

天津大学

本科生毕业设计（论文）



题目：投资者有限关注与股票市场投资决策
——来自龙虎榜的经验证据

学 院 管理与经济学部

专 业 金融学系

年 级 大四

姓 名 郑行健

学 号 3016209449

指导教师 沈德华

独创性声明

本人声明：所呈交的毕业设计（论文），是本人在指导教师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本毕业设计（论文）中不包含任何他人已经发表或撰写过的研究成果。对本毕业设计（论文）所涉及的研究工作做出贡献的其他个人和集体，均已在论文中作了明确的说明。本毕业设计（论文）原创性声明的法律责任由本人承担。

论文作者签名：

年 月 日

本人声明：本毕业设计（论文）是本人指导学生完成的研究成果，已经审阅过论文的全部内容。

论文指导教师签名：

年 月 日

摘要

基于“龙虎榜数据”的准自然实验，我们发现：（1）媒体报道带来的注意力会使投资者的交易行为更频繁；（2）个体投资者倾向于净买入那些能够引起他们注意力的股票；（3）个体投资者会因为他们不理智的注意力净买入行为而获得负的超额收益。最后，我们的实验设计是稳健的，并且异常交易行为和收益在控制了公司异质性和投资者类别上仍然显著。本文对“基于注意力的交易行为与股票收益”的研究支持了 Hong and Stein（1999）的理论模型，表明在新兴市场中媒体报道更容易引起投资者的“过度反应”，而不是 Merton（1987）中提出“投资者关注假说”所认为的关注度能够分散特质风险，导致资产价格上升的理论。

关键词： 有限关注，投资者行为，资产定价

ABSTRACT

Based on a quasi-natural experiment from “the Dragon and Tiger List”, we document three results: (1) Attention brought by media coverage makes investors trade more heavily. (2) Individual investors tend to be net buyers of attention-grabbing stocks. (3) Individual investors earn an abnormal negative return because of their irrational attention-driven buying decisions. Finally, the experiment is robust, and the empirical findings remain significant controlling for exogeneity as well as investor characteristics. Our findings supports the theoretical model proposed by Hong and Stein (1999), indicating attention-based investor trading are more likely to cause overaction instead of Merton’s Investor recognition hypothesis (1987), which proposes that attention can diverse idiosyncratic risks.

KEY WORDS: Limited attention; Investor behavior; Asset pricing

目 录

第一章 绪论.....	1
1.1 研究意义.....	1
1.2 研究目的和研究内容.....	3
1.3 研究方法.....	4
1.4 主要创新点.....	4
第二章 文献综述.....	6
2.1 注意力综述.....	6
2.2 注意力与投资者行为的研究.....	6
2.3 注意力与资产价格的研究.....	8
2.4 基于交易公开信息的投资者关注研究.....	10
2.5 本章小结.....	10
第三章 数据和研究设计.....	12
3.1 数据：龙虎榜.....	12
3.2 研究变量.....	15
3.2.1 资金流变化比.....	15
3.2.2 买卖失衡比.....	16
3.2.3 累计超额收益率.....	17
3.2.4 其他变量.....	17
第四章 信息：有限关注能影响投资者投资决策吗？.....	18
4.1 投资者关注行为.....	18
4.2 不考虑异质性影响的 RDD 主实验.....	19
4.3 基于 Logit 模型和 Z 检验的实验.....	22
4.4 考虑个股异质性影响的 RDD 实验.....	23
第五章 行为：投资者更倾向于净买入还是净卖出？.....	26
5.1 单变量分析.....	26
5.2 多变量分析.....	27
第六章 价格：注意力买入行为与投资者收益.....	31

6.1 基于事件研究法的超额收益.....	31
6.2 通过构建零投资组合研究投资者的盈亏.....	32
6.3 投资者的注意力净买入行为和投资收益之间的回归分析.....	34
第七章 基于 DID 的稳健性检验.....	37
第八章 结论与展望.....	39
参考文献.....	40
附 录.....	44
附录一 我们如何构建 DGTW 指标?	44
附录二 我们的 RDD 模型稳健吗?	45
致 谢.....	49

第一章 绪论

1.1 研究意义

媒体报道（Media coverage）一定会带来投资者注意力的变化，而后者往往通过两种途径影响投资者的行为。第一条是从信息不对称（Information Asymmetry）的角度，学者们认为媒体报道能够降低投资者获取信息的成本（Grossman and Stiglitz, 1980）和并且分散特质风险（Merton, 1987）；换言之，媒体报道因为“信息共享”而对投资者有益；另外一条则是显著性（Salience）的角度，学者们认为有限关注只是让投资者短暂地关注到这只股票（Daniel et al., 1998; Hong and Stein, 1999），加剧了他们非理性投资者的程度，和风险偏好和成本并没有太大关系；换言之，媒体报道因为导致了“过度反应”，而对投资者有害。学界目前对前者（Fang and Peress, 2009; Fang et al., 2014; Tetlock, 2010）和后者（Barber and Odean, 2008; Wang, 2019; Frydman and Wang, 2020）都找到充足的证据证明了这两种逻辑。然而，因为学界目前尚未对找到一个完美的能够识别“投资者关注度”的变量达成共识，故而尚未有文献从两种机制对“媒体报道对投资者行为和资产价格的影响”做出比较研究。

自从 Markowitz（1952）引入了资产组合理论之后，现代金融学从华尔街操盘手口中的技术分析逐步转变成为了一种严谨的科学。在他所建立的理论框架中，每一位投资主体都具有贝叶斯理性，都同质地通过风险-收益的判断准则来进行投资决策，对任何一种资产未来收益的概率分布都能了熟于心，并且拥有者无限的认知资源，能够立即处理所获得的信息。

基于这一分析框架，Treyner（1961），Sharpe（1970），Mossin（1966）和 Lintner（1965）分别从公司金融和投资组合的视角推导出了 CAPM 模型，证明了资产的收益是在无风险收益的基础上由市场组合收益线性决定的。之后，ICAPM，CCAPM 等多种资产定价的模型也如燎原之火一般在之后的几十年内发展起来。

可以说，正是基于这些假设，金融经济学才能被归类成经济学的子学科，在均衡和边际相等的概念之下得到求解。在一个模型中，我们只要得到代表性参与者的效用函数和证券的支付就能求出均衡价格，利用 Arrow-Debru 的分析方式，进而对市场中任意证券进行定价。

自从上世纪 70 年代末，Tversky 和 Kahneman 通过引入前景理论的模型第一次对理性人的假定进行了挑战，之后的几十年内，市场中各种各样的异象，如动量（Jegadeesh 和 Titman, 1993），盈余漂移（Ball 和 Brown, 1968）等诸多

发现均对基于理性人假定而发现的有效市场假说带来了巨大的冲击。在不考虑“套利”的前提下，我们对金融经济学作为一门严谨学科的思考必然会从本质上受到现实世界中投资者不理性的影响，而这也是动摇金融经济学根基最大几个的问题之一。

在实证中，个体投资者往往会受到过度自信（Odean, 1999）、追求刺激（sensation seeking）（Grinblatt 和 Keloharju, 2009）、对标的公司的熟悉程度（Massa 和 Simonov, 2006）等因素的影响而做出完全脱离基本面的非理性投资决策。在众多的影响因素中，近年来的热点研究方向之一聚焦于投资者的有限关注（limited attention）。

“有限关注”强调（个体）投资者在投资时的认知能力是有限的。从一方面来看，投资者无法时刻对瞬息万变的市场保持极高的关注度，因为媒介的缺失或者信息传递速率过慢的原因，他们无法知道市场上所有的信息，进而无法对内在价值做出精确的判断。当市场上一件事情发生时，它往往要先经过媒体报道，或者是口耳相传的谣言方式抵达真正在二级市场上进行交易的投资者，这其中就会存在比较大的时滞。从另外一方面来看，在这个信息爆炸的时代下，信息以多种多样的方式，海量地涌入投资者地视野之中。相较于上世纪 60 年代，如今投资者可以借助专业的电子金融终端如彭博、万得，通过互联网金融论坛如雪球、股吧等媒介接受信息。然而大部分信息都是无效的，在很多情况下甚至是起到负面作用的噪音。投资者原本早已有限的注意力会进一步被这些噪音所消耗，进而无法用充足的注意力去关注真正值得他们关注的问题。

从学术上的意义来看，对现实世界中投资者非理性行为的研究能够丰富我们对市场微观结构和信息结构的理解，同时能够通过放宽对于投资者理性人的假定，将行为金融理论中异质性主体的思想引入传统的金融经济学中，对调和两者之间的矛盾做出一定的贡献。此外，我们还能够从信息分享和过度反应两条路径来研究投资者关注对行为和资产价格的影响。在两者影响机制和结果截然相反的情况下，确定哪一种在中国这个新兴市场中更占主流是一个非常有意义的研究课题。

从现实意义上讲，研究基于有限关注的投资者行为和资产价格能够为广大个体投资者指明未来的投资方向，为我国监管层指出重点。由于我国存在诸多保护个体投资者的制度（如 T+1 交易制度等），我国股市仍然活跃着大量直接交易股份的个体投资者。然而，受限于个体在金融知识上的不足，情绪上波动过于剧烈和信息获取手段过于贫乏，个体投资者往往会（错误地）根据自己在股市中观察到地现象归纳出错误的投资策略（如“跟庄”，“打板战法”等）。从有限关注的角度来看，个体投资者往往容易受到别有用心的媒体操纵和市场热点炒作的影响，无法认识到他们所获得的信息早已是陈旧（stale news）或者错误的事实，进而做出不理性的决策。通过正确认识到有限关注这一概念，个体投资者能够放弃

在市场中“无谓的挣扎”，通过分散化投资股票或者购买指数基金来控制自己的风险敞口，进而形成稳健的投资理念，最终实现财富自由。从监管层的指导方面来看，我国股市由于起步较晚，较多证券法规仍然不完善，无法切实保护好投资者的个人利益。在市场中仍然存在着大量操纵投资者关注的行为：如众多没有投顾资格的自媒体从业者在利用社交平台优势推荐特质波动率高，股价低，公司财务风险高的股票等；不少上市公司也倾向于通过发布言过其实的消息，抓取投资者的关注，以方便他们市值管理。通过正确理解有限关注在监管中的作用，证监会和相关管理机构可以加强对上市公司的约束，通过问询函、问责函等方式警告上市公司的不良行为等，深入整治证券市场上仍然十分泛滥的违规行为，引导市场正确发挥价格发现的作用。

在本文中，我们希望通过深交所每个交易日收盘后都会公布的“交易公开信息”（俗称龙虎榜）来研究注意力在投资者行为和资产价格中的作用。通过对从 2012-2019 年的龙虎榜上榜和那些恰好没上榜公司的研究，我们发现投资者的确会因为某一只股票吸引了他们注意力而增加对这只股票的交易，并且他们往往会过多地交易这只股票。这种注意力引起的异常交易行为大约持续时间在 10 个交易日左右。相对于那些没有上榜的股票，买入那些上榜股票的投资者会因为他们过度反应造成的“追高”行为遭受损失，并且这一关注度折价会持续至少长达 20 个交易日之久。

1.2 研究目的和研究内容

本文的研究目的是从有限关注的视角出发，研究投资者受到注意力冲击情况下的投资决策行为，并且希望进一步探讨投资者可能存在的非理性行为和资产价格之间的关系。

本文的研究内容主要包括如下几个方面：

- 1) 在发生了外生的注意力冲击的情况下，投资者的决策是否会发生变化？其中，我们主要把目光聚焦于投资者的买入卖出行为，不对投资者其他行为做出研究。这是对 Barber 和 Odean（2008）在注意力和投资者行为因果关系上的补充。
- 2) 在上一个问题的基础之上，我们进一步研究投资者的注意力交易行为的方向。更具体地说，在有限的关注度情况下，投资者更倾向于买入还是卖出？这个问题是对 Miller（1977）的可见性假说（visibility hypothesis）的求证。
- 3) 最后，我们希望研究投资者基于有限关注的交易行为的投资收益，换言之，他们买了这些股票之后，是盈利还是损失？这是对 Merton

（1987）的投资者关注假说（Investor recognition hypothesis）和 Hong and Stein（1999）的统一理论模型（Unified theory）直接验证。

1.3 研究方法

1) 基于断点回归和双重差分的准自然实验法：

在所有研究投资者关注的文献中，最令学者们头疼的问题当属如何识别投资者关注这一十分抽象的概念了。在绝大多数情况下，投资者关注都是由一些能够引起上市公司基本面的事件引起的。在这种情况下，我们就不能够轻易识别出投资者因为关注到这只股票和因为基本面发生变化而进行投资决策之间的差异，换言之，研究对象所包含的信息十分复杂以至于这是一个被污染（Dirty）的实验。除了像成交量（Gervais et al., 2001；Zheng 和 Shen, 2019）和价格变化幅度（Barber et al., 2009）之外，学者们曾尝试通过一些离资本市场本身较远，能够作为投资者关注度本身的变量来度量投资者关注，其中比较新颖的有谷歌搜索指数（Da et al., 2011），上市公司广告支出（Lou, 2014）等，新闻标题数量（Yuan, 2015）。与他们不同，我们通过设计了一个准自然实验来区分由基本面发生变化导致的和由注意力冲击自身导致的关注度变化。

在实验中，我们通过一个非常独特的数据集，利用断点回归设计和双重差分实验来区别出注意力冲击给投资者造成的影响。

2) 基于统计的实证方法

分别通过单变量的事件研究法和多变量的因子回归法，研究投资者对于那些能够引起他们注意力的股票的投资行为和交易那些股票之后的收益情况。

3) 文献分析法

通过广泛地阅读牵涉到投资者关注的理论模型和实证检验来理清发展脉络，介绍投资者关注方面文献的新进展，进而总结本文和学界过去对于这一问题理解的异同和本文的边际贡献

1.4 主要创新点

1) 理论分析有创新

我们通过收益率反转证明了“可见性假说”（Merton, 1987）和“风险

分担”的局限性，为“统一理论模型”和“过度反应”（Hong and Stein, 1999）在新兴市场中信息不对称环境下的投资者行为和收益率预测提供了证据。据我们所知，目前仅有 Solomon et al.（2014）以美国市场基金投资者的数据对两种假说进行了探讨，而尚未有文献对不成熟市场中，媒体报道是如何通过影响投资者行为的方式，以影响资产价格的。理解媒体报道对资产价格的作用机制不仅有助于学术界理解信息在市场中传播效率的原理，还有助于我国监管部门完善证券市场在信息“公正、公开、公平”方面的制度设计和日常监督。

2) 研究思路有创新

本文利用一个完整的数据集建立起了一个从投资者关注到投资者行为、再到资产价格的逻辑链。在众多文献中，学者往往将这三部分割裂地对待，如 Baber and Odean（2008）和 Yuan（2015）只研究了注意力是如何影响投资者决策，没有涉及资产价格变化；而 Gervais et al.（2001）和 Peng et al.（2016）等文献只研究了投资者关注是如何影响资产价格的，而缺少行为在其中的作用。相较于之前的文献，我们通过因果推断完整地讲了一个故事。

3) 实验设计有创新

我们基于龙虎榜的数据，我们构建出了一个十分稳健的准自然实验。我们通过比较那些上榜和没上榜的股票分离了由投资者关注导致的交易行为和资产价格的变化，克服了常年困扰实证经济学家多年的，在估计简化模型（Reduced Form）时常见的内生性问题，从而丰富了行为金融方向上的有限关注、个体投资者和资产定价的相关学术成果。

第二章 文献综述

2.1 注意力综述

注意力是认知主体对于信息加工的一种机制，心理学界在研究注意力时往往会从选择、加工、控制和意识等方面入手（Moray, 1970; Solso, 1979）。心理学理论认为，注意力不是人类在认知世界，理解世界时最先开始进行的一步，它往往具体表现在人类接受世界时指向客观事物时的内心活动。

心理学界在研究人类注意力行为的时候，逐步发展出多种理论模型，如选择性注意理论，过滤器理论（Broadbent, 1958），衰减模型（Treisman, 1960），中枢能量理论（Kahneman, 1973），后选择理论（Norman, 1968），多阶段选择理论（Johnston, 1970）等。其中，由卡尼曼（Kahneman, 下同）提出的中枢能量理论则在他和 Tversky（1980, 1992）的共同发展下，同经济学紧密联系在了一起，开创了行为经济学的一个崭新时代。

