

硕士专业
学位论文

中国股票市场投资者学习的影响研究

蓝大镇

廣西大學

二〇二四年六月

分类号 F832

学校代码 10593

密级 公开

学号 2102391034

硕士专业学位论文

中国股票市场投资者学习的影响研究

Effect of Investor Learning in the China Stock Market

作者姓名: 蓝大镇

指导教师: 谢军副教授

合作导师: 薛洪言

专业名称: 金融

研究方向: 金融市场与金融投资

所在学院: 经济学院

论文答辩日期 2024 年 5 月 28 日 学位授予日期 2024 年 6 月 19 日

答辩委员会主席 曾海舰 教授

中国股票市场投资者学习的影响研究

摘 要

在当前数据量和信息量快速增长的时代，投资者需要主动选择学习哪些信息并进行投资决策，投资者主动交易驱动了股票市场的价格发现，因此，理解投资者学习在股票市场中的影响是至关重要的。

本文基于一个已有的信息选择和投资决策的理论模型使用 2000 年至 2021 年的中国股票市场数据估计了个股的学习指数，个股被学习得越透彻，学习指数值越高。首先，本文检验了学习指数对股票收益的预测能力，并对学习指数的预测能力进行分解。其次，本文研究了学习指数与错误定价之间的关系，分析了这种关系产生的原因。最后，本文基于个股的学习指数构建了一个学习因子，考察了学习因子在中国股票市场的定价能力。

本文的实证研究发现：首先，学习指数对股票横截面收益具有负向预测能力，学习指数越高的股票未来的收益越低，并且这种负向预测能力是长期存在且随时间不断衰减的，两阶段分解的结果表明这种负向预测能力不能被信息流相关的变量所解释。其次，在学习指数高的股票中，股票的错误定价更小，这种现象在投资者情绪较高的时期更显著的原因可能是股票被高估的情况通常在投资者情绪较高的时期出现；在被低估的股票中学习指数对股票横截面收益的预测能力更强，这可能是因为在被高估的股票中存在更强的套利限制。最后，本

文构建的学习因子无法被其他多因子模型所解释,相比于其他多因子模型,使用学习因子构建的五因子模型张成的切点投资组合具有最高的夏普比率,能够解释的市场异象也最多。

关键词: 投资者学习 市场异象 错误定价 中国市场 因子定价

Effect of Investor Learning in the China Stock Market

ABSTRACT

In the current era of rapid growth in the amount of data and information, investors need to actively choose what information to learn and make investment decisions, and investor active trading drives the price discovery in the stock market. Therefore, it is crucial to understand the impact of investor learning in the stock market.

Based on an existing theoretical model of information choice and investment decision, this paper estimates a learning index of individual stocks using China stock market data from 2000 to 2021. The more thoroughly a stock has been learned, the higher the learning index value. First, this paper examines the ability of learning index to predict stock returns, and decomposes the ability of prediction of the learning index. Secondly, this paper studies the relationship between learning index and mispricing, and discusses the reasons for this relationship. Finally, this paper constructs a learning factor based on the learning index of individual stocks, and examines the pricing ability of the learning factor in the China stock market.

The empirical research in this paper finds that: First, the learning index has a negative predictive ability for cross-sectional stock returns, and the higher the learning index is, the lower the future returns of stocks, and this negative predictive ability exists for a long time and decreases with time. The results of two-stage decomposition show that this negative predictive ability cannot be explained by the variables related to information flow. Secondly, in the stocks with high learning index, the mispricing of stocks is smaller. The reason why this phenomenon is more

significant in the period of high investor sentiment may be that the situation of overvalued stocks usually occurs in the period of high investor sentiment. The learning index has a stronger ability to predict cross-sectional returns of stocks in undervalued stocks, which may be because there is a stronger arbitrage limit in overvalued stocks. Finally, the learning factor constructed in this paper cannot be explained by other multi-factor models. Compared with other multi-factor models, the tangential portfolio spanned by the five-factor model constructed using learning factors has the highest Sharpe ratio and can explain the largest number of market anomalies.

KEY WORDS: Investor learning; Market anomaly; Mispricing; China Stock Market; Factor pricing

目 录

第一章 绪论.....	1
1.1 研究背景、研究目的与研究意义	1
1.1.1 研究背景.....	1
1.1.2 研究目的.....	2
1.1.3 研究意义.....	2
1.2 研究思路、研究内容与研究方法	3
1.2.1 研究思路.....	3
1.2.2 研究内容.....	5
1.2.3 研究方法.....	5
1.3 创新点与不足.....	6
1.3.1 研究的创新点.....	6
1.3.2 研究的不足之处.....	6
第二章 文献综述	8
2.1 投资者学习的相关研究.....	8
2.2 错误定价的相关研究.....	9
2.3 资产定价模型的相关研究	11
2.4 文献评述.....	13
第三章 理论基础与学习指数构建	14
3.1 理论基础.....	14
3.2 学习指数构建.....	17
第四章 投资者学习与股票横截面收益	19
4.1 样本数据和变量说明.....	19
4.1.1 数据来源与样本选择.....	19
4.1.2 变量说明	19
4.2 数据描述性统计特征.....	21
4.3 学习指数对股票收益的预测能力	24
4.3.1 单变量组合排序.....	24
4.3.2 双变量组合排序.....	27
4.3.3 Fama-Mac Beth 回归.....	34
4.3.4 长期预测能力.....	38
4.3.5 样本外检验.....	42
4.4 学习指数对股票收益预测能力的来源	43
4.4.1 学习指数与信息流变量的关系.....	43
4.4.2 两阶段分解.....	45

4.5 稳健性检验.....	48
4.5.1 更换多因子模型.....	49
4.5.2 更换其他的排序组数.....	50
4.5.3 增加一些额外的控制变量.....	51
4.5.4 更换两阶段分解中使用的信息变量.....	54
第五章 学习指数与错误定价的关系	55
5.1 数据来源和变量说明.....	55
5.1.1 数据来源与样本选择.....	55
5.1.2 变量说明.....	55
5.2 数据描述性统计特征.....	56
5.3 学习指数与错误定价.....	57
5.3.1 双变量排序.....	57
5.3.2 长期持续分析.....	60
5.3.3 子样本分析.....	63
5.3.4 套利限制的影响.....	64
5.4 稳健性检验.....	67
5.4.1 更换错误定价的衡量指标.....	67
5.4.2 其他公司特征正交化的学习指数.....	71
第六章 学习因子及其定价能力检验	73
6.1 数据来源和变量说明.....	73
6.2 CH4-LI 五因子模型.....	73
6.2.1 学习因子的构建.....	73
6.2.2 因子描述性统计.....	74
6.2.3 因子的冗余性检验.....	77
6.3 模型定价能力比较.....	79
6.3.1 夏普比率测试.....	79
6.3.2 解释市场异象.....	80
6.4 稳健性检验.....	85
6.4.1 Bootstrap 方法	85
6.4.2 剔除“壳污染”后的股票	87
第七章 结论、启示与展望	89
7.1 研究结论.....	89
7.2 研究启示.....	90
7.3 研究展望.....	92
参考文献.....	94

第一章 绪论

1.1 研究背景、研究目的与研究意义

1.1.1 研究背景

在过去的十年间,数据量的爆炸性增长和人工智能的快速发展使得数据成为当代经济增长的重要投入品,很多企业比如百度、阿里巴巴、腾讯、美团、滴滴等互联网平台的头部企业早已将数据作为重要的投入品进行生产以获取经济价值,很多行业的头部企业,如智能手机、新能源汽车、软件开发等,也使用数据和人工智能提升自己产品的竞争力,许多国家都将大数据和人工智能作为国家发展战略的重要组成部分。同时,信息选择和信息生产是理解人类经济和金融体系中最重要维度之一,在大数据和人工智能时代,投资者可以轻松地获取大量的金融数据和信息,并且利用强大的分析工具对信息进行处理,加深对资产价格的理解,由此进行投资决策。理解投资者学习行为对资产价格的影响,有助于帮助我们更深入地了解金融市场。

近年来,中国股票市场经历了快速的发展和深刻的变革,中国股票市场的成交量和总市值连年攀升,已经成为全世界最重要的股票交易场所之一。随着中国经济的持续增长和金融市场改革的推进,越来越多的投资者涌入股票市场,中国股票市场的独有特点为投资者学习行为的研究提供了丰富的材料。首先,中国股票市场的信息披露制度相对于发达国家而言不够完善和成熟,信息不对称的程度更高,投资者需要在信息环境非常复杂的情况下选择和学习信息并进行投资决策。其次,中国股票市场的投资者结构不同于成熟的股票市场,散户投资者在中国股票市场中占比很高,散户投资者和机构投资者具有不同的学习偏好和行为特征,对股票市场的影响也各不相同。因此,中国市场与发达国家成熟的股票市场的投资者学习不尽相同,在中国市场研究投资者学习行为,有助于我们了解散户投资者的学习行为。最后,中国股票市场发展时间较短,相对发达国家而言存在更多的套利限制,投资者获取和处理信息后可能会因为流动性不足、做空限制等套利限制而无法构建投资组合。因此,在中国市场研究投资者学习对股票收益的影响有助于帮助人们更深入地理解信息选择与投资决策模型。

