知情交易者的买卖单到达率为 μ 。所有的指令到达之间是相互独立的。如图 2.1 所示，在第一个节点即每个交易日开盘之前，自然决定该交易日与资产真实价值相关的知情事件是否发生，如果知情事件发生，自然再将之区分为好消息和坏消息。给定已区分的事件之后，买卖单到达率都服从泊松分布。当好消息发生时，买单到达率为 $\varepsilon + \mu$ ，卖单到达率为 ε ；当坏消息发生时，买单到达率为 ε ，卖单到达率为 $\varepsilon + \mu$ 。在没有消息发生时，只有非知情交易者交易，因此买卖单的达到率都为 ε 。

二、经典的 PIN 模型的价格决定

在价格决定方面，做市商知道到达每个分支的概率以及每个分支的指令到达过程，但是他不知道自然到底选择了三个分支中的哪一个。假定做市商根据贝叶斯法则利用到达的交易指令来更新他对市场知情事件的信念。由于每个交易日之间相互独立，我们可以分别分析做市商对市场看法的演进。令 $P(t) = (P_n(t), P_b(t), P_g(t))$ 为做市商在时刻 t 关于“没有消息”（ n ）、“坏消息”（ b ）和“好消息”（ g ）的先验信念，因此其在时刻 0 的先验信念为 $P(0) = (1 - \alpha, \alpha\delta, \alpha(1 - \delta))$ 。

为了决定在时刻 t 的报价，做市商根据相关类型指令的到达来更新他的先验信念。例如，时刻 t 的买价 $b(t)$ 为在时刻 t 的指令到达之前的资产交易历史和有人想要卖出一单位资产的条件期望值。令 S_t 表示在时刻 t 到达的卖出指令， B_t 表示在时刻 t 到达的买入指令。令 $P(t|S_t)$ 表示做市商根据过去时间内所有交易历史资料以及在时刻 t 一个卖出指令到达后所更新的条件先验信念，根据贝叶斯法则，若一个卖出指令在时刻 t 到达，做市商对于时刻 t 没有消息发生的事后概率修正为：

$$P_n(t|S_t) = \frac{P_n(t)\varepsilon}{\varepsilon + P_b(t)\mu} \quad (2.1)$$

其中，分子表示没有消息发生时的卖单期望值，分母表示所有情况下卖单的期望值。类似地，对于时刻 t 坏消息发生的事后概率为：

$$P_b(t|S_t) = \frac{P_b(t)(\varepsilon + \mu)}{\varepsilon + P_b(t)\mu} \quad (2.2)$$

对于时刻 t 好消息发生的事后概率为：

$$P_g(t|S_t) = \frac{P_g(t)\varepsilon}{\varepsilon + P_b(t)\mu} \quad (2.3)$$

在任意的时刻 t ，做市商的零期望利润的买价 $b(t)$ ，是做市商根据时刻 t 之前的所有交易历史和 S_t 的信息之下的条件资产期望值。因此，在第 i 天时刻 t 的买价 (bid price) 为：

$$b(t) = \frac{P_n(t)\varepsilon V_i^n + P_b(t)(\varepsilon + \mu)V_i^b + P_g(t)\varepsilon V_i^g}{\varepsilon + P_b(t)\mu} \quad (2.4)$$

同理，可以计算出在第 i 天时刻 t 的卖价 (ask price) 为：

$$a(t) = \frac{P_n(t)\varepsilon V_i^n + P_b(t)\varepsilon V_i^b + P_g(t)(\varepsilon + \mu)V_i^g}{\varepsilon + P_g(t)\mu} \quad (2.5)$$

做市商在时刻 t 之前的条件资产期望值为：

$$E[V_i|t] = P_n(t)V_i^n + P_b(t)V_i^b + P_g(t)V_i^g \quad (2.6)$$

将 (2.6) 分别代入 (2.4) 和 (2.5)，可以得到：

$$\begin{aligned} b(t) &= \frac{P_n(t)\varepsilon V_i^n + P_b(t)(\varepsilon + \mu)V_i^b + P_g(t)\varepsilon V_i^g}{\varepsilon + P_b(t)\mu} \\ &= \frac{E[V_i|t]\varepsilon + P_b(t)\mu V_i^b}{\varepsilon + P_b(t)\mu} \\ &= \frac{\{(\varepsilon + P_b(t)\mu)E[V_i|t]\} - P_b(t)\mu E[V_i|t] + P_b(t)\mu V_i^b}{\varepsilon + P_b(t)\mu} \\ &= E[V_i|t] - \frac{P_b(t)\mu}{\varepsilon + P_b(t)\mu} \{E[V_i|t] - V_i^b\} \end{aligned} \quad (2.7)$$

同理，可以将卖价重新写为：

$$a(t) = E[V_i|t] + \frac{P_g(t)\mu}{\varepsilon + P_g(t)\mu} \{V_i^g - E[V_i|t]\} \quad (2.8)$$

由上述式子可以看出知情交易者的交易指令和非知情交易者的交易指令是如何影响交易价格的，如果市场上没有知情交易者（即 $\mu = 0$ ），则所有交易指令都为非知情交易指令，这使得买卖价格等于时刻 t 的条件资产期望值。反之，如果市场上没有非知情交易者（即 $\varepsilon = 0$ ），则时刻 t 的买价等于坏消息发生时的资产价值（ $b(t) = V_i^b$ ），卖价等于好消息发生时的资产价值（ $a(t) = V_i^s$ ）。在这些价格下，没有知情交易者会进行交易，市场实际上会停滞。通常，知情交易者和非知情交易者同时存在于市场上，因此，做事商报出的买价会低于资产价值的期望值而卖价会高于资产价值的期望值。

这样，我们可以计算出时刻 t 的买卖价差为：

$$\begin{aligned} \sum(t) &= a(t) - b(t) \\ &= \frac{P_g(t)\mu}{\varepsilon + P_g(t)\mu} \{V_i^s - E[V_i|t]\} - \frac{P_b(t)\mu}{\varepsilon + P_b(t)\mu} \{E[V_i|t] - V_i^b\} \end{aligned} \quad (2.9)$$

式 (2.9) 表示在时刻 t 的买卖价差是由来自于知情交易者的买单概率乘以做市商面对知情交易者的预期损失，加上来自于知情交易者的卖单概率乘以做市商面对知情交易者的预期损失所构成。

三、经典的 PIN 模型对信息风险的测度

由式 (2.1)、(2.2)、(2.3) 可以发现，在时刻 t ，由知情交易者发出的交易指令占总交易指令的概率称为知情交易概率（Probability of Informed Trading, PIN），即信息风险，可以由下式导出：

$$\begin{aligned} PIN(t) &= \frac{P_g(t)\mu + P_b(t)\mu}{\{P_g(t)(\varepsilon + \mu) + P_b(t)\varepsilon + P_n(t)\varepsilon\} + \{P_g(t)\varepsilon + P_b(t)(\varepsilon + \mu) + P_n(t)\varepsilon\}} \\ &= \frac{(P_g(t) + P_b(t))\mu}{(\varepsilon + P_g(t)\mu) + (\varepsilon + P_b(t)\mu)} \\ &= \frac{(1 - P_n(t))\mu}{2\varepsilon + (1 - P_n(t))\mu} \end{aligned} \quad (2.10)$$

理论上，给定时刻 t 之前的交易资料，可以利用上述式子估计出任何时间点的知情交易概率。但在实证中，使用开盘的报价来决定 PIN，即估计 $t = 0$ 的 PIN。因此，每个交易日的 PIN 可以表示成：

$$\begin{aligned}
 PIN &= \frac{(1-P_n(0))\mu}{2\varepsilon+(1-P_n(0))\mu} \\
 &= \frac{\alpha\mu}{2\varepsilon+\alpha\mu}
 \end{aligned} \tag{2.11}$$

信息风险 PIN 表示来自知情交易者的交易指令占总交易指令的比例，它与知情交易指令的到达率以及非知情交易指令的到达率有关，同时也和知情事件发生的概率有关。若没有知情事件发生， $P_n(t)=1$ ，或者没有来自于私有消息的交易（ $\mu=0$ ），则 $PIN=0$ ，且买卖之间没有价差。相反，如果所有交易都是知情交易（ $\varepsilon=0$ ），那么 $PIN=1$ ，且买卖价差为 $V_i^s - V_i^b$ 。

四、经典的 PIN 模型的参数估计

对经典的 PIN 模型的参数估计需要估计出参数向量 $\theta=(\alpha, \delta, \varepsilon, \mu)$ ，这比仅仅从独立的泊松过程估计买卖单的到达率要更加复杂，因为我们不能直接观察到任何知情事件的到达。参数 α 和 δ 决定了三种知情事件（没有消息（ n ）、坏消息（ b ）和好消息（ g ））的概率，但是它们都是不可直接观察到的。剩下的参数 ε 和 μ 涉及到非知情交易者和知情交易者的到达率。我们只能观察到买单和卖单的到达率，但是不能直接观察到哪些交易者是非知情交易者，哪些是知情交易者。因此，估计所有这些参数需要一个结构化的模型。经典的 PIN 模型提供了从可观察的变量，即买单和卖单中提取参数信息所必需的结构。

在模型中，每一个交易日区分为好消息、坏消息和没有消息三个分支，其买卖单到达率都服从泊松分布。虽然我们无法得知该交易日属于哪一个分支，但却可以从交易资料中得到可能的结果。当该交易日买单数量较多时，可以预计该日属于好消息层面的概率较大；类似地，当该交易日卖单数量较多时，可以预计该日属于坏消息层面的概率较大；而在没有消息发生的交易日，没有知情交易者进行交易，买卖单之间的数量大致相等。对于每个交易日内情况的三种分支以及交易指令的到达率可以利用一个混合概率模型来描述，将每一分支所属概率的权重乘以该分支的泊松分布，形成一个混合的泊松分布。假设我们考虑一个坏消息发生的交易日，卖单到达率为 $\mu+\varepsilon$ ，即知情交易者与非知情交易者卖单的加总；只

有非知情交易者会买入，因此买单到达率为 ε 。

因此，在坏消息发生日的总时间 T 内观察到的包含 B 个买单和 S 个卖单的交易指令的独立的泊松分布的似然函数为：

$$e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\varepsilon+\mu)T} \frac{[(\varepsilon+\mu)T]^S}{S!} \quad (2.12)$$

类似地，在一个没有消息发生的交易日，包含 B 个买单和 S 个卖单的交易指令的似然函数为：

$$e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (2.13)$$

最后，在一个好消息发生日，包含 B 个买单和 S 个卖单的交易指令的似然函数为：

$$e^{-(\varepsilon+\mu)T} \frac{[(\varepsilon+\mu)T]^B}{B!} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (2.14)$$

从式 (2.12)、(2.13) 和 (2.14) 我们可以看到买单以及卖单数量 (B, S) 是参数 $\theta = (\alpha, \delta, \varepsilon, \mu)$ 在时间 T 内的充分统计量。

因此，在估计参数时只需要知道在一个交易日内买单的总数量和卖单的总数量。当不知道该交易日属于何种消息时，可以根据每一个分支发生的概率进行加权平均得到混合的似然函数。也就是说，没有消息发生、好消息和坏消息发生的概率分别为 $1-\alpha$ 、 $\alpha\delta$ 和 $\alpha(1-\delta)$ ，因此，混合似然函数可以写为：

$$\begin{aligned} L((B, S)|\theta) &= (1-\alpha) \times e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \\ &+ \alpha\delta \times e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\varepsilon+\mu)T} \frac{[(\varepsilon+\mu)T]^S}{S!} \\ &+ \alpha(1-\delta) \times e^{-(\varepsilon+\mu)T} \frac{[(\varepsilon+\mu)T]^B}{B!} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \end{aligned} \quad (2.15)$$

在一个给定的交易日，利用最大似然估计法所估计出来的参数 α 和 δ 为介于 0 到 1 之间的概率，反映知情事件在一个交易日发生的可能性与好消息和坏消息发生的概率。在交易日之间，可以利用每个交易日所累积的总买卖单数量来估计出

日间的这些参数。由于交易日之间彼此独立^①， I 个观察日之内的买卖单数量 $M = (B_i, S_i)_{i=1}^I$ 的似然函数为每天的似然函数的乘积：

$$L(M|\theta) = \prod_{i=1}^I L(\theta|B_i, S_i) \quad (2.16)$$

为了从任意的数据集 M 中估计出参数向量 θ ，只需要最大化方程 (2.16) 中的似然函数。这种方法提供了对一只股票上的知情和非知情交易的直接估计方法。因此，当利用市场上所观察到的交易资料作为输入值，以最大似然法估计参数向量 $\theta = (\alpha, \delta, \varepsilon, \mu)$ ，便能够计算出信息风险 PIN。

Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出 PIN 模型后，该模型立即成为测度信息风险的经典模型，随后大多数涉及信息风险的文献都直接采用该模型，并把该模型广泛应用于解决证券市场微观结构中的各种问题。

§ 2.2 修正的 PIN 模型

正如前文所述，虽然 Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的 PIN 模型目前已成为测度信息风险的经典模型，但也有不同学者对其测度信息风险的准确性提出了质疑。因此，在本节中，我们对 Easley, Kiefer, O'Hara and Paperman (1996)^[1]的经典的 PIN 模型进行修正，以便使得修正后的模型能够更加准确地对信息风险进行测度。

一、对经典的 PIN 模型进行修正的基本思路

Easley, Kiefer, O'Hara and Paperman (1996)^[1]的 PIN 模型假定交易只有两种可能的动机，即根据公司特有的私人信息进行的知情交易和外生的流动性交易。与私人信息相关的交易发生在知情交易者收到私人信息的日子，如果私人信息是好消息，这会导致更大数量的买入指令，如果私人信息是坏消息则导致更大数量的卖出指令。结果是，大量的买入和卖出指令在不同日到达，这造成了买入和卖出指令之间的负相关关系 (Duarte and Young (2008)^[17])。

^① Easley, Kiefer and O'Hara (1997b)^[8]检验了不同交易日之间知情事件的独立性，发现他们不能拒绝不同交易日之间的知情事件相互独立的假设。

而 Hasbrouck and Seppi (2001)^[61]证明了指令流非平衡在不同的股票之间是相关的, 这意味着存在驱动指令流非平衡的共同因素。而且, Chordia, Roll and Subrahmanyam (2002)^[62]的研究表明整个市场的交易指令流非平衡对同时期的指数收益具有显著影响。整个市场的指令流非平衡不太可能是由于信息不对称引起的, 因此可能存在其他造成非平衡的因素。

因此, 本文考虑增加 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的经典的 PIN 模型中的交易动机, 以使得每天的买卖指令增加, 从而使得模型隐含的买入指令和卖出指令之间的相关性与实际数据中买入指令和卖出指令之间的相关性一致(为正)并产生更接近于数据中观测到的买卖指令的方差。

二、驱动整个市场交易增加的因素

本文把在修正的 PIN 模型中所增加的驱动整个市场范围内的交易增加的因素称为市场指令流冲击。

市场指令流冲击的存在至少有两种原因。市场指令流冲击存在的一种可能的原因是使投资者产生分歧的公开信息事件的发生。这种对公开信息分歧使得买单和卖单的数量都增加。在文献中有关对公开信息事件的分歧是交易量的重要来源具有相当长的历史, 这一历史可以追溯到 Bachelier (1900)^[63]。正如 Kandel and Pearson (1995)^[64]所发现的, 观察到同样信息的投资者对于如何解释这些信息会产生分歧——如果投资者对于应该被用来解释这些信息的方式存在不同看法的话。大量的研究提供了对这一假说的实证证据, 包括 Harris and Raviv (1993)^[65]、Kandel and Pearson (1995)^[64]、Bessembinder, Chan and Seguin (1996)^[66]和 Bamber, Barron and Stober (1999)^[67]等。市场指令流冲击存在的另外一种可能的原因是交易者仅仅在某个交易日合作交易以降低交易成本, 具体论述可见 Admati and Pfleiderer (1988)^[68]。

尽管市场指令流冲击的存在超过一种可能的原因, 本文的目的并不是要在这些不同的原因之间进行区分。我们的目的是为了证明在经典的 PIN 模型中加入这一市场指令流冲击, 可以提高模型对实际数据的解释力度, 这对于更加准确地测度信息风险具有重要意义。

三、对经典的 PIN 模型的修正

(一) 修正的PIN模型的基本假定

为了将Easley, Kiefer, O'Hara and Paperman (1996)^[1]的模型应用到指令驱动交易机制下的我国股票市场上, 本文假定存在一个隐形的做市商, 知情交易者和非知情交易者通过提交限价委托单与隐形做市商进行交易。同Easley, Kiefer, O'Hara and Paperman (1996)^[1]一样, 本文假设交易者在有限的天数内进行交易, 但是一天之内的交易是连续的, 假定一天之内的交易存在基于公司特有的私人信息进行的知情交易和由于市场指令流冲击引起的交易以及外生的流动性交易(即非知情交易)。假定由于市场指令流冲击引起的交易者、非知情交易者和知情交易者的到达由相互独立的泊松过程决定^①。来自于前一天的每个资产的所有信息价值在交易开始之前被揭示。

因此, 本文修正后的PIN模型与原来经典的PIN模型存在以下区别。首先, 本文对公司特有的知情事件发生时买者和卖者采用不同的知情交易到达率。本文用 μ^b 表示公司特有的知情事件发生时买者的知情交易到达率, 用 μ^s 表示公司特有的知情事件发生时卖者的知情交易到达率。这样做的原因在于在实际数据中, 买单数量的方差通常总是比卖单数量的方差更大, 对模型做这一改动之后可以使新的模型更加能够与实际数据的特征相一致。其次, 本文在模型中加入了驱动整个市场交易增加的因素, 即市场指令流冲击。加入这一因素后可以使得在每个交易日中当市场指令流冲击事件发生时买单和卖单的数量都增加, 从而使得模型隐含的买单和卖单数量之间的相关性为正, 并且买单和卖单数量的变动也增加, 以便更好地与实际数据的特征相匹配。在修正的PIN模型中, 我们用 λ 表示市场指令流冲击事件发生的可能性。同样, 本文也对市场指令流冲击事件发生时买者和卖者采用不同的交易到达率。本文用 γ_b 表示市场指令流冲击事件发生时买者的交易到达率, 用 γ_s 表示市场指令流冲击事件发生时卖者的交易到达率。

^① Easley, Kiefer and O'Hara (1997b)^[8]检验了不同交易日之间知情事件的独立性, 发现他们不能拒绝不同交易日之间的知情事件相互独立的假设。因此, 我们在修正的 PIN 模型中直接做出这一假定。

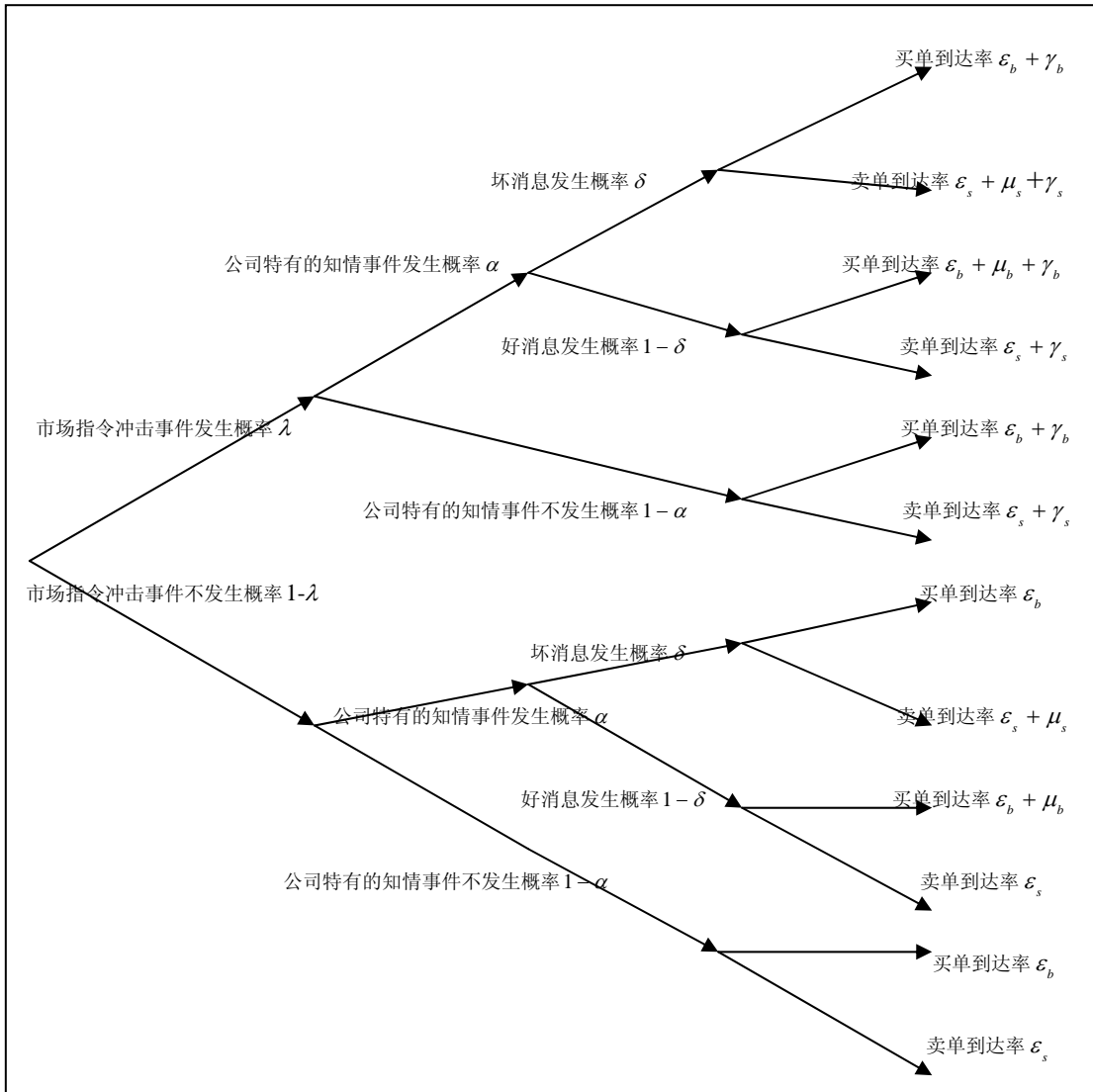


图 2.2 修正的 PIN 模型中股票在一个交易日的交易过程

如图 2.2 所示，在第一个节点即每个交易日开盘之前，自然决定该交易日市场指令流冲击事件是否发生。市场指令流冲击事件发生的概率为 λ ，市场指令流冲击事件没有发生的概率为 $1-\lambda$ 。自然接下来决定与资产真实价值相关的公司特有的知情事件是否发生，公司特有的知情事件发生的概率为 α ，公司特有的知情事件没有发生的概率则为 $1-\alpha$ 。如果知情事件发生，自然再将之区分为坏消息和好消息，其中，坏消息发生的概率为 δ ，好消息发生的概率为 $1-\delta$ 。给定已区分的事件之后，买卖单到达率都服从泊松分布。当市场指令流冲击事件发生、且公司特有的坏消息发生时，买单到达率为 $\varepsilon_b + \gamma_b$ ，卖单到达率为 $\varepsilon_s + \mu_s + \gamma_s$ ；当市场指令流

冲击事件发生、且公司特有的好消息发生时，买单到达率为 $\varepsilon_b + \mu_b + \gamma_b$ ，卖单到达率为 $\varepsilon_s + \gamma_s$ ；当市场指令流冲击事件发生、且公司特有的知情事件没有发生时，买单到达率为 $\varepsilon_b + \gamma_b$ ，卖单到达率为 $\varepsilon_s + \gamma_s$ ；当市场指令流冲击事件没有发生、且公司特有的坏消息发生时，买单到达率为 ε_b ，卖单到达率为 $\varepsilon_s + \mu_s$ ；当市场指令流冲击事件没有发生、且公司特有的好消息发生时，买单到达率为 $\varepsilon_b + \mu_b$ ，卖单到达率为 ε_s ；当市场指令流冲击事件没有发生、且公司特有的知情事件也没有发生时，买单到达率为 ε_b ，卖单到达率为 ε_s 。

（二）修正的 PIN 模型对信息风险的测度

类似地，信息风险可以由修正的 PIN 模型中知情交易者发出的交易指令占总交易指令的概率即新的知情交易概率（New Probability of Informed Trading, NPIN）来测度。在修正的 PIN 模型中，新的基于公司特有的私人信息的知情交易概率为：

$$NPIN = \frac{\alpha \times (\delta \times \mu_s + (1 - \delta) \times \mu_b)}{\alpha \times (\delta \times \mu_s + (1 - \delta) \times \mu_b) + \lambda \times (\gamma_b + \gamma_s) + \varepsilon_b + \varepsilon_s} \quad (2.17)$$

同时，我们也可以利用修正的 PIN 模型估计出市场指令流冲击事件发生时引起的交易概率（Probability of Market Order Flow Shock, PMOS）。在修正的 PIN 模型中，由于市场指令流冲击事件发生时引起的交易概率为：

$$PMOS = \frac{\lambda \times (\gamma_b + \gamma_s)}{\alpha \times (\delta \times \mu_s + (1 - \delta) \times \mu_b) + \lambda \times (\gamma_b + \gamma_s) + \varepsilon_b + \varepsilon_s} \quad (2.18)$$

四、修正的 PIN 模型的参数估计

与估计经典的 PIN 模型相比，对修正的 PIN 模型的参数估计需要估计出参数向量 $\theta = (\lambda, \alpha, \delta, \mu_b, \mu_s, \gamma_b, \gamma_s, \varepsilon_b, \varepsilon_s)$ ，这比估计经典的 PIN 模型更加复杂，因为我们不能直接观察到任何知情事件和市场指令流冲击事件的到达。参数 λ 决定了市场指令流冲击事件发生的概率，参数 α 和 δ 决定了三种知情事件（没有消息（ n ）、坏消息（ b ）和好消息（ g ））的概率，但是它们都是不可直接观察到的。剩下的

参数涉及到市场指令流冲击事件发生时买卖单的到达率以及非知情交易者和知情交易者的买卖单到达率。我们只能观察到买单和卖单的到达率，但是不能直接观察到哪些买卖单是由于市场指令流冲击事件发生时引起的，哪些买卖单是非知情交易者提交的，哪些买卖单是知情交易者提交的。因此，估计所有这些参数需要一个结构化的模型。修正的 PIN 模型提供了从可观察的变量，即买单和卖单中提取参数信息所必需的结构。

模型中，在每一个交易日中根据市场指令流冲击事件是否发生以及公司特有的私人消息为好消息、坏消息和没有消息总共区分为六个分支，每个分支的买卖单到达率都服从泊松分布。虽然我们无法得知该交易日属于哪一个分支，但却可以从交易资料中得到可能的结果。当该交易日买单数量较多时，可以预计该日发生市场指令流冲击事件和属于好消息层面的概率较大；类似地，当该交易日卖单数量较多时，可以预计该日发生市场指令流冲击事件和属于坏消息层面的概率较大。对于每个交易日内情况的六种分支以及交易指令的到达率本文利用一个混合概率模型来描述，将每一分支所属概率的权重乘以该分支的泊松分布，形成一个混合的泊松分布。假设我们考虑一个市场指令流冲击事件发生、且公司特有的坏消息发生时的交易日，买单到达率为 $\varepsilon_b + \gamma_b$ ，即市场指令流买单和非知情交易者的买单的加总；卖单到达率为 $\varepsilon_s + \mu_s + \gamma_s$ ，即市场指令流卖单、知情交易者、和非知情交易者卖单的加总。