基于注意力地认知资源理论（Kahneman, 1973）的行为经济学认为，系统中的个体接受信息的通道容量是有限的，并且注意力是一种可以被消耗的稀缺资源。行为主体在进行理性决策时往往会受到有限注意力的约束，他们的决策往往会滞后于注意力行为，并且他们只有在合理分配注意力资源之后才能对个体禀赋进行新的分配。这一理论也被运用到了行为金融中，在 Merton（1987），Peng and Xiong（2006）基于有限注意力的非对称信息模型中，以及 Gervais et al.（2001），Kaniel et al.（2011）和 Yuan（2015）的模型中，他们都将有限注意力非常巧妙地同投资者行为和资产价格结合在了一起。

当然，心理学界和行为金融学界对卡尼曼的注意力资源理论的批判仍然存在，这种理论的优点在于它的模型十分简单易懂，能够很好地对现实世界中信息接收行为做出描述；然而，它的缺陷则来源于它过于模糊，完全忽视了信息在传播到个体时所出现地信号阻断和信号衰减的现象发生。截止目前，心理学界和经济学界仍未就如何将注意力和行为个体完美结合的达成共识。

2.2 注意力与投资者行为的研究

在模型方面，除了上述提及的 Merton（1987）的不对称信息下的模型之外，Hirshleifer 和 Teoh（2003）也从有限注意力的角度对市场上不同类型的投资者注

意力进行了建模。在他们的模型中，他们分别定义了那些关注度较弱的投资者（*inattentive investors*）和关注度高的投资者（*attentive investors*）。这两组投资者都具有相同的同质偏好，并且会对市场中的信息做出不同程度的反映。根据设定的模型，他们发现当前投资者的行为由他们的注意力大小所线性决定，并且市场中的均衡价格也由注意力大小决定。

此外，也有学者从市场微观结构的角度出发，研究信息不平等所造成投资者行为的变化。其中最著名的当属 Kyle（1985）和 Grossman 和 Stiglitz（1980）的模型。在 Kyle（1985）基于内幕交易者-噪音交易者-做市商三种行为主体构建起的模型中，他发现内幕交易者会因为自己的信息优势而获利，而噪音交易者则会在信息弱势的情况下亏损。类似的，在 Barber 和 Odean（2008）附录中的模型里他们构建起了一个类似的模型证明了个体投资者会因为信息和关注度受限而蒙受大量的损失损失。

近两年内，国内的学者也通过理论模型证明了由于有限关所导致的投资者行为非理性和资产价格变化。如彭叠峰一系列的模型中（2011，2015a，2015b）中，他分别将投资者关注分为了预期关注和非预期关注，将投资者根据他们的投资者关注情况划分出了关注和疏忽两类投资者，并且根据市场出清的均衡条件对关注度和资产的风险溢价关系进行了建模刻画。他的发现类似于 Peng 和 Xiong（2006）的分行业关注，并且他所建立的均衡模型在引入贝叶斯理性的疏忽交易者之后，核心结论仍然稳健。

在投资者行为方面的实证研究中，过去二十年来最亮眼的无疑是 Baber 和 Odean 一系列的文章（2001a，2000，2008，2001b），对于个体投资者非理性交易行为的研究中，Odean 通过独特的数据集证明了我们对于“散户”的所有偏见。

在著名的 Baber 和 Odean（2008）的研究中，他们发现投资者的关注度能够极大地影响个体投资者地买入决策。当投资者在搜寻中意的股票时，他们往往会遇到有限关注的问题。相对于市场上所有的股票，（个体）投资者更倾向于仅仅考虑那些在新闻上出现过的，能够引起他们关注的股票。而由于在 Miller（1977）中所提到的卖空约束的原因，他们的买入行为相较于卖出行为会更加频繁和剧烈。在实证中，他们使用了异常成交量和过去一个交易日的回报率来作为投资者关注度的代理变量，他们发现个体投资者会成为那些能够引起他们注意力股票的净买入方。

在 Seaholes 和 Wu（2007）利用上交所所供的账户级别（*account level*）的交易数据中，他们发现当股票冲击涨停时买卖强度会出现失衡的情况，且买入强度要大于卖出强度，与 Baber et al.（2009），Baber 和 Odean（2008）的发现一致。此外，他们还发现相对于那些完全基于注意力交易的非理性投资者而言，那些理性的投资者会利用前者的过度反映而卖出被高估的证券，从而获利。

在 Engleberg et al. (2011) 基于美国一档家喻户晓的电视节目：疯狂的钱 (Mad Money) 的研究中，他们发现投资者的关注会受到每天晚上股票推荐的影响，并且在荐股环节中，那些第一次得到“买入”评级的股票第二天开盘的收益大约为 2.4%，显著大于那些第一次得到“卖出”评级的股票的隔夜开盘收益。

此外，还有很多学者利用很多很有趣的非公开/半公开数据集 (non-public/semi-public) 数据集对基于投资者关注的资产定价做出了研究。其中，Huang et al. (2009) 利用台湾股票市场的账户交易数据在有限关注的情况下，投资者对市场级别的整体信息更加敏感，从而容易忽略公司层面特有的信息，进而倾向于做出不足的反应。Grullon et al. (2004) 和 Lou (2014) 通过上市公司月度的广告指出费用，利用多元线性回归的方式发现上市公司可以通过增加自己在投资者日常生活中的曝光度引起投资者关注，进而引起投资者对他们公司证券的购买欲望。精于文本分析的 Tetlock (2011) 通过对道琼斯新闻文本内容的分析，利用事件研究法对陈旧信息是否能够吸引投资者的关注进行了分析。他发现，投资者得投资决策不会受到信息时效的影响，相反他们更容易对那些陈旧的二手信息过度反应而推高证券价格以致于脱离基本面，最终出现收益率反转的现象。

除了在权益投资这一方面之外，在对投资者在投资基金的相关研究中也牵涉到了注意力。Baber et al. (2005) 利用美国基金市场的交易数据发现投资者更容易投资那些容易引起他们注意力的基金，并且相对于基金净值的变化情况，投资者可能更加关注基金的前端收费和管理费用。在 Kaniel et al. (2017) 利用华尔街日报风格之王 (Category Kings) 的数据中，他通过一个十分漂亮的断点回归设计验证了投资者更倾向于投资那些上了榜单的基金，并且基金的异质性不会影响他们结果的稳健。

最后，在国内学者的相关研究中，Peng et al. (2016) 利用每日涨幅排名前十和前十至二十的股票构建了一个准自然实验，他们发现那些涨幅更好的股票容易吸引投资者的关注，并且超额关注度带来的不是超额收益，而是折价。在沈德华 (2013) 的博士论文中，他利用爬虫技术对我国互联网上多重信息源比较分析后发现，媒体偏见能够经由投资者有限关注的机制影响个体投资的投资决定。

2.3 注意力与资产价格的研究

如上所述，学界发现由于有限关注的阻挠，信息在传递到投资者时往往会发生阻滞，以至于投资者不能及时有效地对信息做出反应，使得股票在短期内其价值低于真实价值；此外，学者们还普遍发现当信息被投资者接受时，投资者往往会表现出过度反应的现象而剧烈买入，而这则普遍会在短期内推高股价，并且价

格在长期发生反转。基于这样的一种推论，学者们便广泛地展开了通过使用注意力预测资产价格的研究。

学者们曾经对个体投资者是否能够通过基于注意力的交易来提升股票收益率做出研究。其中，Barber et al. (2009) 基于经纪商数据模拟了个体投资者的买入卖出行为，发现个体投资者能够影响市场价格，同时个体投资者在买入行为上常常表现出羊群效应，并且他们买入的股票平均收益要低于她们卖出股票 4.4%。他们基于买入卖出订单构建起了买卖失衡比指标 (buy sell imbalance)，在对不同成交量的失衡比进行回归之后，他们发现那些成交量较小的失衡比能够显著地预测股票的未来收益，并且在未来一周的时间里由买单引起的超额收益都是负的。

在他们的研究基础之上，Kaniel et al. (2012) 和 Kaniel et al. (2016) 也构建起了类似的失衡比指标，并且通过多变量分析发现个体投资者的买入行为会在短期内推高资产的价格，为资产带来超额收益，提供市场中的流动性，然而在长期内他们因为没有及时卖出获利而蒙受损失。

在投资者关注与资产定价这一方面，很多学者也通过各种有趣的数据来探讨关注度强弱能否预测资产未来的收益。其中，Fang 和 Peress 通过研究媒体对股票的报道来识别那些能够引起投资者关注的股票。他们利用 LexisNexis 数据库，通过构建投资组合并进行因子回归的方式研究关注度引起的投资收益在传统风险因子上的敞口和不能被定价模型解释的超额 α 。他们的回归结果展现了媒体报道能够显著地预测股票地收益率，并且超额的 α 能够被投资者关注假说和流动性补偿假说所解释。此外，Grullon et al. (2004) 和 Lou (2014) 分别从上市公司广告指出费用的角度对资产价格进行了预测，并且结果仍然支持 Fang 和 Peress (2009) 的发现。

在定价方面，近十年来兴起的数据源当属利用互联网上的信息来作为投资者关注的代理变量了。这一类研究较早地可以追溯到由 Da et al. (2011) 使用的谷歌指数。他们使用从 2004 年到 2008 年的投资者在谷歌搜索引擎中对上市公司的搜索量构建起了异常关注度指数，并且发现了当投资者的搜索频率能够显著地预测未来 2 个月的股票收益，并且在更长的时间段内该收益会出现反转。除了谷歌搜索，其他学者还通过网络爬虫的方式，对推特 (Bollen et al. 2011)，百度指数 (Zhang et al., 2013; 张永杰, 2013)，互联网论坛 (Li et al. 2018)，互联网新闻 (Shen et al. 2017) 对投资者在网络上的关注进行了建模。他们发现基于互联网信息提取出的投资者关注能够显著地预测股票地未来收益，在信息抵达投资者过程中，媒体偏见和投资者异质性都会对股价变化产生显著的影响。

2.4 基于交易公开信息的投资者关注研究

在所有利用龙虎榜数据的投资者关注实证中，主要有刘杰，等（2019）、刘杰，等（2018）、冯旭南（2016）、康恒（2016）和高昕祎（2019）这五篇。在他们的研究中存在着诸多不足：如刘杰（2019）在实验设计部分可能考虑不周；康恒（2016）在复制龙虎榜时完全没有考虑到由于大宗交易导致的无法完全复制的问题，而将错误的数据纳入了实证范围中；冯旭南（2016）的研究中，他“剔除了由重大事件导致上榜”的数据集仍然没有办法帮助他在面板回归中解决回归内审性的问题。

相对于他们的研究，我们的研究在实验设计上有如下的区别和创新：

- 1) 首先，我们的 RDD 实验能够通过将研究样本限制在上榜和恰好要上榜的股票之间，利用上榜和未上榜之间分类的随机性，构建一个正确的准自然实验。这种研究方法能够在最大程度上解决一只股票由于其自身基本面发生重大变化而吸引投资者关注的问题，从而得到一个清洁的实验。
- 2) 其次，在所有利用龙虎榜的研究中，他们都集中在研究：信息是如何影响资产价格的，但缺少信息是如何通过影响投资者行为、进而传导到资产价格的这样一个机制的分析。在我们的研究中，我以信息-行为-价格为线索，串起了一条完整的逻辑链。

2.5 本章小结

本章重点在于回顾过去几十年来注意力在心理学的发展以及它和行为金融学的结合过程，并且借由文献综述的思路为后文定下主要的行文脉络。我们希望结合有限关注在过去几十年来的发展脉络，从投资者关注到投资者行为，再到资产定价，循序渐进地完成从起因到结果的流程。

此外，在传统的对于注意力和投资者行为、资产定价的研究中，虽然有各种各样的指标，然而学者们往往无法完美地将注意力和重大事件所造成的影响区分开来，他们设计的模型往往也无法解决内生性问题。只有通过（类似后文所设计的）准自然实验才能比较稳健地摆脱这个问题地困扰。

基于以上文献总结，我们对后面的章节做出如下安排：第三章主要介绍了我们的数据和模型，对我们独特的数据做出描述，阐述我们数据的优越性；第四章主要解释为什么我们的断点回归设计（RDD）是一个能解决内生性问题的实验，它是如何研究受到外生注意力冲击投资者的投资决策行为的；第五章主要通过结合事件研究和面板回归的方式，研究注意力影响后投资者具体的投资行为，即他们更倾向于买入还是卖出股票；第六章通过构建零投资组合并且进行因子回归的方式，并采用固定效应面板回归来研究投资者基于注意力的投资行为到底会

给他们的投资组合带来什么样的影响，他们究竟会赚还是会赔？最后一章总结全文。

第三章 数据和研究设计

3.1 数据：龙虎榜

我们的主要数据来自于深交所公布的“证券交易公开信息”，俗称“龙虎榜”。深交所一共有三个板块：主板、中小板和创业板。根据深交所制定的交易规则，对于有价格涨跌幅限制的股票（即存在涨跌停板制度的股票）、封闭式基金，当其收盘价涨跌幅偏离值达到 $\pm 7\%$ ¹、价格振幅²达到 15%、换手率达到 20%时，在当日收盘之后各取前五只股票予以全市场范围进行信息披露，具体披露信息不仅包括披露原因，还有前五大买入卖出席位（具体到各地营业部、交易经纪商）；此外，深交所还会同时根据一只股票在连续三个交易日内的累计涨跌幅偏离、ST 和*ST 股票的累计涨跌幅来对那些有异动的股票予以披露。在众多的披露信息中，当属“日价格涨跌幅偏离”这一特性最为普遍，并且这一类股票往往会触发中国股市 $\pm 10\%$ 的涨跌幅限制（Seaholes 和 Wu, 2007），所以这一种龙虎榜上榜理由不仅数据量大，方便我们进行实验，而且包含的信息更多、最能引起投资者的关注这也符合 Barber and Odean（2008）中对于那些能够引起投资者关注股票的概括：“高收益，高成交，有新闻”。所以我们在之后的研究中就对这一类龙虎榜上榜原因进行研究。

在此，我们随机选取了深交所某日龙虎榜信息（2019 年 11 月 11 日），对深交所三个板块：主板、中小板和创业板的上榜证券和上榜原因予以展示，如表 1。在我们的实验中，考虑到数据的准确性和可得性，我们选取 2012 年 1 月 4 日至 2019 年 11 月 8 日为研究区间对龙虎榜信息进行采集，数据来源是 Tushare 金融数据库。这个数据库由于其开源、准确、便捷的特点被各大基金公司和量化交易平台使用。当然，为了确保数据的准确性，我们在使用数据的时候利用随机抽样的方法，通过比对来自 Tushare 数据库的龙虎榜数据和资金流数据和来自 Wind、Resset 数据库的数据，结果发现数据在 1%的置信水平上准确。

如前文所述，我们在本文中将要采用的是断点回归实验设计，考虑到我们数据的特殊性，这一实验设计要求我们不仅拥有上了龙虎榜的股票的数据、并且还要识别出那些符合上榜条件、但是没有上龙虎榜的股票，换言之，我们需要在对

¹ 即这只股票当日价格变化幅度减去其所属市场指数的当日涨跌幅后所得差值。例如，假设格力电器（000651.SZ）当日上涨 8%，而深圳 A 指（399107）当日涨幅 1%，那么它的涨跌幅偏离值就是 7%。在深圳市场的三个不同板块之中，主板对应的市场指数是 399107，中小板对应的市场指数是 399101，创业板对应的市场指数是 399102。

² 股票当日最高价和最低价之间的差值得绝对值除以这只股票昨日收盘价的百分比比值

表 1 龙虎榜信息示例

主板龙虎榜		
证券代码	证券简称	披露原因
000590	启迪古汉	日价格涨幅偏离值达到 12.26%
000697	炼石航空	日价格涨幅偏离值达到 7.35%
000803	金宇车城	日价格跌幅偏离值达到 -7.74%
000815	美利云	日价格跌幅偏离值达到 -7.74%
000935	四川双马	日价格跌幅偏离值达到 -7.74%
000973	佛塑科技	日价格涨幅偏离值达到 8.30%
中小板龙虎榜		
002201	九鼎新材	日价格涨幅偏离值达到 12.02%
002330	得利斯	日价格涨幅偏离值达到 12.02%
002552	宝鼎科技	日价格涨幅偏离值达到 12.02%
002577	雷柏科技	日价格涨幅偏离值达到 12.02%
002656	摩登大道	日价格涨幅偏离值达到 12.02%
创业板龙虎榜		
300004	南风股份	日价格涨幅偏离值达到 12.44%
300192	科斯伍德	日价格涨幅偏离值达到 12.44%
300364	中文在线	日价格涨幅偏离值达到 12.44%
300541	先进数通	日价格涨幅偏离值达到 12.44%
300598	诚迈科技	日价格跌幅偏离值达到 -7.56%
300643	万通智控	日价格涨幅偏离值达到 12.44%
300759	康龙化成	日价格跌幅偏离值达到 -7.56%
300795	米奥兰特	日价格跌幅偏离值达到 -7.54%