此外,对中国股票市场投资者学习行为的研究,可以为金融市场的监管机构、学术界和金融业界提供参考与指导。对于金融市场的监管机构而言,可以参考投资者学习对股票预期收益的影响,制定更有效的政策和规定,强化信息披露的制度,帮助投资者了解更多的信息,减少信息不对称带来的股票错误定价,保护投

投资者权益,维护市场的公平和稳定。对于学术界而言,信息选择已经是金融经济学领域中金融市场微观结构研究的一个重要组成部分,研究中国股票市场中投资者学习产生的影响,可以对相关的文献进行有益的补充,填补中国市场相关研究的空白。对于金融业界而言,研究中国股票市场中投资者学习的影响,有助于帮助机构投资者理解市场,根据研究成果分析投资者行为构建投资策略,获取超额收益。

1.1.2 研究目的

本文以中国股票市场的上市公司作为主要研究对象,使用实证资产定价的系列研究方法,研究投资者学习对股票横截面收益的影响,并依据投资者学习与股票横截面收益之间的关系探索其在中国股票市场的定价能力,可以帮助市场的参与者意识到信息处理的重要性,引导市场投资者改变投资观念,通过深入研究市场信息,构建独特的投资组合获取超额收益,消除股票的错误定价。具体而言,本文的研究目的主要有以下三点:

(1) 检验投资者学习与股票横截面预期收益之间的关系。本文在中国股票市场估计了一个学习指数,学习指数代表了在截面上投资者对个股的学习程度的大小。本文检验了学习指数对股票横截面预期收益的预测能力,并对这种预测能力进行分解分析。

(2) 分析学习指数与错误定价之间的关系。还没有研究对学习指数与错误定价之间的关系进行讨论,所以本文研究了学习指数对错误定价的影响,同时考察学习指数与股票预期收益的关系在被高估和被低估的股票之间是否存在差异,并结合投资者情绪和套利限制深入分析了这种现象产生的原因。

(3) 探索学习因子在中国股票市场的定价能力。投资者学习可以对股票的未來收益产生影响,这意味着投资者学习可以驱动股票价格共同运动,因此,本文在中国股票市场使用个股的学习指数在构建了一个学习因子,使用学习因子与其他多因子模型结合构建了一个新的多因子模型,可以增强模型在中国市场的定价能力。

1.1.3 研究意义

本文研究了个股的投资者学习,丰富了现有关于中国股票市场个股投资者的学习的相关文献。本文的研究结果可以帮助投资者和金融监管机构加深对中国股票市场的投资者学习及其对股票横截面收益的影响,在当前大数据和人工智能的时代下,理解投资者学习的影响具有重要的理论意义和现实意义。

在理论层面,有关投资者学习的相关文献非常丰富,很多学者都构建了相关的理论模型探讨信息选择与投资决策,但是很少有相关的实证研究讨论投资者学习与股票横截面收益之间的关系。首先,本文参考Cao等(2021)提出的方法,在中国股票市场估计了一个学习指数,在理性预期一般均衡模型中学习指数代表

了个股被投资者学习的程度。在此基础上,本文研究在中国股票市场中,学习指数对股票横截面收益的预测能力,对学习指数预测能力进行分解,这为投资者学习对股票收益影响提供了新的认知,拓展了股票收益预测变量的相关研究。其次,本文对学习指数与错误定价之间的关系进行了探讨,发现投资者学习高的股票错误定价更小,学习指数与股票预期收益之间的负相关性在被低估的股票中更强,这在一定程度上丰富了投资者学习及其影响的内涵,基于投资者情绪和套利限制的分析使得学习指数对股票收益影响的机理更为清晰。最后,对学习因子在中国股票市场的定价能力进行了研究,将学习因子与其他因子结合,构建一个新的具有较强定价能力的多因子模型,对模型进行夏普比率测试,并考察模型在中国股票市场解释市场异象的能力,对现有实证资产定价的多因子模型相关的研究进行了有益的补充。

在现实层面,首先,从普通投资者的角度来说,本研究有助于投资者加深对信息选择与股票收益之间的关系的理解,帮助投资者认识到获取信息和深入学习资产的重要性,从而促使投资者寻找独特的信息进行学习,进而构建投资组合以获得更高的收益。其次,对于金融监管机构而言,当前中国股票市场存在着大量的散户投资者,同时各项制度尚不完善,金融监管机构需要加强对上市公司信息披露的监管,并对财务造假的行为进行严厉打击,从而改善市场中存在的信息不对称的情况,只有这样,才能引导投资者摆脱炒作投机的心态,转向挖掘独特的信息寻找交易机会的投资心态,从而促进股票市场的健康发展。因此,研究投资者学习对股票横截面收益的影响,可以为监管机构提供丰富的证据,帮助监管机构制定更加合理的信息披露制度,推进融资融券制度的发展,从而推动金融体系的改革。最后,从量化投资者的角度而言,学习指数是一个可以预测股票收益的变量,量化投资者可以使用学习指数在股票市场上构建量化投资策略,做多学习指数较低的股票,做空学习指数较高的股票,得到一个零成本的多空组合,获取超额收益,同时改善最终组合的风险收益特征。

1.2 研究思路、研究内容与研究方法

1.2.1 研究思路

本文主要研究中国市场学习指数对股票横截面收益的影响,从学习指数对股票横截面收益的预测、学习指数与错误定价之间的关系、学习因子的定价应用三个方面进行了具体的研究,三个主体内容为:

第一部分内容将学习指数作为股票横截面收益的预测变量进行研究,主要分为对股票横截面收益预测和预测能力分解分析两个部分,检验学习指数对股票横截面收益的预测能力。

第二部分研究学习指数与错误定价之间的关系，主要分为学习指数与错误定价的相关关系和这种现象产生的原因两个部分，验证投资者学习缩小股票错误定价的能力。

第三部分的内容首先验证学习因子的有效性，随后拓展了 Liu 等（2019）的四因子模型，得到一个五因子模型，通过比较模型和其他模型的张成的切点组合的夏普比率的大小，以及对市场异象解释能力，从而比较五因子模型和其他模型的定价能力。