因此，在市场指令流冲击事件发生、且有关公司特有的坏消息发生时的交易日内观察到的包含 B 个买单和 S 个卖单的交易指令的独立的泊松分布的似然函数为：

$$e^{-(\varepsilon_b + \gamma_b)} \frac{(\varepsilon_b + \gamma_b)^B}{B!} e^{-(\varepsilon_s + \mu_s + \gamma_s)} \frac{(\varepsilon_s + \mu_s + \gamma_s)^S}{S!} \quad (2.19)$$

类似地，在市场指令流冲击事件发生、且有关公司特有的好消息发生时的交易日内观察到的包含 B 个买单和 S 个卖单的交易指令的似然函数为：

$$e^{-(\varepsilon_b + \mu_b + \gamma_b)} \frac{(\varepsilon_b + \mu_b + \gamma_b)^B}{B!} e^{-(\varepsilon_s + \gamma_s)} \frac{(\varepsilon_s + \gamma_s)^S}{S!} \quad (2.20)$$

在市场指令流冲击事件发生、且有关公司特有的知情事件没有发生时的交易

日内观察到包含 B 个买单和 S 个卖单的交易指令的似然函数为：

$$e^{-(\varepsilon_b + \gamma_b)} \frac{(\varepsilon_b + \gamma_b)^B}{B!} e^{-(\varepsilon_s + \gamma_s)} \frac{(\varepsilon_s + \gamma_s)^S}{S!} \quad (2.21)$$

在市场指令流冲击事件没有发生、且有关公司特有的坏消息发生时的交易日内观察到的包含 B 个买单和 S 个卖单的交易指令的独立的泊松分布的似然函数为：

$$e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-(\varepsilon_s + \mu_s)} \frac{(\varepsilon_s + \mu_s)^S}{S!} \quad (2.22)$$

类似地，在市场指令流冲击事件没有发生、且有关公司特有的好消息发生时的交易日内观察到的包含 B 个买单和 S 个卖单的交易指令的似然函数为：

$$e^{-(\varepsilon_b + \mu_b)} \frac{(\varepsilon_b + \mu_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!} \quad (2.23)$$

最后，在市场指令流冲击事件没有发生、且有关公司特有知情事件也没有发生时的交易日内观察到包含 B 个买单和 S 个卖单的交易指令的似然函数为：

$$e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!} \quad (2.24)$$

因此，在估计参数时只需要知道在一个交易日内买单的总数量和卖单的总数量。当不知道该交易日属于何种消息时，可以根据每一个分支发生的概率进行加权平均得到混合的似然函数。也就是说，市场指令流冲击事件发生且公司特有的坏消息发生时的概率为 $\lambda\alpha\delta$ ；市场指令流冲击事件发生且公司特有的好消息发生时的概率为 $\lambda\alpha(1-\delta)$ ；市场指令流冲击事件发生且公司特有的知情事件没有发生时的概率为 $\lambda(1-\alpha)$ ；市场指令流冲击事件没有发生且公司特有的坏消息发生时的概率为 $(1-\lambda)\alpha\delta$ ；市场指令流冲击事件没有发生且公司特有的好消息发生时的概率为 $(1-\lambda)\alpha(1-\delta)$ ；市场指令流冲击事件没有发生且公司特有的知情事件也没有发生时的概率为 $(1-\lambda)(1-\alpha)$ 。因此，修正的 PIN 模型的混合似然函数可以写为：

$$\begin{aligned}
 L(\theta|B, S) = & \lambda\alpha\delta e^{-(\varepsilon_b+\gamma_b)} \frac{(\varepsilon_b+\gamma_b)^B}{B!} e^{-(\varepsilon_s+\mu_s+\gamma_s)} \frac{(\varepsilon_s+\mu_s+\gamma_s)^S}{S!} \\
 & + \lambda\alpha(1-\delta) e^{-(\varepsilon_b+\mu_b+\gamma_b)} \frac{(\varepsilon_b+\mu_b+\gamma_b)^B}{B!} e^{-(\varepsilon_s+\gamma_s)} \frac{(\varepsilon_s+\gamma_s)^S}{S!} \\
 & + \lambda(1-\alpha) e^{-(\varepsilon_b+\gamma_b)} \frac{(\varepsilon_b+\gamma_b)^B}{B!} e^{-(\varepsilon_s+\gamma_s)} \frac{(\varepsilon_s+\gamma_s)^S}{S!} \\
 & + (1-\lambda)\alpha\delta e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-(\varepsilon_s+\mu_s)} \frac{(\varepsilon_s+\mu_s)^S}{S!} \\
 & + (1-\lambda)\alpha(1-\delta) e^{-(\varepsilon_b+\mu_b)} \frac{(\varepsilon_b+\mu_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!} \\
 & + (1-\lambda)(1-\alpha) e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!}
 \end{aligned} \tag{2.25}$$

在一个给定的交易日，利用最大似然估计法所估计出来的参数 λ 、 α 和 δ 为介于 0 到 1 之间的概率，反映市场指令流冲击事件在一个交易日发生的可能性、知情事件在一个交易日发生的可能性和公司特有的坏消息发生的概率。在交易日之间，可以利用每个交易日所累积的总买卖单数量来估计出日间的这些参数。由于交易日之间彼此独立， I 个观察日之内的买卖单数量 $M = (B_i, S_i)_{i=1}^I$ 的似然函数为每天的似然函数的乘积：

$$L(M|\theta) = \prod_{i=1}^I L(\theta|B_i, S_i) \tag{2.26}$$

为了从任意的数据集 M 中估计出参数向量 θ ，只需要最大化方程 (2.26) 中的似然函数。这种方法提供了对一只股票上的市场指令流冲击事件发生引起的交易以及知情和非知情交易的直接估计方法。因此，利用市场上所观察到的交易资料作为输入值，以最大似然法估计参数向量 $\theta = (\lambda, \alpha, \delta, \mu_b, \mu_s, \gamma_b, \gamma_s, \varepsilon_b, \varepsilon_s)$ ，便能够计算出新的信息风险 $NPIN$ 和市场指令流冲击事件发生时引起的交易概率 $PMOS$ 。

第3章 中国股票市场信息风险测度实证研究

§ 3.1 中国股票市场交易机制概述

目前对股票信息风险进行研究的国外大部分文献都针对报价驱动市场 (Quote Driven Market) 进行研究。在报价驱动市场上, 每只股票都有做市商 (Market Maker) 为它做市, 也就是提供买卖报价, 并在自己所报的价格水平上接受交易者的买卖。而我国深、沪两市都是订单驱动市场 (Order Driven Market), 是采取电子竞价交易的方式进行交易的。因此, 在本节我们首先对中国股票市场的交易机制进行简单的介绍。

一、中国股票市场的基本交易机制

中国股票市场主要以上海证券交易所和深圳证券交易所为代表, 它们具体的交易机制如下:

深圳证券交易所和上海证券交易所实行的是会员制的组织形式, 所有交易均在交易大厅中进行, 参加交易的人员是固定的, 即证券交易所的会员。在我国目前只允许具有中国证券监督管理委员会颁发的证券经营资格的法人成为交易所的会员, 因此, 我国广大的投资者只有通过证券经纪商才能参与交易所的交易。投资者的股票买卖指令是通过交易所的会员报送到交易所电脑主机进行撮合的。目前的报盘方式有两种, 分为有形席位和无形席位。有形席位是属于一种场内报盘, 即证券公司在柜台接到投资者的委托指令并审查后, 以电话等通讯方式向驻交易所的场内交易员 (因穿红马甲工作而被称为“红马甲”) 转达买卖指令; 该证券公司场内交易员接到指令后, 将其输入交易所的电脑主机。无形席位是属于一种场外报盘, 它无需证券公司派驻场内交易员, 而由投资者在场外通过证券公司的电脑终端直接向交易所电脑主机输入买卖证券的指令。两个交易所都只接受会员的限价申报, 但会员可以接受投资者的限价委托或市价委托, 委托当日有效, 且投资者可以撤销委托的未成交部分。申报价格的最小变动单位 (tick size) 为 0.01 元, 买入申报数量必须为 100 股或其整数倍。

交易所交易在每周一至周五进行（国家法定假日和交易所公告的休市日除外）。上午 9:15~9:25 为集合竞价时间。集合竞价遵循价格优先、同一价格下时间优先的原则，对所有有效委托进行集中处理，并按集中竞价规则产生开盘价。集中竞价中未成交的订单进入后面的连续竞价交易，交易采取价格优先和时间优先的原则，由电脑对有效委托进行逐笔处理，交易时间为上午的 9:30~11:30 和下午的 13:00~15:00。

二、中国股票市场流动性的提供

从上面对我国股票市场交易机制的概述中可以看出，在我国的股票市场中不存在专门的机构为股票提供买卖报价，即不存在做市商，那么流动性由谁来提供呢？实际上，每个投资者既是流动性的“消费者”，又是流动性的提供者。当投资者向交易所的电脑主机提交订单时，交易所电脑主机会自动对订单进行判断，如果目前主机的订单簿上有合适的订单与之匹配，投资者的这笔交易马上就成交，此时投资者就是流动性的“消费者”；而如果主机的订单簿上没有合适的订单可以与之成交，那么他的限价订单就进入订单簿，等待合适的可与之成交的订单。此时，订单簿实际上起了一个蓄水池的作用，把所有未成交的限价买单和卖单都累积起来，从而为连续交易提供了条件。因此，实际上每个提供限价订单的投资者都起到了做市商的作用，只是他们是不自觉地扮演了这个角色^①。所以，我们可以近似地把订单驱动市场描述成一个存在众多“做市商”共同做市的市場，他们的报价相互竞争，这正好解释了为什么大量的实证研究都发现订单驱动市场的交易成本要小于报价驱动市场的交易成本。

因此，在假定做市商为风险中性的及其期望利润为零的情况下，我们可以应用 Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的经典的 PIN 模型以及本文在该模型基础上提出的修正的 PIN 模型对我国股票的信息风险进行测度。

^① 正如 Brockman and Chung (1999)^[70]指出的，在订单驱动市场上，公众的限价订单为市场提供了流动性，同时（最优的买价和最优的卖价之间）形成了买卖价差。

§ 3.2 经典的 PIN 模型的估计结果

在本节,我们首先基于我国股票市场上的交易数据,利用 Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的经典的 PIN 模型来对我国股票的信息风险进行测度。

一、数据及描述性统计

Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的经典的PIN模型的估计需要每只股票每天由买方和卖方发起的交易数量的数据。为了计算每天买方和卖方发起交易的数量,本文使用来自于万得金融高频数据库的分笔数据。万得金融高频数据库的分笔数据包含每个交易日内的每分钟内的每笔成交报价数据。分笔数据中包含证券代码、日期、时间、最新价、本笔成交量、本笔成交金额、买卖方向等共计三十六个字段的数据。由于该数据中包含了买卖方向,因此我们就不需要像国外学者一样在研究中对于没有提供买卖方向的交易数据通常采用Lee and Ready (1991)^[39]的买卖判断准则^①。由于数据的可得性,我们只有2007年一年内共计241个交易日的日内实时交易数据^②。我们选择的样本为上证180指数的成份股^③。

为了研究交易活跃程度不同的股票其信息风险大小之间的区别^④,本文按样本期内日均换手率的高低将上证180指数的成份股股票分为5组,为了避免新股上市首日的高换手率对分组的影响,本文剔除了上证180指数的成份股在2007年中上市的股票,这样样本中只有157只股票。换手率的定义为:

$$\text{换手率} = \frac{\text{成交股数}}{\text{流通在外的股数}} \times 100\%$$

鉴于相邻两组之间的差别不是很大,为了更好地比较交易活跃程度不同的股

^① PIN模型的估计需要判断一笔交易是由买方发起还是由卖方发起,而大部分交易数据都不提供交易的买卖方向,对此国外学者通常采用Lee and Ready (1991)^[39]的买卖判断准则,即将一笔交易的成交价与买卖报价的中点相比较,高于中点的这笔交易为卖,低于中点的交易为买,如果成交价与买卖报价的中点相同,则与上一笔成交价相比较,以此来判断出一笔交易的买卖方向。当然,这种判断准则不一定完全准确。Boehmer, Grammig and Theissen (2007)^[15]的研究表明,如果将买卖方向错误分类的话,会导致下偏的PIN估计值。

^② Easley, Kiefer and O'Hara (1997a)^[7]的研究表明只需要60个交易日的数据就可以充分获得PIN模型的可靠参数估计值。

^③ 基于其客观科学的选择方法,通常认为上证180指数的成份股具有很好的代表性,能够较好地代表上海股票市场的概貌和运行状况。

^④ 大部分学者都根据股票交易量的高低来衡量股票交易活跃程度(如Easley, Kiefer, O'Hara and Paperman (1996)^[1]、Brockman and Chung (2000)^[10]、Barclay and Hendershott (2003)^[32]等),但是本文认为股票交易量的高低还受到股票自身规模的影响,而换手率能够更加准确地反映股票交易的活跃程度。

票其信息风险大小之间的区别，本文选择换手率最低、换手率居中和换手率最高的三组股票作为我们最后的研究样本，每组股票平均在 30 只左右^①。本文剔除了发生在开盘之前的所有交易数据以避免包括发生在开盘期间的竞价交易。

表 3.1 列出了换手率低、换手率中和换手率高的三组股票在 2007 年内的 241 个交易日中平均每日换手率的描述性统计结果。

表 3.1 各组股票在样本期内的日均换手率描述统计 (单位: %)

组别	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
换手率低	1.80	1.98	2.53	0.77	0.60	-0.39	1.63	2.60	0.27
换手率中	3.77	3.69	4.17	3.28	0.28	0.04	1.62	1.98	0.37
换手率高	5.71	5.69	6.69	5.07	0.42	0.54	2.66	1.35	0.51

从表 3.1 中可以看出，三组股票在样本期内的日均换手率上存在较大的差异，换手率最低的一组股票的平均换手率为 1.80%，换手率居中的一组股票的平均换手率为 3.77%，换手率最高的一组股票的平均换手率为 5.71%。在换手率最低的一组股票中，换手率的最小值为 0.77%，最大值为 2.53%，中位数为 1.98%；在换手率居中的一组股票中，换手率的最小值为 3.280%，最大值为 4.17%，中位数为 3.69%；在换手率最高的一组股票中，换手率的最小值为 5.07%，最大值为 6.69%，中位数为 5.69%。

为了检验三组股票的换手率之间的差异在统计上是否显著，本文对三组股票的换手率进行了统计检验。由于不知道各组股票的换手率的分布情况，因此本文选择了 Kruskal-Wallis 非参数检验方法。

Kruskal-Wallis 非参数检验的原假设为：多个独立样本来自的多个总体分布无显著差异。设样本观察值的总数为 N ， $N = \sum_{i=1}^k n_i$ ，其中 n_i 为第 i 组的观察值个数。

Kruskal-Wallis 非参数检验的步骤为：(1) 将样本观察值从小到大进行排列并赋予它们秩，令 $R(X_{ij})$ 表示赋给 X_{ij} 的秩， \bar{R}_i 为赋给第 i 个样本的秩的平均，即：

^① 本文剔除了模型的参数估计不收敛的股票，最后每组股票数量各为 25 只。对各只样本股票的具体说明见附录 1。

$\bar{R}_i = \sum_{j=1}^{n_i} R(X_{ij}) / n_i, i=1,2,\dots,k$ 。如果观察值相等，就赋予其平均秩。(2) 求检验

的统计量 $T = \frac{1}{S^2} \sum_{i=1}^k n_i (\bar{R}_i - \frac{N+1}{2})^2$ ，其中， $S^2 = \frac{1}{N-1} \sum_{i,j} [R(X_{ij}) - \frac{N+1}{2}]^2$ ，该统计量

渐进服从自由度为 k 的 χ^2 分布，然后对检验的统计量与一定置信水平下相应自由度的 χ^2 分布的临界值进行比较，并得出检验结论。当检验的统计量大于一定置信水平下的 χ^2 分布的临界值时，则拒绝原假设，反之则接受原假设。

对各组股票换手率之间的差异进行的 Kruskal-Wallis 非参数检验结果在 1% 的置信度水平下拒绝了原假设，这表明各组股票换手率之间的差异在 1% 的置信度水平下是显著的。

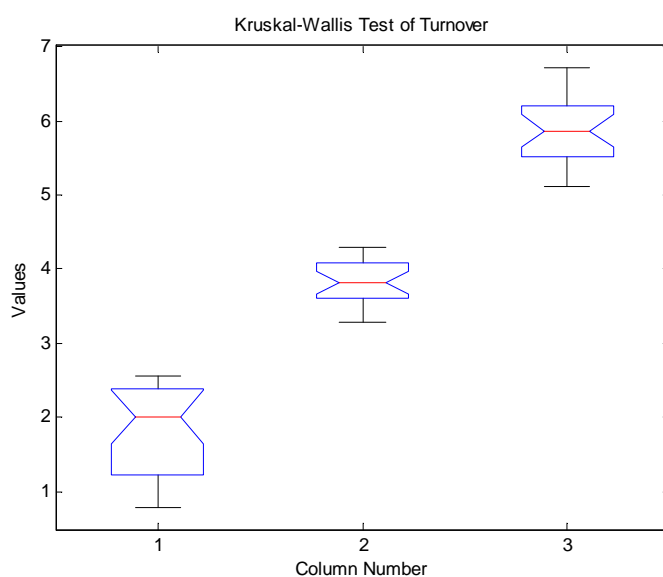


图 3.1 对各组股票换手率的 Kruskal-Wallis 非参数检验

图 3.1 显示了对各组股票换手率之间的差异进行的 Kruskal-Wallis 非参数检验图。从图 3.1 中可以看出，各组股票的换手率之间存在明显的差异。

表 3.2 列出了换手率最低、换手率居中和换手率最高的三组股票在样本期内的每日平均买卖单数量的描述性统计。

表 3.2 各组股票在样本期内的日均买卖单数量的描述性统计

组别	买/卖 单	均值	中位 数	最大 值	最小 值	标准 差	偏度	峰度	J-B	P 值
换手 率低	买单	853.04	798	1267	483	226.57	0.19	2.15	0.90	0.64
	卖单	821.80	785	1136	497	188.87	0.03	2.13	0.80	0.67
换手 率中	买单	1058.08	1088	1270	751	146.59	-0.46	2.33	1.36	0.51
	卖单	989.36	993	1152	692	124.90	-0.66	2.57	2.02	0.36
换手 率高	买单	1051.72	1082	1242	851	103.62	-0.15	2.14	0.86	0.65
	卖单	982.80	998	1132	839	83.59	-0.22	1.88	1.51	0.47

从表 3.2 中可以看出, 换手率最低的一组股票在样本期内的每日买单数量的均值为 867 笔, 每日卖单数量的均值为 832 笔; 换手率居中的一组股票在样本期内的每日买单数量的均值为 1058 笔, 每日卖单数量的均值为 989 笔; 换手率最高的一组股票在样本期内的每日买单数量的均值为 1052 笔, 每日卖单数量的均值为 983 笔。

表 3.3 列出了换手率最低、换手率居中和换手率最高的三组股票在样本期内的每日平均买单数量、卖单数量和样本期内的买单方差、卖单方差以及样本期内的买单和卖单数量之间的相关性在第 95 分位、第 75 分位、第 50 分位、第 25 分位和第 5 分位的统计数。

表 3.3 各组股票的买单和卖单数量、方差及其相关性的分位数统计

组别	类别	第 95 分位	第 75 分位	中位数	第 25 分位	第 5 分位
换手率低	买单数量	1232	1022	798	703	488
	卖单数量	1123	996	785	714	506
	买单方差	71699	60254	48957	38769	33857
	卖单方差	60106	47796	37870	28473	23036
	买卖单数 量相关性	0.69	0.60	0.50	0.43	0.30

换手率中	买单数量	1264	1171	1088	935	780
	卖单数量	1137	1096	993	902	751
	买单方差	78223	60464	52148	47824	41658
	卖单方差	63870	52347	43476	35982	27940
	买卖单数量相关性	0.75	0.68	0.60	0.54	0.43
换手率高	买单数量	1221	1131	1082	967	888
	卖单数量	1100	1052	998	913	850
	买单方差	77211	62894	58943	51057	45115
	卖单方差	59195	53550	45943	40349	34579
	买卖单数量相关性	0.82	0.75	0.70	0.62	0.43

从表 3.3 中可以看出, 换手率最低的一组股票的买单数量方差的取值范围为从第 5 分位的 33857 到第 95 分位的 71699, 卖单数量方差的取值范围为从第 5 分位的 23036 到第 95 分位的 60106, 买单和卖单数量之间的相关性为从第 5 分位的 0.30 到第 95 分位的 0.69; 换手率居中的一组股票的买单数量方差的取值范围为从第 5 分位的 41658 到第 95 分位的 78223, 卖单数量方差的取值范围为从第 5 分位的 27940 到第 95 分位的 63870, 买单和卖单数量之间的相关性为从第 5 分位的 0.43 到第 95 分位的 0.75; 换手率最低的一组股票的买单数量方差的取值范围为从第 5 分位的 45115 到第 95 分位的 77211, 卖单数量方差的取值范围为从第 5 分位的 34579 到第 95 分位的 59195, 买单和卖单数量之间的相关性为从第 5 分位的 0.43 到第 95 分位的 0.82。可见, 换手率不同的各组股票的买单和卖单数量之间的相关性都为正, 而且买单和卖单数量之间的这种正相关性与股票的换手率之间成正相关关系, 即换手率越高的股票其买单和卖单数量之间的正相关性越高。

二、经典的 PIN 模型的参数估计

最大化公式 (2.16) 中的联合似然函数^①可以得到经典的 PIN 模型的参数估计

^① 我们此处最大化的似然函数与公式 (2.16) 有些微的差别, 那就是不再假设非知情交易者的买单到达率和卖

值。然而，对于该最大化问题没有解析解，只能采用数值最大化技术来得到该最大化问题的解。Easley, Hvidkjaer and O'Hara (2005)^[70]推荐使用对联合似然函数的下列因式分解^①来使得数值最大化过程变得更加容易。

$$L\left(\left(B_i, S_i\right)_{i=1}^I \mid \theta\right) = \sum_{i=1}^I \left[-\varepsilon_b - \varepsilon_s + M_i (\ln x_b + \ln x_s) + B_i \ln (\varepsilon_b + \mu) + S_i \ln (\varepsilon_s + \mu) \right] \\ + \sum_{i=1}^I \ln \left[\alpha (1 - \delta) e^{-\mu} x_s^{S_i - M_i} x_b^{-M_i} + \alpha \delta e^{-\mu} x_b^{B_i - M_i} x_s^{-M_i} + (1 - \alpha) x_b^{B_i - M_i} x_s^{S_i - M_i} \right]$$

其中， $M_i = (\min(B_i, S_i) + \max(B_i, S_i)) / 2$ ， $x_s = \frac{\varepsilon_s}{\varepsilon_s + \mu}$ ， $x_b = \frac{\varepsilon_b}{\varepsilon_b + \mu}$ 。采用这种因

式分解的好处是为了提高计算的效率。这对于那些有数量巨大的买单和卖单的股票尤其重要，因为如果没有因式分解，经常会造成大多数计算软件的超值问题。

在估计参数的过程中，为了尽可能避免数值最大化过程产生角点解（Corner Solution）时参数估计的误差问题，我们对每只股票都代入参数空间中的许多不同的初始值来求该最大化问题，以便对于每只股票都找到全局最优解。由于初始值的不同会对参数估计值产生较大的影响，本文对初始值的选取采用 Yan and Zhang (2006)^[71]提出的方法。

经典的 PIN 模型需要估计的参数向量为 $\theta = (\alpha, \delta, \mu, \varepsilon_b, \varepsilon_s)$ ，其中 α 为知情事件发生的概率， δ 为知情事件发生时为坏消息的概率，它们的取值范围都在 0 到 1 之间； μ 为知情交易者的买卖单到达率， ε_b 和 ε_s 分别为非知情交易者的买单到达率和卖单到达率，它们的取值范围在 0 到正无穷。对于有边界的参数 α 和 δ ，可以在其取值范围内选取等距离的值作为初始值。然而，对于没有边界的参数 μ 、 ε_b 和 ε_s ，选取它们的初始值并不容易。

为了解决参数 μ 、 ε_b 和 ε_s 的初始值选取问题，可以采用每日买单和卖单数量的概率分布的矩条件。根据每日买单数量 B 和卖单数量 S 的联合概率分布，可以推

单到达率相同，而是假设非知情交易者的买单到达率为 ε_b ，非知情交易者的卖单到达率为 ε_s ，因为这与实际数据更加匹配。Easley, O'Hara and Saar (2001)^[11]开始将非知情交易者的买单和卖单到达率分别设为 ε_b 和 ε_s ，在随后有关股票信息风险的研究中，学者们都采取了这种假定，如 Straser (2002)^[31]、Brown, Hillegeist and Lo (2004)^[33]和 Vega (2006)^[14]等。

^① Easley, Hvidkjaer and O'Hara (2005)^[70]并没有给出对该因式分解的推导，本文对该因式分解进行了推导，具体推导过程见附录 2。

导出每日买单数量 B 和卖单数量 S 的边际期望值分别^①为：

$$E(B) = \varepsilon_b + \alpha \times (1 - \delta) \times \mu \quad (3.1)$$

$$E(S) = \varepsilon_s + \alpha \times \delta \times \mu \quad (3.2)$$

这两个公式可以帮助我们解决五个参数初始值的选取问题。首先将间隔 $[0,1]$ 进行等分并选择等距离的值作为参数 α 和 δ 的初始值。然后，用每日买单和卖单数量的样本平均值 \bar{B} 和 \bar{S} 来代替上述公式 (3.1) 和 (3.2) 中的 $E(B)$ 和 $E(S)$ 。公式 (3.1) 中等式右边的第二项总是正的 ($\alpha \times (1 - \delta) \times \mu > 0$)，因此 ε_b 必须小于 \bar{B} 。这意味着可以选取 \bar{B} 的部分来作为 ε_b 的初始值。最后，解联立方程 (3.1) 和 (3.2) 可以求出 ε_s 和 μ 的初始值。

简而言之，五个参数的初始值可以由下列公式来选取：

$$\alpha^0 = \alpha_i, \quad \delta^0 = \delta_j, \quad \varepsilon_b^0 = \omega_k \cdot \bar{B}, \quad \mu^0 = \frac{\bar{B} - \varepsilon_b^0}{\alpha^0 \cdot (1 - \delta^0)}, \quad \varepsilon_s^0 = \bar{S} - \alpha^0 \cdot \delta^0 \cdot \mu^0$$

其中， α^0 、 δ^0 、 ε_b^0 、 μ^0 和 ε_s^0 分别为需要估计的五个参数的初始值， α_i 、 δ_j 和 ω_k 为从 $(0.1, 0.3, 0.5, 0.7, 0.9)$ 中的取值。 α_i 、 δ_j 和 ω_k 的组合可以产生 125 个初始值的集合^②。

根据这一初始值的选取方法，本文在参数估计的过程中首先根据样本内的每只股票的每日买单和卖单数量构造出 125 个初始值的集合；然后代入每一组合理的初始值进入最大化问题的求解过程；最后选取似然函数值最大的参数估计值作为模型的最后参数估计值。

表 3.4 列出了换手率最低、换手率居中和换手率最高的三组股票在样本期内的经典的 PIN 模型的各项参数的估计值和信息风险在第 95 分位、第 75 分位、第 50 分位、第 25 分位和第 5 分位的统计数。

^① 对于每日买单数量 B 和卖单数量 S 的边际期望值的具体推导过程见附录 3。

^② Yan and Zhang (2006)^[71] 尝试了令 α_i 、 δ_j 和 ω_k 为从 $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ 中的取值。这样的组合可以产生 729 个初始值的集合，但是他们的研究发现做这一更精细的划分除了增加计算时间外并没有显著的差异。