每只股票的日价格涨跌幅偏离值进行从高到低的排序，得到上榜阈值（ ≥ 5 ）之下的股票，具体复制方法类似 Kaniel and Parham（2017）。在排序时，因为涨跌停幅度限制³，我们会经常地遇到那些涨跌幅偏离值一样的股票。根据深交所交易公开信息规则，在碰到这种问题时我们需要进一步地比较这些股票的成交量，成交量大的更能上榜，成交量小的则因为排序靠后，更不容易上榜。

在这里，我们在图 1 中给出了复制后得到的排名，其中第一列“排名”和深交所所公布的龙虎榜排名一致，而“经调整排名”是我们为了方便进行 RDD 实验，对原先排名进行调整后所得，具体操作在第 4 章会有介绍，且不会影响我们实验的准确性。这就是我们 RDD 实验的驱动变量。

我们在排序时使用的价格涨跌幅是未经复权调整的当日价格涨跌幅，其和日股票交易量数据均数据来源 CSMAR。在复制排名时，我们还遇到的问题是：由于大宗交易的存在，个股真实成交量大于我们所记录的成交量，是因为我们所使用的数据库（包括 CSMAR 和 tushare）中只记录了集合竞价和连续竞价成交的订

³ 根据交易所结算规则，无论股票当日涨跌幅是否达到 10.00%，只要触及当日涨跌停价格限制，均自动强制归类成 10.00%，这样就可能出现某天有许多股票涨跌幅偏离值相同的情况。在目前相似主题的研究中，学者们往往会忽略这一点，具体错误见附录 3。

图 1 根据涨跌幅偏离值、成交量排序复制得到的，2012 年 1 月 4 日中小板市场上股票排名

	排名	经调整排名	股票代码	
	1	-4.5	002355.SZ	全部上榜股票（实验组）
	2	-3.5	002354.SZ	
	3	-2.5	002433.SZ	
	4	-1.5	002389.SZ	
恰好上榜的股票	5	-0.5	002207.SZ	断点/cut point, 设定为0
恰好没上榜的股票	6	0.5	002561.SZ	
	7	1.5	002292.SZ	全部恰好未上榜股票（对照组）
	8	2.5	002362.SZ	
	9	3.5	002151.SZ	
	10	4.5	002383.SZ	
	11	5.5	002517.SZ	
	12	6.5	002006.SZ	
	13	7.5	002533.SZ	
	14	8.5	002371.SZ	
	15	9.5	002193.SZ	

注：最左列是根据涨跌幅偏离值、成交量大小依次排序下来的排名，也是原始排名；中间“经调整排名”是基于初始排名调整之后所得，目的是为了便于之后的 RDD 实验的计算，不会改变原始结果的显著性。

单。根据深交所交易规则，大宗交易的数据只在盘后发布，并且在计算每日龙虎榜时，交易所会将大宗交易发生的成交量考虑在内。这就意味着我们无法准确地复制出从 2012 年到 2019 年以来每一天每一个板块的龙虎榜排名，一定会有因为缺失大宗交易数据而导致的复制错误。不过根据描述性统计表 2 中 Panel A 中的错误比率来看，这一错误数量相对于我们的样本总量而言是非常小的，错误比率最大仅为 3.20%，最小为 1.07%，正确比率平均为 97.89%，远好过 Kaniel and Parham（2017）中使用的 WSJ 数据，并且考虑到我们的数据是日频的，稳健性更佳，所以我们不必过多担心复制错误导致的数据问题。在 Panel A 中，龙虎榜数量是 9 年来所有的龙虎榜榜单的数量，总上榜股票是对所有上了每日龙虎榜的股票的数量加总，是对出现股票数量总数的统计。错误数量是由于缺少大宗交易数据，我们统计出的榜单不符合公布的榜单的数量。

最后，在描述性统计表 2 中的 Panel B 和 Panel C 中分别对日度龙虎榜上榜公司数量和每日上榜公司的收盘价、日价格变化、流通市值、成交量和换手率进行了记录。其中，Panel B 的上涨和下跌分别代表涨跌幅偏离值为正和为负的情况。

表 2 描述性统计

Panel A: 年度龙虎榜总量统计								
	2012	2013	2014	2015	2016	2017	2018	2019
龙虎榜数量	1458	1428	1470	1464	1458	1464	1458	1242
总上榜股票	3279	3898	4301	4528	4349	4696	5472	4346
错误数量	37	33	20	50	27	22	24	24
正确比率	97.46%	97.68%	98.63%	96.58%	98.14%	98.49%	97.87%	97.86%
Panel B: 日度龙虎榜特征统计								
	2012	2013	2014	2015	2016	2017	2018	2019
异常上涨 而上榜	均值	3.87	4.55	4.68	4.96	4.77	4.45	4.78
	方差	1.88	0.92	0.67	0.06	0.56	1.26	0.45
Panel C: 龙虎榜上榜公司统计								
	上榜公司				未上榜公司			
	均值	标准差	Max	Min	均值	标准差	Max	Min
日收盘价	22.00	14.78	52.49	6.55	19.83	12.431	45.1	6.50
流通市值	486421	417095	1384370	87212	473453	401032	1336240	86649
T+1 日价 格变化	0.79	4.93	9.98	-6.32	0.07	3.07	5.59	-4.58
成交量	271251	246298	801715	39542	159538	144659	471717	23805
换手率	16.34	10.24	35.81	4.38	9.40	6.23	21.59	2.38

3.2 研究变量

3.2.1 资金流变化比

我们在这篇文章中首先想验证的问题是：投资者关注是否能影响投资者的决策？对此，我们希望利用断点回归设计（RDD）模型，对投资者在阈值（即图 1 中的断点）上下股票的投资行为做出研究。如果变量在断点处是显著的，那么我们就有理由断定其中的因果推断关系的成立。

在这个 RDD 实验中，我们的被解释变量是资金流变化比率 $Flowchg_{i,q,t+1,F}$ ，这是通过分别加总每一只股票买入或者卖出的小、中、大单成交金额后，再分别除以其上榜前 5 日算数平均值所得到的。具体计算公式如下：

$$Flowchg_{i,q,t+1,F} = \frac{Flow_{i,q,t+1,F} - avgFlow_{i,q,t,F}}{avgFlow_{i,q,t,F}} \quad (1)$$

$$avgFlow_{i,q,t,F} = \frac{\sum_{\tau=t-6}^{t-1} Flow_{i,q,\tau,F}}{5} \quad (2)$$

$$Flow_{i,q,t,F} = \sum order_{i,q,t,F} \quad (3)$$

其中， $order_{i,q,t+1,F}$ 是第 i 只股票在第 $t+1$ 日，根据成交量大小分为第 q 档的

所有买入或者卖出成交量，资金流入或流出方向由 F 表示。参考 Lee and Radhakrishna (2000)，结合中国股市中散户多的特点，我们对所有成交的订单根据交易金额划分为 4 个档次：5 万及以下的是小单、5 万至 20 万的是中单、20 万至 100 万的是大单，当成交额大于 100 万是我们将其归类为特大单。不过在实际计算时，特大单出现的频次比较低，我们推断在市场中大部分特大单都以大宗交易的方式进行。值得一提的是，我们不需要根据买卖中间价和执行价来判断这一订单到底是买单还是卖单，因为中国的深圳和上海两大证券交易所类似于台湾和澳洲的市场（Lee et al., 2004），没有做市商或者专门的代理商，都是电子限价交易系统，在连续竞价时系统会自动根据成交价来判断买卖情况。

$Flow_{i,q,t+1,F}$ 是第 i 只股票在 $t+1$ 日， q 档成交量的总资金流； $avgFlow_{i,q,t,F}$ 是从第 $t-6$ 日至 $t-1$ 日的资金流入流出的平均值； $Flowchg_{i,q,t+1,F}$ 是第 i 只股票在第 $t+1$ 日后的、资金大小归为 q 类的资金流入或者流出的变化情况，如果这个指标为正，那么就意味着在第 $t+1$ 日这只股票资金流强度变大，交易活动也变得更加频繁。

3.2.2 买卖失衡比

在本实验中，我们还对投资者关注会使得投资者更倾向于买入还是卖出某只股票进行了研究。换言之，我们想要用我们的数据集来证明 Baber and Odean (2008) 所推断的：投资者是注意力吸引股票（attention-grabbing stocks）的净买入方。

在第一个部分之中，因为 RDD 仅仅能通过对 LATE 系数的估计判断出断点处的处理效应是否显著，不能够做出进一步的推断，所以我们还需要通过别的方法来研究投资者具体的投资决策，即，他们更倾向于买入还是卖出？

在这里，我们借鉴在研究投资者行为中著名的买卖失衡比（Buy Sell Imbalance，或 BSI）（Barber and Odean, 2008），构建起投资者具体投资行为变量。由于 Barber and Odean (2008)，Barber et al. (2009) 和 Yuan (2015) 等所使用的数据库是来自交易经纪商的投资者个人交割单（即具体到哪位投资者买入卖出了什么证券），在数据质量和可得性方面远超我们所能获得，所以我们只能利用现有可得的数据在一定程度上进行模仿，最终构建出如下的 BSI 指标（4）。在构造方式上，这一指标更接近于 Kelley and Tetlock (2013) 和 Barrot et al. (2016) 的 $Imb[i]$ 。

$$BSI_{i,q,t} = \frac{Flow_{i,q,t,inflow} - Flow_{i,q,t,outflow}}{Flow_{i,q,t,inflow} + Flow_{i,q,t,outflow}} \quad (4)$$

其中，不同于（1）式中 $Flow_{i,q,t,F}$ 由订单的成交金额（dollar value）所计算出，在（4）式中我们还通过加总每一笔成交订单的成交量（trading volume）来计算出 BSI^4 ，即下文中的 Dollar BSI 和 Volume BSI。例如， $BSI_{i,q,t+1}$ 表示在上榜日之后一天时，成交级别是 q 档的买入卖出订单的强度比较。如果 $BSI_{i,q,t+1}$ 为正，那么我们就认为投资者们是这只股票的净买入方，反之为负。

3.2.3 累计超额收益率

我们在本研究最后还对有限关注是如何影响投资者的收益进行了探讨。其中，我们分别采用了 3 种方法对股票收益进行了度量。

第一种方法是考虑现金红利再投资的股票日回报率，即未经风险调整后得到的回报率，数据来源 CSMAR；第二种方法是考虑市场收益的日回报率，具体计算方法是将第一种方法的日回报率减去深证成指的日回报率，得到市场调整后的超额收益率；第三种方法参考 Daniel et al.（1997）的考虑公司特征的回报率（即著名的 DGTW 调整法）。我们通过考虑公司规模、账面市值比和动量强度对公司进行 $5 \times 5 \times 5$ 分组平均后得到组内平均收益率⁵。将实际收益率减去组内平均收益率即是我们要求的超额累计收益率，具体计算同时如下：

$$CAR_{i,t_1} = \prod_{t=t_1}^T (1 + R_{i,t}) - \prod_{t=t_1}^T (1 + R_{p,t}) \quad (5)$$

其中， $R_{i,t}$ 是第 i 支股票第 t 日的考虑现金红利再投资日回报率， $R_{p,t}$ 是第 i 只股票所在分组的投资组合平均回报率。

3.2.4 其他变量

在本文中我们还用到了个股流通市值、日收盘价数据，这些来自 Tushare 数据库，账面市值比、动量数据，包括日度 Fama-French 三因子、五因子、Pastor-Stambaugh 流动性因子数据来自 CSMAR 数据库，日度的 Carhart 四因子数据来自中央财经大学的资产管理研究中心。

⁴ 类似于 Dollar volume 和 Number volume 之间的区别。

⁵ 对于投资组合的分类识别和构建方法参见附录一。

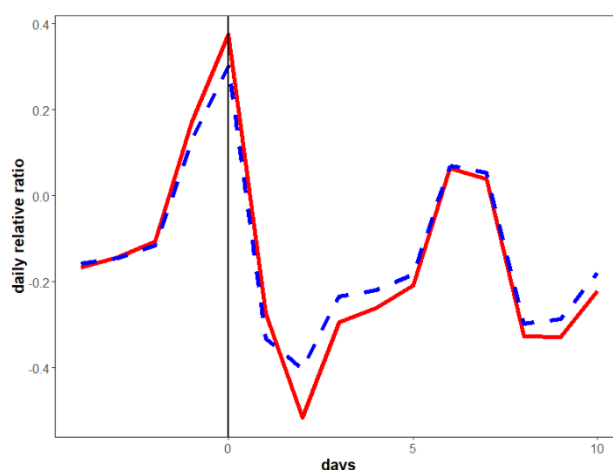
第四章 信息：有限关注能影响投资者投资决策吗？

4.1 投资者关注行为

为了研究注意力对于投资者的影响，我们必须找到合适的、能够引发外生冲击的事件，于是我们把目光投向了“龙虎榜”。相对于那些上龙虎榜的公司，那些差一点点上榜的公司每天盘后必然曝光度相对较小，于是我们就有了一个准自然实验：投资者对于那些曝光度不同的股票他们的投资决策会发生变化吗？

首先为了确定投资者的确是注意到了龙虎榜带来的曝光，我们选取了百度指数作为投资者关注的代理变量（俞庆进，张兵，2012；Zhang et al., 2013）。我们分别根据“公司名称”、“公司简称”和“股票代码”爬取了所有深圳市场上上市公司的百度指数，并且在对各类指数标准化后等权平均，最后计算平均值的日环比数据。对于那些缺失的数据（如一家公司只有公司名称的百度指数，而没有股票代码的百度指数），我们通过调整剩余数据的权重或者跳掉这只股票。最终，我们清洗得到那些：恰好上榜和恰好没上榜的公司，从上榜日前 5 天到上榜日 10 天后的百度指数的环比变化图，如图 1 所示。图一中，红色实线的是那些上榜公司百度指数环比变化图，蓝色虚线代表那些未上榜的公司。我们可以发现在上榜日前后 2 天内，红色实线始终高于蓝色实线，这证明了投资者的确会注意到由龙虎榜带来的曝光度，我们的实验背景是合理、稳健的。我们可以利用这样一个准自然实验开始我们的断点回归设计。

图 2 上榜和恰好没上榜公司的百度指数环比变化图



注：上榜股票是当日在深圳股市某一个板块上榜的股票，恰好没上榜的股票是那些恰好上榜的那只股票后十只的股票，具体选取方法如图 1 所示。

4.2 不考虑异质性影响的 RDD 主实验

考虑到我们数据的特殊性，我们的 RDD 实验主要使用的是一个非参数的局部线性回归（Non-parametric Local Linear Regression，简称为 LLR），下同，这也是在研究局部平均处理效应（LATE effect）时，不知道样本分布的离散数据最常用的模型（Hahn et al., 2001；Imbens and Lemieux, 2008）。我们在附录中使用了全局参数估计的方法进行了 RDD 实验，结果同样稳健。

如前所述，我们模型的被解释变量是资金流变化比率 $Flowchg_{i,q,t+1,F}$ ，解释变量是经调整后的股票排名，而不是标准的排名。因为在数据清洗这一部分，我们根据先日涨跌幅、后成交量规则对每日股票进行排序，并取每日上榜股票数量加上其之后的 10 只股票，例如某一天有 3 只股票上榜，我们就整理出共 13 只股票。但是由于龙虎榜上榜有一个大于 $\pm 7\%$ 的阈值，这意味着每一天股票上榜数量并不是一致的，有可能某天在主板市场有 1 只股票上榜，而另一天在主板市场有 5 只股票上榜。变动的榜单长度使得我们在估计 RDD 模型时无法根据标准排名来设定固定的断点。所以为了将断点固定住，我们先将龙虎榜长度减去 10.5 得到调整后的长度，再将每一个标准排名减去这一调整后的长度就得到了每一只股票某一天的经调整后的排名。经调整后，0 就是固定的断点，上榜的股票排名为负，排名从 -4.5 一直到 -0.5，未上榜的排名从 0.5 一直到 9.5（参考图 1 的例子）。