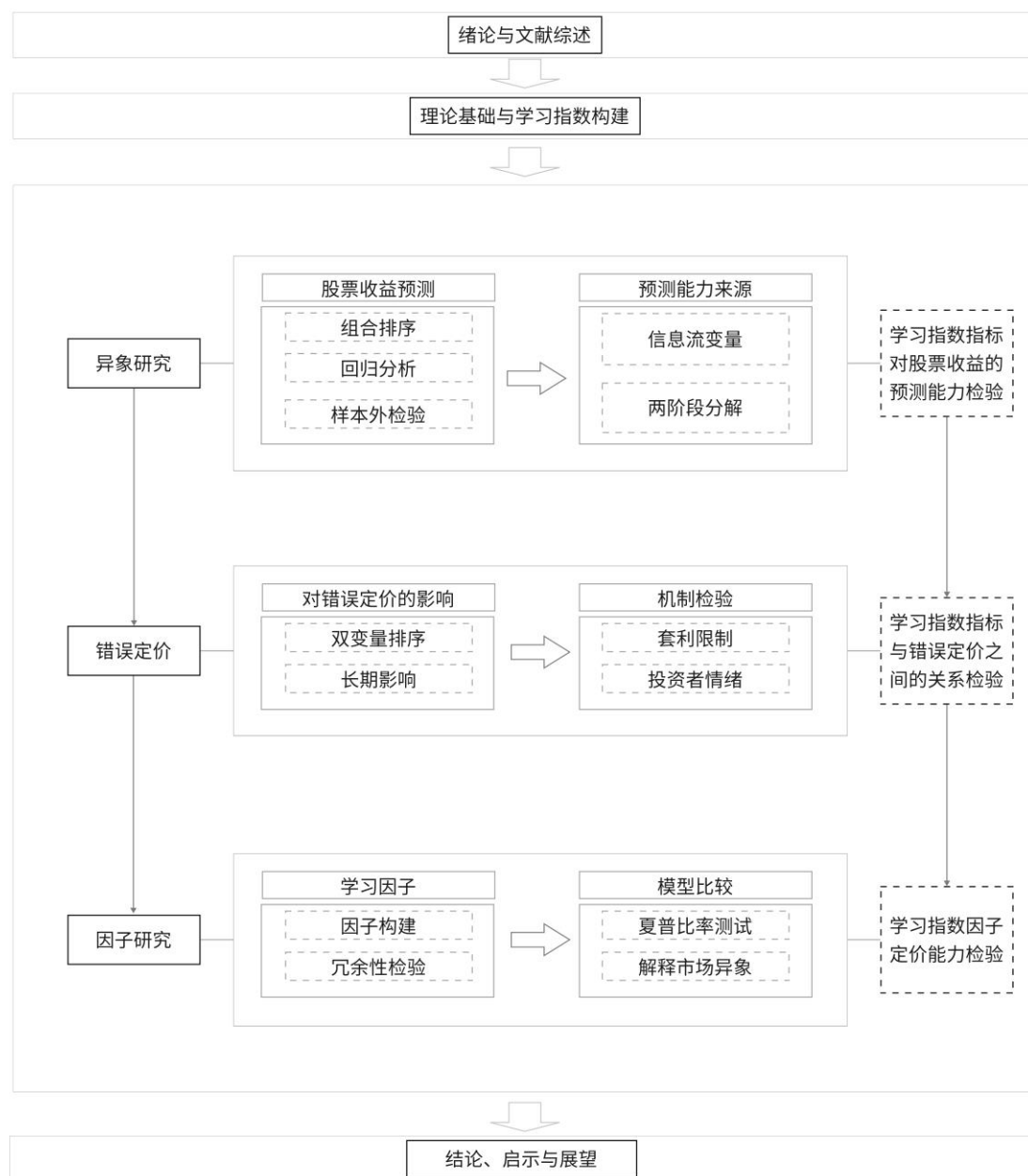


图 1-1 研究思路图

1.2.2 研究内容

结合本文的研究思路，研究内容主要分为七个章节。

第一章：绪论。首先，介绍了本文的研究背景、研究目的和研究意义；其次，介绍了本文的研究思路、研究内容和研究方法；最后探讨了本文的创新点和不足之处。

第二章：文献综述。第一，整理了现有文献关于投资者学习的研究，包括理论研究和实证研究。第二，总结了错误定价相关研究的文献。第三，梳理了现有的实证资产定价多因子模型的相关研究。最后，对相关的文献进行评述，指出文献存在的不足并提出本文研究的方向。

第三章：理论基础和学习指数构建。对投资者学习背后的理论模型进行概述，并以此引出学习指数指标的构建方法，使用中国 A 股市场的股票数据，构建学习指数指标。

第四章：投资者学习对横截面收益预测能力的实证分析。使用单变量组合排序、双变量组合排序、Fama-MacBeth 回归、样本外检验等方法，检验了学习指数与股票预期收益的关系，同时分析学习指数对股票收益的长期预测能力。对学习指数的预测能力进行分解，分析学习指数与 A 股市场常见的信息流相关变量的相关性，使用两阶段分解探讨与学习指数相关的成分对预测能力的贡献大小。

第五章：研究学习指数与错误定价的关系。检验学习指数在被高估和被低估的股票中的对股票收益的预测能力是否不同，并探讨投资者学习是否有助于减少股票的错误定价。同时，还考察了投资者情绪和套利限制在这一关系中扮演的角色。

第六章：使用学习指数构建定价因子进行应用。首先对学习因子进行冗余性分析，以确保学习因子无法被其他多因子模型所解释。随后将学习因子与中国市场的四因子模型结合构建一个新的五因子模型，考察五因子模型在中国市场的定价能力，具体方法包括夏普比率测试和解释市场异象。

第七章：结论、启示与展望。首先，总结本文主要的实证结果，阐述本文的主要发现。其次，从投资者的角度给出投资建议，从学术界的角度给出研究建议，从监管机构的角度给出政策建议。最后，结合本文的不足之处，并分析未来可能存在的研究方向。

1.2.3 研究方法

本文拟采用的研究方法主要是实证资产定价领域常用的研究方法，包括统计分析法、单变量排序法、双变量排序法、Fama-MacBeth 回归分析法、样本外检验、Fama 和 French（1993）的双变量独立排序因子构造法、数值模拟分析以及 Boehmer 等（2021）的两阶段分解法。此外，本文还使用了一些传统的计量经济学方法，包括 OLS 回归和 WLS 回归。使用这些常用的分析方法，可以确保本文

对于中国股票市场学习指数影响研究的稳健性。

1.3 创新点与不足

1.3.1 研究的创新点

现有文献关于中国股票市场学习指数的相关研究相对较少,尤其是学习指数应用拓展方面的相关研究更是匮乏。本文研究了学习指数对股票横截面收益的负向预测能力,并对预测能力进行分解,考察了学习指数与错误定价的关系,探索学习因子在中国股票市场的定价能力。本文的创新点主要集中在以下三个方面。

(1) 检验了学习指数对股票横截面收益的预测能力,并对其进行分解。本文在中国股票市场估计了一个学习指数,检验学习指数对股票横截面收益的预测能力,并对学习指数预测能力进行分解,分析了各个成分的贡献。此前还没有研究对学习指数的预测能力进行分解,并探讨各个成分对学习指数预测能力的贡献程度,本文的分解分析丰富了投资者学习的相关研究,也为中国股票市场横截面收益预测的相关文献增加了新的视角。

(2) 研究了学习指数与错误定价之间的关系。在先前的文献中,尚未对学习指数与错误定价之间的关系进行过研究。本文研究了学习指数与错误定价之间的关系,发现学习指数高的股票中错误定价更小,并且学习指数对股票收益的预测能力在被高估和被低估的股票之间存在差异。研究表明套利限制和投资者情绪可能在这种关系中扮演重要的角色,这一部分的研究为投资者学习和错误定价相关的文献做出了贡献。

(3) 研究了学习因子在中国市场的定价能力。到目前为止,还没有使用学习指数构建定价因子的研究。本文使用学习指数构建了一个学习因子,扩展了现有的多因子模型,提高了模型在中国股票市场的定价效率。中国股票市场中现有的多因子模型大多是直接依据美国市场的经验构造的,不适合中国股票市场。本文使用个股的学习指数在中国股票市场构建了一个学习因子,考察学习因子的冗余性,结果表明,学习因子不能被主流的多因子模型解释。本文将学习因子与 Liu 等(2019)提出的中国市场的四因子模型结合,构建了一个新的五因子模型,与其他的多因子模型相比,新的五因子模型在解释中国市场的市场异象方面效果最好,其张成的切点组合也具有最高的夏普比率,这一部分的研究对中国股票市场多因子模型和市场异象的相关研究做了有益的补充。

1.3.2 研究的不足之处

本文的研究也存在一些不足之处。

第一,本文研究的股票收益频率是实证资产定价领域常用的月度频率,学习指数对股票收益的预测能力是否存在于更低频的季度频率和年度频率中,或者是

否存在于更高频率的周度频率和日度频率中，甚至是否存在于日内的分钟频率中，这些问题值得深入思考和研究，使用不同的数据可能得出不同的结论，可能会发现投资者学习对市场产生不一样的影响。

第二，本文主要站在学术的角度对学习指数与股票横截面收益之间的关系进行研究，研究表明使用个股的学习指数可以构造一个多空组合获得超额收益。但是在实际业界投资中，要将一个研究转化为投资策略需要评估的内容还有很多，包括策略的夏普比率、换手率、交易滑点、交易换手率、边际投资收益、最大回撤等，这些都需要进一步的补充。同时，本文没有对学习指数多空组合在不同时期、不同市场状态的收益情况进行了评估。因子择时一直是业界关注的重点，如何通过择时方法避开学习指数表现差的时期，有助于获取更高的超额收益，这是未来可以进一步进行研究方向之一。

第二章 文献综述

本章首先对“投资者学习的相关研究”、“错误定价的相关研究”、“资产定价模型的相关研究”三个方向的文献进行梳理。随后对现有的相关研究进行评述,总结现有研究的局限性,并引出本文的实证研究问题,为后续章节的实证分析奠定基础。

2.1 投资者学习的相关研究

关于投资者学习的相关文献十分丰富,一类文献通过构建理论模型来解释一些经济现象,另一类文献则直接使用一些独特的数据集对投资者学习的行为及其影响进行实证研究。

在理论模型方面,有很多学者构建了模型描述投资者学习对资产价格产生的影响。Grossman 和 Stiglitz (1980) 构建了一个只有单个资产的信息获取理性预期均衡模型,投资者可以选择是否获取和信息相关的信号,随着更多投资者选择观察信号,信号的价值会下降。Verrecchia (1982) 扩展了 Grossman 和 Stiglitz (1980) 的研究,允许投资者可以选择信号的精确度。Admati (1985) 将模型扩展为具有多个风险资产和无限多投资者的模型,模型假设投资者被赋予异质性信息并根据信息选择投资组合。基于 Grossman 和 Stiglitz (1980) 的研究,Peress (2004) 利用股票市场中的信息成本对股票市场中的有限参与现象进行了解释,Peress (2010) 研究了风险分担和信息生产之间的权衡。Garcia 和 Vanden (2005) 则与 Kacperczyk 等 (2009) 利用信息获取建立了一个关于共同基金作用的理论模型。Kacperczyk 等 (2016) 在 Kacperczyk 等 (2009) 的基础上构建了一个共同的基金经理的注意力分配模型,这个模型可以决定投资组合的选择。Van Nieuwerburgh 和 Veldkamp (2010) 基于 Admati (1985) 求解的均衡资产价格,考虑投资者对信息的选择,提出一个信息选择与投资选择的理性预期一般均衡模型(以下简称 VNV 模型)。在 VNV 模型中,投资者可以自己选择获取哪些资产的信息,并根据模型的均衡结果定义了一个学习指数 (Learning Index), 投资者了解学习指数更高的资产的信息,可以带来更高的期望效用。投资者通过对与资产相关的信息的学习,可以降低资产收益的不确定性,而风险厌恶投资者更喜欢持有他们了解更多的资产,所以 VNV 模型对信息选择、风险和预期收益之间的关系进行预测: 资产信息的增加导致不确定性的减少和预期收益的降低。Mondria (2010) 放宽了 VNV 模型中投资者对风险因子结构的假设,投资者可以选择观察风险因子收益的线性组合作为私人信号。但是由于模型只考虑了两个资产,所以难以进行实证检验。