表 3.4 经典的 PIN 模型中各组股票的参数估计值及信息风险的分位数统计

组别	类别	第 95 分位	第 75 分位	中位数	第 25 分位	第 5 分位
换手率低	α	0.77	0.62	0.50	0.45	0.35
	δ	0.64	0.45	0.34	0.26	0.15
	μ	410.58	374.56	339.15	313.35	284.80
	ε_b	1016.06	906.56	702.89	614.12	422.00
	ε_s	1052.34	892.07	750.32	655.97	464.80
	PIN	0.14	0.12	0.11	0.10	0.09
换手率中	α	0.78	0.73	0.68	0.58	0.45
	δ	0.88	0.22	0.12	0.05	0.02
	μ	502.83	430.00	388.59	347.94	330.20
	ε_b	1106.41	910.41	840.95	796.17	644.08
	ε_s	1111.59	1069.23	925.68	836.50	731.53
	PIN	0.16	0.14	0.12	0.10	0.10
换手率高	α	0.76	0.68	0.63	0.54	0.46
	δ	0.51	0.18	0.06	0.02	0.01
	μ	486.87	452.82	406.54	366.30	323.07
	ε_b	1026.55	837.81	795.52	780.96	700.14
	ε_s	1090.53	1023.61	967.80	877.22	828.03
	PIN	0.16	0.15	0.12	0.10	0.09

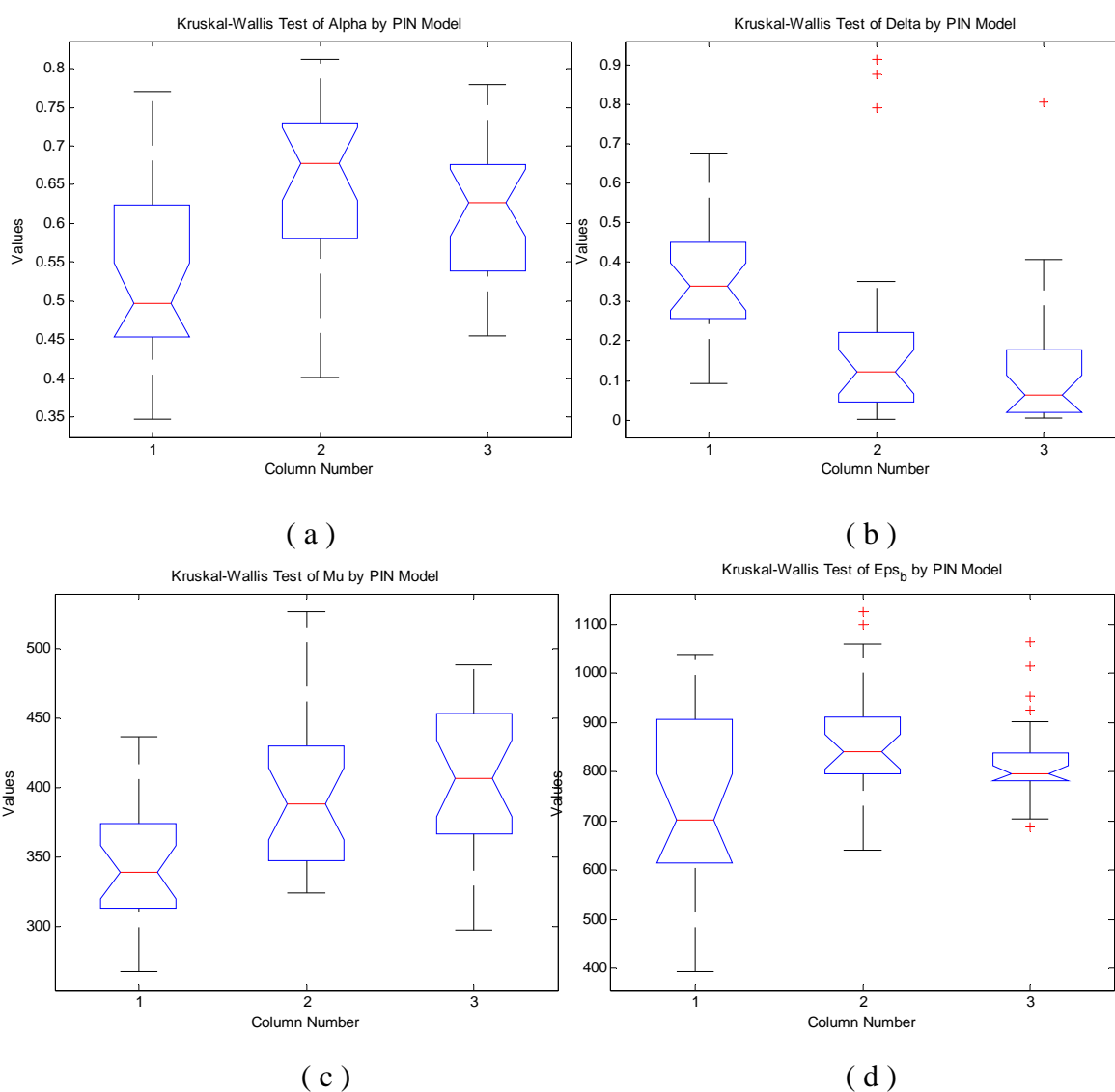
从表 3.4 中可以看出,知情事件发生的概率 α 在换手率最低的一组股票中的中位数为 0.5,其变动范围为从第 5 分位的 0.35 到第 95 分位的 0.77;知情事件发生的概率 α 在换手率居中的一组股票中的中位数为 0.68,其变动范围为从第 5 分位的 0.45 到第 95 分位的 0.78;知情事件发生的概率 α 在换手率最高的一组股票中的中位数为 0.63,其变动范围为从第 5 分位的 0.46 到第 95 分位的 0.76。可以看出知情事件在换手率最低的一组股票中发生的概率低于换手率居中和换手率最高的股票。

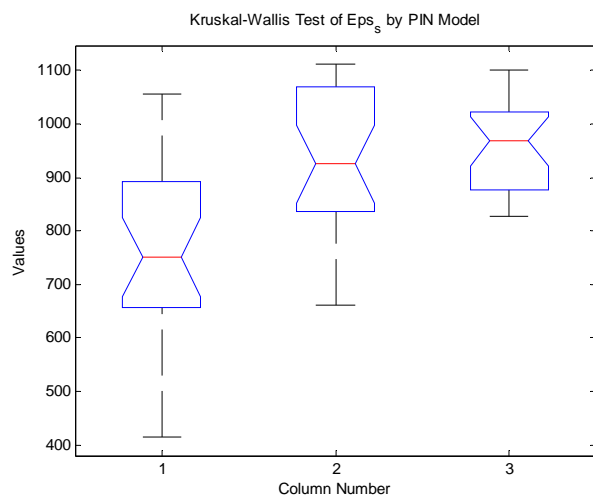
为了比较各组股票由经典的 PIN 模型估计得到的参数的分布在统计上是否存在显著差异,本文同样对各组股票的参数之间的差异进行 Kruskal-Wallis 非参数检验,检验结果见表 3.5。

表 3.5 经典的 PIN 模型估计得到的参数分布的 Kruskal-Wallis 非参数检验

参数	P 值
α	0.0016
δ	1.2761e-06
μ	0.0002
ε_b	0.0252
ε_s	3.9321e-05

从表 3.5 中可以看出，对各组股票由经典的 PIN 模型估计得到的参数进行的 Kruskal-Wallis 非参数检验结果都在 95% 的置信水平下拒绝了各组股票参数的分布相同的原假设，这表明各组股票的参数之间存在显著差异。





(e)

图 3.2 经典的 PIN 模型估计得到的参数 Kruskal-Wallis 非参数检验

图 3.2 显示了各组股票由经典的 PIN 模型估计得到的参数的 Kruskal-Wallis 非参数检验图。从图中可以比较直观地发现各组股票由经典的 PIN 模型估计得到的参数之间确实存在较为显著的差异。

为了进一步比较任意两组股票由经典的 PIN 模型估计出来的参数分布之间是否存在显著的差异，本文还进行了 Mann-Whitney 非参数检验。

Mann-Whitney 非参数检验又称为 Wilcoxon Rank Sum 非参数检验，其基本思想是如果两组样本的总体分布相同，那么两组样本都来自于同一个总体分布，此时任意一组样本的秩和不应太大或太小，即任意一组的秩和 T 与两组样本的平均秩和 $n_1(N+1)/2$ 应该相差不大。其中， $N = n_1 + n_2$ ， $n_1 = \min(n_1, n_2)$ ， n_1 和 n_2 分别为两组样本观察值的个数。

Mann-Whitney 非参数检验的基本步骤是：（1）建立检验的原假设，即两样本的总体分布相同；（2）将两组样本观察值按照从小到大顺序统一编秩，若大小相同的观察值不在同一样本内，取其平均秩次，计算两组样本观察值的秩和，当两组样本观察值的个数 $n_1 \neq n_2$ 时，取样本观察值个数较小的秩和作为检验统计量 T ，当两组样本观察值的个数 $n_1 = n_2$ 时，则取任意一组样本观察值的秩和作为检验统计量 T ；（3）确定 p 值作出检验结论，当样本观察值个数较少（ $n_1 \leq 10$ ， $n_2 - n_1 \leq 10$ ）时，可以直接查 T 界值表，当样本观察值个数较多时，需要采用正态近似法进行计

算, 计算公式为 $u = \frac{|T - n_1(N+1)/2| - 0.5}{\sqrt{n_1 n_2 (N+1)/12}}$, 而当两组样本观察值的相同秩次较多时,

计算出的 u 值偏小, 还应进行校正, 校正公式为 $u_c = u/\sqrt{C}$, 其中 $C = 1 - \frac{\sum t_j^3 - t_j}{N^3 - N}$, t_j 为第 j 个相同秩次的样本观察值个数。

对任意两组股票由经典的 PIN 模型估计出来的参数分布之间是否存在显著的差异进行的 Mann-Whitney 非参数检验结果见表 3.6。

表 3.6 经典的 PIN 模型估计得到的参数分布的 Mann-Whitney 非参数检验

参数	低对中 P 值	低对高 P 值	中对高 P 值
α	0.0011	0.0072	0.2444
δ	1.5460e-04	1.8141e-06	0.0653
μ	0.0017	1.3218e-04	0.3224
ε_b	0.0232	0.0397	0.1624
ε_s	2.4528e-04	5.0092e-05	0.6695

从表 3.6 中可以看出, 根据 Mann-Whitney 非参数检验, 换手率最低的一组股票的知情事件发生概率 α 在 99% 的置信水平下低于换手率居中和换手率最高的股票的知情事件发生概率, 而换手率居中的一组股票的知情事件发生概率与换手率最高的一组股票相比并没有显著差异。包括知情交易者的买卖单到达率 μ 、非知情交易者的买单到达率 ε_b 和非知情交易者的卖单到达率 ε_s 的其他几个参数与参数 α 的检验结果相类似, 那就是换手率最低的一组股票在 95% 的置信水平下低于换手率居中和换手率最高的股票。只有知情事件发生时为坏消息的概率 δ 在任意两组股票之间都存在显著的差异, 换手率最低的一组股票的知情事件发生时为坏消息的概率 δ 在 99% 的置信水平下高于换手率居中和换手率最高的股票的知情事件发生时为坏消息的概率, 而换手率居中的一组股票的知情事件发生时为坏消息的概率 δ 在 90% 的置信水平下高于换手率最高的一组股票。

三、经典的 PIN 模型估计得到的信息风险与交易活跃程度之间的关系

同样，从表 3.4 中可以看出，由经典的 PIN 模型估计得到的信息风险的测度 PIN 值在换手率最低的一组股票中的中位数为 0.11，取值范围为从第 5 分位的 0.09 到第 95 分位的 0.14；PIN 值在换手率居中的一组股票中的中位数为 0.12，取值范围为从第 5 分位的 0.10 到第 95 分位的 0.16；PIN 值在换手率最高的一组股票中的中位数为 0.12，取值范围为从第 5 分位的 0.09 到第 95 分位的 0.16。可见，换手率居中和换手率最高的两组股票的 PIN 值都高于换手率最低的一组股票，这意味着换手率居中和换手率最高的两组股票的信息风险都大于换手率最低的一组股票的信息风险。为了更好地比较交易活跃程度不同的股票在信息风险大小上的差异，本文列出换手率最低、换手率居中和换手率最高的三组股票利用经典的 PIN 模型估计出来的信息风险大小 (PIN) 的描述性统计，结果见表 3.7。

表 3.7 经典的 PIN 模型估计得到的各组股票信息风险的描述性统计

组别	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
换手率 低	0.1090	0.1070	0.1444	0.0855	0.0168	0.6618	2.7195	1.9071	0.3854
换手率 中	0.1237	0.1193	0.1669	0.0975	0.0211	0.4029	1.9027	1.9307	0.3808
换手率 高	0.1238	0.1206	0.1651	0.0794	0.0253	-0.0235	1.8690	1.3345	0.5131

从表 3.7 中可以看出，由经典的 PIN 模型估计得到的换手率最低、换手率居中和换手率最高的三组股票的信息风险 (PIN) 的均值分别为 0.1090、0.1237 和 0.1238。可见，换手率最高的一组股票的信息风险总体上高于换手率居中和换手率最低的一组股票的信息风险，即股票的信息风险与交易活跃程度呈正相关的关系。而换手率最高的一组股票的信息风险的波动程度也大于换手率最低和换手率居中的股票。

Easley, Kiefer, O'Hara and Paperman (1996)^[1]利用纽约证券交易所 (NYSE) 从

1990年10月1日到1990年12月23日的股票交易数据进行的研究发现交易量最大的一组股票具有的信息风险的平均值为0.16，交易量居中的一组股票具有的信息风险的平均值为0.21，交易量最小的一组股票具有的信息风险的平均值为0.22，即他们发现交易量大的股票具有的信息风险低于交易量小的股票具有的信息风险。Easley, Hvidkjaer and O'Hara (2005)^[70]用每年的交易数据估计了纽约证券交易所(NYSE)的股票从1993年到2000年的信息风险，发现美国股票具有的信息风险总体上呈逐渐下降的趋势，股票具有的信息风险的平均值从1983年的0.22下降到了2000年的0.19。Yan and Zhang (2006)^[71]利用纽约证券交易所(NYSE)和美国证券交易所(AMEX)从1993年到2004年每个季度的股票交易数据进行的研究也发现美国股票具有的信息风险呈逐渐下降的趋势，所有股票具有的信息风险的中位数从1993年第一季度的0.169下降到了2004年第4季度的0.115，这种趋势在纽约证券交易所和美国证券交易所从2001年实行以分为报价单位^①后尤其明显。

Brockman and Chung (2000)^[10]利用香港证券交易所从1996年5月1日到1997年8月29日的股票交易数据进行的研究发现交易量最大的一组股票具有的信息风险的平均值为0.27，交易量居中的一组股票具有的信息风险的平均值为0.27，交易量最小的一组股票具有的信息风险的平均值为0.52，即他们也发现交易量大的股票具有的信息风险低于交易量小的股票具有的信息风险。Lu and Wong (2008)^[49]利用台湾证券交易所从1997年1月到2005年12月的股票交易数据估计了台湾股票具有的信息风险，发现台湾股票具有的信息风险存在较大的波动，样本股票具有的信息风险的中位数的变动范围为从2005年的0.16到2000年的0.18。

在国内学者的研究中，杨之曙和姚松瑶(2004)^[55]利用上海证券交易所从2002年7月1日到2004年6月30日之间的股票交易数据进行的研究发现交易量最大的一组股票具有的信息风险的平均值为0.12，交易量居中的一组股票具有的信息风险的平均值为0.15，交易量最大的一组股票具有的信息风险的平均值为0.16。韩立岩、郑君彦和李东辉(2008)^[60]利用上海证券交易所2004年的股票交易数据进行的研究发现交易量最大的一组股票具有的信息风险的平均值为0.14，交易量

^① 纽约证券交易所和美国证券交易所过去采用1美元为报价和交易单位，2001年1月29日，两个交易所都采用1美分为报价和交易单位。

居中的一组股票具有的信息风险的平均值为 0.18，交易量最大的一组股票具有的信息风险的平均值为 0.18。也就是说，他们的研究也发现交易量大的股票具有的信息风险低于交易量小的股票具有的信息风险。

将本文利用经典的 PIN 模型对我国股票具有的信息风险进行的实证研究结果与上述学者的研究结果相比较可以得出以下结论。首先，同美国股票市场具有的信息风险的变动情况类似，我国股票具有的信息风险总体上也呈逐渐下降的趋势，近年来我国股票市场的信息不对称程度已经降低了很多；其次，本文以换手率高作为对交易活跃程度的衡量得出的结论为股票的信息风险与交易活跃程度呈正相关的关系，这与其他学者发现股票的信息风险与交易活跃程度呈负相关关系的结论不同。造成这种不同的原因可能有两个方面的原因，一是本文选取换手率作为交易活跃程度的衡量指标，而其他学者大多选择交易量作为交易活跃程度的衡量指标；二是本文没有考虑价格对交易活跃程度的影响，其他学者大多选择价格接近的股票作为样本股票。

为了进一步从统计上检验各组股票的信息风险之间的差异，本文对各组股票由经典的 PIN 模型估计得到的信息风险进行 Kruskal-Wallis 非参数检验，检验结果见表 3.8。

表 3.8 经典的 PIN 模型估计得到的信息风险 Kruskal-Wallis 非参数检验

名称	P 值
PIN	0.0255

从表 3.8 中可以看出，Kruskal-Wallis 非参数检验的结果表明我们可以在 95% 的置信水平下拒绝各组股票之间的信息风险相同的原假设，这意味着换手率不同的各组股票之间的信息风险确实存在显著的差异。图 3.3 显示了由经典的 PIN 模型估计得到的各组股票信息风险之间差异的 Kruskal-Wallis 非参数检验图。

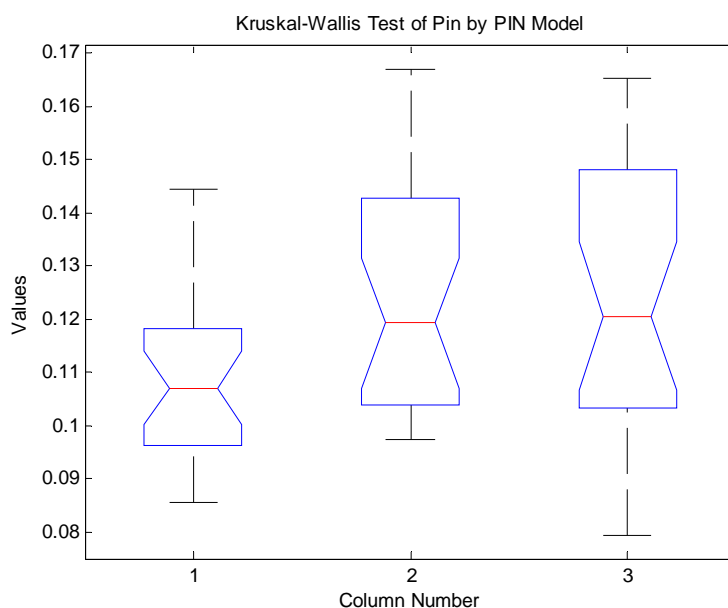


图 3.3 经典的 PIN 模型估计得到的各组股票信息风险的 Kruskal-Wallis 非参数检验

从图 3.3 中可以看出，换手率最低的股票与换手率居中以及换手率最高的股票之间的信息风险存在较为明显的差异，而换手率居中的股票与换手率最高的股票之间信息风险的差异从图上看则不是那么明显。

为了进一步考察由经典的 PIN 模型估计出来的任意两组股票具有的信息风险之间的差异，本文再对任意两组股票具有的信息风险进行 Mann-Whitney 非参数检验，检验结果见表 3.9。

表 3.9 经典的 PIN 模型估计得到的信息风险的 Mann-Whitney 非参数检验

名称	低对中 P 值	低对高 P 值	中对高 P 值
PIN	0.0153	0.0257	1.0000

从表 3.9 中可以看出，根据 Mann-Whitney 非参数检验，由经典的 PIN 模型估计出来的换手率最低的一组股票具有的信息风险在 95% 的置信水平下小于换手率居中和换手率最高的股票具有的信息风险，这意味着交易活跃程度低的股票具有的信息风险低于交易活跃程度高的股票具有的信息风险。然而，换手率居中的股票和换手率最高的股票具有的信息风险之间并不存在显著的差异。

四、经典的 PIN 模型中隐含的买卖单之间的相关性同实际数据的比较

虽然 Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的 PIN 模型目前已成为测度信息风险的经典模型,但也有不同学者对其测度信息风险的准确性提出了质疑。Venter and De Jongh (2006)^[16]从统计的角度发现 PIN 模型隐含的买卖指令之间的负相关关系与实际数据中买卖指令之间的正相关关系并不相符。Duarte and Young (2008)^[17]进一步证实了该结论。本文接下来对此结论进行验证。

利用对 PIN 模型估计出来的参数,可以计算出经典的 PIN 模型隐含的买单和卖单数量的均值、方差及它们之间的相关性。表 3.10 列出了经典的 PIN 模型隐含的换手率最低、换手率居中和换手率最高的三组股票的每日平均买单数量、卖单数量和买单方差、卖单方差以及买单和卖单数量之间的相关性在第 95 分位、第 75 分位、第 50 分位、第 25 分位和第 5 分位的统计数。

表 3.10 经典的 PIN 模型隐含的买卖单、方差及买卖单之间的相关性在不同分位数下的统计

组别	类别	第 95 分位	第 75 分位	中位数	第 25 分位	第 5 分位
换手率低	买单数量	1230	1020	797	701	483
	卖单数量	1125	997	789	714	508
	买单方差	39919	30749	24561	19595	12971
	卖单方差	31029	22785	17422	10883	7461
	买卖单数量相关性	-0.19	-0.26	-0.30	-0.45	-0.55
	换手率中	买单数量	1263	1183	1086	933
卖单数量		1139	1095	993	903	751
买单方差		53277	37723	31453	26827	10343
卖单方差		37429	15971	9852	4784	2420
买卖单数量相关性		-0.15	-0.21	-0.29	-0.37	-0.49
换手率高		买单数量	1221	1129	1082	967
	卖单数量	1100	1052	1003	913	850
	买单方差	54477	43636	39195	33164	19249
	卖单方差	28347	11228	4589	2792	1056
	买卖单数量相关性	-0.11	-0.16	-0.20	-0.27	-0.42

从表 3.10 中可以看出，经典的 PIN 模型在换手率最低、换手率居中和换手率最高的三组股票中隐含的买单和卖单之间的相关性都为负数，且在换手率最低的一组股票中，经典的 PIN 模型中隐含的买单和卖单之间的这种负相关性最高，随着换手率的增加，经典的 PIN 模型中隐含的买单和卖单之间的这种负相关性逐渐降低。然而，对比表 3.3 中实际数据的买单和卖单之间的正相关性，可以看出，经典的 PIN 模型隐含的买卖单之间的负相关性与实际数据中买卖单之间的正相关性并不能很好地相匹配。这一结果证实了 Venter and De Jongh (2006)^[16]和 Duarte and Young (2008)^[17]的结论。

实际上，可以推导出经典的 PIN 模型中隐含的买单和卖单数量之间的协方差^①为：

$$COV(B, S) = -\delta \times (1 - \delta) \times (\alpha \mu)^2 \leq 0$$

这个结果与经典的 PIN 模型本身是密切相关的。经典的 PIN 模型利用买卖单数量之间的非平衡来测度信息风险，信息相关的交易发生在知情交易者收到私人信息的日子，如果私人信息是好消息，这会导致更大数量的买入指令，如果私人信息是坏消息则导致更大数量的卖出指令。结果是，大量的买入和卖出指令在不同日到达，这造成了买入和卖出指令之间的负相关关系。

Duarte and Young (2008)^[17]的实证研究还发现经典的 PIN 模型隐含的买卖指令的方差与实际数据中买卖指令相对较大的方差不能很好地匹配。比较表 3.3 中利用实际数据计算出来的买单和卖单数量的方差和表 3.10 中利用经典的 PIN 模型估计出来的参数计算出来的模型隐含的买单和卖单数量的方差，我们可以发现经典的 PIN 模型中隐含的买单和卖单数量的方差远远小于实际数据中买单和卖单数量的方差。这一结果也证实了 Duarte and Young (2008)^[17]的结论。

§ 3.2 修正的 PIN 模型的估计结果

根据前面的实证结果，可以发现经典的 PIN 模型不能满足实际数据中买卖指令之间的正相关关系，此外，经典的 PIN 模型也不能匹配实际数据中买卖指令相对较大的方差。

^① 对于经典的 PIN 模型中隐含的买单和卖单数量之间的协方差公式的推导见附录 3。

Hasbrouck and Seppi (2001)^[61]证明了指令流非平衡在不同的股票之间是相关的，这意味着存在驱动指令流非平衡的共同因素。而且，Chordia, Roll and Subrahmanyam (2002)^[62]的研究表明整个市场的交易非平衡对同时期的指数收益具有显著影响。整个市场的指令非平衡不太可能是由于信息不对称引起的，因此可能存在其他造成非平衡的因素。

因此，本文在 Easley, Kiefer, O' Hara and Paperman (1996)^[1]的经典的 PIN 模型的基础上，增加由于市场指令流冲击引起的交易动机，对 PIN 模型进行了修正，以使得每天的买卖指令增加，从而使得模型隐含的买入和卖出指令之间的相关性与实际数据中买入和卖出指令之间的相关性一致（为正）并产生更接近于数据中观测到的方差。