最后，我们的 LLR 回归式如式（6）：

$$\begin{aligned} Flowchg_{i,q,t+1,F} &= \alpha + \beta \times (rank_{i,t,adj} - cut\ point) + \delta \times D_i \\ &+ \gamma (cut\ point - rank_{i,t,adj}) \times D_i + \varepsilon_i \end{aligned} \quad (6)$$

其中， D_i 是表示某只股票是否上榜的虚拟变量，当股票没有上榜了的话 $D_i = 1$ ，反之等于 0。跟所有的 RDD 文献一样，我们的目标就是估计此 LLR 式子中虚拟变量的系数 δ ，即局部平均处理效应。 $rank_{i,t,adj}$ 是经调整后的排名， $cut\ point$ 是 0。

在式（6）中，驱动变量 $rank_{i,t,adj}$ ，结局变量是 $Flowchg_{i,q,t+1,F}$ 。在这一准自然实验中，因为一只股票是否恰好上榜取决于榜单的长度和是否达到披露要求，所以驱动变量不会被人为干预所影响，并且，在 0 处的断点也是完全外生的。最后，这是一个精确断点的 RDD 实验，因为一只股票一旦达到上榜要求且被披露，那么它就一定会被投资者所看到，正如图 2 所展示的那样。

我们对这个 LLR 式子采用加权最小二乘法（WLS）回归而不是普通最小二乘法（OLS），估计权重由 Imbens and Kalyanaraman（2012）的 IK 估计法决定，

LLR 估计时的带宽 (bandwidth) 也由这个方法决定。此外, 在估计时我们的估计核函数时三角形核函数, 因为根据 Lee and Lemieux (2010) 的推导, 这个是在进行 RDD 实验时最有效且无偏的方式。不过在估计带宽时我们也采用了由 Calonico et al.(2014) 所提出的估计带宽的 CCT 法, 当然估计结果同样稳健。在估计时, 所有的标准误都经过在一阶 HAC⁶ 上调整过, 并且我们还对 LLR 在两倍带宽、一半带宽时同样进行了估计, 回归结果均稳健。

表 3 断点处检验系数

	<i>Flowchg_{i,q,t+1,F}</i>					
	Small inflow	Small outflow	Medium inflow	Medium outflow	Large inflow	Large outflow
Panel A : RDD 检验						
RDD	-0.23 (-3.11)	-0.20 (-3.12)	-0.18 (-2.28)	-0.24 (-3.08)	-0.46 (-3.29)	-0.39 (-2.72)
RDD w/c	-0.26 (-3.52)	-0.22 (-3.47)	-0.21 (-2.64)	-0.26 (-3.45)	-0.51 (-3.72)	-0.45 (-3.18)
Panel B : Logit 检验						
Logit	0.26 (60.10)	0.24 (50.14)	0.24 (57.00)	0.23 (52.67)	0.11 (46.31)	0.12 (49.97)
Logit w/c	0.02 (22.16)	0.01 (17.83)	0.01 (19.61)	0.01 (16.49)	0.00 (12.63)	0.00 (12.94)
Panel C : Z 检验						
Z test	0.52 (9.69)	0.32 (7.47)	0.49 (8.26)	0.42 (7.94)	0.93 (8.33)	0.99 (8.40)
Z fitted at 0.5	0.91 (14.45)	0.57 (11.55)	0.96 (14.65)	0.86 (13.37)	1.41 (10.50)	2.04 (14.44)

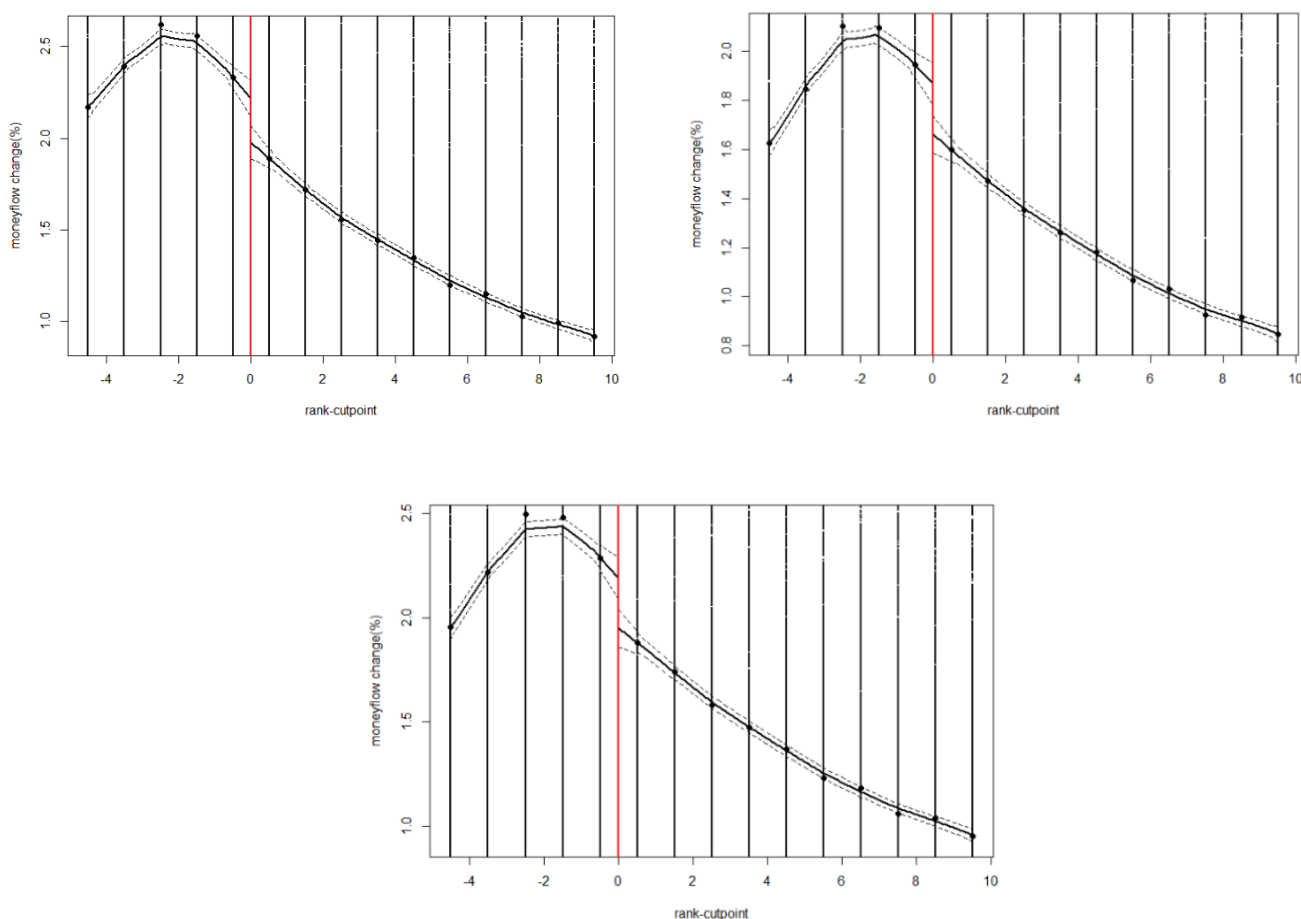
注: 每一列分别是成交量分类为小、中、大单, 流入或者流出方向的股票资金流变化情况 $Flowchg_{i,q,t+1,F}$ 。例如, Small Inflow 是由成交金额归类为小单的, 资金流方向为流入的资金流变化比率 $Flowchg_{i,small,t+1,inflow}$ 。第一行是由 (5) 式 LLR 通过加权最小二乘回归估计得到, 第二行是在 (5) 式中加入控制变量 (with controls), 控制变量包括股票的流通市值、账面市值比、 β 值、收盘价、成交量。第三行是对 (6) 式 logistic 回归的估计, 其中被解释变量是代表股票是否上榜的虚拟变量的对数, 被解释变量是资金流变化比率。第四行是对加上了控制变量后的 (6) 式进行回归的结果。第五行是一个对于断点处的 Z 检验, 即检验: 即将上榜的那支股票 (即 $rank_{i,t,adj} = 0.5$) 和恰好上榜的那支股票 (即 $rank_{i,t,adj} = -0.5$) 之间资金流变化情况的比较。第六行也是一个 Z 检验, 只不过我们首先通过线性拟合的方法, 根据经调整后排名为从 9.5 至 1.5 股票的资金流变化情况, 拟合出即将上榜的那支股票 (即 $rank_{i,t,adj} = 0.5$) 的 $\widehat{Flowchg_{i,q,t+1,F}}$, 再将其同恰好上榜股票的实际的 $Flowchg_{i,q,t+1,F}$ 进行比较所得的结果。表中, 括号内是 t 值, 且所有系数均显著并且以加粗的方式标出。

式 (5) 的回归结果在表 3 中第一行所展示, 断点回归示意图如 3 所示。在回归中我们分别将不同类型的订单当作被解释变量予以回归 (即成交量档次为小、中、大的, 资金方向为流入、流出的变化比率, 资金流入用 inflow 表示, 流出用 outflow 表示)。可以发现所有的系数均显著为负。当我们的研究对象是个体投资者时 (small inflow/out flow), 我们能够发现因为股票上榜带来的投资者关注使得

⁶ 即 Newy-West (1987) 的方法, 下同

资金流入变化率提升了 0.23%，而资金流出变化率而言提升了 0.2%，换言之，关注度能够引起投资者股票决策行为的变化。

图 3 RDD 实验中对（5）式的回归拟合，断点处在 0 点



注：从第一行左上启至第二行，分别是对式（6）中结果变量（即普通回归中的被解释变量）未 Small inflow、Medium inflow 和 Large inflow 时的非参数回归结果图。

此外，在图 3 中，可以发现在上榜股票中，股票排名和资金流变化呈现的是“倒 U 型”的关系。造成这种图像的原因是由于龙虎榜是主要根据涨跌幅顺序进行排名的，对于那些排名第一的或者靠前的股票而言，往往会经历多次涨停板，这就导致了在上龙虎榜第二天之后仍然无法买入这些股票。根据我们的统计，对于那些排名第一的股票，他们第二天触及涨跌、跌停（俗称“天地板”）或盘中至少有触及停牌的股票概率达到了 13%，暂停交易所带来的资金流骤降导致了 $t+1$ 日他们资金流变化比会显著地低于排名第三、第四的股票。

不过这种“倒 U 型”关系不会影响我们对于 LATE 系数的估计，因为在估计 LLR 式中，根据 IK 带宽算法则得出的平均带宽长度不会超过 2，这意味着在

回归时不会考虑排名小于-2.5 的股票（经调整后的排名），所以我们的 RDD 实验仍然稳健。

因为分类策略的随机性，一个稳健的 RDD 实验一般是不需要加入其他协变量来控制模型的内生性的，不过我们在这里也加入了有一些控制变量进行回归，如表中第二行 RDD w/c 所示，控制变量包括股票的流通市值、账面市值比、 β 值，收盘价，成交量。可以发现所有结果也均显著，代表龙虎榜上榜能够影响投资者之后的投资决策，无论是个体投资者还是机构投资者，他们的投资决策行为都得到了极大地加强。在表 3 的回归结果中，无论是加入协变量与否，LATE 系数的变化都在 ± 0.06 以内，这符合我们对于 RDD 实验的理解：控制变量并不会影响我们对 MATE/LATE 的估计（Lee and Lemieux, 2010）。

4.3 基于 Logit 模型和 Z 检验的实验

我们参考 Kaniel and Parham（2017）和 Peng et al.（2016），还采用了另外 2 个模型来验证断点位置处理效应的显著性和稳健性，分别是 Logit 模型，和 Z 检验模型。在 Logit 模型中，我们的被解释变量是代表某只股票是否上榜的虚拟变量，解释变量则是股票的资金流变化比率，具体表述如式（7）。我们在表 3 Panel b 分别汇报了是否加入控制变量的模型后估计结果。

$$\text{Logit}(D_i) = \beta_0 + \beta_1 \times \text{Flowchg}_{i,q,t+1,F} + \beta_3 \times \sum \text{controls} + \varepsilon_i \quad (7)$$

其中， D_i 代表了一只股票是否被上榜的虚拟变量， $\text{Flowchg}_{i,q,t+1,F}$ 是资金流变化比率，控制变量包括流通市值、账面市值比、 β 值、收盘价、成交量。

在回归结果中，系数均显著为正，这意味着股票上榜与否和投资者的投资决策有着显著的相关度。值得说明的是，虽然 Logit 模型经常用在因果推断中，但是在这里我们并非如此，我们只想对实验在断点处的处理效应进行识别，即判断断点前后资金流变化程度是否显著。

类似地，表 3 中 Panel c 中（即第 5、6 行）是对断点前后资金流变化比率的 Z 检验。这个是最简单、也是最纯粹地检验，当然其包含的噪声也最多。

第五行的 Z 检验就是将断点前（即 $\text{rank}_{i,t,adj}$ 为 -0.5 的股票）的资金流变化比减去断点后（即 $\text{rank}_{i,t,adj}$ 为 0.5 的股票）的资金流变化比的差值同 0 做 Z 检验。结果显示上榜的股票平均而言资金流变化比要大于那些没有上榜股票的 0.52%，这也证明了在有限关注的情况下，投资者的选择集（Choice Set）会受到

极大的影响，那些关注度高的股票相对于那些关注度低的而言更容易出现关注度溢价。

表 3 中第六行的 Z 检验是将断点前股票的实际资金流变化和拟合出的断点后资金流变化相互进行检验后得到的结果。其中，我们通过线性拟合第 1.5 只直到第 9.5 只股票的数据，预测出排名第 0.5 只的股票资金流变化，并且将预测出排名第 0.5 的变化和实际排名-0.5 名的资金流变化做 Z 检验。我们可以发现结果同样稳健，对于无论是小单、中单还是大单，资金流方向不论是流入还是流出的资金流变化比而言，由于投资者注意力分配不均所导致的处理效应十分显著，都证明了投资者关注会影响他们的投资决策。

最后，对于我们的结果，我们只汇报了那些触发“偏离涨跌幅”为正的股票，而没有汇报那些因为异常下跌而上榜的股票（即忽略了涨跌幅偏离值超过-7%的股票）。这是因为相较于那些异常下跌的股票，异常上涨的股票上榜数量更多。从 2012 年 1 月至 2019 年 11 月，每一个市场上自己的龙虎榜中，由于异常上涨而上榜的股票平均数量达 4.59 只，而异常下跌的平均数量仅有 2.44 只。在利用 LLR 估计断点处 LATE 系数时，如果断点前数据量过少的话可能会出现带宽不足的问题，导致无法估计出准确、可靠的结果⁷。

结合上述分析我们得到的结论为学界对于“有限关注下的投资者行为”研究提供了更多的证据，在无论是传统的定价文献还是市场微观结构的文献中，未来的学者们都很有必要将投资者的注意力作为不可或缺的变量纳入模型中。

4.4 考虑个股异质性影响的 RDD 实验

接下来我们研究股票的异质性特征，如市值、当日成交量或换手率等特性是如何影响投资者的决策的。根据我们之前的推论，那些市值小的、平时成交量较低股票在受到外生注意力冲击时的更容易受到投资者的关注。

为了量化股票的异质性是如何经由注意力渠道影响资金流的，我们在(7)式中加入交叉项和代表异质性特征的协变量，如下：

$$\begin{aligned}
 Flowchg_{i,q,t+1,F} &= \alpha + \beta \times (rank_{i,t,adj} - cut\ point) + \delta \times D_i \\
 &+ \gamma (cut\ point - rank_{i,t,adj}) \times D_i \\
 &+ \Gamma_0 \times Covariates_{i,t+1} + \Gamma_1 Covariates_{i,t+1} \times D_i
 \end{aligned} \tag{8}$$

⁷ 对于我们的 RDD 模型是否稳健，我们在附录 2 里进行了进一步的说明。

其中, $rank_{i,t,adj}$ 是经调整后的排名, $cut\ point$ 是 0, D_i 是代表股票是否上榜的虚拟变量, $Covariates_{i,t+1}$ 是我们加入的协变量, 用来研究个股异质性对于资金流变化比率 $Flowchg_{i,q,t+1,F}$ 的影响。