在实证方面,想要检验投资者学习相关理论的最好的办法就是使用某一个信息获取的变量进行相关的研究。Gargano 和 Rossi (2018)使用了一个新颖的经纪账户数据集,这个数据集包含了个人投资者在经纪网站上的网络活动数据,从投资者的角度研究信息获取与投资收益之间的关系。这个研究给投资者学习的相关研究提供了新的见解,但是这个数据集有样本周期过短的问题(只有 18 个月)。Biais 等(2010)认为,价格包含了与不知情投资者价值相关的信息,并发现基于事前信息的价格决定的投资组合优于被动指数。Banerjee (2011)提出了一个模型,该模型嵌套了理性预期和意见分歧,并发现投资者表现出理性的预期,并将他们的信念置于价格之上。Burlacu 等(2012)根据 Admati (1985)的工作,制定了一种信息精度和供应不确定性的衡量标准,并研究了其与预期收益之间的关系。Cao 等(2021)在 VNV 模型的基础上,实证估计了股票的学习指数,学习指数反映了股票市场上的投资者对股票的学习程度。同时,还发现学习指数对股票未来收益和波动率具有显著的预测能力,并且发现学习指数对股票横截面收益率具有长期预测能力,学习指数还与市场上的信息流代理变量相关,这些发现都有力验证了 VNV 模型的猜想。

大量的文献对理论模型进行了拓展研究,但是很少有研究对模型进行实证验证,实证研究的文献更多专注于使用市场上某些代表信息获取的数据和变量,研究投资者学习产生的影响。Cao 等(2021)基于 VNV 模型在美国市场估计了一个学习指数,并分析了学习指数对股票收益的预测能力。然而 Cao 等(2021)没有对学习指数的预测能力进行分解,研究学习指数的预测能力的每个成分对预测能力的贡献大小,本文使用一个两阶段分解的方法,可以分析学习指数的各个相关部分对预测能力的贡献大小,填补这一领域的空白。

2.2 错误定价的相关研究

有效市场假说认为能够影响价格的信息出现时,投资者会立刻根据信息进行交易,股票的价格会迅速反应,所以股票的价格反映了所有的已知信息,没有变量可以对股票收益进行预测,也就无法获得超额收益,市场中也就不会存在错误定价。但是大量的研究表明,有很多可以预测股票收益的异象变量,股票市场中可能存在着错误定价。Xie (2005)发现投资者会高估应计项目,导致股票的错误定价。Kakimoto 和 Akazawa (2007)发现流动性较差的股票普遍持续定价过高,这与分析师的分歧有关。Fairfield 和 Yohn (2011)发现投资者普遍高估了应计利润,最终也会反映到股票的价格上。同时,也有学者从不同的角度对预测股票收益的市场异象变量进行了解释。例如 Barberis 等(1998)与 Hong 和 Stein (1999)从投资者心理的角度给出了部分市场异象存在的解释、Shleifer 和

Vishny (1997) 则从市场微观结构的角度出发, 对部分市场异象存在的原因进行了解释。

对于错误定价的研究难点主要在于错误定价难以衡量。对于错误定价最直观的考虑就是计算企业内在价值和市场价值的差异, 如果企业的市场价值高于企业的内在价值, 那么股票被高估, 如果企业的市场价值低于企业的内在价值, 那么股票被低估。这也是学术界长期以来对错误定价的衡量方式, 大量的文献使用相关的方法估计错误定价。Berger 等 (1995) 与 Rhodes-Kropf (2005) 各自提出使用本公司和同行业公司的财务数据与市场价值衡量股票的错误定价的方法。Feltham 和 Ohlson (1995) 建立了一个剩余收益估值模型, 使用公司的财务数据计算公司的内在价值, 与市场价值进行比较, 得到股票的错误定价程度。游家兴和吴静 (2012) 使用上述三种方法在中国股票市场衡量了股票的错误定价。张静和王生年 (2016) 参考 Frankel 和 Lee (1998) 的假设, 修改了使用剩余收益模型对错误定价进行了估计。这种估计错误定价的方法很容易理解, 但是公司的内在价值难以合理估计, 同时, 财务价值只是公司价值的一部分。

由于错误定价是很多异象产生的根源, Stambaugh 等 (2015) 利用美国股票市场上 11 个不能被 Fama-French 三因子模型解释的异象, 构建错误定价的代理变量, 使用多个异象可以减少单个异象对指标的影响, 保证指标的稳健性。相比之下, Cao 和 Han (2016) 只使用了 4 个市场异象, 没有办法很好的捕捉错误定价, 所以没有观察到套利不对称的现象。Gu 等 (2018) 根据中国市场的实际情况, 在 11 个异象中选择了 6 个可得且显著的市场异象, 在中国市场构建错误定价指标。尹玉刚等 (2018) 也根据中国市场的情况, 选择了 6 个异象在中国市场构建了错误定价的代理变量。Hai 等 (2020) 则使用了 5 个异象在中国市场构造错误定价指标。Zhong 和 Gray (2016) 则在澳大利亚股票市场选择了 7 个市场异象估计股票的错误定价。近年来, 使用市场异象构建错误定价的方式已经被学术界广泛接受。

很多研究者从错误定价的角度对一些市场异象进行了解释。Stambaugh 等 (2015) 的使用错误定价指标解释了特质波动率异象产生的原因是套利的不对称性。尹玉刚等 (2018) 与 Gu 等 (2018) 则在中国市场检验套利不对称性对特质波动率异象的解释能力。Hai 等 (2020) 与 Zhong 和 Gray (2016) 则分别在中国市场和澳大利亚市场检验了博彩效应是否与风险补偿相关。Han 等 (2022) 发现交易量会放大股票的错误定价程度, 在被低估 (被高估) 的股票中, 交易量与预期收益正相关 (负相关), 这是因为投资者分歧导致的。

有很多研究关注错误定价与其他变量之间的关系, 但是还没有研究关注学习指数与错误定价之间的关系, 研究学习指数在被低估的股票和被高估的股票之中对股票收益的预测能力是否存在差异。同时, 根据 VNV 模型的理解, 投资者根

据信息进行决策，应该能够使得股票价格趋向内在价值，缩小股票的错误定价，还没有学者对这个领域进行研究，本文的研究可以对这一领域的文献进行有益的补充。

2.3 资产定价模型的相关研究

由 VNV 模型推导出的学习指数，能够负向预测股票的横截面收益，这就意味着 CAPM 模型的存在定价偏误，学习指数很有可能是一个可能在横截面上定价资产的全新因子，因此，对资产定价模型的相关研究进行梳理，可以为后续使用学习指数构造因子奠定理论基础。

虽然 Sharpe (1964)、Lintner (1965) 和 Mossin (1966) 的资本资产定价模型 (CAPM) 一直是资产定价文献中的主导范式，但是越来越多的学者发现其他的变量对股票横截面收益具有预测能力。Banz (1981) 发现美国市场，小市值股票比大市值股票预期月度收益高 0.4%，差异非常显著。Rosenberg 等 (1985) 账面市值比较高的股票，预期收益也显著更高。Fama 和 French (1992) 由此构建了一个包含市场因子、规模因子、价值因子的三因子模型，对股票收益在横截面上的共同变动进行解释。但是随着研究的推进，又有学者发现很多无法被三因子模型所解释的市场异象。Fama 和 French (2006) 表明净资产收益率代表的公司盈利能力与股票未来收益呈显著正相关。Titman 等 (2004) 发现资产增长率代表的公司投资大小对股票横截面收益具有预测能力。因此，Fama 和 French (2015) 在三因子模型的基础上增加了盈利因子和投资因子，构造了一个五因子模型，期望能够解释更多的市场异象。然而，Jegadeesh 和 Titman (1993) 在美国股票市场发现的动量效应（即过去一段时间总收益率较高的股票，在未来一段时间收益也会更高）一直无法被 Fama-French 多因子模型所解释。Fama 和 French (2018) 在五因子模型的基础上，增加了动量因子构建了一个六因子模型，但是同时也表达了将动量因子加入模型的担忧，因为动量因子的背后没有一个很好的理论基础，可能只是数据挖掘的结果，而且加入动量因子只能增强模型对动量类异象的解释能力，无法增强模型对其他类别的异象的解释能力。要想检验市场的有效性，就要先有一个足够合理的多因子模型，所以除了多因子模型的集大成者 Fama-French 之外，还有很多学者在这个领域做出了贡献。包括：Hou 等 (2015) 根据托宾 q 理论提出 q 因子模型，Stambaugh 和 Yuan (2017) 提出错误定价四因子模型，Daniel 等 (2020) 提出行为三因子模型。