本节首先对修正的 PIN 模型进行参数估计，然后考察修正的 PIN 模型隐含的买卖指令之间的相关性是否与实际数据中买卖指令的相关性一致。

一、修正后的 PIN 模型的参数估计

最大化公式 (2.26) 中的联合似然函数可以得到修正的 PIN 模型参数估计值。然而，同经典的 PIN 模型的估计问题类似，对于该最大化问题也没有解析解，只能采用数值最大化技术来得到该最大化问题的解。本文采用同 Easley, Hvidkjaer and O'Hara (2005)^[70]类似的办法，通过对联合似然函数进行因式分解来使得数值最大化过程变得更加容易。通过因式分解后，联合似然函数的形式^①为：

$$\ln(L(B_i, S_i)) = \ln \left(\begin{aligned} & \lambda \alpha \delta e^{-\gamma_b - \mu_s - \gamma_s} (x_{b\mu})^{B_i - M_i} (x_{b\gamma})^{-M_i} (x_{b\mu\gamma})^{B_i - M_i} (x_{s\mu})^{S_i - M_i} (x_{s\gamma})^{S_i - M_i} (x_{s\mu\gamma})^{-M_i} \\ & + \lambda \alpha (1 - \delta) e^{-\mu_b - \gamma_b - \gamma_s} (x_{b\mu})^{B_i - M_i} (x_{b\gamma})^{B_i - M_i} (x_{b\mu\gamma})^{-M_i} (x_{s\mu})^{S_i - M_i} (x_{s\gamma})^{-M_i} (x_{s\mu\gamma})^{S_i - M_i} \\ & + \lambda (1 - \alpha) e^{-\gamma_b - \gamma_s} (x_{b\mu})^{B_i - M_i} (x_{b\gamma})^{-M_i} (x_{b\mu\gamma})^{B_i - M_i} (x_{s\mu})^{S_i - M_i} (x_{s\gamma})^{-M_i} (x_{s\mu\gamma})^{S_i - M_i} \\ & + (1 - \lambda) \alpha \delta e^{-\mu_s} (x_{b\mu})^{B_i - M_i} (x_{b\gamma})^{B_i - M_i} (x_{b\mu\gamma})^{B_i - M_i} (x_{s\mu})^{-M_i} (x_{s\gamma})^{S_i - M_i} (x_{s\mu\gamma})^{S_i - M_i} \\ & + (1 - \lambda) \alpha (1 - \delta) e^{-\mu_b} (x_{b\mu})^{-M_i} (x_{b\gamma})^{B_i - M_i} (x_{b\mu\gamma})^{B_i - M_i} (x_{s\mu})^{S_i - M_i} (x_{s\gamma})^{S_i - M_i} (x_{s\mu\gamma})^{S_i - M_i} \\ & + (1 - \lambda) (1 - \alpha) (x_{b\mu})^{B_i - M_i} (x_{b\gamma})^{B_i - M_i} (x_{b\mu\gamma})^{B_i - M_i} (x_{s\mu})^{S_i - M_i} (x_{s\gamma})^{S_i - M_i} (x_{s\mu\gamma})^{S_i - M_i} \end{aligned} \right) \\ + \left\{ \begin{aligned} & -\varepsilon_b - \varepsilon_s + M_i \left[\ln(x_{b\mu}) + \ln(x_{b\gamma}) + \ln(x_{b\mu\gamma}) + \ln(x_{s\mu}) + \ln(x_{s\gamma}) + \ln(x_{s\mu\gamma}) \right] \\ & + B_i \left[\ln(\varepsilon_b + \mu_b) + \ln(\varepsilon_b + \gamma_b) + \ln(\varepsilon_b + \mu_b + \gamma_b) \right] - 2B_i \ln(\varepsilon_b) \\ & + S_i \left[\ln(\varepsilon_s + \mu_s) + \ln(\varepsilon_s + \gamma_s) + \ln(\varepsilon_s + \mu_s + \gamma_s) \right] - 2S_i \ln(\varepsilon_s) \end{aligned} \right\}$$

^① 本文给出了对该因式分解的推导过程，具体推导过程见附录 4。

$$\text{其中, } M_i = (\min(B_i, S_i) + \max(B_i, S_i)) / 2, \quad x_{b\mu} = \frac{\varepsilon_b}{\varepsilon_b + \mu_b}, \quad x_{b\gamma} = \frac{\varepsilon_b}{\varepsilon_b + \gamma_b},$$

$$x_{b\mu\gamma} = \frac{\varepsilon_b}{\varepsilon_b + \mu_b + \gamma_b}, \quad x_{s\mu} = \frac{\varepsilon_s}{\varepsilon_s + \mu_s}, \quad x_{s\gamma} = \frac{\varepsilon_s}{\varepsilon_s + \gamma_s}, \quad x_{s\mu\gamma} = \frac{\varepsilon_s}{\varepsilon_s + \mu_s + \gamma_s}。$$

为了比较修正的PIN模型与经典的PIN模型之间对数据的拟合能力的高低, 本文首先进行似然比检验^① (Likelihood Ratio Test)。修正的PIN模型与经典的PIN模型相比, 增加了 λ 、 μ_s 、 γ_b 和 γ_s 这四个参数, 因此, 对于每只股票, 检验的原假设为:

$$H_0: \lambda = \mu_s = \gamma_b = \gamma_s = 0$$

在原假设下, 没有市场范围内的指令流冲击事件发生, 基于公司特有的私人信息进行交易的买单和卖单的到达率相同。我们用 $\log(L_0)$ 表示经典的PIN模型的对数似然函数值, $\log(L_1)$ 表示修正的PIN模型的对数似然函数值。似然比检验的统计量为:

$$LR = -2(\log(L_0) - \log(L_1)) \sim \chi_{\alpha}^2(n)$$

其中, α 为置信水平, n 为自由度。对修正的PIN模型与经典的PIN模型的似然比检验结果^②见表3.11。

表 3.11 修正的 PIN 模型与经典的 PIN 模型的似然比检验

组别	检验统计量	临界值	拒绝原假设的频率
换手率低	6120	13.28	99%
换手率中	6584	13.28	99%
换手率高	7200	13.28	99%

从表 3.13 中可以看出, 在换手率最低、换手率居中和换手率最高这所有三组

^① 修正的 PIN 模型与经典的 PIN 模型属于嵌套模型 (Nested Models), 对于这种嵌套模型优劣的比较通常采用似然比检验 (Likelihood Ratio Test), 似然比检验的统计量渐进服从 χ^2 分布。

^② 因为修正的 PIN 模型与经典的 PIN 模型相比增加了 4 个参数, 因此似然比检验的统计量服从的 χ^2 分布的自由度为 4。本文对所有三组股票中的每只股票都分别进行了似然比检验, 每只股票的检验结果都在 99% 的置信水平下拒绝原假设。限于篇幅的限制, 本文在这里只给出各组似然比检验统计量的平均值。

股票中，似然比检验的统计量都远远大于对应的临界值，我们可以在 99% 的置信水平下拒绝原假设 H_0 。从似然比检验的结果可以看出，修正的 PIN 模型在对数据的拟合能力上明显优于经典的 PIN 模型。

本文接下来估计修正的 PIN 模型的参数。修正的 PIN 模型与经典的 PIN 模型相比需要估计的参数更多，估计过程也更加复杂，修正的 PIN 模型需要估计的参数向量为 $\theta = (\lambda, \alpha, \delta, \mu_b, \mu_s, \gamma_b, \gamma_s, \varepsilon_b, \varepsilon_s)$ ，其中 λ 为市场指令流冲击事件发生的概率， α 为知情交易事件发生的概率， δ 为知情交易事件发生时为坏消息的概率， μ_b 为知情交易者的买单到达率， μ_s 为知情交易者的卖单到达率， γ_b 为市场指令流冲击事件发生时的买单到达率， γ_s 为市场指令流冲击事件发生时的卖单到达率， ε_b 为非知情交易者的买单到达率， ε_s 为非知情交易者的卖单到达率。由于初始值的不同会对参数估计值产生较大的影响，本文对初始值的选取采用同 Yan and Zhang (2006)^[71]类似的方法。

对修正的 PIN 模型的参数估计结果见表 3.12。表 3.12 列出了换手率最低、换手率居中和换手率最高的三组股票在样本期内的修正的 PIN 模型的各项参数的估计值和信息风险在第 95 分位、第 75 分位、第 50 分位、第 25 分位和第 5 分位的统计数。

表 3.12 修正的 PIN 模型中各组股票的参数估计值及信息风险的分位数统计

组别	参数	第 95 分位	第 75 分位	中位数	第 25 分位	第 5 分位
换手率低	α	0.68	0.65	0.64	0.58	0.44
	δ	0.91	0.79	0.72	0.68	0.62
	λ	0.71	0.67	0.65	0.61	0.41
	ε_b	899.16	798.56	625.52	541.99	338.85
	γ_b	376.46	351.92	310.60	295.04	257.98
	μ_b	665.02	640.52	581.45	340.72	209.96
	ε_s	936.55	785.27	627.49	559.16	332.54

股票信息风险测度研究

	γ_s	275.40	182.15	153.89	133.20	48.42
	μ_s	308.01	237.18	205.15	190.75	159.75
	NPIN	0.15	0.12	0.10	0.08	0.07
	PMOS	0.28	0.20	0.17	0.14	0.11
	α	0.67	0.64	0.60	0.58	0.53
	δ	0.90	0.77	0.71	0.69	0.66
	λ	0.71	0.67	0.65	0.62	0.57
	ε_b	937.05	890.08	818.19	744.24	601.47
	γ_b	398.99	373.78	355.63	319.62	302.62
换手率中	μ_b	672.55	635.61	560.53	300.62	234.94
	ε_s	958.75	859.05	801.45	720.53	577.51
	γ_s	211.55	188.47	151.39	106.02	64.46
	μ_s	313.58	238.04	196.13	171.75	140.96
	NPIN	0.11	0.09	0.08	0.07	0.05
	PMOS	0.21	0.18	0.15	0.14	0.11
	α	0.69	0.65	0.63	0.59	0.50
	δ	0.84	0.74	0.71	0.67	0.64
	λ	0.73	0.67	0.65	0.63	0.50
	ε_b	928.77	867.36	810.04	747.02	670.47
	γ_b	388.06	374.60	362.53	338.83	309.69
换手率高	μ_b	678.02	315.82	287.47	259.80	215.81
	ε_s	916.45	806.70	757.04	708.22	631.12
	γ_s	214.33	189.83	153.31	121.54	44.58
	μ_s	287.43	248.46	216.40	173.62	155.76
	NPIN	0.12	0.08	0.07	0.06	0.06
	PMOS	0.20	0.18	0.16	0.14	0.12

从表 3.12 中可以看出, 知情事件发生的概率 α 在换手率最低的一组股票中的中位数为 0.64, 其变动范围为从第 5 分位的 0.44 到第 95 分位的 0.68; 知情事件发生的概率 α 在换手率居中的一组股票中的中位数为 0.60, 其变动范围为从第 5 分位的 0.53 到第 95 分位的 0.67; 知情事件发生的概率 α 在换手率最高的一组股票中的中位数为 0.63, 其变动范围为从第 5 分位的 0.50 到第 95 分位的 0.69。可以看出换手率低的股票发生知情事件的概率高于换手率居中和换手率最高的股票。市场指令流冲击事件发生的概率 λ 在换手率最低的一组股票中的中位数为 0.65, 其变动范围为从第 5 分位的 0.41 到第 95 分位的 0.71; 市场指令流冲击事件发生的概率 λ 在换手率居中的一组股票中的中位数为 0.65, 其变动范围为从第 5 分位的 0.57 到第 95 分位的 0.71; 市场指令流冲击事件发生的概率 λ 在换手率最高的一组股票中的中位数为 0.65, 其变动范围为从第 5 分位的 0.50 到第 95 分位的 0.73。可以看出无论是换手率最低的股票还是换手率居中和换手率最高的股票发生市场指令流冲击事件的概率都是大致相同的。

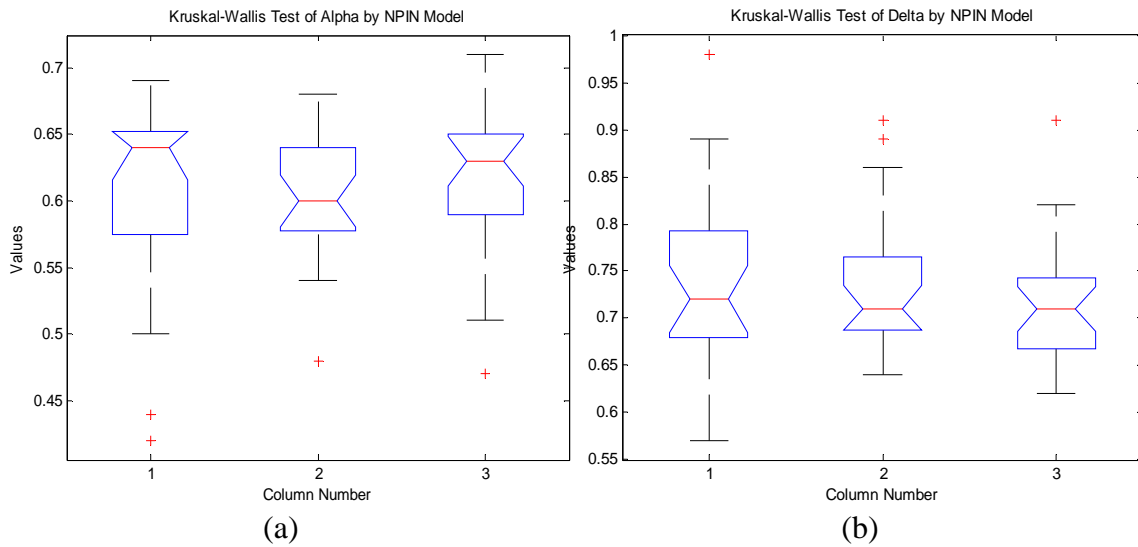
接下来考察各组股票由修正的 PIN 模型估计得到的参数分布之间是否存在显著的差异。本文首先采用 Kruskal-Wallis 非参数检验的方法。检验结果见表 3.13。

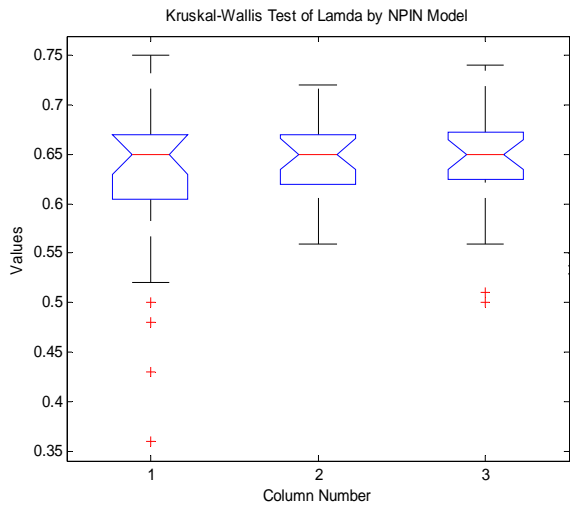
表 3.13 修正的 PIN 模型估计得到的参数分布的 Kruskal-Wallis 非参数检验

参数	P 值
α	0.2986
δ	0.6213
λ	0.8416
ε_b	0.0003
γ_b	0.0015
μ_b	0.0014
ε_s	0.0074
γ_s	0.9367
μ_s	0.5891

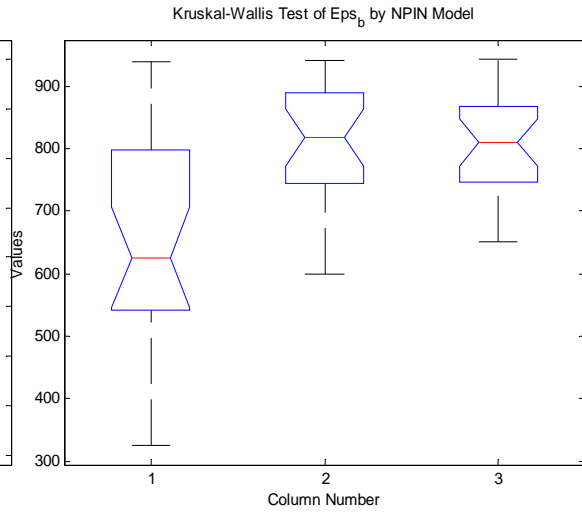
从表 3.13 中可以看出，对各组股票由修正的 PIN 模型估计得到的参数进行的 Kruskal-Wallis 非参数检验结果表明，知情事件发生的概率 α 、知情事件发生时为坏消息的概率 δ 和市场指令流冲击事件发生的概率 λ 在换手率不同的各组股票之间并不存在显著的差异。而包括非知情交易者的买单到达率 ε_b 和卖单到达率 ε_s 、知情交易者的买单到达率 μ_b 和市场指令流冲击事件发生时的买单到达率 γ_b 都在 99% 的置信水平下拒绝了换手率不同的各组股票之间的分布相同的原假设，这表明这些参数在各组股票之间存在显著差异。需要指出的是，检验结果并不能拒绝各组股票之间市场指令流冲击事件发生的概率 λ 相同的原假设，这并不矛盾，因为市场指令流冲击事件发生时会引起所有股票的交易都增加。同样，检验结果也不能拒绝各组股票之间当市场指令流冲击事件发生时引发的卖单到达率 γ_s ，这意味着当有市场恐慌事件发生时，会引发所有股票的抛售行为。

图 3.4 显示了各组股票由修正的 PIN 模型估计得到的参数的 Kruskal-Wallis 非参数检验图。

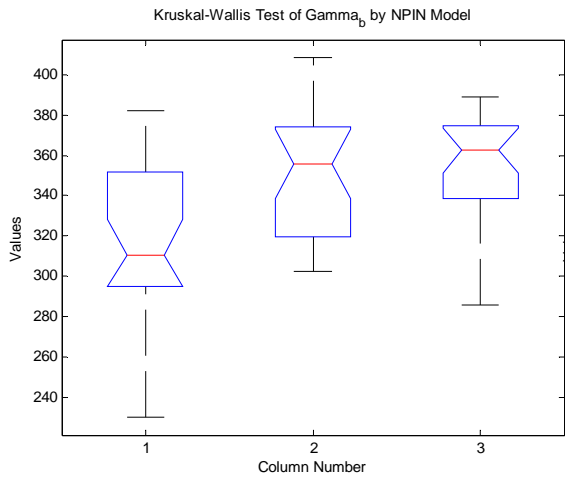




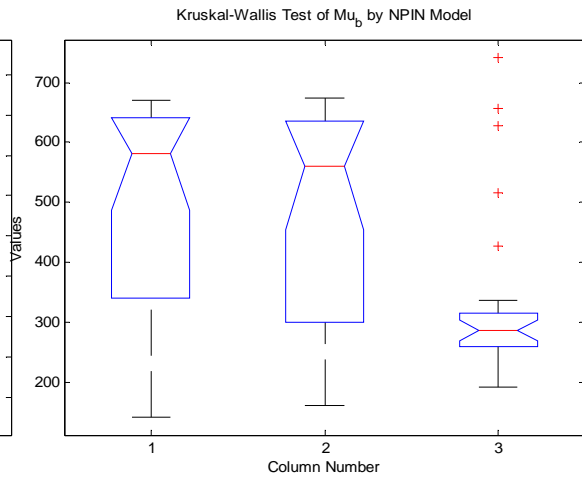
(c)



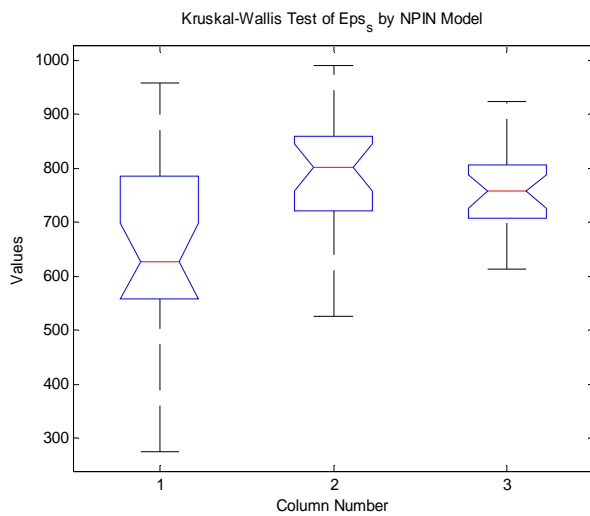
(d)



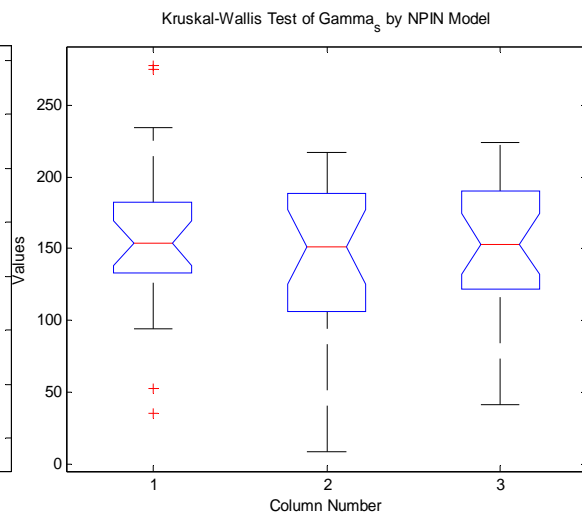
(e)



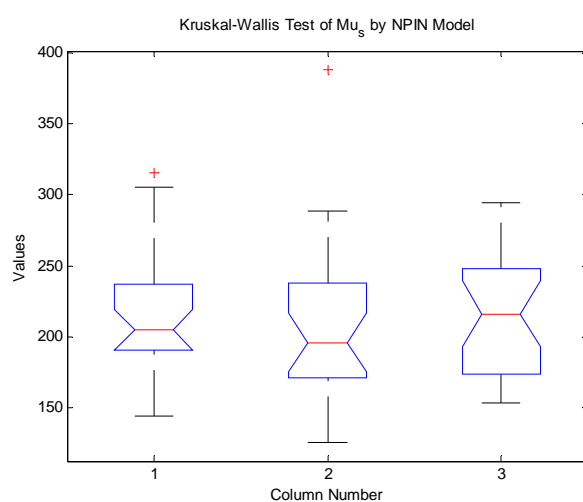
(f)



(g)



(h)



(i)

图 3.4 修正的 PIN 模型估计得到的参数 Kruskal-Wallis 非参数检验

从图 3.4 中可以看出，除了与市场指令流冲击事件相关的参数外，各组股票由修正的 PIN 模型估计得到的参数之间存在较为显著的差异。

为了进一步比较任意两组股票由经典的 PIN 模型估计出来的参数分布之间是否存在显著的差异，本文再进行 Mann-Whitney 非参数检验，检验结果见表 3.14。

表 3.14 修正的 PIN 模型估计得到的参数分布的 Mann-Whitney 非参数检验

参数	低对中 P 值	低对高 P 值	中对高 P 值
α	0.2890	0.8228	0.1218
δ	0.9612	0.4031	0.4081
λ	0.6191	0.6190	0.9845
ε_b	5.1446e-04	6.3805e-04	0.6980
γ_b	0.0043	0.0011	0.5157
μ_b	0.8310	9.0701e-04	0.0038
ε_s	0.0059	0.0199	0.1870
γ_s	0.7415	0.9227	0.8009
μ_s	0.3417	0.8009	0.4492

从表 3.14 中可以看出, 根据 Mann-Whitney 非参数检验, 知情事件发生的概率 α 、知情事件发生时为坏消息的概率 δ 和市场指令流冲击事件发生的概率 λ 在换手率不同的各组股票之间并不存在显著的差异, 这与前面的 Kruskal-Wallis 非参数检验结果相同。换手率最低的一组股票的非知情交易者的买单到达率 ε_b 和卖单到达率 ε_s 、市场指令流冲击事件发生时的买单到达率 γ_b 与换手率居中和换手率最高的股票之间在 99% 的置信水平下存在显著差异, 而非知情交易者的买单到达率 ε_b 和卖单到达率 ε_s 、市场指令流冲击事件发生时的买单到达率 γ_b 在换手率居中和换手率最高的股票之间并不存在显著差异。换手率最高的一组股票的知情交易者的买单到达率 μ_b 与换手率最低和换手率居中的股票之间在 99% 的置信水平下存在显著差异。而知情交易者的卖单到达率 μ_s 和市场指令流冲击事件发生时的卖单到达率 γ_s 在换手率不同的各组股票之间并不存在显著差异。

二、修正的 PIN 模型与经典的 PIN 模型估计得到的信息风险比较

从表 3.12 还可以看出, 在修正的 PIN 模型中, 信息风险 NPIN 的值在换手率最低的一组股票中的中位数为 0.10, 取值范围为从第 5 分位的 0.07 到第 95 分位的 0.15; 信息风险 NPIN 的值在换手率居中的一组股票中的中位数为 0.08, 取值范围为从第 5 分位的 0.05 到第 95 分位的 0.11; 信息风险 NPIN 的值在换手率最高的一组股票中的中位数为 0.07, 取值范围为从第 5 分位的 0.06 到第 95 分位的 0.12。这意味着在修正的 PIN 模型中, 信息风险的大小同交易活跃程度呈负相关关系, 即随着换手率的增加, 信息风险的大小逐渐降低。

为了更好地比较经典的 PIN 模型和修正的 PIN 模型中交易活跃程度不同的股票在信息风险大小上的差异, 本文列出换手率最低、换手率居中和换手率最高的三组股票分别利用经典的 PIN 模型和修正的 PIN 模型估计出来的信息风险大小和由于市场范围内的指令流冲击事件引起的交易概率大小的描述性统计, 结果见表 3.15。

表 3.15 经典的 PIN 模型和修正的 PIN 模型估计得到的各组股票的信息风险比较

组别	类别	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
换手率低	PIN	0.1090	0.1070	0.1444	0.0855	0.0168	0.6618	2.7195	1.9071	0.3854
	NPIN	0.1016	0.1000	0.1700	0.0600	0.0275	0.4931	2.7545	1.0757	0.5840
	PMOS	0.1720	0.1700	0.3300	0.1100	0.0507	0.2048	4.9207	9.8911	0.0071
换手率中	PIN	0.1237	0.1193	0.1669	0.0975	0.0211	0.4029	1.9027	1.9307	0.3808
	NPIN	0.0808	0.0800	0.1200	0.0500	0.0178	0.0589	2.6018	0.1796	0.9141
	PMOS	0.1568	0.1500	0.2200	0.0800	0.0315	-0.1821	2.9938	0.1382	0.9332
换手率高	PIN	0.1238	0.1206	0.1651	0.0794	0.0253	-0.0235	1.8690	1.3345	0.5131
	NPIN	0.0752	0.0700	0.1200	0.0500	0.0173	1.3094	4.4146	9.2281	0.0099
	PMOS	0.1592	0.1600	0.2100	0.1100	0.0243	0.0771	2.5212	0.2636	0.8765

从表 3.15 中可以看出，在经典的 PIN 模型中，信息风险的大小（PIN 值）随着交易活跃程度（换手率）的增加而上升，信息风险大小的波动（PIN 值的标准差）随着交易活跃程度（换手率）的增加而下降。在修正的 PIN 模型中，信息风险的大小（NPIN 值）随着交易活跃程度（换手率）的增加而下降，信息风险大小的波动（NPIN 值的标准差）也随着交易活跃程度（换手率）的增加而下降。

这一结果与经典的 PIN 模型中得出的信息风险的大小同交易活跃程度呈正相关关系的结论不同，这主要是因为修正的 PIN 模型中，我们加入了市场指令流冲击这一因素。

从表 3.12 中可以看出，市场指令流冲击事件发生的概率 PMOS 的值在换手率最低的一组股票中的中位数为 0.17，取值范围为从第 5 分位的 0.11 到第 95 分位的 0.28；在换手率居中的一组股票中的中位数为 0.15，取值范围为从第 5 分位的 0.11 到第 95 分位的 0.21；在换手率最高的一组股票中的中位数为 0.16，取值范围为从第 5 分位的 0.12 到第 95 分位的 0.20。从表 3.15 中可以看出，市场指令流冲击事

件发生时引起的交易概率（PMOS 值）在交易活跃程度最低时（换手率最低）最高，在交易活跃程度一般时（换手率居中）最低，市场指令流冲击事件引起的交易概率大小的波动（PMOS 的标准差）也随着交易活跃程度（换手率）的增加而下降。在各组股票中，NPIN 值都低于对应的 PIN 值，而 PMOS 值都明显大于对应的 NPIN 和 PIN 值，这意味着交易在更大程度上是由于市场指令流冲击事件引起的，在经典的 PIN 模型中仅仅把交易动机区分为知情交易和流动性交易是不合理的，当市场指令流冲击事件发生时会在更大程度上促使投资者进行交易。在换手率最低的一组股票中，市场指令流冲击事件发生时引起的交易概率是最高的，这表明经典的 PIN 模型由于没有考虑到市场指令流冲击事件发生时会在很大程度上引起投资者进行交易，从而高估股票具有的信息风险。

为了进一步从统计上检验各组股票的信息风险之间的差异，本文对各组股票由经典的 PIN 模型估计得到的信息风险进行 Kruskal-Wallis 非参数检验，检验结果见表 3.16。