在此, 我们加入了多个协变量, 并且将其与代表股票是否上榜的虚拟变量相乘得到交叉项, 这些协变量主要包括流通市值 (SIZE)、上榜日之前 5 日的换手率平均值 (TOVER), β (此处未汇报结果) 和当日上榜股票数量 (Distraction) 等。其中, 当某一张龙虎榜中上榜股票数量等于 4 或 5 时, 取 Distraction 为 1, 当上榜股票数量等于 1 或者 2 时, 取 Distraction 为 0。由于篇幅限制, 我们这里只汇报其中 3 类协变量。在估计 LLR 式时, 我们对于模型的具体设定和之前并没有太大变化。表 4 中, 我们在 Panel A 对不同异质性分别进行了检验, 并分别汇报了断点处的 LATE 估计量, 协变量和交互项的系数。

因为我们的研究对象是个体投资者, 所以我们只汇报了当订单成交金额属于最小一个类别 (Small inflow/outflow), 即由个体投资者驱动的买入卖出订单情况。从回归结果中可以发现, 无论是对于何种异质性特征, 虚拟变量 δ 的估计都显著, 并且各协变量的交互项也显著。这意味着在发生外生的注意力冲击的情况下, 当某一只股票原本的曝光度越低时, 他们对于投资者的吸引力也越强。同时我们可以发现, 不管是资金流入还是资金流出而言, 当龙虎榜上榜股票数量越多时, 资金流变化比率越少。这支持了 Peng et al. (2016) 的注意力分散假说, 即投资者的注意力是稀缺资源, 他们会将关注度分散到那些能够吸引它的股票上。Panel B 汇报了我们同时控制多个协变量时的结果, 从表中可以发现, 所有回归系数依然稳健, 支持了我们之前的观点。

表 4 考虑了协变量的 RDD 实验 (个体投资者)

Panel A:添加独立的协变量作为交叉项检验							
		$Flowchg_{i,small,t+1,F}$					
		Small inflow			Small outflow		
		SIZE	TOVER	Distraction	SIZE	TOVER	Distraction
δ		-0.27	0.57	6.32	0.24	0.49	4.75
		(12.68)	(22.93)	(44.07)	(12.80)	(22.79)	(38.54)
Γ_0		-0.00	-0.01	0.37	-0.00	-0.02	0.26
		(-26.68)	(-7.83)	(77.15)	(-29.36)	(-26.69)	(60.09)
Γ_1		-0.00	-0.26	-0.41	-0.00	-0.02	-0.31
		(-4.20)	(-24.03)	(-42.21)	(-4.64)	(-24.12)	(-36.68)
Panel B: 同时添加多个协变量进行交叉项检验							
Coefficients							
		α	δ	$\Gamma_1 SIZE$	$\Gamma_1 TOVER$	$\Gamma_1 Distraction$	N
outcome	Small	-3.03	6.14	-0.00	-0.02	-0.38	59791
	inflow	(-47.21)	(43.21)	(-6.58)	(-19.66)	(-38.30)	
	Small	-1.97	4.59	-0.00	-0.02	-0.28	60301
	outflow	(-31.44)	(37.68)	(-29.75)	(-20.50)	(-32.96)	

注：Panel A 是我们对（8）进行的回归估计，其中，结局变量是成交量被归类为小单的资金流变化比率 $Flowchg_{i,small,t+1,F}$ ，方向 F 分别包括流入 inflow 和流出 outflow。在估计（8）式中，我们分别加入流通市值（SIZE）、上榜日之前 5 日的换手率平均值(TOVER)和代表当日上榜股票数量的虚拟变量（Distraction）。当某一张龙虎榜上榜股票数量等于 4 或者 5 时，Distraction 取 1，当上榜股票数量等于 1 或者 2 时，Distraction 取 0。Panel A 中 Γ_1 是虚拟变量和协变量交叉项的系数，也是我们主要的研究对象，它代表了股票的异质特征（如市值）是如何影响投资者买入卖出这支股票的行为。Panel B 是在 Panel A 的基础上，同时加入多个协变量和交叉项后回归所得的结果。 $\Gamma_1 Covariates_{i,t+1}$ 是在同时加入了多个协变量的情况下， $Covariates_{i,t+1} \times D_i$ 交叉项的系数。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

在短期内（上榜一天后），代表上榜的虚拟变量和注意力分散交互项系数再次显著，交互项系数并且显著为负，这代表了投资者在短期内的确会受到注意力分散的影响，当上榜股票数量越多，他们的净买入行为更不明显。同时，在这个回归中，我们可以发现 10 天之后交互项系数都不显著，这证明了注意力分散不仅会从强度上影响投资者的投资决策，还会对这一强度的持续时间起到关键作用。

第五章 行为：净买入还是净卖出？

5.1 单变量分析

在上一部分，我们通过 RDD 实验验证了注意力对投资者决策行为的影响。但是受限于模型限制，我们并不能直接地判断出投资者在遭受注意力冲击后，更倾向于买入还是卖出行为。在 Baber and Odean（2008）对投资者行为建模研究中，他们发现投资者更倾向于成为那些能够引起他们注意力股票（attention-grabbing stocks）的净买入方，在 Fang and Peress（2009），Da et al.（2011）和 Yuan（2015）的不同实验中，他们也间接证明了这一点。在这里，我们感兴趣的是在中国市场这一结论是否仍然成立？投资者会更多地买入那些引起他们注意力地股票吗？如果是的话，这一效应强弱是否能持续很长时间？

我们通过构建买卖失衡比指标（BSI）来研究投资者具体的投资行为，具体定义方式参考式（4）。我们分别对由成交金额和成交量构建起的 BSI 进行单变量分析，通过汇报那些上榜和没上榜股票 1 日至 50 日之后的 BSI 均值和均值之间的差值来判断股票的买入卖出方向。

单边量分析结果汇报在表 5。我们在 Panel A 我们通过订单金额计算出了每一只股票的每一天的买卖失衡比（BSI），Panel B 是通过订单成交金额计算得出的。以 Panel A 为例，第一列和第二列分别是全部上榜股票和没有上榜股票各自平均的 BSI 值，第三列是将上榜股票的 BSI 减去没有上榜股票的 BSI 计算所得。第 4 列和第五列是恰好在断点前后的两只股票（即刚好上榜和差一点上榜的两只股票）各自 BSI 值，第六列是其差值。在这里，我们汇报了从第 0 日股票上了龙虎榜，直到第 50 日一类股票的 BSI 均值。可以发现，不管是上榜了还是未上榜的股票，当日 BSI 指标都是显著负的，这是因为当一只股票触发涨跌幅偏离值时，它往往也会触发涨停。此时会有大量的买单积攒未被成功执行交易，而卖单总是能成功成交。从第一行和第三行，5 日及之后的 BSI 均显著为正来看，我们不能直接认为上榜引起了投资者的净买入，因为对于那些未上榜的股票而言在 50 日候他们的 BSI 值也均显著为正。这意味着一只股票的上榜与否很大程度上都取决于更深层次、非注意力相关的原因。

但是，当我们把上榜股票和断点处股票的 BSI 值减去未上榜股票和断点下股票的 BSI 值之后，可以从第三列和第 6 列发现，上榜意味着投资者的净买入。对于那些在断点前后的股票而言，上榜后第一天它的 BSI 值平均会显著提升 0.02，5 日后仍会显著地保持这一水准，直到 10 日之后上榜股票和未上榜股票的 BSI 差值才不显著。这一结论证实了“投资者是那些吸引他们注意力股票的净买

入者”这一发现，为投资者关注假说提供了经验证据。此外，投资者关注的持续时间相对地会保持 1~2 周时间长度，之后就会变得不显著。这一发现也证实了对于投资者而言，注意力是有限资源，他们只会被最近的龙虎榜事件所吸引，而会忽视 2 个礼拜之前的事件。

在利用成交量计算出的 BSI 指标中，Panel B 的研究发现同 Panel A 基本一致，在此我们不多做赘述。

表 5 上榜日之后几天内，不同股票的 BSI 均值

		Panel A: Dollar BSI					
		Day (0)	Day (1)	Day (5)	Day (10)	Day (20)	Day (50)
全榜单的股票+后 10 只股票	上榜	-0.21 (-85.56)	0.05 (22.02)	0.04 (31.45)	0.06 (51.34)	0.06 (54.92)	0.06 (52.27)
	未上榜	-0.19 (-75.29)	-0.02 (-8.15)	0.02 (9.05)	0.04 (37.88)	0.05 (62.36)	0.05 (65.08)
	上榜-未上榜	-0.02 (-6.55)	0.07 (23.30)	0.03 (14.80)	0.02 (12.71)	0.00 (4.60)	0.00 (1.63)
	仅断点上下两只股票	-0.26 (-60.92)	-0.00 (-1.45)	0.02 (7.96)	0.05 (21.28)	0.06 (31.45)	0.06 (30.88)
仅断点上下两只股票	未上榜	-0.24 (-56.11)	-0.02 (-5.65)	0.01 (4.07)	0.04 (20.23)	0.06 (30.32)	0.05 (28.30)
	上榜-未上榜	-0.02 (-5.01)	0.02 (4.13)	0.02 (3.32)	0.00 (0.62)	0.00 (1.02)	0.00 (1.50)
		Panel B: Volume BSI					
		Day (0)	Day (1)	Day (5)	Day (10)	Day (20)	Day (50)
全榜单的股票+后 10 只股票	上榜	-0.21 (-85.25)	0.05 (23.04)	0.05 (32.62)	0.06 (52.96)	0.06 (26.08)	0.06 (53.46)
	未上榜	-0.19 (-75.45)	-0.01 (-7.70)	0.02 (9.96)	0.04 (39.40)	0.05 (63.87)	0.05 (66.83)
	上榜-未上榜	-0.02 (-6.14)	0.07 (23.73)	0.03 (14.87)	0.01 (12.87)	0.00 (4.57)	0.00 (1.60)
	仅断点上下两只股票	-0.26 (-60.92)	-0.00 (-1.03)	0.02 (8.50)	0.04 (22.12)	0.06 (32.08)	0.06 (31.13)
仅断点上下两只股票	未上榜	-0.23 (-55.96)	-0.24 (-5.27)	0.02 (4.64)	0.05 (21.08)	0.06 (30.87)	0.06 (29.27)
	上榜-未上榜	-0.02 (-5.04)	0.02 (4.16)	0.02 (3.28)	0.00 (0.63)	0.00 (1.02)	0.00 (1.71)

注：我们在 Panel A 和 Panel B 分别汇报了由成交金额和成交量计算得出的 BSI 均值。以 Panel A 为例，上半部分的数据集包括由全部上龙虎榜的股票和上榜股票后 10 支股票组成的全样本（参考图 1）。相对地，下半部分数据集仅包括断点前后股票的 BSI，即 $rank_{i,t,adj}$ 等于 ± 0.5 的股票。我们分别汇报了那些上榜或者没上榜的股票在上榜日当天、一天后、5 天后、10 天后、20 天后和 50 天后的 Dollar BSI 和 Volume BSI 的均值。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

5.2 多变量分析

在上一部分，我们采用单变量分析的方法，通过研究上榜日后不同天数内平均 BSI 的变化来验证投资者是否会成为那些能吸引他们注意力的股票的净买入方。在这一部分，我们通过对上榜日后不同天数内一只股票的 BSI 进行固定效应面板回归，进一步探究到底什么因素能影响投资者的净买入行为。

在这里，我们对于每一只处在断点上下的股票（即那些刚好上榜和刚好没上榜的股票）的买卖失衡比和其他可能影响到投资者关注和股票本身的特性进行了研究，具体的面板回归设计如式（9）。

$$BSI_{i,t+1}^{100} = \beta_0 + \beta_1 \times D_i + \beta_i \times \sum Controls_{i,t} + Year_i + Market_j \quad (9)$$

其中， $BSI_{i,t+1}^{100}$ 是这只股票在 $t+1$ 日的由小单成交金额计算出的买卖失衡比，并且被放大了 100 倍⁸， D_i 是代表这只股票是否上榜的虚拟变量，控制变量包括代表输赢家的虚拟变量、分析师跟踪人数、财务杠杆、 β 系数、当天同时上榜数量多或者少的虚拟变量、公司透明度、非流动性指标和换手率。其中，输赢家代表动量的虚拟变量。具体计算方法是我们对某一只股票从前 13 个月到上 1 个月共 12 个月的累计收益率进行排序之后，取收益率靠前 30% 的股票作为赢家，且赢家的虚拟变量为 1。代表股票流动性因素的是 Amihud（2002）的非流动性指标，通过将考虑现金红利再投资的个股回报率除以个股日成交金额计算得出。其他数据均来自与 CSMAR 数据库。在我们的面板回归中，我们分别先后控制了年份 $Year_i$ 和所属市场因素 $Market_j$ （即这只股票属于主板、中小板还是创业板），并且所有残差都经过 Newey-West（1987）调整。

表 6 不同天数之后投资者净买入效应（BSI）所受的影响因素

	1day	1day	10days	10days	50days	50days
Dummy	5.07 (26.04)	4.96 (25.19)	0.42 (2.75)	0.36 (2.29)	0.16 (1.01)	0.11 (0.65)
Winner	-0.27 (-1.36)	-0.29 (-1.45)	0.25 (1.56)	0.30 (1.85)	0.27 (1.62)	0.28 (1.63)
AnaCov	-0.04 (-3.25)	-0.07 (-5.59)	-0.15 (-1.53)	-0.03 (-3.61)	0.02 (2.15)	0.01 (0.72)
Leverage	-0.42 (-2.84)	-0.52 (-3.47)	-0.24 (-2.07)	-0.39 (-3.23)	-0.10 (-0.86)	-0.30 (-2.36)
Beta	0.32 (1.27)	0.07 (0.29)	0.35 (1.73)	0.24 (1.18)	-0.05 (-0.26)	-0.12 (-0.58)
Distraction	0.76 (2.23)	-0.67 (-1.94)	0.16 (0.58)	-0.58 (-2.12)	0.67 (2.34)	0.01 (0.25)
Opacity	0.10 (0.36)	-0.43 (-1.57)	-1.45 (-0.66)	-0.43 (-1.98)	-0.50 (-2.19)	-0.77 (-3.34)
Illiqd	-0.14 (-1.69)	-0.71 (-7.85)	-0.24 (-3.36)	-0.66 (-9.29)	-0.11 (-1.44)	-0.45 (-5.98)
Turnover	2.09 (19.89)	1.58 (15.44)	0.28 (3.41)	0.01 (0.14)	0.27 (3.11)	0.04 (0.50)
Fixed year	True	True	True	True	True	True
Fixed market		True		True		True
N	37147	37147	36921	36921	35945	35945

⁸ 我们并没有取自然对数，这是因为原始数据中 BSI 有零值，我们不想简单地清洗掉那些原本有意义的数。放大 100 倍之后，原先只在小数点三位后有意义的系数能够在表中显示了。

R2	0.036	0.035	0.01	0.00	0.00	0.00
----	-------	-------	------	------	------	------

注：被解释变量分别是一只股票 1 天、10 天和 50 天之后并倍乘 100 倍之后的 BSI，解释变量式代表公司是否恰好上龙虎榜的虚拟变量 *Dummy*，控制变量代表输赢家的虚拟变量、分析师跟踪人数、财务杠杆、 β 系数、当天同时上榜数量多或者少的虚拟变量、公司透明度、非流动性指标和换手率。我们分别在年份和深交所板块上对固定效应进行了控制，所有残差也经过 Newey-west（1987）调整。表中，括号内是 *t* 值，且所有系数均显著并且以加粗的方式标出。

从表 6 中可以发现，在 1 天和 10 天之后，所有代表股票是否上榜的虚拟变量均显著，而 50 天之后的结果并不显著，这证明了我们之前对 BSI 的单变量分析，即一只投资者会在关注到上榜股票买入这只股票，并且这一现象大约能够持续 10 个交易日左右。

除了在第一部分进行的 RDD 实验之外，我们还在这里对注意力分散假说进行了再检验，具体方法是在回归中加入代表上榜股票数量孰多孰少的虚拟变量。当某一天有 4 只或以上股票上榜时，投资者的注意力就会被这 4 只给分散掉，我们便取虚拟变量为 1，即分散效应较强；当上榜数量小于 2 只时，我们取虚拟变量为 0。此外，我们还对 BSI 乘以 100 倍地放大了，这样我们的系数在汇报时就不会因为过小而无法在表中正常表示。我们的面板回归式如（10）式所示：

$$BSI_{i,t+1}^{100} = \beta_0 + \beta_1 \times D_i + \beta_2 \times Distraction_i + \beta_3 \times D_i \times Distraction_i + \beta_i \times \sum Controls_{i,t} + Year_i + Market_i \quad (10)$$