很多在美国股票市场中对股票横截面收益具有预测能力的变量，在中国 A 股市场中也有显著的预测能力。美国股票市场中存在的规模效应 (吴世农和许年行，2004；汪炜和周宇，2002；田利辉等，2014)、账面市值比效应 (吴世农和许年

行, 2004; 陈莹, 2008) 在 A 股市场中也同样存在, 杨炘和陈展辉 (2003) 使用规模效应和账面市值比效应构建了中国 A 股的 Fama-French 三因子模型。更进一步, 高春亭和周孝华 (2016)、李志冰等 (2017)、在中国 A 股构建了 Fama-French 五因子模型, 发现盈利因子、投资因子都对股票收益具有显著预测能力。而中国 A 股市场的动量效应是否存在则一直存在争议, 很多研究都没有发现显著的动量效应 (王永宏和赵学军, 2001; 刘博和皮天雷, 2007; 潘莉和徐建国, 2011)。同时, 鲁臻和邹恒甫 (2007) 发现中国 A 股市场存在 6 个月的动量效应。因此, 一直没有学者在中国市场将动量效应纳入多因子模型。除此之外, 很多学者也发现在中国市场照搬美国市场构建的一些多因子模型定价能力较弱。因此, Liu 等 (2019) 根据中国市场的实际情况, 提出了一个更适合中国市场的四因子模型 (CH4), 包括了市场因子、规模因子、价值因子、情绪因子, 并且发现中国市场的四因子模型定价能力显著优于其他的多因子模型, 为中国市场的异象检验提供了基准。

除了被纳入多因子模型的变量外, 还有很多变量具有多因子模型无法解释的超额收益, 具有对股票横截面收益的预测能力。例如: 股票非流动性 ILLIQ (Amihud, 2002)、短期回报反转 STR (Jegadeesh, 1990)、长期回报反转 LTR (Debondt 和 Thaler, 1985)、特质波动率 IVOL (Ang 等, 2006)。在中国股票市场中, 很多学者也发现了类似的效应, 吴文锋等 (2003) 发现均值调整后的非流动性与股票横截面收益正相关, 杨华蔚和韩立岩 (2009)、左浩苗等 (2011) 发现特质波动率也对股票横截面收益存在影响。此外, 还有很多学者针对 A 股市场的特点进行研究, 发现很多可以预测股票横截面收益的变量。张然等 (2017) 使用分析师盈余预测修正和投资评级修正作为异象变量, 在股票市场上构建多空组合, 多空组合的三因子模型调整后超额收益分别为 1.34% 和 0.92%。林虎等 (2013) 使用日度交易数据构造出的换手率波动的指标对于未来收益率有着稳定并且显著的解释能力。谢谦等 (2019) 从 12 个衡量公司盈利能力的指标中使用偏最小二乘和组合预测法, 提取了一个衡量上市公司综合盈利水平的指标, 能够显著预测未来股票收益。张然等 (2022) 通过分析相关上市公司在电商平台的线上销售数据, 发现线上销售增长可以预测未来股票收益。

尽管被发现具有股票横截面收益的预测能力的变量很多, 但是被纳入多因子模型的变量很少, 大多数变量对股票横截面收益的预测能力可以被多因子模型所解释, 这就意味着这些变量无法对多因子模型的定价能力具有边际贡献; 除此之外, 想要将一个变量纳入多因子模型的框架之中, 变量对股票的收益预测能力还需要有坚实的理论基础, 很多变量都是数据挖掘的结果, 没有相应的理论基础。Cao 等 (2020) 发现学习指数无法被主流的多因子模型所解释, 同时学习指数的背后有一个坚实的理论模型。目前还没有一篇文献将学习指数作为股票市场的定

价因子纳入多因子模型，本文的研究可以填补这一领域的空白，同时为因子定价研究相关的文献进行有益的补充。

2.4 文献评述

通过对相关文献的归纳和梳理，可以发现投资者学习相关的研究越来越得到理论界和实务界的讨论。目前对于学习指数的实证研究主要还是 Cao 等（2020）在美国市场实证估计了学习指数，并研究了学习指数在美国市场对股票横截面收益的预测能力。以上学者的研究至少有以下三个方面的不足。

首先，还没有学者对学习指数的预测能力进行分解，研究和学习指数相关的成分对学习指数预测能力的贡献大小。

其次，还没有学者对错误定价与学习指数之间的关系进行研究。根据学习指数的经济含义，投资者会对和公司相关的信息进行了解，从经济含义上，学习指数应该可以降低股票的错误定价程度。

最后，对学习指数的研究主要还集中在研究学习指数与股票横截面收益之间的关系上，还没有学者使用学习指数构建因子，研究学习因子在股票市场中的定价能力。学习指数如果对股票横截面收益具有负向的预测能力，就说明他是一个多因子模型无法解释的市场异象。此外，学习指数背后有一个坚实的理论模型，学习指数就有成为一个解释股票横截面收益的因子的可能性。

第三章 理论基础与学习指数构建

本章首先介绍了学习指数指标的理论基础,根据理论模型推导出学习指数指标和股票预期收益之间的关系,随后介绍学习指数指标的估计方法,结合理论模型中学习指数的表达式给出学习指数指标的估计步骤。本部分的内容为后续学习指数与股票横截面预期收益的检验、学习指数与错误定价的关系的考察、学习因子的定价应用等三个方面的实证分析提供了具体的方向和理论基础。

3.1 理论基础

本文估计的学习指数来源于 Van Nieuwerburgh 和 Veldkamp (2010) 提出的理性预期一般均衡模型。模型假设市场上存在 N 个风险资产、收益为 r 的无风险资产和连续的无原子投资者 (continuum of atomless investor)。风险资产的平均供给是 $\bar{x} + x$; 其中, \bar{x} 是正的常数向量, $x \sim N(0, \sigma_x^2 I)$ 。 x 是对风险资产供给的各种冲击,这些冲击导致了价格噪音,阻碍了价格的完美发现,这使得投资者产生了对资产进行学习的动机。

模型分为三期展开:

- ① 第一期: 所有的投资者都是相同的,事前投资者获取未知的资产支付 f 的信息,资产支付 f 的服从均值向量为 μ 和协方差矩阵为 Σ 的正态分布。投资者的学习决策包括,学习哪些风险资产、对这些资产学习多少。
- ② 第二期: 投资者可以观察到一个 $N \times 1$ 阶的信息信号向量 $\eta = f + e_n$, 其中 $e_n \sim N(0, \Sigma_\eta)$ 。假设投资者都观察到了带有独立噪音的无偏的信息信号,噪音可以被视为投资者处理信息和解释信息时所犯的误差。信号实现的信息、先验信念的信息、资产价格的信息构成了后验信念 $\hat{\mu}$ 和 $\hat{\Sigma}$ 。投资者基于后验信念构建投资组合,选择自己的投资组合分配 q , 并且投资组合受预算线 $W = W_0 r + q'(f - pr)$ 的约束, 其中 p 是市场出清时的资产价格向量。
- ③ 第三期, 投资者获得资产支付并实现效用。

具体而言, 假设投资者效用函数满足均值-方差偏好、绝对风险厌恶系数为 ρ , 效用函数的具体表达式如下,

$$U_1 = E_1[\rho E_2[W]] - \frac{\rho^2}{2} V_2[W] \quad (3-1)$$

假设资产支付 f 是独立的, 这个假设是不失一般性的, 因为可以将相关的资产通过主成分分析 (PCA) 转化为不相关的资产的线性组合。而进行主成分分解后得到的资产可以解释为主成分 (PC)、风险因子或者阿罗德布鲁证券。具体而

言,可以对协方差矩阵进行特征分解($\Sigma = \Gamma \Lambda \Gamma'$, Γ 代表了资产在风险因子上的因子载荷, Λ 代表了风险因子的方差)。

类似的可以假设信息信号 η 也是独立的,这个假设同样也是不失一般性的,因为 Van Nieuwerburgh 和 Veldkamp (2010) 讨论了放宽这一假设的情形,不改变模型的定性结果。这个假设可以将信息信号的协方差矩阵 Σ_η 从 $N(N+1)/2$ 个不同元素变为只有对角线上的 N 个元素,这个假设意味着后验信念的方差和先验信念的方差具有相同的特征向量。这一假设的经济解释是,投资者根据给定的风险因子的相关结构,选择通过学习降低风险的程度。因此,对于独立的资产和信息信号,投资者的信息选择相当于为每项资产选择后验方差(特征值)。