表 3.16 修正的 PIN 模型估计得到的信息风险 Kruskal-Wallis 非参数检验

名称	P 值
NPIN	0.0006
PMOS	0.6788

从表 3.16 中可以看出，Kruskal-Wallis 非参数检验的结果表明可以在 99% 的置信水平下拒绝各组股票之间的信息风险相同的原假设，这意味着换手率不同的各组股票之间的信息风险确实存在显著的差异。而在 10% 的置信水平下也不能拒绝各组股票之间由市场指令流冲击事件发生引起的交易概率不存在显著差异的原假设，这与我们的直觉是相符的，市场指令流冲击事件发生时会引起所有股票的交易增加。

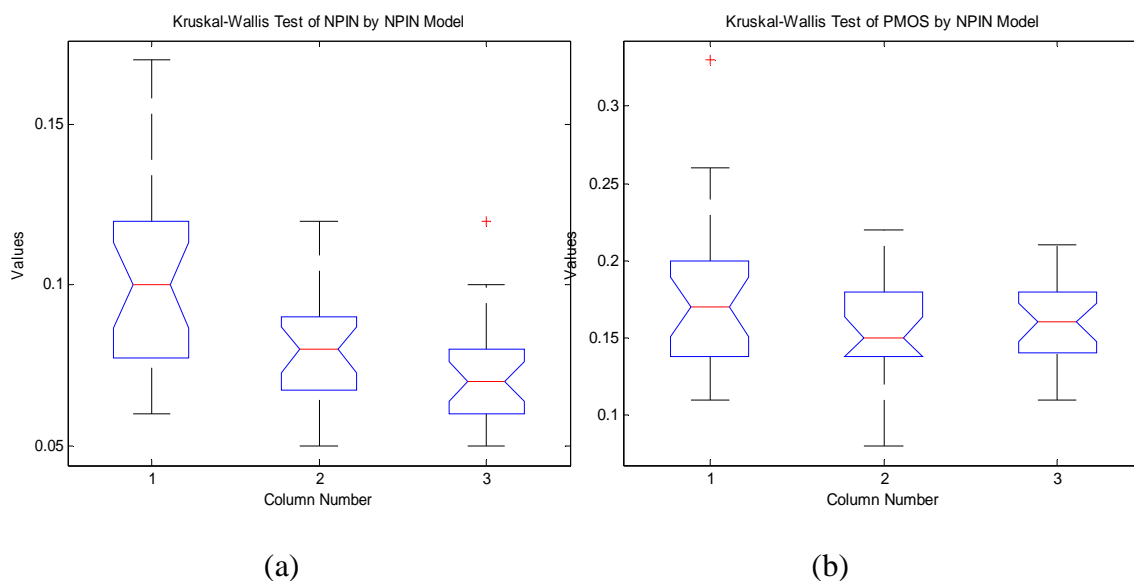


图 3.5 修正的 PIN 模型估计得到的各组股票信息风险和市场指令流冲击事件发生时引起的交易概率的 Kruskal-Wallis 非参数检验

图 3.5 显示了由修正的 PIN 模型估计得到的各组股票信息风险和市场指令流冲击事件发生时引起的交易概率之间差异进行的 Kruskal-Wallis 非参数检验图。从图 3.5 (a)和图 3.5 (b)中可以看出，信息风险在换手率最低的股票与换手率居中以及换手率最高的股票之间存在较为明显的差异，而市场指令流冲击事件发生时引起的交易概率在换手率不同的各组股票之间并不存在显著差异。

为了进一步比较任意两组股票由修正的 PIN 模型估计出来的信息风险之间是否存在显著的差异，我们再进行 Mann-Whitney 非参数检验，检验结果见表 3.17。

表 3.17 修正的 PIN 模型估计得到的信息风险的 Mann-Whitney 非参数检验

名称	低对中 P 值	低对高 P 值	中对高 P 值
NPIN	0.0065	4.0078e-04	0.1310
PMOS	0.4185	0.5196	0.8146

从表 3.17 中可以看出，根据 Mann-Whitney 非参数检验，由修正的 PIN 模型估计出来的换手率最低的一组股票具有的信息风险在 99%的置信水平下大于换手率居中和换手率最高的股票具有的信息风险，这意味着交易活跃程度低的股票具有的信息风险高于交易活跃程度高的股票具有的信息风险。然而，换手率居中的

股票和换手率最高的股票具有的信息风险之间并不存在显著的差异。而各组股票由于市场指令流冲击事件发生时引起的交易概率之间都不存在显著的差异。

三、修正的 PIN 模型隐含的买卖单之间的相关性与实际数据比较

正如前文所述，经典的 PIN 模型中隐含的买单和卖单之间的负相关性与实际数据揭示的正相关性不相符合，并且经典的 PIN 模型中隐含的买单和卖单的方差也远远低于实际数据揭示的买单和卖单的方差。我们希望修正后的 PIN 模型中隐含的买单和卖单的方差以及它们之间的相关性能够更好地与实际数据揭示的特征相匹配。利用对修正的 PIN 模型估计出来的参数，可以计算出修正的 PIN 模型中隐含的买单和卖单数量的均值、方差及它们之间的相关性，结果见表 3.18。

表 3.18 修正的 PIN 模型中隐含的各组股票的买卖单均值、方差及它们之间的相关性

组别	类别	第 95 分位	第 75 分位	中位数	第 25 分位	第 5 分位
换手率低	买单数量	1273	1058	908	760	590
	卖单数量	1089	972	822	739	507
	买单方差	98441	74718	55449	40154	30604
	卖单方差	22720	18810	17710	15206	12103
	买卖单数量相关性	0.42	0.26	0.19	0.13	0.05
换手率中	买单数量	1262	1198	1116	1075	880
	卖单数量	1108	1054	976	886	765
	买单方差	89917	79497	55421	39508	35745
	卖单方差	20816	18429	15723	14336	11322
	买卖单数量相关性	0.49	0.24	0.14	0.05	0.01
换手率高	买单数量	1253	1154	1129	1038	933
	卖单数量	1035	979	957	890	837
	买单方差	65259	49988	42572	39361	31557
	卖单方差	22161	19502	17155	15725	14278
	买卖单数量相关性	0.51	0.31	0.16	0.08	0.002

从表 3.18 中可以看出，修正的 PIN 模型在换手率最低、换手率居中和换手率最高的三组股票中隐含的买单和卖单之间的相关性都为正数，且这种买单和卖单之间的正的相关性随着交易活跃程度的增加而增加，在换手率最高的一组股票中修正的 PIN 模型中隐含的买单和卖单之间的这种正相关性最高。对比表 3.3 中实际数据的买单和卖单之间的正相关性，可以看出，修正的 PIN 模型隐含的买单和卖单之间的这种正相关性与实际数据中买单和卖单之间的正相关性能够很好地相匹配。

同样，可以推导出修正的 PIN 模型中隐含的买单和卖单数量之间的协方差^①为：

$$COV(B,S) = \lambda(1-\lambda)\gamma_b\gamma_s - \alpha^2\delta(1-\delta)\mu_b\mu_s$$

这时，本文通过对经典的 PIN 模型进行修正，可以使得修正后的 PIN 模型隐含的买单和卖单数量之间的协方差为正，从而使得修正后的 PIN 模型中隐含的买单和卖单之间的相关性能够与实际数据中买单和卖单之间的正相关性相匹配。

此外，从表 3.18 中还可以看出，各组股票在修正的 PIN 模型中隐含的买单和卖单的方差都明显大于经典的 PIN 模型中隐含的买单和卖单的方差，这与实际数据中买单和卖单较大的方差更加匹配。

^① 本文给出了对于修正的 PIN 模型中买单和卖单数量的均值以及买单和卖单数量之间的协方差公式的推导，具体推导过程见附录 5。

第4章 中国股票市场信息风险性质研究

§ 4.1 中国股票市场信息风险时间序列特征研究

为了考察中国股票市场信息风险随时间而演变的特征，本文首先估计出信息风险的时间序列，然后考察中国股票市场信息风险的时间序列特征。本文分别用经典的 PIN 模型和修正的 PIN 模型估计出股票信息风险的时间序列，然后对两个模型估计出的股票信息风险的时间序列特征进行比较。

一、数据及估计方法

PIN模型的估计需要每只股票每天由买方和卖方发起的交易数量的数据。同前一章中估计PIN模型采用的数据来源相同，为了计算每天买方和卖方发起交易的数量，我们使用来自于万得金融高频数据库的分笔数据。为了同前一章中采用全年的交易数据估计出来的信息风险进行比较，我们选择的样本为包含在换手率最低、换手率居中和换手率最高的三组股票中的钢铁行业的股票作为我们的研究样本^①。之所以选择同一个行业的股票是因为我们想知道市场指令流冲击事件发生时引起的交易概率（PMOS）对股票收益率的影响，而钢铁行业在样本中的股票数量最多。同样，我们也剔除了发生在开盘之前的所有交易数据以避免包括发生在开盘期间的竞价交易。

由于数据长度的限制，要想估计出中国股票市场信息风险的时间序列，只能采用滚动（rolling）的方法，本文选取 120 个交易日作为滚动（rolling）的时间窗口^②。具体而言，就是对于样本内的每只股票，先用最前面的 120 个交易日的数据估计出对应的PIN模型的参数值，并计算出相应的信息风险，然后去除掉第一个交易日的数据，加入第 121 个交易日的数据，然后估计出对应的PIN模型的参数值，同样计算出相应的信息风险，这样以此类推，我们可以估计出样本内每只股票的信息风险的时间序列。

^① 我们采用的行业分类标准为中国证监会的行业分类标准，钢铁行业的证监会行业分类名称为制造业/黑色金属冶炼及压延加工业。

^② 虽然 Easley, Kiefer and O'Hara (1997a)^[7]的研究表明只需要 60 个交易日的数据就可以充分获得 PIN 模型的可靠参数估计值，但在参数估计的过程中本文发现采用 60 个交易日作为滚动（rolling）的时间窗口会使得参数的估计值非常不稳定，有时甚至估计不出合理的参数值，为了保证参数估计的可靠性，本文采用 120 个交易日作为滚动（rolling）的时间窗口。

二、由经典的 PIN 模型估计得到的股票信息风险时间序列

本文首先采用滚动(rolling)的方法估计出经典的 PIN 模型的参数的时间序列。参数估计结果见图 4.1。

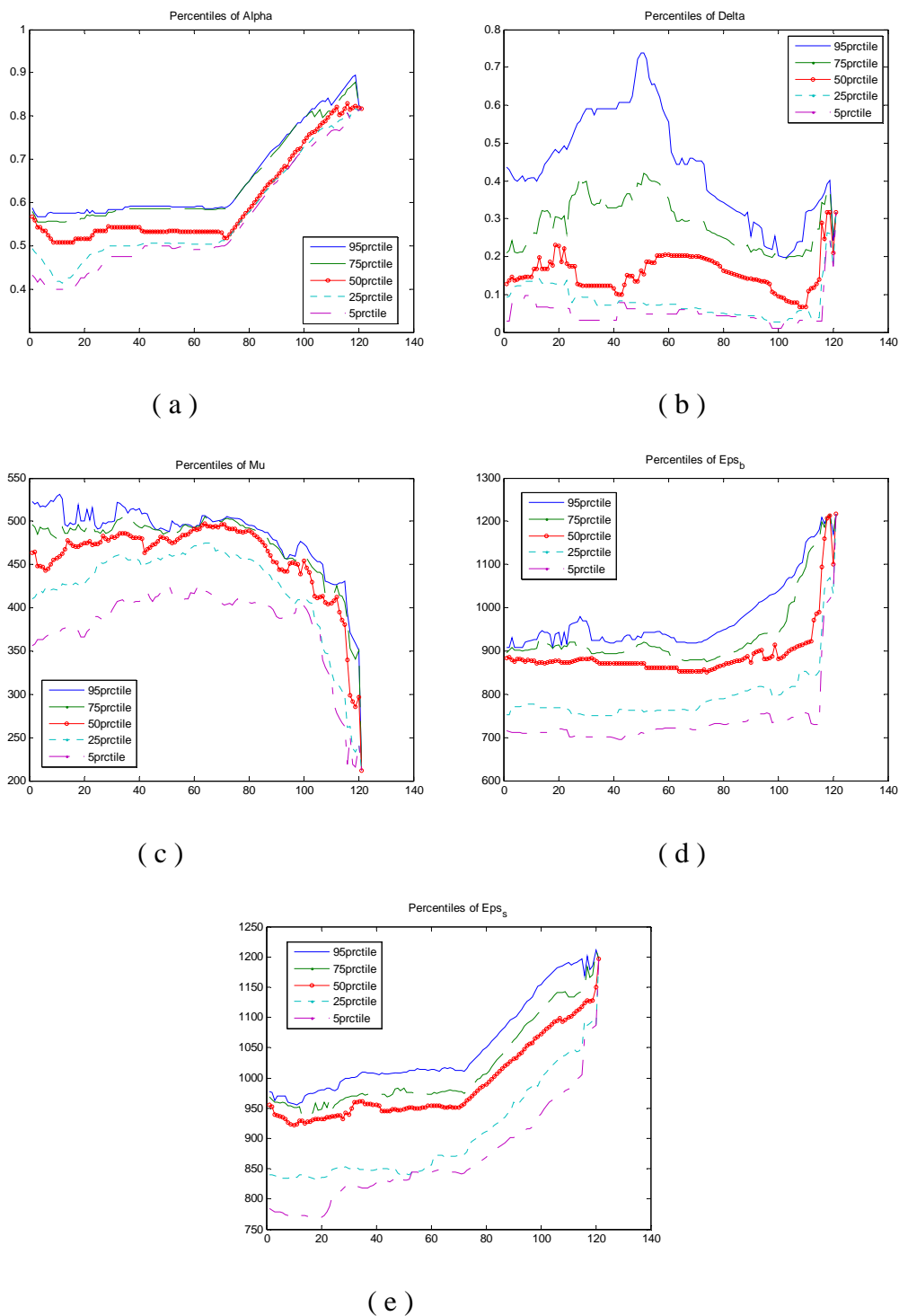


图 4.1 经典的 PIN 模型参数估计结果的时间序列

图 4.1(a)、4.1(b)、4.1(c)、4.1(d)和 4.1(e)分别显示了钢铁行业股票由经典的 PIN 模型估计出来的参数 α 、 δ 、 μ 、 ε_b 和 ε_s 在 95 分位、75 分位、50 分位、25 分位和 5 分位的时间序列。从图中可以看出， α 的估计值在前面比较平稳，后面则逐渐上升，变动范围在 0.40 到 0.90 之间； δ 的估计值波动比较剧烈，变动范围在 0.01 到 0.74 之间； μ 的估计值前面比较平稳，后面则逐渐下降，变动范围在 212 到 531 之间； ε_b 的估计值前面比较平稳，后面则逐渐上升，变动范围在 694 到 1217 之间； ε_s 的估计值呈逐渐上升的趋势，变动范围在 768 到 1212 之间。

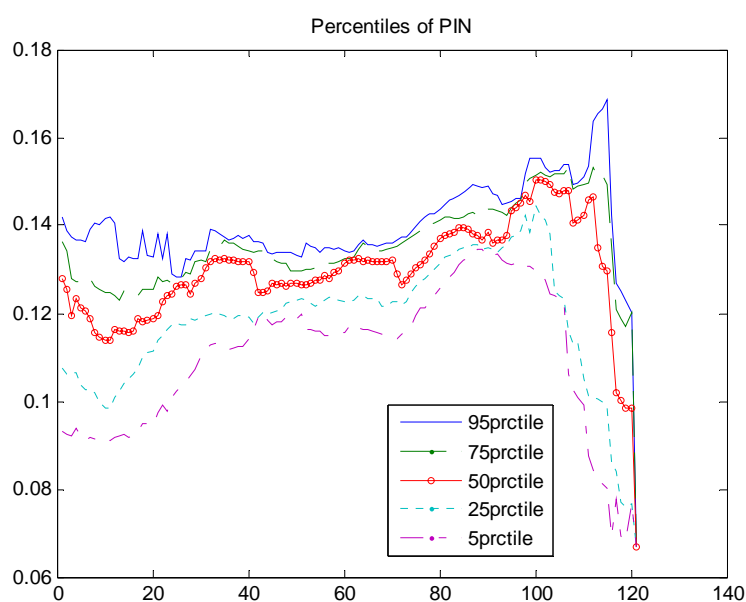


图 4.2 经典的 PIN 模型估计出的信息风险时间序列

图 4.2 显示了钢铁行业股票由经典的 PIN 模型估计出来的参数计算出来的信息风险在 95 分位、75 分位、50 分位、25 分位和 5 分位的时间序列。从图中可以看出，钢铁行业股票的信息风险在前面时间段里比较平稳，后面时间段里则逐渐下降，变动范围在 0.07 到 0.17 之间。

表 4.1 显示了钢铁行业股票由经典的 PIN 模型估计出来的参数的横截面平均值的时间序列的描述性统计。

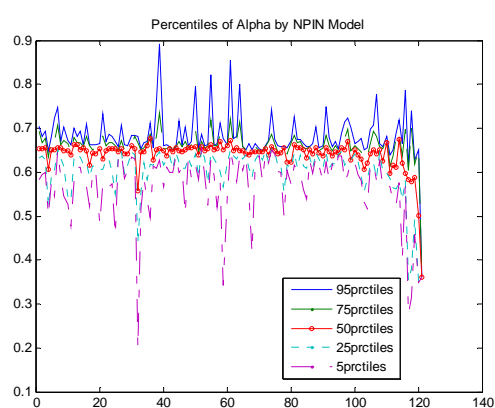
表 4.1 钢铁行业股票由经典的 PIN 模型估计得到的参数描述性统计

变量	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
α	0.5935	0.5421	0.8217	0.4999	0.0993	1.0722	2.6655	22.569	0.0000
δ	0.1908	0.2023	0.2698	0.1045	0.0452	-0.3570	2.0250	6.9989	0.0302
μ	454.50	463.82	484.15	356.56	28.140	-1.8674	6.1089	113.15	0.0000
ε_b	855.48	841.47	988.41	826.85	36.633	2.1010	6.7973	153.70	0.0000
ε_s	956.11	924.73	1101.6	890.55	64.088	1.0551	2.7105	21.740	0.0000

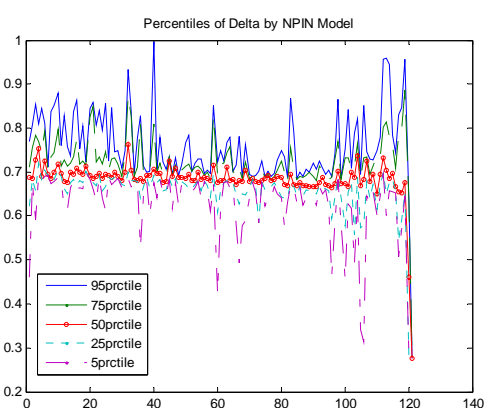
从表 4.1 中可以看出,钢铁行业股票的知情事件发生的概率 α 的平均值为 0.59, 知情事件发生为坏消息的概率 δ 的平均值为 0.19, 知情交易者的买卖单到达率 μ 的平均值为 454.50, 非知情交易者的买单到达率 ε_b 和卖单到达率 ε_s 分别为 855.48 和 956.11。

三、由修正的 PIN 模型估计得到的股票信息风险时间序列

我们接下来采用滚动 (rolling) 的方法估计出修正的 PIN 模型的参数的时间序列。参数估计结果见图 4.3。



(a)



(b)

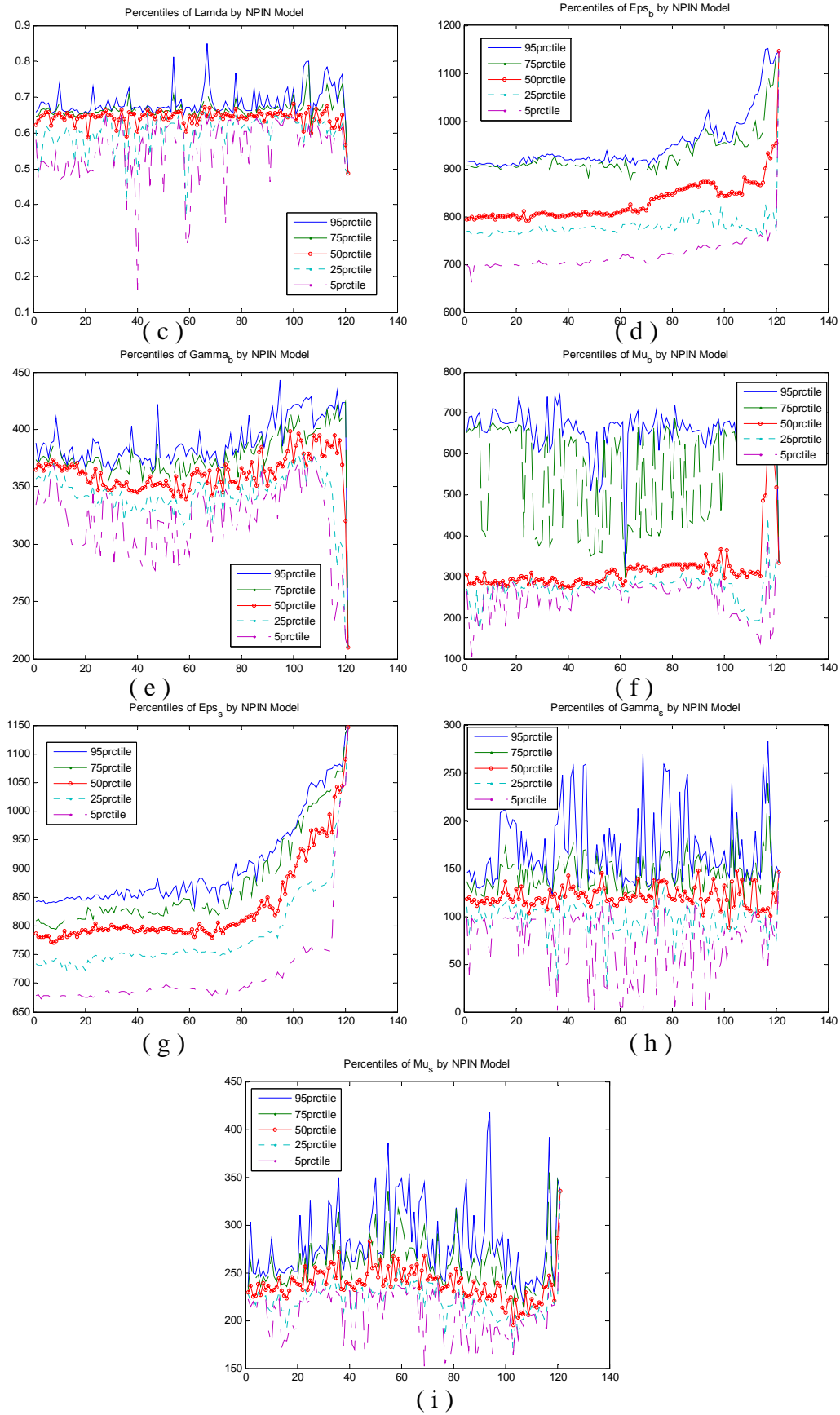


图 4.3 修正的 PIN 模型参数估计结果时间序列

图 4.3(a)、4.3(b)、4.3(c)、4.3(d)、4.3(e)、4.3(f)、4.3(g)、4.3(h)和 4.3(i)分别显示了钢铁行业股票由修正的 PIN 模型估计出来的参数 α 、 δ 、 λ 、 ε_b 、 γ_b 、 μ_b 、 ε_s 、 γ_s 和 μ_s 在 95 分位、75 分位、50 分位、25 分位和 5 分位的时间序列。从图中可以看出， α 的估计值波动比较剧烈，变动范围在 0.20 到 0.89 之间； δ 的估计值波动也比较剧烈，变动范围在 0.28 到 0.995 之间； λ 的估计值波动也比较剧烈，变动范围在 0.16 到 0.85 之间； ε_b 的估计值前面比较平稳，后面则逐渐上升，变动范围在 661 到 1151 之间； γ_b 的估计值前面比较平稳，后面则逐渐上升，变动范围在 210 到 428 之间； μ_b 的估计值波动比较剧烈，变动范围在 105 到 743 之间； ε_s 的估计值呈逐渐上升的趋势，变动范围在 672 到 1147 之间； γ_s 的估计值波动比较剧烈，变动范围在 1 到 283 之间； μ_s 的估计值波动比较剧烈，变动范围在 153 到 418 之间。

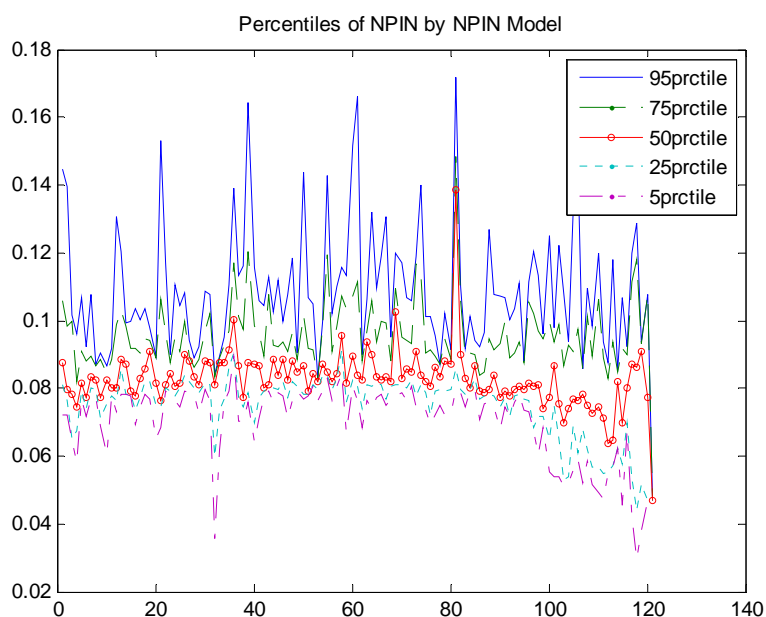


图 4.4 修正的 PIN 模型估计出的信息风险时间序列

图 4.4 显示了钢铁行业股票由修正的 PIN 模型估计出来的参数计算出来的信息风险在 95 分位、75 分位、50 分位、25 分位和 5 分位的时间序列。从图中可以看出，钢铁行业股票的信息风险在整个时间段里波动都比较剧烈，变动范围在 0.03

到 0.17 之间。

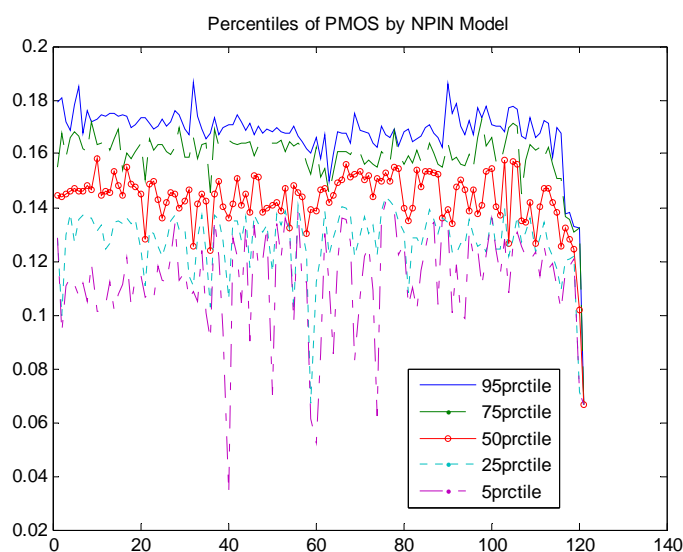


图 4.5 修正的 PIN 模型估计出的市场指令流冲击事件发生时引起的交易概率时间序列

图 4.5 显示了钢铁行业股票由修正的 PIN 模型估计出来的参数计算出来的市场范围内的指令流冲击事件引起的交易概率在 95 分位、75 分位、50 分位、25 分位和 5 分位的时间序列。从图中可以看出，钢铁行业股票由市场范围内的指令流冲击事件引起的交易概率在整个时间段里波动都比较剧烈，变动范围在 0.03 到 0.19 之间。