在短期内（上榜一天后），代表上榜的虚拟变量和注意力分散交互项系数再次显著，交互项系数并且显著为负，这代表了投资者在短期内的确会受到注意力分散的影响，当上榜股票数量越多，他们的净买入行为更不明显。同时，在这个回归中，我们可以发现 10 天之后交互项系数都不显著，这证明了注意力分散不仅会从强度上影响投资者的投资决策，还会对这一强度的持续时间起到关键作用。

表 7 对注意力分散假说的再检验

	1day	1day	10days	10days	50days	50days
D_i	6.75 (7.12)	6.82 (7.11)	0.29 (0.39)	0.34 (0.44)	-0.39 (-0.49)	-0.36 (-0.45)
D_i	-1.75	-1.93	0.13	0.02	0.58	0.49
$\times Distraction_i$	(-1.82)	(-1.98)	(0.17)	(0.02)	(0.71)	(0.60)
<i>Controls</i>	YES	YES	YES	YES	YES	YES
Fixed year	True	True	True	True	True	True
Fixed market		True		True		True
N	37147	37147	36921	36921	35945	35945
R2	0.036	0.035	0.01	0.00	0.00	0.00

注：类似于表 6，我们对某一只股票不同天数后的 BSI 进行了面板回归。不同于表 6 的地方在于，我们在式（9）中

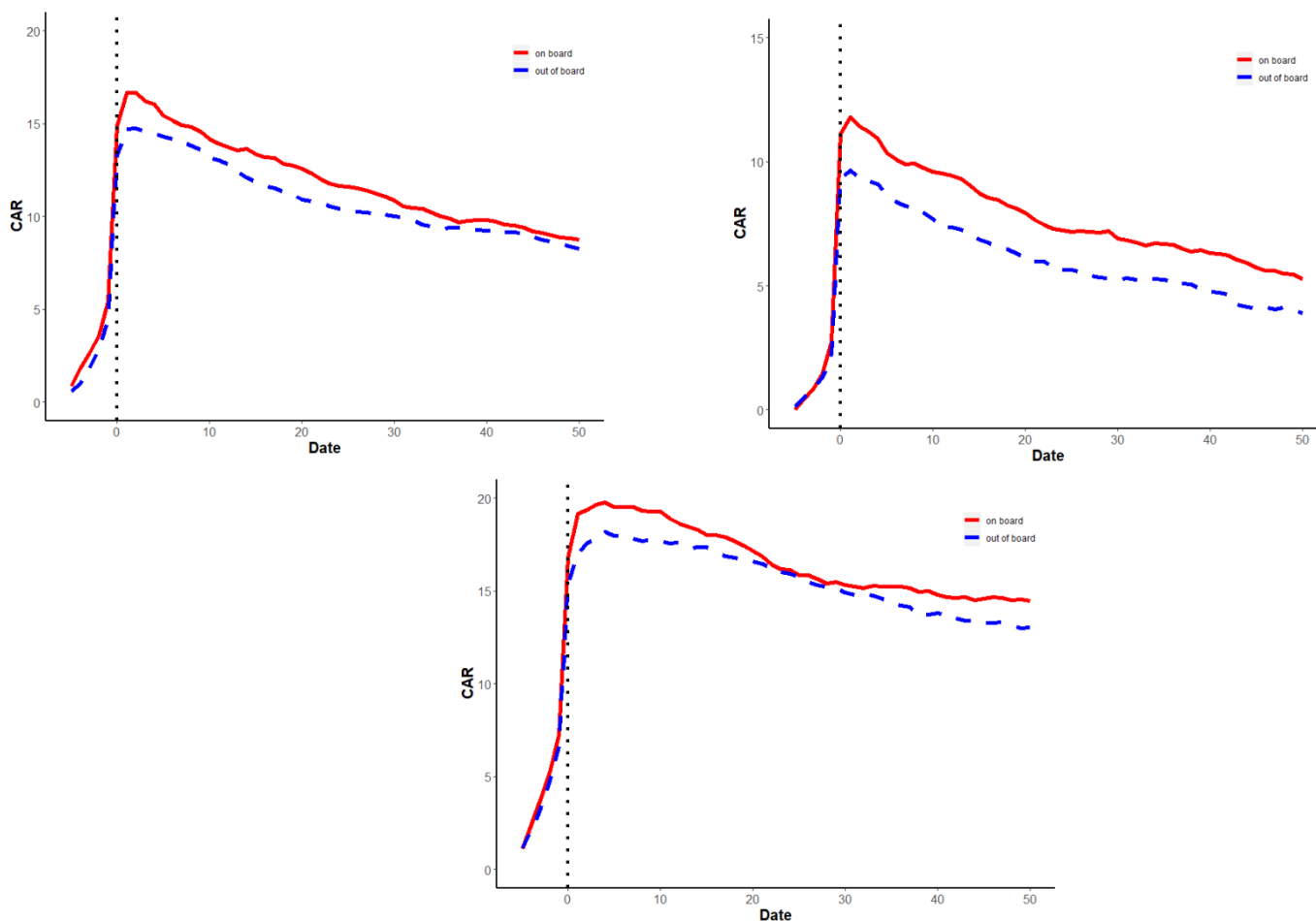
加入了 Dummy 和 Distraction 相乘的交叉项，并且在第二行这一行汇报了这个交叉项的系数。我们分别在年份和深交所板块上对固定效应进行了控制，所有残差也经过 Newy-west（1987）调整。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

第六章 价格：注意力买入行为与投资者收益

6.1 基于事件研究法的超额收益

在之前的部分我们分别验证了投资者会净买入那些能够引起他们注意力的股票，我们最后想要探究的是：投资者会因为这种非理性行为亏损还是盈利？这也是我们对 Merton（1987）和 Hong and Stein（1999）的假说的直接检验。根据 Merton（1987），由风险偏好相对提升所导致的净买入行为之后并没有理由导致收益反转；而根据 Hong and Stein（1999），由动量交易者的净买入行为则会推升泡沫越来越大，最终破裂。前者是“信息共享”的理论机制，后者则是“过度反应”的主要依据。

图 4 上榜日后的平均超额累计收益



注：从第一行左上起至第二行，分别是深圳主板、中小板和创业板股票上榜日前 5 日至上班日 50 日后的累计超额收益。红线代表的是那些上榜的股票，蓝线代表那些恰好没上榜的股票

可以发现，在最初，那些恰好上榜的股票的平均累计收益率要大于那些恰好没上榜的股票；随着时间的推移，他们之间收益率的差值逐步减少。我们推断这可能是因为投资者关注度的逐步消散所导致的。

所以，直观的来看，两者收益率差值的逐步减小意味着的确存在所谓的“未报道溢价”（No-coverage Premium, Fang and Peress, 2009），也从侧面验证了投资者关注假说的成立。但是，图 4 中我们能够很明显的发现投资者在买入并持有这些能够引起他们注意力股票之后，在长期会遭受十分严峻的损失。对于买入并持有在主板中能吸引他们注意力的股票会使得他们在未来 50 个交易日后亏损 6%，买入中小板股票会亏损约 4.5%，买入创业板的股票会亏损 2%。虽然买入那些能引起他们注意力的股票在短期的确能够获利 2% 左右，但是在长期收益率则会发生极大的反转。我们认为这种反转是由于投资者的过度反应（over reaction）所导致的，并且这种非理智行为所表现出的动量效应在我们独特的数据集中得到了充分的放大。我们推论，龙虎榜事件给一只股票带了巨大的关注度，而这种过度关注导致了投资者在短期内产生对于这只股票的过度反应，进而引起股票价格的过分高估。而在 5 个交易日之后，投资者的过度反应逐步消退，收益率也开始反转，导致“过度关注弱势”现象的发生（Peng et al., 2016；饶育蕾，等，2011）

6.2 通过构建零投资组合研究投资者的盈亏

在 Fang 和 Peress（2009）对于媒体效应和股票横截面预期收益的实证检验中，他们通过买入那些引起被媒体报道过的股票、卖空那些没有被媒体报道过的股票，构建了一个零投资组合，并将这个组合的收益率序列对常见的风险因子进行回归。他们的结果发现，投资者的非理性买入行为会有一个显著的溢价，虽然这种溢价能在很大程度上被风险因子所解释。

类似地，在这里我们通过在第 $t+1$ 日买入在 t 日上榜的股票，卖出那些恰好没有上榜的股票，在主板、中小板、创业板之间等权构建投资组合，分别按 5 日、一个月、50 日进行调仓操作，在调仓中所有股票权重保持不变。对于构建出的投资组合的收益率序列对各风险因子回归后所得的截距显著的话，我们就能够断定投资者的注意力买入行为对他们的收益产生了重要的影响。在这里，我们采用了 5 个模型，分别是 Jensen 的市场模型，Fama-French 的 3 因子和 5 因子模型，Carhart 的动量模型以及加入 Pastor-Stambaugh 流动性因子的 Fama-French3 因子模型。

其中，SMB 是小公司股票收益率减去大公司股票收益率，HML 是高账面市值比公司收益率减去低账面市值比公司，UMD 是过去 13 个月到过去 2 个月内

高高收益股票收益率减去低收益股票收益率，RMW 是高盈利能力股票收益率减去低盈利能力股票收益率，CMA 是高投资使用水平股票减去低投资使用水平股票的收益率，PS 的计算是将深圳市场所有股票月度的流动性因子等权平均所得的市场流动性。

从表 8 中可以看出，在构造期较短的时候，所有 Alpha 系数均显著为负，这意味着在控制了诸多风险因子之后，短期投资者的注意力买入行为会使得他们亏损。此外，我们还能发现所有风险模型的 R^2 均在 0.2 左右，F 值均显著，这意味

表 8 短期和长期的龙虎榜影响

Panel A: 构造时长 5 日					
	Jensen	FF3	Carhart	PS	FF5
Alpha	-0.01 (-2.02)	-0.00 (-2.38)	-0.01 (-2.41)	-0.01 (-2.28)	-0.00 (-2.28)
RMRF	1.07 (19.51)	0.92 (14.57)	0.94 (16.25)	0.08 (13.69)	0.89 (13.79)
SMB		1.01 (7.60)	0.91 (6.52)	1.07 (7.90)	0.65 (3.20)
HML		-0.11 (-0.60)	0.03 (0.21)	-0.06 (-0.343)	-0.27 (-1.00)
UMD			0.02 (0.24)		
PS				0.00 (0.23)	
RMW					-0.48 (-1.18)
CMA					0.69 (2.02)
R^2	0.18	0.21	0.21	0.21	0.21
F	380.70	148.20	111.30	107.20	88.60
Panel B: 构造时长 1 月					
	Jensen	FF3	Carhart	PS	FF5
Alpha	-0.00 (-0.85)	-0.00 (-1.28)	-0.00 (-1.01)	-0.00 (-1.38)	-0.00 (-1.13)
RMRF	1.21 (20.26)	1.15 (16.84)	1.14 (18.16)	1.16 (16.67)	1.10 (15.83)
SMB		0.47 (3.79)	0.33 (2.43)	0.48 (3.90)	0.32 (1.71)
HML		-0.58 (-3.30)	-0.21 (-1.49)	-0.56 (-3.09)	-0.88 (-3.56)
UMD			-0.09 (-1.07)		
PS				-0.00 (-0.16)	
RMW					-0.00 (-0.02)
CMA					0.77 (2.58)
R^2	0.20	0.22	0.21	0.22	0.22
F	410.50	142.70	107.10	106.90	89.20

续表 8 短期和长期的龙虎榜影响

Panel C: 构造时长 50 日					
	Jensen	FF3	Carhart	PS	FF5
Alpha	0.00 (1.57)	0.00 (1.23)	0.00 (1.25)	0.00 (1.12)	0.00 (1.22)
RMRF	1.12 (25.54)	1.05 (20.96)	1.01 (22.08)	1.04 (20.26)	1.04 (20.12)
SMB		0.79 (7.60)	0.73 (6.52)	0.78 (7.27)	0.59 (3.73)
HML		-0.04 (-0.27)	0.17 (1.45)	-1.12 (-0.31)	0.01 (3.73)
UMD			0.11 (1.62)		
PS				-1.43 (-0.31)	
RMW					-0.49 (-1.52)
CMA					0.14 (0.54)
R ²	0.27	0.28	0.29	0.29	0.28
F	652.40	232.20	179.20	168.70	141.00

注：我们按流通市值加权买入那些上龙虎榜的股票，卖出那些没有上榜的股票，以 5 个交易日、1 个月、50 个交易日为长度构建零投资组合，并且将投资组合的收益率对风险因子进行回归。主要的因子回归有：市场因子（RMRF），市值因子（SMB），账面市值比因子（HML），动量因子（UMD），流动性因子（PS），价值因子（RMW），投资因子（CMA）。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

着我们的模型非常稳健。在所有因子中，市场因子最为显著，解释能力也最强，这也符合 Fang and Peress（2009）以及冯旭南（2016）的发现。在中长期，如果我们以月频和 50 日长度进行调仓的话，可以发现此时模型的截距都不再显著，这同我们之前的发现一致，即投资者的关注度是有限资源，这一关注强度并不能持续很长时间，在前 5 日内投资者注意力比较集中，而长期注意力则会慢慢消散。

6.3 投资者的注意力净买入行为和投资收益之间的回归分析

根据 Peng, 等（2016）和饶育蕾, 等（2011）等众多涉及中国资本市场上的投资者关注和股票横截面收益的相关研究，中国市场上存在着显著的“过度关注弱势的现象”。这种“过度关注弱势”背后的逻辑实际上是“注意力驱动”(attention-driven)和过度反应（overreaction）所共同导致的。在短期，投资者由于受到注意力冲击会过分地关注一只股票，进而过分高估其内在价值。在 Peng, 等（2016）的事件研究中，他们发现由投资者过度关注而导致的交易行为中，他们的盈利只能持续 2 天，之后累计收益就会显著为负，变成亏损的状态。

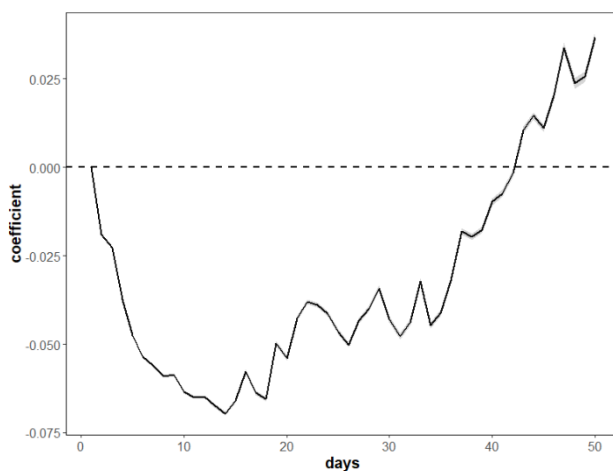
在这里，我们模仿 Kaniel, 等（2016），通过面板回归的方法，研究由投资者的注意力净买入和他们投资收益的关系，具体回归式如式（11）：

$$\begin{aligned}
 Ret_{[x,y]} = & \beta_0 + \beta_1 \times D_1 + \beta_2 \times BSI_{i,t+1} + \beta_3 \times D_1 \times \\
 & BSI_{i,t+1} + \beta_i \times \sum Controls_{i,t} + \alpha_1 \times Ret_{[-1,0]} + \alpha_1 \times \\
 & Ret_{[-5,-1]} + \alpha_2 \times Ret_{[-26,-5]} + FE
 \end{aligned} \tag{11}$$

其中，被解释变量是上榜日第 x 天起，这只股票持有至第 y 天时的累计收益，解释变量分别有代表这只股票是否上榜的虚拟变量 D ，这只股票在上榜日之后一天的买卖失衡比 $BSI_{i,t+1}$ ，这两者的交叉项，包括市值、ROE、动量、分析师关注度、换手率等控制变量 $Controls_{i,t}$ 以及包括年份和市场类型的固定效应 FE 。除此之外，我们还控制了这只股票过去 1 天，过去 1 周和过去 1 个月的不重叠的（non-overlapping）收益率，以此在最大程度上控制回归式中的共线性。

我们首先取 $x=1$ ， $y=1\sim50$ 对上式进行回归，并且控制年份变量，对残差经 Newy-west（1987）调整。对上式进行估计后，绘出了如下的图 5。其中，横轴是代表持有的天数，纵轴是回归式中交叉项 $D_1 \times BSI_{i,t+1}$ 的系数。

图 5 用买卖失衡比指标预测未来收益的系数图



注：我们取 $x=1$ ， $y \in [1,50]$ 对式（11）进行面板回归。横轴是代表持续时长 y 的天数，纵轴是交叉项的系数 β_3 。我们分别在年份和深交所板块上对固定效应进行了控制，所有残差也经过 Newy-west（1987）调整。