模型使用逆向归纳法求解。在第二期,投资者的最优选择是基于后验信念 $\hat{\mu}$ 和 $\hat{\Sigma}$ 的多元化投资组合,可以将投资组合分配 q 求解为:

$$q^* = \frac{1}{\rho} \hat{\Sigma}^{-1} (\hat{\mu} - pr) \quad (3-2)$$

根据 Admati (1985), 均衡资产价格 p 是资产支付 f 和供给冲击 x 的线性函数,

$$pr = A + Bf + Cx \quad (3-3)$$

其中,系数矩阵 A, B, C 是平均投资者的后验信念、风险厌恶水平、资产供应的函数。因为 VNV 模型和 Admati (1985) 使用相同的假设,所以资产价格 p 是平均投资者的后验均值和方差的函数。因此,后验信念可以用贝叶斯更新表示如下:

$$\begin{aligned} \hat{\mu} &\equiv E[f | \mu, \eta, p] \\ &= (\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1})^{-1} (\Sigma^{-1} \mu + \Sigma_\eta^{-1} \eta + \Sigma_p^{-1} B^{-1} (rp - A))^{-1} \end{aligned} \quad (3-4)$$

$$\hat{\Sigma} \equiv V[f | \mu, \eta, p] = (\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1})^{-1} \quad (3-5)$$

另一方面,将(3-2)代入(3-1)可得:

$$U_1 = E_1 \left[\frac{1}{2} (\hat{\mu} - pr)' \hat{\Sigma}^{-1} (\hat{\mu} - pr) \right] + rW_0 \quad (3-6)$$

因为在第一期,资产价格 p 是资产支付 f 都是未知的,所以 $\hat{\mu} - pr$ 是一个随机变量。方程(3-6)等号右边服从非中心的卡方分布的随机变量,因此计算(3-6)的期望值为:

$$U_1 = \frac{1}{2} \text{Trace}(\hat{\Sigma}^{-1} V_1 [\hat{\mu} - pr]) + \frac{1}{2} E_1 [\hat{\mu} - pr]' \hat{\Sigma}^{-1} E_1 [\hat{\mu} - pr] + rW_0 \quad (3-7)$$

其中:

$$E_1 [\hat{\mu} - pr] = (I - B) \mu - A$$

$$V_1 [\hat{\mu} - pr] = \Sigma - \hat{\Sigma} + B \Sigma B' + C C' \sigma_x^2 - 2 B \Sigma$$

同时，信息选择受到两个约束：

(1) 信息容量限制。VNV 模型使用的是基于熵的学习，这意味着学习的总量（即， Σ 和 $\hat{\Sigma}$ 之间的距离）不能超过容量 K 。

(2) 非负的学习。这意味着投资者不能忘记包含在先验信念或价格中的信息。

因此，第一期中的信息选择问题可以描述为在两个约束的条件下最大化期望效用 U_1 ：

$$\begin{aligned} \max_{\hat{\Sigma}_1, \dots, \hat{\Sigma}_N} & \frac{1}{2} \left[-N + \sum_{i=1}^N LIF_i \frac{\Sigma_i}{\hat{\Sigma}_i} \right] \\ \text{subject to} & \prod_{i=1}^N \frac{\Sigma_i}{\hat{\Sigma}_i} = K \text{ and } (\hat{\Sigma}_i^{-1} \geq \Sigma_i^{-1} + \Lambda_{pi}^{-1}) \forall i \end{aligned} \quad (3-8)$$

其中：

i 代表风险因子 i ，

$$LIF_i = (((I - B)\mu - A)' \Gamma_i)^2 \Lambda_i^{-1} + (1 - \Lambda_{Bi})^2 + \Lambda_i^{-1} \Lambda_{Ci}^2 \sigma_x^2$$

在第一期中，投资者最优的信息选择是将所有的学习能力分配给学习指数最高的风险因子。而风险因子 i 的学习指数为：

$$LIF_i = (((I - B)\mu - A)' \Gamma_i)^2 \Lambda_i^{-1} + (1 - \Lambda_{Bi})^2 + \Lambda_i^{-1} \Lambda_{Ci}^2 \sigma_x^2 \quad (3-9)$$

风险因素 i 的学习指数可以拆解为以下三项的和，

第一项： $((I - B)\mu - A)' \Gamma_i)^2 \Lambda_i^{-1}$ ，可以看作 $E[\Gamma'_i(f - pr)]$ 和 $\frac{E[\Gamma'_i(f - pr)]}{Var[\Gamma'_i f]}$ 的乘积。 $E[\Gamma'_i(f - pr)]$ 意味着对于预期收益更高的风险因子，学习指数更高。 $\frac{E[\Gamma'_i(f - pr)]}{Var[\Gamma'_i f]}$ 乘以 $\frac{1}{\rho}$ 就是投资组合的份额，这意味着，预期被持有更多的风险因子，学习指数更高。

第二项： $(1 - \Lambda_{Bi})^2$ ，反映风险因子 i 的支付 f 与风险因子 i 的价格 p 的信息相关的预期定价误差。 Λ_{Bi} 是矩阵 B 的第 i 个特征值，它捕捉了收益和价格之间的关系。当 Λ_{Bi} 较低时，支付 f 变动导致价格 p 变动较小，使得关于支付 f 的信息更有价值。

第三项： $\Lambda_i^{-1} \Lambda_{Ci}^2 \sigma_x^2$ ，反映与价格 p 对风险因子 i 的供应冲击敏感性相关的预期定价误差。 Λ_i 和 Λ_{Ci} 分别是先验协方差矩阵 Σ 和 C 的第 i 个特征值。 σ_x^2 是风险因子供给冲击的方差，假设对于所有的主成分(PC)都是相同的。若先验不确定性不变，较高的 Λ_{Ci} 表明供应冲击对价格的影响更大，因此产生知情投资者可以利用的定价错误，因此可能导致更高的学习指数。

由于 VNV 模型中的学习选择是基于风险因子而不是资产，所以将风险因子的学习指数乘以相关资产的特征向量矩阵（代表了风险因子的载荷），来生成相关资产的学习指数。

$$LI = \Gamma(LIF) \quad (3-10)$$

投资者更喜欢了解其他投资者不知道的信息,因为随着越来越多的投资者了解某一个风险因子,该风险因子的预期收益就会降低,这就降低了解该风险因子的价值。因为在模型中,投资者可以通过对资产的学习降低资产的不确定性,而风险厌恶投资者更喜欢持有他们了解更多的资产,所以该模型对信息选择、风险和预期收益之间的关系进行预测:资产信息的增加导致不确定性的减少和预期收益的降低。

3.2 学习指数构建

本文的核心研究变量股票的学习指数。本文参考 Cao 等(2021)提出的方法计算学习指数。由于中国股票市场投资者机构化程度不高,存在较多散户投资者,散户投资者更关注最近的信息,所以本文使用 1 年期的滚动窗口估计个股的学习指数。

本文使用 1 年期的周度收益率滚动窗口构建价格 p 、支付 f 、协方差矩阵 Σ 的估计。周度收益率计算为从本周三到下周三的持有期总收益率。在每个月末,执行以下步骤:

第一步:构建个股的价格 p 和支付 f 的时间序列。每支股票的价格在第一周被设定为 1, 股票价格 p 根据后续周收益率的时间序列而变化,使用下一周的股票价格 p 作为股票支付 f 的代理变量,使用对数价格作为研究对象保持与 VNV 模型的设定的一致。超额收益 $f - pr = \ln(\frac{P_{t+1}}{P_t}) - \ln(1 + r_f)$, 其中 r_f 是无风险利率。

第二步:将相关的资产转换为不相关的资产。将股票支付 f 标准化(均值为 0, 方差为 1),使用标准化后的股票支付 f 来估计 f 的先验协方差矩阵 Σ , 再将协方差矩阵 Σ 分解为对角特征值矩阵 Λ 和特征向量矩阵 Γ : $\Sigma = \Gamma \Lambda \Gamma'$ 。构造主成分价格 $\Gamma'p$, 主成分支付 $\Gamma'f$, 主成分超额收益 $\Gamma'(f - pr)$ 。

第三步:估计主成分的学习指数。根据模型的定义,主成分学习指数是三项的简单累加: $LIF_i = (((I - B)\mu - A)' \Gamma_i)^2 \Lambda_i^{-1} + (1 - \Lambda_{Bi})^2 + \Lambda_i^{-1} \Lambda_{Ci}^2 \sigma_x^2$ 。在主成分水平上对均衡价格方程 $\Gamma'pr = \Gamma'A + \Lambda_B \Gamma'f$ 进行回归,计算出来的回归系数和收益系数 Λ_B 可以用于计算第二项与第三项。 Λ_{Bi} 是 Λ_B 的对角线上第 i 个的元素。三项的计算分别如下:

$$\begin{aligned} \textcircled{1} \quad & (((I - B)\mu - A)' \Gamma_i)^2 \Lambda_i^{-1} = \frac{\text{squared average excess return}}{\text{variance payoffs}} = \frac{[E(\Gamma_i'(f - pr))]^2}{\text{Var}(\Gamma_i'f)} \\ \textcircled{2} \quad & (1 - \Lambda_{Bi})^2 \\ \textcircled{3} \quad & \Lambda_i^{-1} \Lambda_{Ci}^2 \sigma_x^2 = \frac{1 - R^2}{R^2} \Lambda_{Bi}^2 \end{aligned}$$