表 4.2 显示了钢铁行业股票由修正的 PIN 模型估计出来的参数的横截面平均值的时间序列的描述性统计。

表 4.2 钢铁行业股票由修正的 PIN 模型估计出来的参数描述性统计

变量	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
α	0.6384	0.6400	0.7013	0.5244	0.0237	-0.9361	7.3825	108.83	0.0000
δ	0.6945	0.6913	0.7879	0.6077	0.0300	0.3377	3.6213	4.0346	0.1330
λ	0.6348	0.6372	0.7034	0.5195	0.0285	-1.0411	5.7776	57.741	0.0000
ε_b	838.35	827.98	889.71	814.16	20.649	0.9737	2.5754	19.034	0.0000

γ_b	360.85	358.28	395.80	337.61	14.257	0.5543	2.5410	6.8981	0.0318
μ_b	401.90	405.34	467.32	281.72	35.840	-0.6159	2.8236	7.4185	0.0245
ε_s	811.76	787.42	954.47	765.83	52.239	1.5457	4.0693	51.272	0.0000
γ_s	120.92	118.45	157.30	85.380	15.429	0.1440	2.3596	2.3621	0.3070
μ_s	240.17	236.69	287.42	193.35	17.625	0.2580	2.7255	1.6367	0.4412

从表 4.2 中可以看出, 由修正的 PIN 模型估计出来的钢铁行业股票的知情事件发生的概率 α 的平均值为 0.64, 知情事件发生时为坏消息的概率 δ 的平均值为 0.69, 市场指令流冲击事件发生的概率 λ 的平均值为 0.63, 非知情交易者的买单到达率 ε_b 和卖单到达率 ε_s 的平均值分别为 838.35 和 811.76, 市场指令流冲击事件发生时的买单到达率 γ_b 和卖单到达率 γ_s 的平均值分别为 360.85 和 120.92, 知情交易者的买单到达率 μ_b 和卖单到达率 μ_s 的平均值分别为 401.90 和 240.17。

四、对信息风险信息时间序列的序列相关检验

为了更深入地对股票信息风险信息的时间序列特征进行考察, 本文对分别由经典的 PIN 模型和修正的 PIN 模型估计得到的股票信息风险信息的时间序列进行了序列相关检验。

序列相关检验的思路为: 如果股票信息风险信息的时间序列为随机游走过程, 那么该序列任意滞后阶数的自相关系数都与零无显著差异, 且其 Q 统计量应该不显著; 反之, 如果该序列存在序列相关, 那么股票的信息风险就不是随机游走过程。

Box and Pierce (1970)^[72]提出的 Q 统计量可以表示为: $Q^*(m) = T \sum_{l=1}^m \hat{\rho}_l^2$, 它渐进服从 $\chi^2(m)$ 分布。Ljung and Box (1978)^[73]对 $Q^*(m)$ 统计量进行了修正, 从而使得在有限样本里的检验效果更好。修正后的 Q 统计量为: $Q(m) = T(T+2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{T-l}$,

它也渐进服从 $\chi^2(m)$ 分布。检验的原假设为： $\hat{\rho}_1 = \hat{\rho}_2 = \dots = \hat{\rho}_m = 0$ ，即不存在序列相关。

表4.3列出了由经典的PIN模型估计得到的股票信息风险时间序列的序列相关检验结果。

表 4.3 经典的 PIN 模型估计得到的股票信息风险时间序列的序列相关检验

滞后阶数	自相关函数	偏自相关函数	Q 统计量	P 值
1	0.9546	0.9546	110.502	0.0000
2	0.9114	-0.0468	212.158	0.0000
3	0.8634	-0.0662	304.358	0.0000
4	0.8114	-0.0590	386.796	0.0000
5	0.7644	0.0082	460.962	0.0000
6	0.7118	-0.1040	526.258	0.0000
7	0.6602	-0.0254	583.488	0.0000
8	0.6066	-0.0442	632.962	0.0000
9	0.5514	-0.0344	675.190	0.0000
10	0.4946	-0.0632	710.742	0.0000
15	0.2694	-0.0584	822.272	0.0000
20	0.1220	0.0324	872.484	0.0000

从表 4.3 中可以看出，由经典的PIN模型估计得到的股票信息风险时间序列的滞后各阶的样本自相关函数都显著异于零，与滞后各阶的自相关函数相对应的Q统计量都相当大，而相伴概率（P值）都为0，即使在1%的显著性水平下，也必须拒绝由经典的PIN模型估计得到的股票信息风险的时间序列不存在序列相关的原假设^①。这表明由经典的PIN模型估计得到的股票信息风险存在很强的自相关性。

表4.4列出了由修正的PIN模型估计得到的股票信息风险时间序列的序列相关检验结果。

^① 我们对每只股票由经典的PIN模型估计得到的信息风险的时间序列都进行了序列相关检验，检验结果都在1%的置信水平下拒绝不存在序列相关的原假设。这里给出的检验结果为钢铁行业所有股票的平均值。

表 4.4 修正的 PIN 模型估计得到的股票信息风险时间序列的序列相关检验

滞后阶数	自相关函数	偏自相关函数	Q 统计量	P 值
1	0.4905	0.4905	33.6687	0.0133
2	0.3988	0.1838	58.4046	0.0313
3	0.4233	0.1830	82.1713	0.0030
4	0.3830	0.1193	103.560	0.0030
5	0.3593	0.0708	121.623	0.0013
6	0.3383	0.0350	136.737	0.0000
7	0.3073	0.0538	149.840	0.0000
8	0.2978	0.0613	162.970	0.0000
9	0.2958	0.0280	174.655	0.0000
10	0.2545	0.0203	184.189	0.0000
15	0.1005	-0.0660	210.708	0.0000
20	0.0068	-0.0518	213.822	0.0000

从表 4.4 中可以看出，由修正的 PIN 模型估计得到的股票信息风险时间序列的滞后各阶的样本自相关函数都显著异于零，与滞后各阶的自相关函数相对应的 Q 统计量都相当大，而相伴概率（P 值）都几乎为 0，即使在 5% 的显著性水平下，也必须拒绝由经典的 PIN 模型估计得到的股票信息风险的时间序列不存在序列相关的原假设^①。这表明由修正的 PIN 模型估计得到的股票信息风险存在很强的自相关性。这一结果与由经典的 PIN 模型估计得到的股票信息风险的时间序列特征相同。

表 4.5 列出了对由修正的 PIN 模型估计得到的市场指令流冲击事件引起的交易概率时间序列的序列相关检验结果。

表 4.5 修正的 PIN 模型估计得到的市场指令流冲击事件发生时引起的交易概

^① 我们对每只股票由修正的 PIN 模型估计得到的信息风险的时间序列都进行了序列相关检验，检验结果除了马钢股份（600808）外其余所有股票都在 1% 的置信水平下拒绝不存在序列相关的原假设。这里给出的检验结果为钢铁行业除了马钢股份（600808）之外所有股票的平均值。

率时间序列的序列相关检验

滞后阶数	自相关函数	偏自相关函数	Q 统计量	P 值
1	0.3057	0.3057	13.052	0.0210
2	0.1073	0.0030	17.003	0.0327
3	0.1053	0.0700	20.428	0.0497
4	0.2243	0.1630	28.347	0.0607
5	0.2967	0.2003	39.826	0.0010
6	0.0783	-0.0797	44.338	0.0007
7	0.0903	0.0747	46.777	0.0013
8	0.0600	-0.0603	48.143	0.0023
9	0.1677	0.0917	52.271	0.0003
10	0.2580	0.1253	62.748	0.0000
15	0.1127	0.0097	71.233	0.0000
20	0.09	0.0180	76.448	0.0000

从表 4.5 中可以看出, 由修正的PIN模型估计得到的股票市场指令流冲击事件发生时引起的交易概率的时间序列滞后各阶的样本自相关函数都显著异于零, 并且滞后各阶的自相关函数相对应的Q统计量都相当大, 而相伴概率(P值)都几乎为 0, 即使在 5%的显著性水平下, 也必须拒绝由经典的PIN模型估计得到的股票信息风险的时间序列不存在序列相关的原假设^①。这表明由修正的PIN模型估计得到的股票市场指令流冲击事件发生时引起的交易概率存在很强的自相关性。

§ 4.2 中国股票市场信息风险与股票收益率关系研究

Hasbrouck (1991)^[27]发现信息冲击会对交易有永久性的影响。Chordia, Roll and Subrahmanyam (2002)^[62]的研究发现在控制交易量和流动性之后, 股票市场收益率

^① 本文对每只股票由修正的PIN模型估计得到的市场指令流冲击事件发生时引起的交易概率的时间序列都进行了序列相关检验, 检验结果除了承德钒钛(600357)和马钢股份(600808)外的其余所有钢铁行业股票都在1%的置信水平下拒绝不存在序列相关的原假设。这里给出的检验结果为钢铁行业除了承德钒钛(600357)和马钢股份(600808)之外所有股票的平均值。

仍然受到整个市场同时和滞后的指令流非平衡的影响。Chordia and Subrahmanyam (2004)^[74]对个股的研究则进一步证实了该结论,即个股收益率受到同时和滞后的指令流非平衡的影响^①。无论是经典的PIN模型还是修正的PIN模型,其实质都是利用股票的指令流非平衡来对信息风险进行测度。由于交易的结果最终会反映在价格上,那么股票的信息风险同股票收益率之间存在什么样的关系呢?本文接下来对该问题进行考察。

由于前文已经分别用经典的PIN模型和修正的PIN模型估计出来钢铁行业股票的信息风险的时间序列,因此接下来本文对钢铁行业股票信息风险同股票收益率之间的关系进行检验。

一、数据及描述性统计

本文在对钢铁行业股票信息风险同股票收益率之间的关系进行检验时,采用的数据为钢铁行业所有股票在2007年的每日开盘价和每日收盘价、上证综合指数的每日开盘价和每日收盘价。数据来源为万得金融数据库。

同Chordia and Subrahmanyam (2004)^[74]的研究相同,本文采用的收益率为每日的开盘收盘收益率^②,其定义为:

$$R_{it} = \ln\left(\frac{P_{cit}}{P_{oit}}\right), \quad R_{mt} = \ln\left(\frac{P_{cmt}}{P_{omt}}\right)$$

其中, R_{it} 为第 i 只股票在第 t 个交易日的开盘收盘收益率; P_{cit} 为第 i 只股票在第 t 个交易日的收盘价; P_{oit} 为第 i 只股票在第 t 个交易日的开盘价; R_{mt} 为指数在第 t 个交易日的开盘收盘收益率; P_{cmt} 为指数在第 t 个交易日的收盘价; P_{omt} 指数在第 t 个交易日的开盘价。

表 4.6 列出了上证综合指数开盘收盘收益率和钢铁行业股票横截面平均开盘收盘收益率的描述性统计。

表 4.6 指数和股票开盘收盘收益率描述性统计

^① 在 Chordia, Roll and Subrahmanyam (2002)^[62]和 Chordia and Subrahmanyam (2004)^[74]中,指令流非平衡的定义都为买单数量减去卖单数量。

^② 采用每日的收盘价计算的收益率并不影响本文的结论。

种类	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
上证综指	0.0030	0.0057	0.0553	-0.0953	0.0207	-0.9287	5.7113	108.91	0.0000
钢铁行业股票	0.0043	0.0063	0.0906	-0.0858	0.0251	-0.2120	4.1510	14.731	0.0006

表 4.7 列出了由经典的 PIN 模型估计出来的钢铁行业股票的信息风险 (PIN) 和修正的 PIN 模型估计出来的钢铁行业股票的信息风险 (NPIN) 以及市场指令流冲击事件发生时引起的交易概率 (PMOS) 的描述性统计。

表 4.7 信息风险和市场指令流冲击事件发生时引起的交易概率描述性统计

种类	均值	中位数	最大值	最小值	标准差	偏度	峰度	J-B	P 值
PIN	0.1284	0.1276	0.1470	0.1132	0.0084	0.1684	2.4519	1.9832	0.3710
NPIN	0.0864	0.0861	0.1237	0.0686	0.0075	1.0397	7.5077	118.08	0.0000
PMOS	0.1441	0.1448	0.1543	0.1159	0.0061	-1.4097	6.4674	95.699	0.0000

从表 4.7 中可以看出,由经典的 PIN 模型估计出来的钢铁行业股票的信息风险 (PIN) 的平均值为 0.13,由修正的 PIN 模型估计出来的钢铁行业股票的信息风险 (NPIN) 的平均值为 0.09,市场指令流冲击事件发生时引起的交易概率的平均值为 0.14。

二、PIN、NPIN 和 PMOS 对股票收益率的影响

我们首先考察由经典的 PIN 模型估计出来的信息风险 (PIN) 对股票收益率的影响。Chordia and Subrahmanyam (2004)^[74]在研究个股收益率同指令流非平衡之间的关系时,利用个股收益率与当期和滞后期的指令流非平衡进行回归,发现不仅是当期的指令流非平衡对股票收益率有显著影响,而且滞后期的指令流非平衡对个股收益率也有影响。本文采用同 Chordia and Subrahmanyam (2004)^[74]类似的方法,对下列方程进行回归:

$$R_{it} - R_{mt} = c_i + \sum_{k=0}^q b_{ik} PIN_{i,t-k} + \varepsilon_i \quad (4.1)$$

之所以在回归中加入经典的PIN模型估计出来的信息风险（PIN）的滞后项，是因为前文对信息风险（PIN）的时间序列的相关性进行检验得到的结论为信息风险（PIN）存在很强的自相关性。本文在回归时根据Akaike信息准则（AIC）和Schwartz的贝叶斯信息准则（SBIC），选择最优的滞后阶数为3。回归结果^①见表4.8。

表 4.8 PIN 对股票收益率的影响回归结果

变量	系数	T 值
c	0.0032 ^{**}	2.40
PIN	0.1434 ^{***}	2.86
PIN_{t-1}	-0.0247	-0.36
PIN_{t-2}	-0.0195	-0.26
PIN_{t-3}	-0.1278 ^{**}	-2.30
R^2	0.16	

注：*表示在 10%的置信度水平下显著；**表示在 5%的置信度水平下显著；***表示在 1%的置信度水平下显著。

从表 4.8 中可以看出，当期的股票信息风险PIN的系数在 1%的置信水平下显著，滞后第三期的系数也在 5%的置信水平下显著，这意味着不仅当期的股票信息风险对股票收益率有着显著的影响，而且滞后期的股票信息风险还对股票收益率有影响。当期的信息风险对股票收益率有显著为正的影响，而滞后期的信息风险对股票收益率的影响为负，这同Chordia and Subrahmanyam (2004)^[74]的研究结论类似^②。

^① 该结果为钢铁行业各只股票的时间序列回归系数的横截面平均值。如无特别说明，以下各个回归的结果类似。

^② Chordia and Subrahmanyam (2004)的研究发现当期的指令流非平衡对个股收益率有显著为正的影响，滞后期的指令流非平衡对个股收益率有显著为负的影响。

接下来我们考察由修正的 PIN 模型估计得到的股票信息风险 (NPIN) 对股票收益率的影响。本文也采用同 Chordia and Subrahmanyam (2004)^[74]类似的方法, 对下列方程进行回归:

$$R_{it} - R_{mt} = c_i + \sum_{k=0}^q b_{ik} NPIN_{i,t-k} + \varepsilon_i \quad (4.2)$$

我们同样在回归方程中分别加入 NPIN 的滞后项, 因为从前面对 NPIN 的时间序列的检验中, 发现它存在很强的自相关性。本文在回归时根据 Akaike 信息准则 (AIC) 和 Schwartz 的贝叶斯信息准则 (SBIC), 选择最优的滞后阶数为 2。回归结果见表 4.9。

表 4.9 NPIN 对股票收益率的影响回归结果

变量	系数	T 值
c	-0.0029 **	-2.56
$NPIN$	0.0125	0.96
$NPIN_{t-1}$	0.0084	0.61
$NPIN_{t-2}$	0.0085	0.64
R^2	0.03	

注: *表示在 10%的置信度水平下显著; **表示在 5%的置信度水平下显著; ***表示在 1%的置信度水平下显著。

从表 4.9 中可以看出, 当采用由修正的 PIN 模型估计得到的信息风险对股票收益率进行回归时, 无论是当期的股票信息风险还是滞后期的信息风险的系数都不显著, 这一结论与表 4.8 中的结论是矛盾的。

接下来我们再考察市场指令流冲击事件发生时引起的交易概率 (PMOS) 对股票收益率的影响。类似地, 我们对下列方程进行回归:

$$R_{it} - R_{mt} = c_i + \sum_{k=0}^q b_{ik} PMOS_{i,t-k} + \varepsilon_i \quad (4.3)$$

在回归时同样根据 Akaike 信息准则 (AIC) 和 Schwartz 的贝叶斯信息准则

(SBIC), 选择最优的滞后阶数为 4。回归结果见表 4.10。

表 4.10 PMOS 对股票收益率的影响回归结果

变量	系数	T 值
c	-0.0236***	-8.54
$PMOS$	0.0173	1.32
$PMOS_{t-1}$	0.0263*	1.71
$PMOS_{t-2}$	0.0442***	2.71
$PMOS_{t-3}$	0.0418**	2.55
$PMOS_{t-4}$	0.0396**	2.41
R^2	0.39	

注: *表示在 10%的置信度水平下显著; **表示在 5%的置信度水平下显著; ***表示在 1%的置信度水平下显著。

从表 4.10 中可以看出, 除了当期的市场指令流冲击事件发生时引起的交易概率 (PMOS) 的系数不显著外, 一直到滞后第四期的系数都在 10%的置信水平下显著, 这意味着市场指令流冲击事件发生时引起的交易概率 (PMOS) 对股票收益率有着持续而且显著的影响。

我们前面发现采用经典的 PIN 模型估计得到的信息风险 (PIN) 和修正的 PIN 模型估计得到的信息风险 (NPIN) 对股票收益率进行回归时得到的结论是矛盾的。为了考察造成这一矛盾的原因, 我们首先在方程 (4.1) 中加入市场范围内的指令流冲击事件引起的交易概率 (PMOS), 对下列方程进行回归:

$$R_{it} - R_{mt} = c_i + \sum_{k=0}^q b_{1ik} PIN_{i,t-k} + \sum_{k=0}^q b_{2ik} PMOS_{i,t-k} + \varepsilon_i \quad (4.4)$$

在回归时仍然根据 Akaike 信息准则 (AIC) 和 Schwartz 的贝叶斯信息准则 (SBIC), 选择 PIN 的最优滞后阶数为 3, PMOS 的最优滞后阶数为 4。回归结果见表 4.11。

表 4.11 PIN 和 PMOS 对股票收益率的影响回归结果

变量	系数	T 值
<i>c</i>	-0.0180 ^{***}	-5.98
<i>PIN</i>	0.0899 ^{**}	2.14
<i>PIN</i> _{<i>t</i>-1}	0.0317	0.52
<i>PIN</i> _{<i>t</i>-2}	-0.0541	-0.82
<i>PIN</i> _{<i>t</i>-3}	-0.0928 ^{**}	-1.93
<i>PMOS</i>	0.0254 [*]	1.96
<i>PMOS</i> _{<i>t</i>-1}	0.0277 [*]	1.89
<i>PMOS</i> _{<i>t</i>-2}	0.0370 ^{**}	2.20
<i>PMOS</i> _{<i>t</i>-3}	0.0331 ^{**}	1.99
<i>PMOS</i> _{<i>t</i>-4}	0.0296 [*]	1.86
<i>R</i> ²	0.48	

注：*表示在 10%的置信度水平下显著；**表示在 5%的置信度水平下显著；***表示在 1%的置信度水平下显著。

从表 4.11 中可以看出，加入市场指令流冲击事件发生时引起的交易概率（PMOS）后，由经典的 PIN 模型估计得到的股票信息风险（PIN）的同期和滞后第三期对股票收益率的影响仍然是显著的，这与表 4.8 中得到的结论相同。而市场指令流冲击事件发生时引起的交易概率（PMOS）对股票收益率也有着显著的影响，这种影响不仅反映在同期，而且在滞后几期这种影响仍然是显著的。

比较分别用经典的 PIN 模型估计得到的股票信息风险（PIN）和用修正的 PIN 模型估计得到的股票信息风险（NPIN）可以发现，用修正的 PIN 模型估计得到的股票信息风险（NPIN）在数值上总是小于经典的 PIN 模型估计得到的信息风险（PIN），我们把它们之间的差定义为变量 NSO，即 $NSO = PIN - NPIN$ 。我们预计 NSO 应该与经典的 PIN 模型估计得到的信息风险（PIN）高度相关。表 4.12 显示

了经典的 PIN 模型估计得到的信息风险 (PIN)、修正的 PIN 模型估计得到的股票信息风险 (NPIN) 和市场指令流冲击事件发生时引起的交易概率 (PMOS) 以及 NSO 之间的相关关系。

表 4.12 PIN、NPIN、PMOS 和 NSO 之间的相关性

	PIN	NPIN	PMOS	NSO
PIN	1.0000 (---)			
NPIN	0.6187 (0.00)	1.0000 (---)		
PMOS	0.2160 (0.02)	-0.0460 (0.62)	1.0000 (---)	
NSO	0.6151 (0.00)	-0.2389 (0.01)	0.3131 (0.00)	1.0000 (---)

注：括号内为相关系数的 P 值。

从表 4.12 中可以看出，经典的 PIN 模型估计得到的信息风险 (PIN) 同修正的 PIN 模型估计得到的股票信息风险 (NPIN) 和市场指令流冲击事件发生时引起的交易概率 (PMOS) 以及 NSO 之间的具有很强的正相关关系，而市场指令流冲击事件发生时引起的交易概率 (PMOS) 和 NSO 之间也具有很强的正相关关系，修正的 PIN 模型估计得到的股票信息风险 (NPIN) 和市场指令流冲击事件发生时引起的交易概率 (PMOS) 以及 NSO 之间的相关性都为负。

然后我们考察 NSO 对股票收益率的影响，我们对下列方程进行回归：

$$R_{it} - R_{mt} = c_i + \sum_{k=0}^q b_{ik} NSO_{i,t-k} + \varepsilon_i \quad (4.5)$$

在回归时仍然根据 Akaike 信息准则 (AIC) 和 Schwartz 的贝叶斯信息准则 (SBIC)，选择 NSO 的最优滞后阶数为 1。回归结果见表 4.13。

表 4.13 NSO 对股票收益率的影响回归结果

变量	系数	T 值
c	0.0016 ^{**}	2.18
NSO	-0.0184	-1.50
NSO_{t-1}	-0.0238 [*]	-1.90
R^2	0.08	

注：^{*}表示在 10%的置信度水平下显著；^{**}表示在 5%的置信度水平下显著；^{***}表示在 1%的置信度水平下显著。

从表 4.12 中可以看出，经典的 PIN 模型估计得到的信息风险（PIN）和修正的 PIN 模型估计得到的股票信息风险（NPIN）之间的差 NSO 的滞后一期对股票收益率有显著为负的影响。

我们接下来把 NSO 加入方程（4.1）中，对下列方程进行回归：

$$R_{it} - R_{mt} = c_i + \sum_{k=0}^q b_{1ik} PIN_{i,t-k} + \sum_{k=0}^q b_{2ik} NSO_{i,t-k} + \varepsilon_i \quad (4.6)$$

在回归时同样根据 Akaike 信息准则（AIC）和 Schwartz 的贝叶斯信息准则（SBIC），选择 PIN 的最优滞后阶数为 3， NSO 的最优滞后阶数为 4。回归结果见表 4.14。

表 4.14 PIN 和 NSO 对股票收益率的影响回归结果

变量	系数	T 值
c	0.0005	0.37
PIN	0.0481	0.88
PIN_{t-1}	0.0361	0.50
PIN_{t-2}	-0.0031	-0.04
PIN_{t-3}	-0.0404	-0.67
NSO	-0.0243 [*]	-1.90

NSO_{t-1}	-0.0200	-1.60
NSO_{t-2}	-0.0263*	-1.92
NSO_{t-3}	-0.0248*	-1.78
NSO_{t-4}	-0.0258*	-1.98
R^2		0.29

注：*表示在 10%的置信度水平下显著；**表示在 5%的置信度水平下显著；***表示在 1%的置信度水平下显著。

从表 4.14 中可以看出，在方程（4.1）中加入经典的 PIN 模型估计得到的信息风险（PIN）和修正的 PIN 模型估计得到的股票信息风险（NPIN）之间的差 NSO 后，同期以及滞后各期的经典的 PIN 模型估计得到的信息风险（PIN）对股票收益率的影响不再显著，而经典的 PIN 模型估计得到的信息风险（PIN）和修正的 PIN 模型估计得到的股票信息风险（NPIN）之间的差 NSO 对股票收益率的影响除了滞后一期不显著外，其余各期都在 10%的置信度水平下显著为负。

综合上述结果，我们可以得到以下结论，即信息风险对股票收益率并没有显著的影响。由经典的 PIN 模型估计得到的信息风险（PIN）对股票收益率有影响是因为在该模型下估计得到的信息风险（PIN）中包含了与市场指令流冲击事件发生时引起的交易概率相关的信息，而市场指令流冲击事件发生时引发的交易概率对股票收益率有着持续显著的影响。

我们很好奇市场范围内的指令流冲击事件发生时引发的交易概率到底与什么有关？我们推测有可能是市场流动性。因为当市场范围内的指令流冲击事件发生时，市场流动性会由于市场交易量的广泛增加而提高。换手率是一种衡量流动性的近似指标，接下来我们通过考察换手率与市场指令流冲击事件发生时引起的交易概率之间的关系来验证这一推测。

表 4.15 显示了经典的 PIN 模型估计得到的信息风险（PIN）、修正的 PIN 模型估计得到的股票信息风险（NPIN）和市场指令流冲击事件发生时引起的交易概率（PMOS）以及换手率（TURNOVER）之间的相关关系。

表 4.15 PIN、NPIN、PMOS 和换手率之间的相关性

	PIN	NPIN	PMOS	TURNOVER
PIN	1.0000 (---)			
NPIN	0.6187 (0.00)	1.0000 (---)		
PMOS	0.2160 (0.02)	-0.0460 (0.62)	1.0000 (---)	
TURNOVER	0.0523 (0.57)	0.0942 (0.30)	0.1584 (0.08)	1.0000 (---)

注：括号内为相关系数的 P 值。

从表 4.15 中可以看出，市场指令流冲击事件发生时引起的交易概率（PMOS）和换手率（TURNOVER）之间存在显著为正的相关性，这证实了之前的推测，即市场指令流冲击事件发生时引起的交易概率（PMOS）同市场流动性有关。

因此，综上所述我们可以得到以下结论，那就是股票信息风险对股票收益率并没有显著的影响，对股票收益率有着持续显著影响的因素是流动性。

第5章 结论及未来进一步的研究方向

§5.1 本文主要结论

对股票信息风险的准确测度对于资产定价和市场绩效的衡量有着重要意义。本文对股票信息风险的测度及信息风险同股票收益率之间的关系进行了研究，得出了以下结论。

一、股票信息风险的测度

1、Easley, Kiefer, O'Hara and Paperman (1996)^[1]提出的 PIN 模型为直接测度股票信息风险的经典模型，本文对利用经典的 PIN 模型测度我国股票信息风险进行的实证研究表明，经典的 PIN 模型隐含的买单和卖单数量之间的负相关关系与实际数据中买单和卖单数量之间的正相关关系并不相符，而且经典的 PIN 模型隐含的买单和卖单数量的方差与实际数据中买卖指令相对较大的方差也不是很匹配。这证实了 Venter and De Jongh (2006)^[16]和 Duarte and Young (2008)^[17]的结论。