在回归式中，我们需要重点关注的是交叉项的系数 β_3 ，这个系数代表着投资者关注引起的净买入行为是如何影响这只股票未来收益率的。如果 β_3 显著为负，那么“过度关注弱势”的假说就能成立，即由注意力引起的买入行为会招致投资者的亏损。从这张图中这一假说得到了验证，从持有上榜股票第一天开始之后，回归系数便显著为负，并这一趋势一直持续到 10 个交易日之后，证明了投资者会因为这种非理性买入行为而亏损。而从 15 个交易日之后，投资者的亏损便逐渐减弱，并且在持有 40 个交易日后 β_3 转负为正，体现出了一定的反转，这也符

合饶育蕾，等（2011）上对于投资者关注所引起的超额收益会在中长期引起反转的发现，即“过度关注弱势”。

进一步地，在下表 6 中，我们进一步记录了回归结果。参考 Kaneil，等（2016），我们分别取 $x=1$ ， $y=10$ 和 $x=11$ ， $y=50$ 两组来进行回归。回归结果基本不变：注意力对投资者的未来收益产生显著的影响，在短期投资者会发生显著的亏损，而再长期收益率会反转为正，但并不显著。

表 9 注意力引起的买入行为和股票收益

	$Ret_{[1,10]}$	$Ret_{[1,10]}$	$Ret_{[10,50]}$	$Ret_{[10,50]}$
$BSI_{i,t+1}$	-0.10 (-7.84)	-0.10 (-8.29)	-0.02 (-0.66)	-0.04 (-1.13)
$D_1 \times BSI_{i,t+1}$	-0.06 (-3.35)	-0.06 (-3.38)	0.07 (1.34)	0.06 (1.13)
$Ret_{[-1,0]}$	0.02 (0.59)	0.03 (0.88)	-0.28 (-2.20)	-0.24 (-2.96)
$Ret_{[-5,-1]}$	-0.06 (-2.67)	-0.05 (-2.24)	-0.28 (-3.68)	-0.22 (-2.96)
$Ret_{[-26,-5]}$	-0.07 (-6.46)	-0.05 (-5.34)	-0.08 (-2.51)	-0.04 (-1.28)
Controls	YES	YES	YES	YES
Fixed year	True	True	True	True
Fixed market		True		True
N	5025	5025	5025	5025
R2	0.05	0.05	0.01	0.01

注：我们在 1、2 两列取 $x=1$ ， $y=10$ 对式（11）进行估计，在 3、4 两列取 $x=10$ ， $y=50$ 对式（10）进行估计，第二行汇报了交叉项的系数 β_3 。我们分别在年份和深交所板块上对固定效应进行了控制，所有残差也经过 Newy-west（1987）调整。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

第七章 基于 DID 的稳健性检验

我们在第四章研究信息是如何影响有限关注的投资者行为时，通过一个 RDD 实验，对那些恰好上榜和恰好没上榜的公司资金流入流出变化比进行了比较研究。这个 RDD 实验利用了龙虎榜数据的特殊性，构建起了一个准自然实验。在所有利用准自然实验进行的研究中，除了 RDD 实验进行因果推断之外，在实证金融研究中学者也常常通过设计双重差分实验（DID）来研究事件发生前后，处理组和对照组之间的差别，并且结果的显著性通常不会发生太大的变化。在这里我们通过一个 DID 实验，重复了我们在第四章中对于投资者有限关注的研究。

类似 Li et al. (2019)，我们的 DID 实验主要是一个基于 t 检验的单变量分析。同 RDD 实验所采用的数据相比，我们在此除了计算出龙虎榜上榜后资金流变化比 $Flowchg_{i,q,t+1,F}$ 外，还计算了龙虎榜上榜日前 10 日资金流变化比的平均值，具体计算方式如式（12）：

$$avgFlowchg_{i,q,t,F} = \frac{1}{10} \times \sum_{i=0}^{10} Flowchg_{i,q,t-5+i,F} \quad (12)$$

其中， $avgFlowchg_{i,q,t,F}$ 就是在第 t 日之前 10 日，第 i 只股票成交量大小为 q 档的，资金流入流出方向为 F 的资金流变化比的平均值，这个也就是在 DID 实验中事件发生前的变量。我们之所以取事前平均值而不是前一日的资金流变化比 $Flowchg_{i,q,t-1,F}$ ，是因为在图二和图三中，我们发现在龙虎榜上榜日之前 1 日，投资者关注和资产价格往往会表现出一个较大的偏移，这可能是由于信息泄露或者投资者预期到了导致公司基本面发生变化的预期。

在这这个 DID 实验中，事件日就是股票上榜日，处理组和对照组分别是那些上了榜的和那些恰好没有上榜的公司，划分标准参考图 1 所示（同 RDD 实验保持一致）。由于篇幅所限，我们只展示了成交量被归为小单时（即个体投资者的交易）所引起的资金流变化比率，即 RDD 实验中的 small inflow 和 small outflow。当我们把目光放在由中单和大单驱动的资金流变化率时，主要结果仍然保持显著，意味着有限关注的确影响投资者的投资行为。

我们在表 10 中分别对 $avgFlowchg_{i,q,t,F}$ 和 $Flowchg_{i,q,t+1,F}$ 的平均值进行了汇报，并且汇报了单变量（均值）双重差分之后的 t 值。结果显示，在事件发生日之后，经过双重差分后的控制组和对照组的差值显著，这意味着增加股票上榜，增加曝光度对的确会加剧个体投资者的交易行为；而在事件发生前，两组差值并不显著，这意味着我们的差分结果稳健。

表 10 基于 DID 的单变量分析

		before	after	After-before
Small inflow				
	Treated	0.06	0.32	0.23
	Control	0.01	0.23	0.20
	Treated-control	0.04	0.08	0.03 (12.99)
Small outflow				
	Treated	0.11	0.53	0.41
	Control	0.06	0.42	0.35
	Treated-control	0.04	0.10	0.06 (12.00)

注：事件日是龙虎榜上榜日，实验组和控制组分别是上榜和恰好未上榜的两组股票，表中汇报了 $avgFlowchg_{i,q,t,F}$ 和 $Flowchg_{i,q,t+1,F}$ 的平均值，并汇报了事件发生后控制组和对照组是否相等的 t 值。表中，显著的系数用加粗标出。

第八章 结论与展望

本文研究媒体报道对投资者行为的影响，以：信息、行为、价格的思路为线索，串起了在信息在金融市场种运动的过程。我们对在新兴市场中，媒体报道究竟导致投资者“信息共享”还是“过度反应”进行了研究。

首先，我们证明了媒体报道带来的注意力变动会影响投资者行为。基于来自龙虎榜独特的数据，我们统计出上榜和恰好未上榜的公司在上榜日后小、中、大单的资金流变化比率。通过一个稳健的 RDD 实验，研究了由于媒体报道带来的注意力冲击，投资者是否会真的在金融市场中做出不同的投资决策。我们的回归结果发现一家公司上榜之后，他们会在未来受到更多的投资者关注，并且会有发生更频繁的买入卖出行为。这一 RDD 实验在控制了公司的异质性特征和投资者的特征之后依然稳健。

其次，我们证明投资者是注意力吸引股票（Attention-grabbing Stocks）的净买入方。在基于事件研究的单变量分析之中，我们利用买卖失衡比指标，发现龙虎榜上榜会在短期和中期内加大投资者的买入卖出行为。

最后，我们证明了中国市场上显著存在的过度反应现象，而非信息共享。我们通过因子回归和面板回归的方式对“投资者基于注意力的净买入行为”究竟对他们有益还是有害做出了研究。其中，我们通过加权构建买入引起注意力，卖空不引人注意股票的零投资组合，并且对其进行因子回归后发现在短期投资者会蒙受损失，换言之，市场上存在未被关注的溢价（No-coverage Premium），这支持了 Merton（1987）和 Hong and Stein（1999）的理论模型。在事件被媒体报道平均 3 日后，由过度反应所导致的收益率反转现象则支持了 Hong and Stein（1999）的统一理论，这一现象却不能由 Merton（1987）的模型所解释。

总结全文，这篇文章是我们对于媒体报道是如何影响投资者行为和资产价格的一个深入探讨，它不仅通过一个清洁的准自然实验丰富了学界对注意力相关问题的认知，同时深化了我们对媒体报道对于资本市场信息传播影响的理解。本文也具有清晰的政策含义，指出了交易所公布“交易公开信息”的弊端，为监管部门规范市场中媒体报道，上市公司信息披露提供了参考依据。

在未来，就龙虎榜这一个独特的数据集而言，我们希望能够进一步地通过研究“有影响力地交易席位”（即“知名游资”）来深挖半公开数据集给我们带来的优势，同时我们也希望能通过高频数据集来研究，投资者是在日内的注意力强弱分布，他们的注意力能够被预测吗等一系列问题。就有限关注这个细分领域而言，未来还可以通过很多很有意思的数据、更广阔、学科交叉程度更大的方向来进一步探索。

参考文献

- [1] Ball, R. and Brown, P., 1968. An empirical evaluation of accounting income numbers. *Journal of accounting research*, pp.159-178.
- [2] Barber, B.M. and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2), pp.785-818.
- [3] Barber, B.M. and Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, 55(2), pp.773-806.
- [4] Barber, B.M. and Odean, T., 2001a. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1), pp.261-292.
- [5] Barber, B.M. and Odean, T., 2001b. The internet and the investor. *Journal of Economic Perspectives*, 15(1), pp.41-54.
- [6] Barrot, J.N., Kaniel, R. and Sraer, D., 2016. Are retail traders compensated for providing liquidity?. *Journal of Financial Economics*, 120(1), pp.146-168.
- [7] Bollen, J., Mao, H. and Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1), pp.1-8.
- [8] Calonico, S., Cattaneo, M.D. and Titiunik, R., 2014. Robust nonparametric confidence intervals for regression - discontinuity designs. *Econometrica*, 82(6), pp.2295-2326.
- [9] Daniel, K., Grinblatt, M., Titman, S. and Wermers, R., 1997. Measuring mutual fund performance with characteristic - based benchmarks. *The Journal of finance*, 52(3), pp.1035-1058.
- [10] Daniel, K., Hirshleifer, D. and Subrahmanyam, A., 1998. Investor psychology and security market under - and overreactions. *the Journal of Finance*, 53(6), pp.1839-1885.
- [11] Engelberg, J., Sasseville, C. and Williams, J., 2012. Market madness? The case of mad money. *Management Science*, 58(2), pp.351-364.
- [12] Fang, L. and Peress, J., 2009. Media coverage and the cross - section of stock returns. *The Journal of Finance*, 64(5), pp.2023-2052.

- [13] Frydman, C. and Wang, B., 2020. The impact of salience on investor behavior: Evidence from a natural experiment. *The Journal of Finance*, 75(1), pp.229-276.
- [14] Gervais, S., Kaniel, R. and Mingelgrin, D.H., 2001. The high - volume return premium. *The Journal of Finance*, 56(3), pp.877-919.
- [15] Grinblatt, M. and Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. *The Journal of Finance*, 64(2), pp.549-578.
- [16] Grossman, S.J. and Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. *The American economic review*, 70(3), pp.393-408.
- [17] Grullon, G., Kanatas, G. and Weston, J.P., 2004. Advertising, breadth of ownership, and liquidity. *The Review of Financial Studies*, 17(2), pp.439-461.
- [18] Hirshleifer, D. and Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. *Journal of accounting and economics*, 36(1-3), pp.337-386.
- [19] Hong, H. and Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, 54(6), pp.2143-2184.
- [20] Imbens, G. and Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3), pp.933-959.
- [21] Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3), pp.881-898.
- [22] Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1), pp.65-91.
- [23] Kahneman, D., 1973. *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- [24] Kahneman, D. and Tversky, A., 1980. Prospect theory. *Econometrica*, 12.
- [25] Kaniel, R., Liu, S., Saar, G. and Titman, S., 2012. Individual investor trading and return patterns around earnings announcements. *The Journal of Finance*, 67(2), pp.639-680.
- [26] Kaniel, R. and Parham, R., 2017. WSJ Category Kings—The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, 123(2), pp.337-356.
- [27] Kelley, E.K. and Tetlock, P.C., 2013. How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, 68(3), pp.1229-1265.

- [28] Lee, C.M. and Radhakrishna, B., 2000. Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets*, 3(2), pp.83-111.
- [29] Lee, C.M. and Ready, M.J., 1991. Inferring trade direction from intraday data. *The Journal of Finance*, 46(2), pp.733-746.
- [30] Lee, D.S. and Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2), pp.281-355.
- [31] Li, J.J., Massa, M., Zhang, H. and Zhang, J., 2019. Air pollution, behavioral bias, and the disposition effect in China. *Journal of Financial Economics*.
- [32] Li, X., Shen, D. and Zhang, W., 2018. Do Chinese internet stock message boards convey firm-specific information?. *Pacific-Basin Finance Journal*, 49, pp.1-14.
- [33] Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4), pp.587-615.
- [34] Lou, D., 2014. Attracting investor attention through advertising. *The Review of Financial Studies*, 27(6), pp.1797-1829.
- [35] Markowitz, H., 1959. Portfolio selection: Efficient diversification of investments (Vol. 16). New York: John Wiley.
- [36] Massa, M. and Simonov, A., 2006. Hedging, familiarity and portfolio choice. *The Review of Financial Studies*, 19(2), pp.633-685.
- [37] Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information. *The journal of finance*, 42(3), pp.483-510.
- [38] Moray, N., 1970. Towards a quantitative theory of attention. *Acta Psychologica*, 33, pp.111-117.
- [39] Mossin, J., 1966. Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pp.768-783.
- [40] Newey, W.K. and West, K.D., 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.
- [41] Odean, T., 1999. Do investors trade too much?. *American economic review*, 89(5), pp.1279-1298.
- [42] Peng, D., Rao, Y. and Wang, M., 2016. Do top 10 lists of daily stock returns attract investor attention? Evidence from a natural experiment. *International Review of Finance*, 16(4), pp.565-593.
- [43] Peng, L. and Xiong, W., 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), pp.563-602.

- [44] Seasholes, M.S. and Wu, G., 2007. Predictable behavior, profits, and attention. *Journal of Empirical Finance*, 14(5), pp.590-610.
- [45] Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), pp.425-442.
- [46] Shen, D., Li, X. and Zhang, W., 2017. Baidu news coverage and its impacts on order imbalance and large-size trade of Chinese stocks. *Finance Research Letters*, 23, pp.210-216.
- [47] Solso, R.L., 1979. Twenty-five years of recommended readings in psychology. *American Psychologist*, 34(8), p.703.
- [48] Tetlock, P.C., 2010. Does public financial news resolve asymmetric information?. *The Review of Financial Studies*, 23(9), pp.3520-3557.
- [49] Treynor, J.L., 1961. Toward a theory of market value of risky assets.
- [50] Tversky, A. and Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), pp.297-323.
- [51] Wang, B., 2017. Ranking and salience. Available at SSRN 2922350.
- [52] Zhang, W., Shen, D., Zhang, Y. and Xiong, X., 2013. Open source information, investor attention, and asset pricing. *Economic Modelling*, 33, pp.613-619.
- [53] Zheng, X. and Shen, D., 2019. The High-Volume Return Premium: Does it Really Exist in the Chinese Stock Market?. *Asia-Pacific Financial Markets*, pp.1-18.
- [54] 冯旭南.注意力影响投资者的股票交易行为吗?——来自“股票交易龙虎榜”的证据[J].*经济学(季刊)*,2017,16(01):255-274.
- [55] 高昕祎. 投资者有限关注对个股市场表现的影响[D].西南财经大学,2018.
- [56] 康恒. 基于有限注意的股票龙虎榜效应研究[D].山东财经大学,2016.
- [57] 刘杰,陈佳,刘力.投资者关注与市场反应——来自中国证券交易所交易公开信息的自然实验[J].*金融研究*,2019(11):189-206.
- [58] 刘杰,阮晨晗,朱英伦.交易公开信息中的“有影响力交易席位”研究[J].*证券市场导报*,2018(05):32-40.
- [59] 彭叠峰. 基于投资者关注的资产定价研究[D].中南大学,2011.
- [60] 饶育蕾,彭叠峰,成大超.媒体注意力会引起股票的异常收益吗?——来自中国股票市场的经验证据[J].*系统工程理论与实践*,2010,30(02):287-297.
- [61] 沈德华. 复杂信息环境、资产价格与金融决策[D].天津大学,2016.
- [62] 张永杰,张维,金曦,熊熊. 互联网知道的更多么?——网络开源信息对资产定价的影响. *系统工程理论与实践*,2011,(04):577-586.