第四步：估计个股的学习指数。使用主成分学习指数与特征矩阵估计所有股票当月的学习指数： $LI = \Gamma \times LIF$ 。

由于特征向量中可能存在负数，为了防止出现负的学习指数，使用特征向量元素的平方作为权重计算股票的学习指数。同时，特征向量是标准化的正交矩阵，平方和是 1，这相当于每一个主成分对股票学习指数的贡献值，在每一个横截面上，只选取主成分大于 1 的主成分。最后，股票的数量和主成分的数量随时间的变化而变化，使用股票数量和主成分数量的比值来缩放学习指数，这相当于将所有股票的学习指数在每个横截面内乘以一个常数，并且不影响学习指数在横截面上的解释力。

第四章 投资者学习与股票横截面收益

根据第三章的分析,本文可以在中国股票市场实证估计个股的学习指数,因此本章使用个股的学习指数验证投资者学习与股票预期收益之间的关系。本章首先检验了学习指数对股票横截面的预测能力,随后对学习指数对股票横截面收益的预测能力进行分解分析,为学习指数背后的理论模型(VNV模型)提供来自中国股票市场的实证证据。

4.1 样本数据和变量说明

4.1.1 数据来源与样本选择

选取中国A股市场的所有上市公司作为实证研究的对象。所有使用的股票日度交易数据、月度交易数据,财务报表数据、Fama-French多因子模型数据,无风险利率数据、新闻数据、公告数据、分析师数据、机构投资者持有数据等均来自国泰安数据库。Liu等(2019)提出的中国市场三因子(CH3)、四因子(CH4)的数据来源于Robert F. Stambaugh的个人网站。

为了减少市场微观结构问题和小市值股票的影响,在每个月末,剔除公司规模小于沪深主板股票市值20%分位数的股票和价格小于5元的股票。为了排除次新股的影响,剔除上市不到1年的股票。为了防止结果受到长时间停牌后的股票收益的影响,在每个月末,样本中包含的股票在过去一年至少有120天的交易记录,过去20个正常交易日至少有15天的交易记录。为了保持财务报表的一致性,剔除了金融行业股票。本章主要分析的时间区间从2000年1月1日到2021年12月31日。中国A股市场始于1990年上海和深圳交易所的建立,关注2000年之后的时期主要是为了确保财务报表数据的一致性,中国有关财务报告各个方面的规章制度的实施直到1999年才基本成型。

为了防止极端值对研究结果造成影响,本文对所有的研究变量在每个时间截面上进行截尾,剔除了小于0.1%分位数和大于99.9%分位数的数据,以保证研究结果的稳健性。

4.1.2 变量说明

本章的核心变量是股票的学习指数(LI),学习指数的构建方法已经在第三章第二节中给出,本节不再赘述。本章的研究重点是研究学习指数与股票横截面收益之间的关系,为了确保学习指数对股票横截面收益的预测能力不被其他公司特征所解释,本章使用一系列常见的公司特征作为控制变量进行研究,这些变量的具体定义以及计算方式定义如下。

(1) 市场风险(BETA): 参考Sharpe(1964), Lintner(1965)和以及Mossin

(1966), 在每个月末, 使用过去 250 天的股票日收益率对市场日收益率进行回归, 得到的 β 系数就是市场风险指标 $BETA$ 。回归方程如下,

$$r_{i,d} - r_{f,d} = \alpha_i + \beta_i^*(r_{m,d} - r_{f,d}) + \varepsilon_{i,d} \quad (4-1)$$

(2) 公司规模 ($SIZE$): 参考 Banz (1981), $SIZE$ 计算为股票流通市值的自然对数, 计算公式如下,

$$SIZE = \log(ME_{i,t}) \quad (4-2)$$

其中, $ME_{i,t}$ 是股票的流通市值, 以百万为单位

(3) 账面市值比 (BM): 参考 Fama 和 French (1992), 每年的 6 月末, 在中国市场构建公司的账面市值比, 账面市值比计算为上年末账面价值除以上一年 12 月的股票总市值。计算公式如下,

$$BM_{i,t} = \frac{BV_{i,t-1}}{MV_{i,t-1}} \quad (4-3)$$

其中, $BV_{i,t-1}$ 为公司在 $t-1$ 年末的普通股权益的账面价值, $MV_{i,t-1}$ 为 $t-1$ 年 12 月末的总市值, 为了避免前瞻偏差, 所以在每年的 6 月末计算一次该指标, 转换为月度频率。

(4) 盈利能力 ($PROF$): 参考 Hou 等 (2015), 盈利能力计算为至少四个月前结束的最近一个财政季度的季度净利润除以滞后一期的账面价值, 计算公式如下,

$$PROF_{i,t} = \frac{GP_{i,t-1}}{BA_{i,t-2}} \quad (4-4)$$

其中, $GP_{i,t-1}$ 为至少四个月前的最近一个季度净利润, $BA_{i,t}$ 为滞后一期的账面价值。为了避免前瞻偏差, 滞后四个月以保证财务数据的可得性。

(5) 投资能力 (INV): 参考 Cooper 等 (2008) 将投资能力计算为公司的总资产的年度变化, 计算公式如下,

$$INV_{i,t} = \frac{TA_{i,t-1}}{TA_{i,t-2}} - 1 \quad (4-5)$$

其中, $TA_{i,t-1}$ 为上一年度的总资产, $TA_{i,t-2}$ 为前两年度的总资产。为了避免前瞻偏差, 所以在 6 月末计算该指标, 数据每年更新一次, 并转化为月度频率。

(6) 动量 (MOM): 参考 Jegadeesh 和 Titman (1993) 的做法, 计算为股票在 $t-11$ 期到 $t-1$ 期的累积收益率, 计算公式如下,

$$MOM_{i,t} = \prod_{\tau=1}^{11} (1 + r_{i,t-\tau}) - 1 \quad (4-6)$$

其中, $r_{i,t-\tau}$ 是公司在 $t-\tau$ 月的股票收益率。

(7) 非流动性 (*ILLIQ*): 参考 Amihud (2002) 的做法, 首先计算股票每日收益率的绝对值除以当日成交额, 随后求该指标在当月的算数平均值, 计算公式如下,

$$ILLIQ_{i,t} = \frac{1}{N} \sum_{d=1}^N \frac{|ret_{i,d}|}{volume_{i,d}} \quad (4-7)$$

(8) 短期回报反转 (*STR*): 参考 Jegadeesh (1990) 的做法, 将短期回报反转定义为股票过去一个月的收益率, 计算公式如下,

$$STR_{i,t} = r_{i,t} \quad (4-8)$$

(9) 长期回报反转 (*LTR*): 参考 De Bond 和 Thaler (1985) 的做法, 将长期回报反转定义为股票在 $t-59$ 到 $t-12$ 的累积收益率, 公式计算如下,

$$MOM_{i,t} = \prod_{\tau=12}^{59} (1 + r_{i,t-\tau}) - 1 \quad (4-9)$$

(10) 特质波动率 (*IVOL*): 参考 Ang 等 (2006) 的定义计算特质波动率, 在中国市场将特质波动率计算为过去一个月个股日超额收益率对市场日超额收益率回归的残差标准差, 公式计算如下,

$$\begin{aligned} r_{i,d} - r_{f,d} &= \alpha_i + \beta_i * (r_{m,d} - r_{f,d}) + \varepsilon_{i,d} \\ IVOL &= \sqrt{\sum (\varepsilon_{i,d} - \bar{\varepsilon}_{i,d})^2} \end{aligned} \quad (4-10)$$

4.2 数据描述性统计特征

学习指数 (*LI*) 是本文的核心研究变量, 股票的学习指数越大, 代表投资者对该股票的学习程度越大。表 4-1 的第一行给出了 2000 年至 2021 年间学习指数的横截面描述性统计的时间序列均值。从表 4-1 中可见, *LI* 的均值是 1.112, *LI* 的 75% 分位数和 25% 分位数之间的差异较大, 说明不同的股票之间学习指数的大小存在很大的差异, 也就是说在不同的股票之间投资者对股票的学习程度存在较大的差异。本章中使用的其他常见公司特征控制变量的描述性统计特征也在表 4-1 中给出。