2、本文在 Easley, Kiefer, O'Hara and Paperman (1996)^[1]的基础上，对经典的 PIN 模型进行了修正，提出了修正的 PIN 模型，与经典的 PIN 模型相比，修正的 PIN 模型在交易动机中增加了由于市场指令流冲击事件发生时引起的交易。本文利用修正的 PIN 模型对我国股票信息风险进行的实证研究表明，修正的 PIN 模型隐含的买单和卖单数量之间的相关性为正，这能够较好地匹配实际数据中买单和卖单数量之间的正相关关系。此外，修正的 PIN 模型也能够较好地匹配实际数据中买单和卖单数量较大的方差。

3、本文选取换手率作为交易活跃程度的衡量指标，利用经典的 PIN 模型对不同的股票具有的信息风险进行的实证研究表明，股票信息风险的大小同交易活跃程度之间呈正相关关系，换手率越高的股票总体上信息风险越高。然而，本文利用修正的 PIN 模型对交易活跃程度不同的股票具有的信息风险进行的实证研究却得出了相反的结论，即股票信息风险的大小同交易活跃程度之间呈负相关关系，换手率越高的股票其信息风险越低。这主要是因为经典的 PIN 模型忽视了市场指

令流冲击事件发生时引起的交易动机，从而倾向于高估交易活跃程度高的股票具有的信息风险。

二、股票信息风险的时间序列特征及其与股票收益率之间的关系

1、本文采用滚动（rolling）的方法估计出了股票信息风险的时间序列，发现股票的信息风险是一个时变的过程，而且存在很强的序列相关性。

2、本文对股票信息风险和股票收益率之间的关系进行的实证研究表明，股票信息风险对股票收益率并不存在显著的影响，而与市场流动性相关的市场指令流冲击事件发生时引起的交易概率对股票收益率有着持续显著的影响。

§5.2 未来进一步的研究方向

本文的一个主要工作是通过在经典的 PIN 模型中加入市场指令流冲击事件发生时引起的交易动机，对经典的 PIN 模型进行了修正，利用我国股票的交易数据进行的实证研究表明，修正的 PIN 模型与经典的 PIN 模型相比能够匹配实际数据中买单和卖单数量之间的正相关关系，而且也能够更好地匹配实际数据中买卖指令相对较大的方差。然而，修正的 PIN 模型隐含的买单和卖单数量之间的正相关关系有时非常低，与实际数据中买单和卖单数量之间较大的正相关关系并不能完全的匹配。此外，虽然修正的 PIN 模型隐含的买单数量的方差与经典的 PIN 模型中隐含的买单数量的方差相比能够更好地与实际数据中卖单数量的方差相匹配，但是修正的 PIN 模型隐含的卖单数量的方差与实际数据中卖单数量较大的方差之间仍然不能完全的匹配。

因此，未来的一个研究方向是在修正的 PIN 模型的基础上，进一步对信息风险的测度模型进行完善，使得模型隐含的买卖单数量之间的相关性和买卖单数量的方差能够更好地与实际数据中的买卖单数量之间的相关性和买卖单数量的方差相匹配。

本文的另一个工作是通过滚动（rolling）的方法构造了股票信息风险的时间序列，对股票信息风险和股票收益率之间的关系进行的实证研究发现股票信息风险对股票收益率并不存在显著的影响，而与市场流动性相关的市场指令流冲击事件

发生时引起的交易概率对股票收益率有着持续显著的影响。然而，限于高频数据可获得性的限制，本文没有对股票信息风险是否是股票横截面收益的显著决定因素进行标准的资产定价检验。

因此，未来的另一个研究方向是在更长时间段内的高频数据可获得的情况下，对股票信息风险是否是股票横截面收益率的显著决定因素进行标准的资产定价检验。

附录 1 样本股票的详细情况

本附录显示了本文在利用经典的 PIN 模型和修正的 PIN 模型来测度我国股票的信息风险时包含在样本内的各只股票的详细情况。为了更好地比较交易活跃程度不同的股票其信息风险大小之间的区别，我们选择换手率最低、换手率居中和换手率最高的三组股票作为我们最后的研究样本，每组股票数量为 25 只。表中平均换手率、平均收盘价、平均成交量、平均成交金额分别为各只股票从 2007 年 1 月 4 日到 2007 年 12 月 28 日的每日平均换手率、每日平均收盘价格、每日平均成交量和每日平均成交金额。2007 年流通市值为各只股票 2007 年末的流通市值。

证券名称	证券代码	平均换手率	平均收盘价	平均成交量	平均成交金额	2007 年流通市值
A 栏：换手率最低的一组股票						
贵州茅台	600519	0.78	132.17	2708924.05	325408261.60	93824122.18
福耀玻璃	600660	0.92	13.08	4154051.87	104358730.29	16544658.95
中国船舶	600150	0.95	148.10	1043195.71	142218214.59	33659476.20
烟台万华	600309	1.00	32.12	6336537.92	246133545.83	31331376.96
浦发银行	600000	1.10	38.49	22475107.47	782615556.02	186818418.44
海油工程	600583	1.13	19.41	4819645.42	196732379.17	24766864.45
上海机场	600009	1.26	33.35	12382174.27	413749311.20	37412260.38
招商银行	600036	1.52	27.92	71747396.25	1764198204.17	186688917.57
中信证券	600030	1.54	64.43	30472815.56	1873003684.44	249202417.74
东方电气	600875	1.59	63.64	1444092.13	86966365.74	8784300.00
金地集团	600383	1.90	18.07	8713964.85	262805481.17	24439220.28
歌华有线	600037	1.98	25.71	9852594.14	286427029.29	18337499.91
天坛生物	600161	2.01	23.22	2858620.42	94432400.00	6790095.00
上海汽车	600104	2.05	18.80	30645670.09	561577358.97	43235260.04
保利地产	600048	2.21	25.88	6527658.97	347122923.08	34091565.04
华能国际	600011	2.24	11.64	26540200.00	395711581.59	41078840.55
百联股份	600631	2.24	18.32	13052605.83	243039862.50	15459573.19

股票信息风险测度研究

大秦铁路	601006	2.37	16.45	49382878.42	772509435.69	66669030.89
同仁堂	600085	2.39	26.29	4633689.63	141204788.38	6779408.24
潞安环能	601699	2.40	25.92	4374955.42	142574858.33	15251125.00
山东黄金	600547	2.42	47.59	1850107.11	150114008.37	13105174.50
宝钢股份	600019	2.45	13.25	121185379.92	1560352297.07	97857279.83
中远航运	600428	2.45	12.53	8026683.04	187740452.17	14003286.71
广船国际	600685	2.54	52.19	4080746.47	183137904.56	13134630.47
青岛啤酒	600600	2.56	26.16	6029988.29	144631033.47	9227470.07
B 栏：换手率居中的一组股票						
大同煤业	601001	3.29	24.86	9723995.85	212451734.44	10584000.00
兰花科创	600123	3.44	24.71	6928868.47	218006560.17	11383536.75
中国石化	600028	3.53	14.95	140080762.19	2090268298.32	199132237.92
上港集团	600018	3.54	8.98	88233150.21	824408561.70	72650965.49
国电电力	600795	3.56	6.91	39778028.45	552300784.48	29944612.58
赣粤高速	600269	3.59	14.05	20786592.92	286177320.83	11373283.85
申能股份	600642	3.60	14.63	49539756.02	741881481.33	24559863.28
浦东金桥	600639	3.63	18.64	8193401.25	154323362.50	5101689.05
四川长虹	600839	3.64	7.97	47166651.70	399849059.32	11294658.73
申华控股	600653	3.66	5.95	53325048.13	338711294.61	8921092.79
雅戈尔	600177	3.70	20.94	39621760.85	964284557.45	30213981.46
生益科技	600183	3.72	13.20	20195999.17	267819312.50	9600387.18
新潮中宝	600208	3.81	8.74	9653682.02	139662640.35	4350365.31
建发股份	600153	3.89	10.65	11647291.60	202406588.24	7758256.00
中国联通	600050	3.92	6.96	328709847.72	2162138423.24	113335862.71
中华企业	600675	3.95	17.52	20703978.75	392038108.33	12443387.67
方正科技	600601	4.01	6.90	55739856.49	479678221.76	13138563.59
中国国航	601111	4.06	13.73	52267746.64	657886521.01	35370160.00
中国银行	601988	4.07	5.73	211195345.87	1208907462.81	34423518.34
世茂股份	600823	4.17	18.28	11059026.52	163319173.91	6868957.63
北京城建	600266	4.17	24.05	9767252.50	226473408.33	6911730.00
福建高速	600033	4.19	7.63	22674666.67	179602408.33	5160067.67

附 录

华夏银行	600015	4.26	14.45	68426309.40	884593764.96	32315256.00
中炬高新	600872	4.28	6.78	25152301.71	180236128.21	4691522.74
中化国际	600500	4.30	16.82	22391257.02	376967378.72	15343742.43
C 栏：换手率最高的一组股票						
包钢股份	600010	5.11	6.69	123711969.58	849804137.50	21381471.02
华电国际	600027	5.16	7.53	39837207.50	299348079.17	10804312.55
马钢股份	600808	5.20	8.27	46226553.75	382448408.33	11998419.63
新黄浦	600638	5.43	19.04	17893193.28	352623432.77	10696370.50
西单商场	600723	5.44	10.40	14024057.50	150609308.33	5104456.02
酒钢宏兴	600307	5.44	14.38	17788760.00	224354744.68	9409424.45
大众交通	600611	5.54	11.79	15844504.15	310904145.23	7527992.76
承德钒钛	600357	5.55	12.81	30208219.07	370160906.78	11129407.13
紫江企业	600210	5.64	7.32	53237680.25	394375306.72	10300092.95
兖州煤业	600188	5.71	15.55	20558253.53	321627361.00	7934400.00
西宁特钢	600117	5.71	10.97	19498020.26	209404577.59	8107222.48
健康元	600380	5.87	8.78	12746703.36	183052247.90	5794551.00
华北制药	600812	5.87	6.82	44209410.04	305981878.66	8892172.85
北大荒	600598	5.88	10.29	28358546.03	297438552.30	7351776.00
江西铜业	600362	5.94	33.52	16748972.37	521806407.90	14397259.25
好当家	600467	5.95	9.81	12714415.42	169580016.67	5785933.25
招商轮船	601872	6.05	11.06	49421669.71	548303904.56	20293977.51
大连控股	600747	6.06	6.73	39663187.62	263768573.39	6764927.16
北辰实业	601588	6.17	11.20	59831413.75	657358104.17	21240000.00
海欣股份	600851	6.34	10.93	33614067.08	390180012.50	7494158.89
香溢融通	600830	6.41	10.66	16607430.21	160960510.64	6567895.66
川投能源	600674	6.47	9.64	13798984.98	166880624.41	7193232.73
海南航空	600221	6.57	7.94	37088444.35	305020020.92	15745672.28
浙江医药	600216	6.60	9.66	16765735.32	154764311.93	5303221.99
金健米业	600127	6.72	8.17	26760191.70	213739186.72	3558194.28

附录 2 对经典的 PIN 模型进行参数估计时采用的因式分解的推导

本附录对在经典的 PIN 模型进行参数估计时采用的因式分解进行推导。

在经典的 PIN 模型中，似然函数为：

$$\begin{aligned} L((B, S) | \theta) &= (1 - \alpha) \times e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!} \\ &\quad + \alpha \delta \times e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-(\varepsilon_s + \mu)} \frac{(\varepsilon_s + \mu)^S}{S!} \\ &\quad + \alpha(1 - \delta) \times e^{-(\varepsilon_b + \mu)} \frac{(\varepsilon_b + \mu)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!} \end{aligned}$$

因为似然函数中各项都有 $\frac{1}{B!} \times \frac{1}{S!}$ ，因此我们可以将这项作为公因式提出来，当我们将似然函数取对数后，该项为一个常数，并不会影响参数的估计，因此我们可以不考虑这项。

令 $M = (\max(B, S) + \min(B, S)) / 2$ ，我们对似然函数取对数后，同时乘以和除以下列这项：

$$e^{-\varepsilon_b - \varepsilon_s} (\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M}$$

因此，有

$$\begin{aligned} \ln(L) &= \ln \left((1 - \alpha) e^{-\varepsilon_b} (\varepsilon_b)^B e^{-\varepsilon_s} (\varepsilon_s)^S + \alpha \delta e^{-\varepsilon_b} (\varepsilon_b)^B e^{-(\varepsilon_s + \mu)} (\varepsilon_s + \mu)^S \right. \\ &\quad \left. + \alpha(1 - \delta) e^{-(\varepsilon_b + \mu)} (\varepsilon_b + \mu)^B e^{-\varepsilon_s} (\varepsilon_s)^S \right) \\ &= \ln \left(\begin{array}{l} (1 - \alpha) e^{-\varepsilon_b} (\varepsilon_b)^B e^{-\varepsilon_s} (\varepsilon_s)^S \\ + \alpha \delta e^{-\varepsilon_b} (\varepsilon_b)^B e^{-(\varepsilon_s + \mu)} (\varepsilon_s + \mu)^S \\ + \alpha(1 - \delta) e^{-(\varepsilon_b + \mu)} (\varepsilon_b + \mu)^B e^{-\varepsilon_s} (\varepsilon_s)^S \end{array} \right) \times \frac{e^{-\varepsilon_b - \varepsilon_s} (\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M}}{e^{-\varepsilon_b - \varepsilon_s} (\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M}} \\ &= \ln \left(\begin{array}{l} (1 - \alpha) (\varepsilon_b)^{B-M} (\varepsilon_b + \mu)^{-(B-M)} (\varepsilon_s)^{S-M} (\varepsilon_s + \mu)^{-(S-M)} \\ + \alpha \delta e^{-\mu} (\varepsilon_b)^{B-M} (\varepsilon_b + \mu)^{-(B-M)} (\varepsilon_s)^{-M} (\varepsilon_s + \mu)^M \\ + \alpha(1 - \delta) e^{-\mu} (\varepsilon_b)^{-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu)^M (\varepsilon_s + \mu)^{-(S-M)} \end{array} \right) \\ &\quad + \left(-\varepsilon_b - \varepsilon_s + \ln \left((\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M} \right) \right) \end{aligned}$$

$$= \ln \left(\begin{array}{l} (1-\alpha) \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \\ + \alpha \delta e^{-\mu} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{-M} \\ + \alpha (1-\delta) e^{-\mu} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{-M} \end{array} \right) \\ + \left[-\varepsilon_b - \varepsilon_s + \ln \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^M \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^M (\varepsilon_b + \mu)^B (\varepsilon_s + \mu)^S \right]$$

$$= \ln \left(\begin{array}{l} (1-\alpha) \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \\ + \alpha \delta e^{-\mu} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{-M} \\ + \alpha (1-\delta) e^{-\mu} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{-M} \end{array} \right) \\ + \left[-\varepsilon_b - \varepsilon_s + M \ln \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right) + M \ln \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right) + B \ln(\varepsilon_b + \mu) + S \ln(\varepsilon_s + \mu) \right]$$

令 $x_s = \frac{\varepsilon_s}{\varepsilon_s + \mu}$, $x_b = \frac{\varepsilon_b}{\varepsilon_b + \mu}$, 则有

$$\ln(L) = \left[-\varepsilon_b - \varepsilon_s + M \ln(x_s) + M \ln(x_b) + B \ln(\varepsilon_b + \mu) + S \ln(\varepsilon_s + \mu) \right] \\ + \ln \left((1-\alpha) (x_s)^{S-M} (x_b)^{B-M} + \alpha \delta e^{-\mu} (x_b)^{B-M} (x_s)^{-M} + \alpha (1-\delta) e^{-\mu} (x_s)^{S-M} (x_b)^{-M} \right)$$

$$\begin{aligned}
 \ln(L) &= \ln \left((1-\alpha)e^{-\varepsilon_b} (\varepsilon_b)^B e^{-\varepsilon_s} (\varepsilon_s)^S + \alpha\delta e^{-\varepsilon_b} (\varepsilon_b)^B e^{-(\varepsilon_s+\mu)} (\varepsilon_s + \mu)^S \right. \\
 &\quad \left. + \alpha(1-\delta) e^{-(\varepsilon_b+\mu)} (\varepsilon_b + \mu)^B e^{-\varepsilon_s} (\varepsilon_s)^S \right) \\
 &= \ln \left(\begin{array}{l} (1-\alpha)e^{-\varepsilon_b} (\varepsilon_b)^B e^{-\varepsilon_s} (\varepsilon_s)^S \\ + \alpha\delta e^{-\varepsilon_b} (\varepsilon_b)^B e^{-(\varepsilon_s+\mu)} (\varepsilon_s + \mu)^S \\ + \alpha(1-\delta) e^{-(\varepsilon_b+\mu)} (\varepsilon_b + \mu)^B e^{-\varepsilon_s} (\varepsilon_s)^S \end{array} \right) \times \frac{e^{-\varepsilon_b-\varepsilon_s} (\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M}}{e^{-\varepsilon_b-\varepsilon_s} (\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M}} \\
 &= \ln \left(\begin{array}{l} (1-\alpha)(\varepsilon_b)^{B-M} (\varepsilon_b + \mu)^{-(B-M)} (\varepsilon_s)^{S-M} (\varepsilon_s + \mu)^{-(S-M)} \\ + \alpha\delta e^{-\mu} (\varepsilon_b)^{B-M} (\varepsilon_b + \mu)^{-(B-M)} (\varepsilon_s)^{-M} (\varepsilon_s + \mu)^M \\ + \alpha(1-\delta) e^{-\mu} (\varepsilon_b)^{-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu)^M (\varepsilon_s + \mu)^{-(S-M)} \end{array} \right) \\
 &\quad + \left(-\varepsilon_b - \varepsilon_s + \ln \left((\varepsilon_b)^M (\varepsilon_s)^M (\varepsilon_b + \mu)^{B-M} (\varepsilon_s + \mu)^{S-M} \right) \right) \\
 &= \ln \left(\begin{array}{l} (1-\alpha) \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \\ + \alpha\delta e^{-\mu} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{-M} \\ + \alpha(1-\delta) e^{-\mu} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{-M} \end{array} \right) \\
 &\quad + \left[-\varepsilon_b - \varepsilon_s + \ln \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^M \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^M (\varepsilon_b + \mu)^B (\varepsilon_s + \mu)^S \right] \\
 &= \ln \left(\begin{array}{l} (1-\alpha) \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \\ + \alpha\delta e^{-\mu} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{B-M} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{-M} \\ + \alpha(1-\delta) e^{-\mu} \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right)^{S-M} \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right)^{-M} \end{array} \right) \\
 &\quad + \left[-\varepsilon_b - \varepsilon_s + M \ln \left(\frac{\varepsilon_b}{\varepsilon_b + \mu} \right) + M \ln \left(\frac{\varepsilon_s}{\varepsilon_s + \mu} \right) + B \ln(\varepsilon_b + \mu) + S \ln(\varepsilon_s + \mu) \right]
 \end{aligned}$$

其中, $M = (\max(B, S) + \min(B, S)) / 2$, 令 $x_s = \frac{\varepsilon_s}{\varepsilon_s + \mu}$, $x_b = \frac{\varepsilon_b}{\varepsilon_b + \mu}$, 有

$$\begin{aligned}
 \ln(L) &= \left[-\varepsilon_b - \varepsilon_s + M \ln(x_s) + M \ln(x_b) + B \ln(\varepsilon_b + \mu) + S \ln(\varepsilon_s + \mu) \right] \\
 &\quad + \ln \left((1-\alpha) (x_s)^{S-M} (x_b)^{B-M} + \alpha\delta e^{-\mu} (x_b)^{B-M} (x_s)^{-M} + \alpha(1-\delta) e^{-\mu} (x_s)^{S-M} (x_b)^{-M} \right)
 \end{aligned}$$

附录 3 经典的 PIN 模型隐含的买单和卖单的期望值、 方差和相关性的推导

在本附录中，我们推导出用来计算经典的 PIN 模型中隐含的买单和卖单的期望值、方差和相关性的公式。

经典的 PIN 模型中买单数量的期望值等于图 2.1 中每个分支的概率乘以该分支上买单到达率的加总：

$$\begin{aligned} E(B) &= \alpha \times \delta \times \varepsilon_b + \alpha \times (1 - \delta) \times (\varepsilon_b + \mu) + (1 - \alpha) \times \varepsilon_b \\ &= \varepsilon_b + \alpha \times (1 - \delta) \times \mu \end{aligned}$$

类似地，经典的 PIN 模型中卖单数量的期望值为：

$$\begin{aligned} E(S) &= \alpha \times \delta \times (\varepsilon_s + \mu) + \alpha \times (1 - \delta) \times \varepsilon_s + (1 - \alpha) \times \varepsilon_s \\ &= \varepsilon_s + \alpha \times \delta \times \mu \end{aligned}$$

买单数量的方差为 $\sigma^2(B) = E(B^2) - E^2(B)$ ，同上面计算 $E(B)$ 的方式相同，我们可以计算出 $E(B^2)$ ，利用上面计算出来的 $E(B)$ 的表达式，我们可以得到：

$$\begin{aligned} \sigma^2(B) &= (\varepsilon_b)^2 + \alpha \times (1 - \delta) \times \mu^2 + 2\alpha \times (1 - \delta) \varepsilon_b \times \mu \\ &\quad - (\alpha \times \delta \times \varepsilon_b + \alpha \times (1 - \delta) \times (\varepsilon_b + \mu) + (1 - \alpha) \times \varepsilon_b)^2 \\ &= \alpha \times (1 - \delta) (1 - \alpha \times (1 - \delta)) \times \mu^2 \end{aligned}$$

同理可以计算出卖单数量的方差为：

$$\begin{aligned} \sigma^2(S) &= (\varepsilon_s)^2 + \alpha \times \delta \times \mu^2 + 2\alpha \times \delta \times \varepsilon_s \times \mu \\ &\quad - (\alpha \times \delta \times (\varepsilon_s + \mu) + \alpha \times (1 - \delta) \times \varepsilon_s + (1 - \alpha) \times \varepsilon_s)^2 \\ &= \alpha \times \delta \times (1 - \alpha \times \delta) \times \mu^2 \end{aligned}$$

买单和卖单数量之间的协方差为 $COV(B, S) = E(B \times S) - E(B) \times E(S)$ ，以同样的方式计算 $E(B \times S)$ 并利用上面计算出来的 $E(B)$ 和 $E(S)$ 的表达式，我们可以得到：

$$COV(B, S) = -\delta \times (1 - \delta) \times (\alpha \mu)^2$$

附录 4 对修正的 PIN 模型进行参数估计时采用的因式分解的推导

在本附录中，我们对修正的 PIN 模型进行参数估计时采用的因式分解进行推导。

在修正的 PIN 模型中，似然函数为：

$$\begin{aligned}
 L(\theta|B,S) = & \lambda\alpha\delta e^{-(\varepsilon_b+\gamma_b)} \frac{(\varepsilon_b+\gamma_b)^B}{B!} e^{-(\varepsilon_s+\mu_s+\gamma_s)} \frac{(\varepsilon_s+\mu_s+\gamma_s)^S}{S!} \\
 & + \lambda\alpha(1-\delta) e^{-(\varepsilon_b+\mu_b+\gamma_b)} \frac{(\varepsilon_b+\mu_b+\gamma_b)^B}{B!} e^{-(\varepsilon_s+\gamma_s)} \frac{(\varepsilon_s+\gamma_s)^S}{S!} \\
 & + \lambda(1-\alpha) e^{-(\varepsilon_b+\gamma_b)} \frac{(\varepsilon_b+\gamma_b)^B}{B!} e^{-(\varepsilon_s+\gamma_s)} \frac{(\varepsilon_s+\gamma_s)^S}{S!} \\
 & + (1-\lambda)\alpha\delta e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-(\varepsilon_s+\mu_s)} \frac{(\varepsilon_s+\mu_s)^S}{S!} \\
 & + (1-\lambda)\alpha(1-\delta) e^{-(\varepsilon_b+\mu_b)} \frac{(\varepsilon_b+\mu_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!} \\
 & + (1-\lambda)(1-\alpha) e^{-\varepsilon_b} \frac{(\varepsilon_b)^B}{B!} e^{-\varepsilon_s} \frac{(\varepsilon_s)^S}{S!}
 \end{aligned}$$

因为似然函数中各项都有 $\frac{1}{B!} \times \frac{1}{S!}$ ，因此我们可以将这项作为公因式提出来，当我们将似然函数取对数后，该项为一个常数，并不会影响参数的估计，因此我们可以不考虑这项。

同样，令 $M = (\max(B,S) + \min(B,S)) / 2$ ，我们对似然函数取对数后同时乘以和除以下列这项：

$$\begin{aligned}
 & e^{-\varepsilon_b-\varepsilon_s} (\varepsilon_b)^M (\varepsilon_b)^{M-B} (\varepsilon_b)^{M-B} (\varepsilon_s)^M (\varepsilon_s)^{M-S} (\varepsilon_s)^{M-S} (\varepsilon_b+\mu_b)^{B-M} (\varepsilon_b+\gamma_b)^{B-M} \\
 & \times (\varepsilon_b+\mu_b+\gamma_b)^{B-M} (\varepsilon_s+\mu_s)^{S-M} (\varepsilon_s+\gamma_s)^{S-M} (\varepsilon_s+\mu_s+\gamma_s)^{S-M}
 \end{aligned}$$