附 录

附录一 我们如何构建 DGTW 指标？

在文中，我们汇报了依据 Daniel, 等（1997）年根据公司个体风险特征对所有在深交所交易的 A 股收益率进行 $5 \times 5 \times 5$ 共计 125 个细分投资组合的分组的方法。相较于 Hirshleifer et al.,（2009）的 5×5 分类方法，加入动量特征的三组分类方法更加被广为使用（如冯旭南，2014，2016；Daniel, 等，1997），更能够反映公司在财务和交易方面的特征。对此，我们首先选取已经上市两年的股票，并且保留其中 ST、*ST 股票，因为他们也反映了一类公司的风险特征。此外，我们将这 $5 \times 5 \times 5$ 的分组日期选在每年的 6 月 31 号，按年度重新分组。因为中国的年报强制披露定在上一个会计年度结束之日起 4 个月内，这意味着对于绝大部分公司在 6 月份时应该就能获得上一年度的财务数据，而这也是最初 DGTW 的基准分组方法规定的日期。

在选出的所有股票中，我们首先依据这些股票上一年度末的流通市值进行从小到大分组，共计 5 组。接下来，我们把其中每一细分组的股票进一步按照账面市值比从小到大划分出 5 组，账面市值比的计算方式类似 Fama 和 French（1993）年的方式，由每一只股票上一年年末的账面市值除以其上一年年末最后一个交易日的市值，如果这只股票停牌了，那么就按照它停牌前的市值来计算。这样，我们就得到了 5×5 共计 25 个小组。最后，我们按照每只股票的动量对每一个已经做过细分后、进一步细分的组别从小到大划分出 5 组。有关动量的计算方法类似 Jegadeesh（1990），我们计算从上一年度 5 月份一直到这一年度的 4 月份股票的考虑现金红利再投资的 12 个月累计收益率，跳过最近的 1 个月份（即计算从前 13 个月直到上上个月的累计收益率），以此避免反转效应。最终我们能得到每年 $5 \times 5 \times 5$ 共计 125 个细分组合，对他们按市值进行加权计算收益率就是我们想要的平均收益率。

值得一提的是，在用这种方法计算 CAR 时，如果龙虎榜事件发生在 6 月底左右时，我们会根据之前定义的分类方法，在 7 月第一个交易日重新分类这只股票的风险组别，并根据新的基准收益继续计算超额收益。

此外，我们在未汇报的结果中使用经市场风险调整的收益率。对于市场收益，我们采用 3 个月的 shibor 利率为基准。经重复试验后，我们的主要结果仍然稳健。

附录二 我们的 RDD 模型稳健吗？

1) 全局参数估计

一般来说，RDD 实验有参数的和非参数形式 2 种，在使用参数方法估计断点处处理效应时，对其中基于驱动变量的多项式估计往往会出现多项式拟合优度不足或者过拟合的问题。在正文中我们使用非参数模型对原始数据进行了估计，在未汇报的全局参数估计中，我们通过设定如下的线性回归式来验证断点处的效应是否显著。

$$\begin{aligned} Flowchg_{i,q,t+1,F} &= \alpha + \beta \times (rank_{i,t,adj} - cut\ point) + \delta \times D_i \\ &+ \gamma \times f(cut\ point - rank_{i,t,adj}) \times D_i + \varepsilon_i \end{aligned} \quad (1)$$

其中，绝大部分变量设定都和正文中的（6）式相同，除了 $f(cut\ point - rank_{i,t,adj})$ 代表“经调整的排名”的多项式，多项式的具体设定有：线性多项式、二次多项式（即经调整的排名的 2 次方）⁹。在对 1 式的估计中，我们通过比较 AIC 准则发现，一次项（线性）的参数估计 AIC 值最小，模型估计最优。在对于 1 阶驱动变量的回归中，我发现 δ 仍然显著，这证明了断点处的平均处理效应存在。

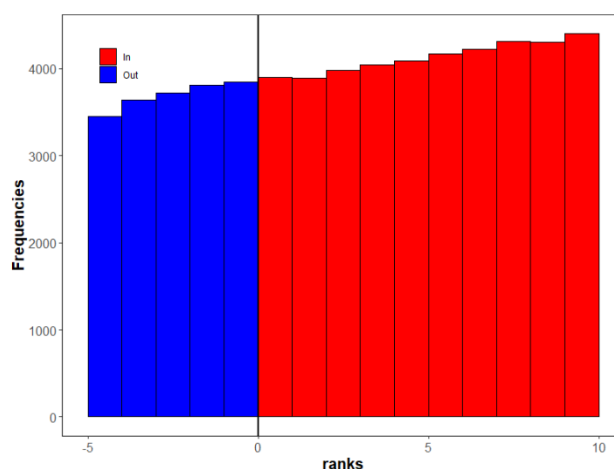
⁹ 我们只选取了一次和二次项进行了参数估计，因为根据 Gelman and Imbens（2014），全样本参数估计在高次项估计时会将过多权重赋给取值较大的配置变量；此外，断点处的处理效应估计值对全参数范围内的高阶多项式阶数十分敏感。

2) 对于非参数估计的稳健性检验

一个好的 RDD 实验必须要有图表展示（Lee and Lemieux, 2010），这也是我们在之前所做的。我们在附录进一步通过展示原始数据结构，不同断点处的 LLR 模型回归系数来检验实验的稳健性。

首先，就原始数据来看，频次十分稳定，在断点处并未有显著的跳跃。值得一提的是，这里的数据是经过在 0.05 分位上对极端值清洗后的结果，也是我们在进行 RDD 回归时使用的数据。

图 1 断点前后股票数量



此外，McCrary 密度检验（McCrary, 2008）得到检验 p 值为 0.92，不拒绝断点处密度不一致的原假设。在未汇报的密度图中，断点前后密度函数非常平滑，这意味着数据密度并未发生了太大变化，且均保持在置信区间之内。

之后检验是一个试错实验，即把断点设置在非零处，来检验此时 LLR 回归式中 LATE 系数是否显著。如果一个 RDD 是稳健的，那么断点应该且仅应该在外生给定处显著（即 0 处）。在没有加入控制变量的 RDD 实验中，当我们设置断点为 1 时，结果在由中单驱动的资金流入和由大单驱动的资金流出这两次实验中显著，而在其他的所有断点处均不显著。这不能说是一个令人满意的结果，不过这并不影响整个实验整体的信度。同样，在加入了控制变量的 RDD 实验中，结果也基本类似。

表 1 调整断点之后的 RDD 稳健性检验

		Cut point at				
		0	1	2	3	4
RDD	Small inflow	-0.23 (-3.11)	0.12 (1.46)	-0.01 (-0.27)	0.17 (0.26)	0.03 (0.51)
	Small outflow	-0.20 (-3.12)	0.10 (1.30)	-0.00 (-0.16)	0.00 (0.13)	0.02 (0.41)
	Medium inflow	-0.18 (-2.28)	0.10 (-1.85)	-0.02 (-0.34)	0.03 (0.47)	0.01 (0.22)
	Medium outflow	-0.24 (-3.08)	0.14 (1.5)	-0.03 (-0.55)	0.02 (0.39)	0.02 (0.29)
	Large inflow	-0.46 (-3.29)	0.23 (1.47)	0.01 (0.07)	-0.00 (-0.07)	-0.03 (-0.38)
	Large outflow	-0.39 (-2.71)	0.25 (1.88)	-0.11 (-0.86)	0.13 (1.06)	-0.10 (-0.88)
RDD w/c	Small inflow	-0.26 (-3.52)	0.13 (1.41)	-0.01 (-0.13)	0.00 (0.05)	0.04 (0.68)
	Small outflow	-0.22 (-3.47)	0.11 (1.46)	-0.00 (-0.05)	-0.00 (-0.00)	0.02 (0.53)
	Medium inflow	-0.21 (-2.64)	0.11 (1.71)	-0.01 (-0.19)	0.01 (0.27)	0.02 (0.36)
	Medium outflow	-0.26 (-3.45)	0.15 (1.55)	-0.03 (-0.44)	0.01 (0.25)	0.02 (0.39)
	Large inflow	-0.51 (-3.71)	0.24 (1.33)	0.02 (0.22)	-0.01 (-0.18)	-0.03 (-0.40)
	Large outflow	-0.45 (-3.18)	0.27 (2.04)	-0.09 (-0.71)	0.10 (0.88)	-0.08 (-0.77)

注：我们通过调整断点的位置对正文中 RDD 实验进行了稳健性检验。正文中，我们的断点是 0，在这里我们分别将断点设置在 0, 1, 2, 3, 4，对正文中的 LLR 式进行了估计，并汇报代表局部平均处理效应的虚拟变量的系数。表格上半部分是对正文中（6）式的估计，下半部分是加入控制变量后的估计。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

最后，我们还基于三个基本面变量进行了有效性检验（Validity Test）。参照 Lee and Lemieux（2010），我们将正文中（6）式中的结果变量从资金流变化比率替换成每股收益、市值和净资产，其余针对 LLR 的设定，如驱动变量、估计核函数和带宽估计方法等均不发生变化。如果我们的模型设定在断点处真的是随机的话，这一稳健性估计应该不显著。

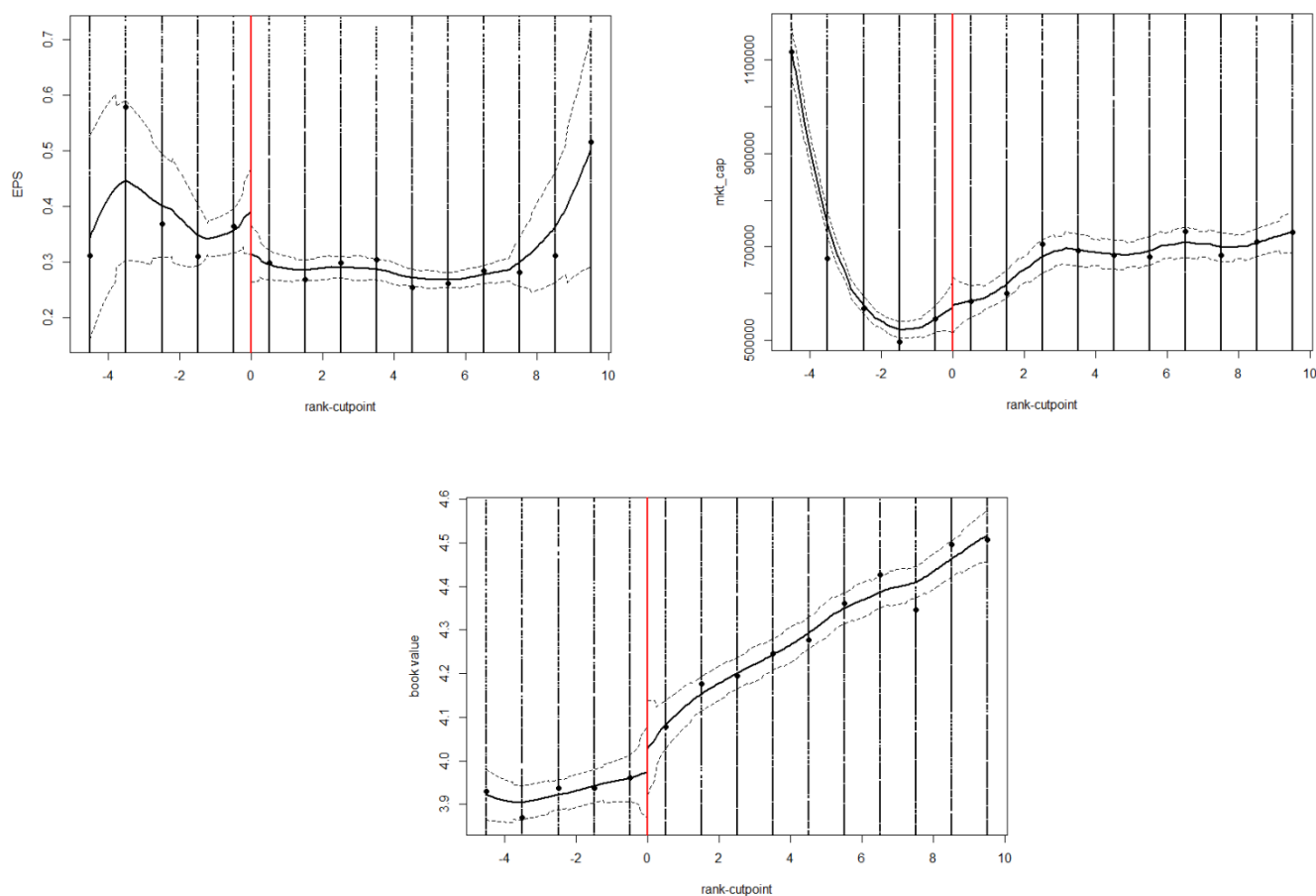
表 2 有效性检验

	EPS		Mkt_cap		Book value	
	plain	w/c	plain	w/c	plain	w/c
δ	-0.07 (-1.338)	-0.07 (-1.31)	4891 (0.11)	44535 (1.22)	0.05 (0.65)	0.11 (1.55)
Number	59793	59793	59793	59793	59793	59793

注：我们针对正文中（6）式，重新利用基本面变量进行了估计。我们将原先的资金流变化比率替换成了代表个股基本面的变量，即每股收益、流通市值和账面市值。表中汇报了没加控制变量（plain LLR）和加了控制变量（LLR with control, w/c）时，代表个股是否上榜的虚拟变量系数 δ 。表中，括号内是 t 值，且所有系数均显著并且以加粗的方式标出。

在表 2 中，我们首先针对每一个基本变量，分别在加了控制变量和没加控制变量时进行了估计，6 个回归结果同时也在图 2 中展示。不管是单从回归系数上还是图像上来看，断点处由于股票上榜而导致的高曝光度和公司基本面并没有显著的因果关系，这也支持了我们这个 RDD 实验是稳健的论证。

图 2 使用基本面变量作为驱动变量时的 RDD 实验



注：从左上至右下分别是我们将结果变量替换为：每股盈余、流通市值和净资产后，对 LLR 回归所得的结果。

致 谢

这篇文章并不是 idea 或者 model 先行，而是数据先行的。最初只想着利用龙虎榜的数据做一个不大不小的实证，然而在有了那个 RDD 实验之后，思路也进一步清晰了起来，发现有更多可以深挖的内容，直到最后看到 Hong and Stein (1999) 的那个模型，才最终在逻辑上自洽。回头望去，整篇文章的各种模型和方法仍然是拙劣的拼凑和模仿，仍然有许多不足和提高的空间，真正的创新和对于学术界的贡献更是可以忽略不计。

在成文过程中对我帮助最大的，也是我在过去四年来最感谢的，是沈德华老师。除了他在学术方面对我的巨大帮助之外，我最要感激的是德华老师对我的批评和指正，帮我逐步走上正路。此外，我还想特别感谢的是张永杰老师，没有永杰老师给我提供的机会和帮助，我绝对不可能静下心来文献、做 model。也正是永杰老师在一次午餐时分享的对于做学术的“三个目标”，为我指明了未来的学术道路。在我院的学习时光中，我还十分幸运地遇到了许多有学识、负责任的好老师，像冯绪老师、曹桂全老师，苏云鹏老师、王冠英老师、余凤燕老师等，是您帮助我从各方面打下了金融学的基础知识。

在过去 4 年中，每一个老师，每一位同学都或多或少地在交往中指出了我性格上的弱点，对此，我深表感谢。我希望在接下来的 5 年中，能通过反复自省以找到内心的平静。每日三省吾身，我又狂妄自满了吗？又浮躁焦虑了吗？我憧憬着未来某一天突然发现，我能压制住自己内心的欲望，站上更高地台阶。

最后，在致谢和自省的同时，我也希望自己未来能踏踏实实地继续把学术做下去，同时仍然能做一个好人。大学四年，失去了很多，却收获了更多，在未来漫长的岁月中，我希望我关心的人和我自己将来能永远快快乐乐。

郑行健¹⁰

2020 年 5 月 12 日夜于上海植物园

¹⁰ 最后，请允许我遵循师门惯例，在致谢部分加一个小小的脚注：虽然我体型适中，但是在全文中先后注了不少水，惭愧啊。一想到我之后的学术生涯中还必须要再灌上几十倍的水，更是无地自容。两位老师还请见谅。