因此，有

$$\ln(L) = \ln \left(\begin{aligned} & \left(\lambda \alpha \delta e^{-(\varepsilon_b + \gamma_b)} (\varepsilon_b + \gamma_b)^B e^{-(\varepsilon_s + \mu_s + \gamma_s)} (\varepsilon_s + \mu_s + \gamma_s)^S \right. \\ & + \lambda \alpha (1 - \delta) e^{-(\varepsilon_b + \mu_b + \gamma_b)} (\varepsilon_b + \mu_b + \gamma_b)^B e^{-(\varepsilon_s + \gamma_s)} (\varepsilon_s + \gamma_s)^S \\ & + \lambda (1 - \alpha) e^{-(\varepsilon_b + \gamma_b)} (\varepsilon_b + \gamma_b)^B e^{-(\varepsilon_s + \gamma_s)} (\varepsilon_s + \gamma_s)^S \\ & + (1 - \lambda) \alpha \delta e^{-\varepsilon_b} (\varepsilon_b)^B e^{-(\varepsilon_s + \mu_s)} (\varepsilon_s + \mu_s)^S \\ & + (1 - \lambda) \alpha (1 - \delta) e^{-(\varepsilon_b + \mu_b)} (\varepsilon_b + \mu_b)^B e^{-\varepsilon_s} (\varepsilon_s)^S \\ & \left. + (1 - \lambda) (1 - \alpha) e^{-\varepsilon_b} (\varepsilon_b)^B e^{-\varepsilon_s} (\varepsilon_s)^S \right) \\ & \times \frac{e^{-\varepsilon_b - \varepsilon_s} (\varepsilon_b)^M (\varepsilon_b)^{M-B} (\varepsilon_b)^{M-B} (\varepsilon_s)^M (\varepsilon_s)^{M-S} (\varepsilon_s)^{M-S} (\varepsilon_b + \mu_b)^{B-M} (\varepsilon_b + \gamma_b)^{B-M}}{e^{-\varepsilon_b - \varepsilon_s} (\varepsilon_b)^M (\varepsilon_b)^{M-B} (\varepsilon_b)^{M-B} (\varepsilon_s)^M (\varepsilon_s)^{M-S} (\varepsilon_s)^{M-S} (\varepsilon_b + \mu_b)^{B-M} (\varepsilon_b + \gamma_b)^{B-M}} \\ & \times \frac{(\varepsilon_b + \mu_b + \gamma_b)^{B-M} (\varepsilon_s + \mu_s)^{S-M} (\varepsilon_s + \gamma_s)^{S-M} (\varepsilon_s + \mu_s + \gamma_s)^{S-M}}{(\varepsilon_b + \mu_b + \gamma_b)^{B-M} (\varepsilon_s + \mu_s)^{S-M} (\varepsilon_s + \gamma_s)^{S-M} (\varepsilon_s + \mu_s + \gamma_s)^{S-M}} \end{aligned} \right)$$

$$= \ln \left(\begin{aligned} & \left(\lambda \alpha \delta e^{-\gamma_b - \mu_s - \gamma_s} (\varepsilon_b)^{-M} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_s)^{-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu_b)^{-(B-M)} \right. \\ & \times (\varepsilon_b + \gamma_b)^M (\varepsilon_b + \mu_b + \gamma_b)^{-(B-M)} (\varepsilon_s + \mu_s)^{-(S-M)} (\varepsilon_s + \gamma_s)^{-(S-M)} (\varepsilon_s + \mu_s + \gamma_s)^M \\ & + \lambda \alpha (1 - \delta) e^{-\mu_b - \gamma_b - \gamma_s} (\varepsilon_b)^{-M} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_s)^{-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu_b)^{-(B-M)} \\ & \times (\varepsilon_b + \gamma_b)^{-(B-M)} (\varepsilon_b + \mu_b + \gamma_b)^M (\varepsilon_s + \mu_s)^{-(S-M)} (\varepsilon_s + \gamma_s)^M (\varepsilon_s + \mu_s + \gamma_s)^{-(S-M)} \\ & + \lambda (1 - \alpha) e^{-\gamma_b - \gamma_s} (\varepsilon_b)^{-M} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_s)^{-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu_b)^{-(B-M)} \\ & \times (\varepsilon_b + \gamma_b)^M (\varepsilon_b + \mu_b + \gamma_b)^{-(B-M)} (\varepsilon_s + \mu_s)^{-(S-M)} (\varepsilon_s + \gamma_s)^M (\varepsilon_s + \mu_s + \gamma_s)^{-(S-M)} \\ & + (1 - \lambda) \alpha \delta e^{-\mu_s} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_s)^{-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu_b)^{-(B-M)} \\ & \times (\varepsilon_b + \gamma_b)^{-(B-M)} (\varepsilon_b + \mu_b + \gamma_b)^{-(B-M)} (\varepsilon_s + \mu_s)^M (\varepsilon_s + \gamma_s)^{-(S-M)} (\varepsilon_s + \mu_s + \gamma_s)^{-(S-M)} \\ & + (1 - \lambda) \alpha (1 - \delta) e^{-\mu_b} (\varepsilon_b)^{-M} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu_b)^M \\ & \times (\varepsilon_b + \gamma_b)^{-(B-M)} (\varepsilon_b + \mu_b + \gamma_b)^{-(B-M)} (\varepsilon_s + \mu_s)^{-(S-M)} (\varepsilon_s + \gamma_s)^{-(S-M)} (\varepsilon_s + \mu_s + \gamma_s)^{-(S-M)} \\ & + (1 - \lambda) (1 - \alpha) (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_b)^{B-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_s)^{S-M} (\varepsilon_b + \mu_b)^{-(B-M)} \\ & \left. \times (\varepsilon_b + \gamma_b)^{-(B-M)} (\varepsilon_b + \mu_b + \gamma_b)^{-(B-M)} (\varepsilon_s + \mu_s)^{-(S-M)} (\varepsilon_s + \gamma_s)^{-(S-M)} (\varepsilon_s + \mu_s + \gamma_s)^{-(S-M)} \right) \\ & + \ln \left(\begin{aligned} & e^{-\varepsilon_b - \varepsilon_s} (\varepsilon_b)^M (\varepsilon_b)^{M-B} (\varepsilon_b)^{M-B} (\varepsilon_s)^M (\varepsilon_s)^{M-S} (\varepsilon_s)^{M-S} (\varepsilon_b + \mu_b)^{B-M} (\varepsilon_b + \gamma_b)^{B-M} \\ & \times (\varepsilon_b + \mu_b + \gamma_b)^{B-M} (\varepsilon_s + \mu_s)^{S-M} (\varepsilon_s + \gamma_s)^{S-M} (\varepsilon_s + \mu_s + \gamma_s)^{S-M} \end{aligned} \right) \end{aligned} \right)$$

$$\begin{aligned}
 & \left(\lambda \alpha \delta e^{-\gamma_b - \mu_s - \gamma_s} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^{-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^{B-M} \right. \\
 & \times \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^{-M} \\
 & + \lambda \alpha (1 - \delta) e^{-\mu_b - \gamma_b - \gamma_s} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^{-M} \\
 & \times \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^{-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^{S-M} \\
 & + \lambda (1 - \alpha) e^{-\gamma_b - \gamma_s} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^{-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^{B-M} \\
 & \times \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^{-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^{S-M} \\
 & + (1 - \lambda) \alpha \delta e^{-\mu_s} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^{B-M} \\
 & \times \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^{-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^{S-M} \\
 & + (1 - \lambda) \alpha (1 - \delta) e^{-\mu_b} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^{-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^{B-M} \\
 & \times \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^{S-M} \\
 & + (1 - \lambda) (1 - \alpha) \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^{B-M} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^{B-M} \\
 & \times \left. \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^{S-M} \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^{S-M} \right) \\
 & + \ln \left(\begin{aligned}
 & e^{-\varepsilon_b - \varepsilon_s} \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b} \right)^M \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b} \right)^M \left(\frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b} \right)^M (\mathcal{E}_b)^{-2B} (\mathcal{E}_b + \mu_b)^B (\mathcal{E}_b + \gamma_b)^B (\mathcal{E}_b + \mu_b + \gamma_b)^B \\
 & \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s} \right)^M \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s} \right)^M \left(\frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s} \right)^M (\mathcal{E}_s)^{-2S} (\mathcal{E}_s + \mu_s)^S (\mathcal{E}_s + \gamma_s)^S (\mathcal{E}_s + \mu_s + \gamma_s)^S
 \end{aligned} \right) \\
 & \text{令 } x_{b\mu} = \frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b}, x_{b\gamma} = \frac{\mathcal{E}_b}{\mathcal{E}_b + \gamma_b}, x_{b\mu\gamma} = \frac{\mathcal{E}_b}{\mathcal{E}_b + \mu_b + \gamma_b}, x_{s\mu} = \frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s}, x_{s\gamma} = \frac{\mathcal{E}_s}{\mathcal{E}_s + \gamma_s}, \\
 & x_{s\mu\gamma} = \frac{\mathcal{E}_s}{\mathcal{E}_s + \mu_s + \gamma_s}, \text{ 则有}
 \end{aligned}$$

$$\begin{aligned}
 \ln(L) = & \ln \left(\begin{aligned}
 & \lambda \alpha \delta e^{-\gamma_b - \mu_s - \gamma_s} (x_{b\mu})^{B-M} (x_{b\gamma})^{-M} (x_{b\mu\gamma})^{B-M} (x_{s\mu})^{S-M} (x_{s\gamma})^{S-M} (x_{s\mu\gamma})^{-M} \\
 & + \lambda \alpha (1-\delta) e^{-\mu_b - \gamma_b - \gamma_s} (x_{b\mu})^{B-M} (x_{b\gamma})^{B-M} (x_{b\mu\gamma})^{-M} (x_{s\mu})^{S-M} (x_{s\gamma})^{-M} (x_{s\mu\gamma})^{S-M} \\
 & + \lambda (1-\alpha) e^{-\gamma_b - \gamma_s} (x_{b\mu})^{B-M} (x_{b\gamma})^{-M} (x_{b\mu\gamma})^{B-M} (x_{s\mu})^{S-M} (x_{s\gamma})^{-M} (x_{s\mu\gamma})^{S-M} \\
 & + (1-\lambda) \alpha \delta e^{-\mu_s} (x_{b\mu})^{B-M} (x_{b\gamma})^{B-M} (x_{b\mu\gamma})^{B-M} (x_{s\mu})^{-M} (x_{s\gamma})^{S-M} (x_{s\mu\gamma})^{S-M} \\
 & + (1-\lambda) \alpha (1-\delta) e^{-\mu^b} (x_{b\mu})^{-M} (x_{b\gamma})^{B-M} (x_{b\mu\gamma})^{B-M} (x_{s\mu})^{S-M} (x_{s\gamma})^{S-M} (x_{s\mu\gamma})^{S-M} \\
 & + (1-\lambda)(1-\alpha) (x_{b\mu})^{B-M} (x_{b\gamma})^{B-M} (x_{b\mu\gamma})^{B-M} (x_{s\mu})^{S-M} (x_{s\gamma})^{S-M} (x_{s\mu\gamma})^{S-M}
 \end{aligned} \right) \\
 & + \left\{ \begin{aligned}
 & -\varepsilon_b - \varepsilon_s + M \left[\ln(x_{b\mu}) + \ln(x_{b\gamma}) + \ln(x_{b\mu\gamma}) + \ln(x_{s\mu}) + \ln(x_{s\gamma}) + \ln(x_{s\mu\gamma}) \right] \\
 & + B \left[\ln(\varepsilon_b + \mu_b) + \ln(\varepsilon_b + \gamma_b) + \ln(\varepsilon_b + \mu_b + \gamma_b) \right] - 2B \ln(\varepsilon_b) \\
 & + S \left[\ln(\varepsilon_s + \mu_s) + \ln(\varepsilon_s + \gamma_s) + \ln(\varepsilon_s + \mu_s + \gamma_s) \right] - 2S \ln(\varepsilon_s)
 \end{aligned} \right\}
 \end{aligned}$$

附录 5 修正的 PIN 模型隐含的买单和卖单的期望值、 方差和相关性的推导

在本附录中，我们推导出用来计算修正的 PIN 模型中隐含的买单和卖单的期望值、方差和相关性的公式。

修正的 PIN 模型中买单数量的期望值等于图 2.2 中每个分支的概率乘以该分支上买单到达率的加总：

$$\begin{aligned} E(B) &= \lambda\alpha\delta(\varepsilon_b + \gamma_b) + \lambda\alpha(1-\delta)(\varepsilon_b + \mu_b + \gamma_b) + \lambda(1-\alpha)(\varepsilon_b + \gamma_b) \\ &\quad + (1-\lambda)\alpha\delta\varepsilon_b + (1-\lambda)\alpha(1-\delta)(\varepsilon_b + \mu_b) + (1-\lambda)(1-\alpha)\varepsilon_b \\ &= \varepsilon_b + \alpha(1-\delta)\mu_b + \lambda\gamma_b \end{aligned}$$

类似地，修正的 PIN 模型中卖单数量的期望值为：

$$\begin{aligned} E(S) &= \lambda\alpha\delta(\varepsilon_s + \mu_s + \gamma_s) + \lambda\alpha(1-\delta)(\varepsilon_s + \gamma_s) + \lambda(1-\alpha)(\varepsilon_s + \gamma_s) \\ &\quad + (1-\lambda)\alpha\delta(\varepsilon_s + \mu_s) + (1-\lambda)\alpha(1-\delta)\varepsilon_s + (1-\lambda)(1-\alpha)\varepsilon_s \\ &= \varepsilon_s + \alpha\delta\mu_s + \lambda\gamma_s \end{aligned}$$

买单数量的方差为 $\sigma^2(B) = E(B^2) - E^2(B)$ ，同上面计算 $E(B)$ 的方式相同，我们可以计算出 $E(B^2)$ ，利用上面计算出来的 $E(B)$ 的表达式，我们可以得到：

$$\begin{aligned} \sigma^2(B) &= (\varepsilon_b)^2 + \alpha(1-\delta)(\mu_b)^2 + \lambda(\gamma_b)^2 + 2\alpha(1-\delta)\varepsilon_b\mu_b + 2\lambda\varepsilon_b\gamma_b + 2\alpha(1-\delta)\lambda\mu_b\gamma_b \\ &\quad - (\varepsilon_b + \alpha(1-\delta)\mu_b + \lambda\gamma_b)^2 \\ &= \alpha(1-\delta)(1-\alpha(1-\delta))(\mu_b)^2 + \lambda(1-\lambda)(\gamma_b)^2 \end{aligned}$$

卖单数量的方差为：

$$\begin{aligned} \sigma^2(S) &= (\varepsilon_s)^2 + \alpha\delta(\mu_s)^2 + \lambda(\gamma_s)^2 + 2\alpha\delta\varepsilon_s\mu_s + 2\lambda\varepsilon_s\gamma_s + 2\lambda\alpha\delta\mu_s\gamma_s \\ &\quad - (\varepsilon_s + \alpha\delta\mu_s + \lambda\gamma_s)^2 \\ &= \alpha\delta(1-\alpha\delta)(\mu_s)^2 + \lambda(1-\lambda)(\gamma_s)^2 \end{aligned}$$

买单和卖单数量之间的协方差为 $COV(B, S) = E(B \times S) - E(B) \times E(S)$ ，以同样的方式计算 $E(B \times S)$ 并利用上面计算出来的 $E(B)$ 和 $E(S)$ 的表达式，我们可以得到：

$$\begin{aligned}
 COV(B, S) &= E(B \times S) - E(B) \times E(S) \\
 &= \varepsilon_b \varepsilon_s + \alpha \delta \varepsilon_b \mu_s + \lambda \varepsilon_b \gamma_s + \alpha(1 - \delta) \mu_b \varepsilon_s + \lambda \alpha(1 - \delta) \mu_b \gamma_s \\
 &\quad + \lambda \gamma_b \varepsilon_s + \lambda \alpha \delta \gamma_b \mu_s + \lambda \gamma_b \gamma_s \\
 &\quad - (\varepsilon_b + \alpha(1 - \delta) \mu_b + \lambda \gamma_b) (\varepsilon_s + \alpha \delta \mu_s + \lambda \gamma_s) \\
 &= \lambda(1 - \lambda) \gamma_b \gamma_s - \alpha^2 \delta(1 - \delta) \mu_b \mu_s
 \end{aligned}$$

[参考文献]

- [1] Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman, 1996, Liquidity, information and infrequently traded stocks, *Journal of Finance* 51, 1405-1436.
- [2] Bagehot, W., 1971, The only game in town, *Financial Analysts Journal* 27, 12-14.
- [3] Copeland, T., and D. Galai, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457-1469.
- [4] Glosten, L., and P. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.
- [5] Easley, D., and M. O'Hara, 1987, Price, Trade Size, and Information in Security Markets, *Journal of Financial Economics* 19, 69-90.
- [6] Jaffe, J. F., and R. L. Winkler, 1976, Optimal speculation against an efficient market, *Journal of Finance* 31, 49-91.
- [7] Easley, D., N. M. Kiefer, and M. O'Hara, 1997a, One Day in the Life of a Very Common Stock, *Review of Financial Studies*, 3, 805-835.
- [8] Easley, D., N. M. Kiefer, and M. O'Hara, 1997b, The information content of the trading process, *Journal of Empirical Finance* 4, 159-186.
- [9] Easley, D., M. O'Hara, and J. Paperman, 1998, Financial analysts and information-based trade, *Journal of Financial Markets* 1, 175-201.
- [10] Brockman, P., and D. Y. Chung, 2000, Informed and Uninformed Trading in an Electronic, Order-Driven Environment, *Financial Review* 35, 125-146.
- [11] Easley, D., M. O'Hara, and G. Saar, 2001, How Stock Splits Affect Trading: A Microstructure Approach, *Journal of Financial and Quantitative Analysis* 36.
- [12] Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is Information Risk a Determinant of Asset Returns?, *The Journal of Finance* 57, 2185-2221.
- [13] Easley, D., and M. O'Hara, 2004, Information and the Cost of Capital, *The Journal of Finance* 59, 1553-1583.
- [14] Vega, C., 2006, Stock price reaction to public and private information, *Journal of Financial*

- Economics* 82, 103-133.
- [15] Boehmer, E., J. Grammig, and E. Theissen, 2007, Estimating the probability of informed trading—does trade misclassification matter?, *Journal of Financial Markets* 10, 26-47.
- [16] Venter, J. H., and D. C. J. De Jongh, 2006, Extending the EKOP model to estimate the probability of informed trading, *Studies in Economics and Econometrics* 30, 25-39.
- [17] Duarte, J. and L. Young, 2008, Why is PIN Priced?, *Journal of Financial Economics* forthcoming.
- [18] Morse, D., and N. Ushman, 1983, The effect of information announcements on the market microstructure, *Accounting Review* 247-258.
- [19] Glosten, L. R., and L. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123-142.
- [20] Foster, F. D., and S. Viswanathan, 1990, A theory of the interday variations in volume, variance, and trading costs in securities markets, *Review of Financial Studies* 593-624.
- [21] McInish, T. H., and R. A. Wood, 1992, An analysis of intraday patterns in bid/ask spreads for NYSE stocks, *Journal of Finance* 47, 753-764.
- [22] Foster, F. D., and S. Viswanathan, 1993, The effect of public information and competition on trading volume and price volatility, *Review of Financial Studies* 23-56.
- [23] Collier, M., and T. L. Yohn, 1997, Management forecasts and information asymmetry: An examination of bid-ask spreads, *Journal of Accounting Research* 181-191.
- [24] Benston, G. J., and R. L. Hagerman, 1974, Determinants of bid-asked spreads in the over-the-counter market, *Journal of Financial Economics* 1, 353-364.
- [25] Stoll, H., 1978, The supply of dealer services in securities markets, *Journal of Finance* 33, 1133-1151.
- [26] Stoll, H. R., 1989, Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests, *Journal of Finance* 44, 115-134.
- [27] Hasbrouck, J., 1991, The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 571-595.
- [28] Chiang, R., and P. C. Venkatesh, 1988, Insider holdings and perceptions of information asymmetry: A note, *Journal of Finance* 1041-1048.

- [29] Sarin, A., K. A. Shastri, and K. Shastri, 2000, Ownership structure and stock market liquidity, *unpublished paper, University of Pittsburgh*.
- [30] Easley, D., and M. O'Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance* 47, 577-605.
- [31] Straser, V., 2002, Regulation Fair Disclosure and Information Asymmetry, *Working Paper, University of Notre Dame*.
- [32] Barclay, M. J., and T. Hendershott, 2003, Price discovery and trading after hours, *Review of Financial Studies* 16, 1041-1073.
- [33] Brown, S., S. A. Hillegeist, and K. Lo, 2004, Conference calls and information asymmetry, *Journal of Accounting and Economics* 37, 343-366.
- [34] Chung, K. H., M. Li, and T. H. McInish, 2005, Information-based trading, price impact of trades, and trade autocorrelation, *Journal of Banking and Finance* 29, 1645-1669.
- [35] Aktas, N., E. de Bodt, F. Declerck, and H. Van Oppens, 2007, The PIN anomaly around M&A announcements, *Journal of Financial Markets* 10, 169-191.
- [36] Chen, Q., I. Goldstein, and W. Jiang, 2007, Price informativeness and investment sensitivity to stock price, *Review of Financial Studies* 20, 619.
- [37] Aslan, H., D. Easley, S. Hvidkjaer, and M. O'Hara, 2007, Firm characteristics and informed trading: Implications for asset pricing, *Unpublished Working Paper*.
- [38] Albuquerque, R. U. I., E. V. A. De Francisco, and L. B. Marques, 2008, Marketwide Private Information in Stocks: Forecasting Currency Returns, *The Journal of Finance* 63, 2297-2343.
- [39] Lee, C., and M. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733-746.
- [40] Nyholm, K., 2002, Estimating the Probability of Informed Trading, *The Journal of Financial Research* 25, 485-505.
- [41] Nyholm, K., 2003, Inferring the private information content of trades: a regime-switching approach, *Journal of Applied Econometrics* 18, 457-470.
- [42] Handa, P., R. Schwartz, and A. Tiwari, 2003, Quote setting and price formation in an order driven market, *Journal of Financial Markets* 6, 461-489.
- [43] Ma, T., M. Hsieh, and J. Chen, 2007, The Probability of Informed Trading and the Performance

- of Stock in an Order-Driven Market, *Asia-Pacific Journal of Financial Studies* 36, 896.
- [44] Wang, J., 1993, A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies* 60, 249-282.
- [45] Hughes, J., 2005, Private Information, Diversification, and Asset Pricing, *Working Paper, University of Southern California*.
- [46] Lambert, R., C. Leuz, and R. E. Verrecchia, 2007, Accounting Information, Disclosure, and the Cost of Capital, *Journal of Accounting Research* 45, 385-420.
- [47] Fama, E. F., and K. R. French, 1992, The Cross-Section of Expected Stock Returns, *The Journal of Finance* 47, 427-465.
- [48] Hughson, E., and Kang, M., 2007, Does More Informed Trading necessarily lead to Higher Expected Returns?, *Working Paper, University of Colorado*.
- [49] Lu, Y. C., and W. K. Wong, 2008, Probability of Information-based Trading as a Pricing Factor in Taiwan Stock Market, *Working Paper, Ming Chuan University*.
- [50] Li, H. T., J. B. Wang, C. C. Wu, and Y. He, 2008, Are Liquidity and Information Risks Priced in Treasury Bond Market, *Journal of Finance* forthcoming.
- [51] Mohanram P., and S. Rajgopal, 2006, Is Information Risk (PIN) Priced?, *Working Paper, Columbia Business School*.
- [52] Carhart, M. M., 1997, On persistence in mutual fund performance, *Journal of Finance* 57-82.
- [53] Daniel, K., and S. Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 1-33.
- [54] Kubota, K., and H. Takehara, 2008, Information based trade, the PIN variable, and portfolio style differences: Evidence from Tokyo stock exchange firms, *Pacific-Basin Finance Journal*.
- [55] 杨之曙、王松瑶, 2004, 沪市买卖价差和信息性交易实证研究, [J], 金融研究, 第4期, 45-56.
- [56] 王春峰、董向征、房振明, 2005, 信息交易概率与中国股市价格行为关系的研究, [J], 系统工程, 第2期, 62-67.
- [57] 李朋、刘善存, 2006, 信息性交易概率分解与买卖价差研究, [J], 南方经济, 第2期, 13-22.
- [58] 杨栋、郭玉清, 2007, 指令驱动型市场、非公开信息与中国证券收益率, [J], 南开经济

- 研究, 第 4 期, 106-117.
- [59] 王展翔, 2007, 关于我国股市知情交易水平及其性质的实证研究, [J], 当代财经, 第 2 期, 43-48.
- [60] 韩立岩、郑君彦、李东辉, 2008, 沪市知情交易概率特征与风险定价能力, [J], 中国管理科学, 第 2 期, 16-24.
- [61] Hasbrouck, J., and D. J. Seppi, 2001, Common factors in prices, order flows, and liquidity, *Journal of Financial Economics* 59, 383-411.
- [62] Chordia, T., R. Roll, and A. Subrahmanyam, 2002, Order imbalance, liquidity, and market returns, *Journal of Financial Economics* 65, 111-130.
- [63] Bachelier, L., 1900, Theorie de la Speculation (Thesis), *Annales Scientifiques de Ecole Normale Supérieure* 3, 21-86. (English Translation in Cootner, P. H., 1964, Random Character of Stock Market Prices, *MIT Press*, 17-78).
- [64] Kandel, E., and N. D. Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy*, 831-872.
- [65] Harris, M., and A. Raviv, 1993, Differences of opinion make a horse race, *Review of Financial Studies* 473-506.
- [66] Bessembinder, H., K. Chan, and P. J. Seguin, 1996, An empirical examination of information, differences of opinion, and trading activity, *Journal of Financial Economics* 40, 105-134.
- [67] Bamber, L. S., O. E. Barron, and T. L. Stober, 1999, Differential interpretations and trading volume, *Journal of Financial and Quantitative Analysis* 369-386.
- [68] Admati, A. R., and P. Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 3-40.
- [69] Brockman, P., and D. Y. Chung, 1999, An analysis of depth behavior in an electronic, order-driven environment, *Journal of Banking and Finance* 23, 1861-1886.
- [70] Easley, D., N. M. Kiefer, and M. O'Hara, 2005, Factoring Information into Returns, *Working Paper, Cornell University*.
- [71] Yan, Y. X., and S. J. Zhang, 2006, An Improved Estimation Method and Empirical Properties of the Probability of Informed Trading, *Working Paper, University of Pennsylvania*.
- [72] Box, G. E. P., and D. Pierce, 1970, Distribution of Residual Autocorrelations in Autoregressive

- Integrated Moving Average Time Series Models, *Journal of the American Statistical Association*, 65, 1509-1526.
- [73] Ljung, G. M., and G. E. P. Box, 1978, On a measure of lack of fit in time series models, *Biometrika* 65, 297-303.
- [74] Chordia, T., and A. Subrahmanyam, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485-518.

后 记

光阴似箭，岁月如梭，十年的大学生活已接近尾声。回首这十年的时光，虽然有求学的艰辛，但也有挑战的乐趣。论文能够得以顺利完成，要感谢的人太多太多，是他们一直以来对我的关心和帮助，才让我能够不断前行。借此机会，我希望我能够对所有那些给予我关心和帮助的人表示最诚挚的谢意。

首先要衷心感谢我的导师郑振龙教授。能够成为郑老师的弟子，亲身体会郑老师的言传身教，我感到非常荣幸。郑老师渊博的学识、严谨的治学态度、刻苦钻研的精神和崇高的人格魅力是激励我不断进步的动力。郑老师不仅教给了我许多宝贵的知识，更在人生态度上给予我许多富有意义的启迪，让我一生都受益无穷。郑老师不仅在学习上给予我悉心的指导，在生活上也对我关怀备至，让我感受到在家的温暖。郑老师对我的关怀和帮助，我将永远铭记于心。

同时要深深感谢金融系的张亦春教授、林宝清教授、江曙霞教授、朱孟楠教授、邱崇明教授、郑鸣教授、陈国进教授、李晓峰教授、何孝星教授、郑荣鸣教授、陈善昂副教授、黄华副教授等金融系其他未能提及的老师，他们在学习上给予我的指导和帮助，让我获益良多。

接下来要感谢我的各位师兄师姐们，他们是林海、陈蓉、康朝锋、陈惠玲、邱文华、陈淼鑫、王保合、郑泽星、黄兴李、冯玲、唐革榕、马喜德、俞琳、张睿、张蕾、秦洪远、包苏昱、刘晓曙、何凯浩和胡炜。他们榜样的力量，永远鞭策着我不断前进。他们在学习上对我的帮助和支持，让我更快的进步。

特别感谢师弟郑文旭对论文写作提供的宝贵数据，师弟王瑞锋在数据处理过程中提供的帮助，以及师兄刘晓曙、秦洪元、林海和师弟张贻军在论文写作过程中的热心帮助和宝贵建议。同时也要感谢在论文写作过程中给我关心和鼓励的朋友们，他们是蔡庆丰、李鹏、陈智文和李丕东等。还要感谢陈鹏军、王磊、原鹏飞、刘猛、黄蕙舟、黄文彬、姚佳、王玉英等其他同班同学，与你们的同学之谊和共同度过的时光让我难以忘怀。

最后要感谢我的父母，是你们无私的养育之恩成就了我的学业。还要感谢我的亲戚朋友对我的帮助和鼓励。我只有不断努力，才能不辜负你们的期望。

杨伟

2009年4月于厦大