表 4-2 统计了基于学习指数 (*LI*) 大小排序形成的 5 个投资组合中公司特征变量的均值。可以观察到, 在 5 个投资组合之间, *LI* 的均值从 0.327 到 2.456 单调递增。*BETA*、*SIZE*、*BM*、*PROF*、*MOM* 和 *ILLIQ* 随着 *LI* 的增加而减小且至少在 10% 的水平上显著, *INV* 和 *LTR* 随着 *LI* 的增加而增加, 且显著, *IVOL* 和 *STR* 随 *LI* 的增加而增加但不显著。表 4-3 展示了变量之间的相关系数矩阵, 即各个变量的在横截面上的相关系数的时间序列均值。表 4-3 显示, 在所有公司特征变量中, *LI* 与 *BETA* 相关性最高, 相关系数达到 -0.182, 其次是 *BM*, 相关系数为 -0.091, 随后是 *SIZE*、*MOM* 和 *LTR*, *LI* 和他们的相关系数的绝对值约为 0.05。因此, 更低

市场贝塔值 (*BETA*)、更低市场规模 (*SIZE*)、更低股票动量 (*MOM*)、更高长期回报反转 (*LTR*) 的股票具有更高的学习指数。另外, 市场规模 (*SIZE*) 和股票非流动性 (*ILLIQ*) 存在较强的负相关性, 相关系数达到-0.527。

表 4-1 描述性统计

变量	均值	标准差	最小值	P25	中位数	P75	最大值
<i>LI</i>	1.112	0.853	0.113	0.529	0.877	1.432	6.553
<i>BETA</i>	1.088	0.242	0.261	0.945	1.103	1.252	1.715
<i>SIZE</i>	8.041	0.834	6.031	7.466	7.915	8.511	11.487
<i>BM</i>	0.410	0.232	-0.122	0.246	0.367	0.524	1.645
<i>PROF</i>	0.067	0.162	-2.285	0.042	0.074	0.108	1.670
<i>INV</i>	0.275	1.279	-0.470	0.029	0.122	0.270	31.548
<i>MOM</i>	0.258	0.443	-0.504	-0.032	0.159	0.436	3.131
<i>IVOL</i>	0.021	0.008	0.006	0.015	0.020	0.026	0.058
<i>ILLIQ</i>	0.141	0.157	0.005	0.054	0.101	0.180	1.829
<i>STR</i>	0.019	0.107	-0.409	-0.046	0.006	0.070	0.606
<i>LTR</i>	0.972	1.434	-0.701	0.090	0.591	1.374	12.254

表 4-2 学习指数组合公司特征

组合	<i>LI</i>	<i>BETA</i>	<i>SIZE</i>	<i>BM</i>	<i>PROF</i>	<i>INV</i>	<i>MOM</i>	<i>ILLIQ</i>	<i>STR</i>	<i>LTR</i>	<i>IVOL</i>
1(Low)	0.327	1.126	8.146	0.437	0.072	0.246	0.406	0.154	0.019	0.937	0.021
2	0.594	1.119	8.054	0.427	0.069	0.270	0.281	0.144	0.020	0.949	0.021
3	0.882	1.103	8.023	0.414	0.067	0.266	0.234	0.139	0.021	0.949	0.021
4	1.301	1.077	8.005	0.402	0.066	0.275	0.204	0.134	0.021	0.981	0.021
5(High)	2.456	1.018	7.979	0.373	0.063	0.319	0.162	0.132	0.020	1.051	0.021
5-1(HML)	2.129*** (16.23)	-0.108*** (-10.74)	-0.167*** (-3.52)	-0.064*** (-5.50)	-0.008*** (-2.97)	0.073*** (3.05)	-0.244*** (-4.06)	-0.022* (-1.87)	0.001 (0.65)	0.114* (1.70)	0.000 (0.07)

表 4-3 相关系数矩阵

变量	<i>LI</i>	<i>BETA</i>	<i>SIZE</i>	<i>BM</i>	<i>PROF</i>	<i>INV</i>	<i>MOM</i>	<i>ILLIQ</i>	<i>STR</i>	<i>LTR</i>	<i>IVOL</i>
<i>LI</i>	1.000										
<i>BETA</i>	-0.182	1.000									
<i>SIZE</i>	-0.052	-0.089	1.000								
<i>BM</i>	-0.091	0.099	0.100	1.000							
<i>PROF</i>	-0.021	-0.024	0.154	-0.016	1.000						
<i>INV</i>	0.012	-0.024	0.054	-0.032	0.026	1.000					
<i>MOM</i>	-0.053	-0.107	0.166	-0.006	-0.032	-0.004	1.000				
<i>ILLIQ</i>	0.029	-0.111	-0.527	-0.072	-0.065	-0.039	-0.134	1.000			
<i>STR</i>	0.001	-0.022	0.039	0.019	-0.015	-0.014	-0.028	-0.037	1.000		
<i>LTR</i>	0.055	-0.102	0.289	-0.316	0.134	0.166	-0.094	-0.154	-0.032	1.000	
<i>IVOL</i>	-0.003	0.081	-0.106	-0.123	-0.077	-0.013	0.196	-0.048	0.324	-0.007	1.000

4.3 学习指数对股票收益的预测能力

4.3.1 单变量组合排序

26页, 学习指数是月度数据

本小节使用单变量排序法研究学习指数 LI 对股票横截面收益的预测能力。单变量组合排序的具体步骤如下:

(1)在每个月末,根据学习指数 LI 的大小从低到高将所有股票划分为5组。使用5个组内的股票分别构建投资组合,构建组合的权重有等权重和流通市值两种,随后持有组合一个月,并计算这些组合直到下个月末的超额收益,以及最高 LI 组合 $High$ 与最低 LI 组合 Low 的收益差异(也就是做多最高 LI 组合与做空最低 LI 组合得到的多空组合 HML 的收益)。

(2)在每个月末重复上述步骤,可以得到5个 LI 排序组合以及多空组合 HML 的超额收益的时间序列。

(3)计算各组的超额收益的时间序列均值、风险(多因子模型)调整后的超额收益以及 Newey-West (1987) 调整后的 t 统计量。

由于部分学者发现 Fama 和 French(1993)的三因子模型以及 Fama 和 French (2015)的五因子模型并不适合中国市场,对市场异象的解释能力较弱(Lin, 2017; Hu 等, 2019),所以 Liu 等(2019)构造了一个更适合中国股票市场的三因子模型(CH3)、四因子模型(CH4)。为了进一步检验学习指数对股票横截面收益之间的关系的稳健性,除了按照学术界对异象检验的惯例使用 CAPM、FF3 和 FF5 计算风险调整后的超额收益,本文还针对中国市场的研究,使用 CH3 和 CH4 计算组合的风险调整后的超额收益,风险调整超额收益也就是风险因子(主流的多因子模型)无法解释的异常收益(α)。

表 4-4 依次展示了单变量组合排序的平均月度超额收益以及 CAPM、FF3、FF5、CH3 和 CH4 调整后的平均月度超额收益以及对应的 Newey-West (1987) 调整后的 t 统计量。Panel A 展示的是等权重投资组合, Panel B 展示的是流通市值加权的投资组合。例如,在 Panel A 中,最高学习指数 LI 组合的下个月月度平均超额收益是 0.754%,最低 LI 组合的下个月月度平均超额收益 1.147%。最高 LI 投资组合和最低 LI 投资组合的平均超额收益的差异是-0.393%, t 统计量为-2.45,在 5%的水平上显著。也就是说,做多最低 LI 组合的同时做空最高 LI 组合得到的多空组合 HML 平均每个月可以获得-0.393%的超额收益。这个结果表示低 LI 组合超额收益比高 LI 组合平均要高,也就是说较高的 LI 的股票平均而言有较低的预期收益。剩余的 5 列展示了使用主流的因子模型调整后的超额收益,第二列展示了在控制了 CAPM 模型的市场因子后,各个 LI 组合的风险调整后收益,多空组合 HML 的风险调整后收益是-0.309%,在 10%的水平上显著。剩余几列分别代表了

控制 Fama-French 三因子、Fama-French 五因子、中国市场三因子、中国市场四因子后的组合风险调整后超额收益。多空组合 *HML* 的超额收益分别是 -0.431%、-0.658%、-0.815%、-0.943%，也都至少在 5% 的水平上显著，这说明即使在控制了主流的多因子模型之后，多空组合 *HML* 的超额收益依旧与 0 有显著差异，也就是说由学习指数排序得到的 *HML* 组合具有主流多因子模型无法解释的超额收益。在表 4-4 的 Panel B 中，最高 *LI* 组合的下个月月度平均超额收益是 0.525%，最低 *LI* 投资组合的下个月月度平均超额收益是 1.069%。多空组合 *HML* 的平均差异是 -0.544%，使用 CAPM、Fama-French 三因子、Fama-French 五因子、中国市场三因子、中国市场四因子调整后的超额收益分别是 -0.429%、-0.594%、-0.929%、-1.103%、-1.286%，相应的，多空组合 *HML* 组合的超额收益也都至少在 10% 的水平上显著。也就是说，市值加权投资组合和等权重投资组合的结果在定性上是相似的。

总的来说，表 4-4 的结果说明高 *LI* 股票比低 *LI* 股票具有更低未来收益，这与 Cao 等(2021)在美国市场发现一致，学习指数对股票收益具有负向预测能力，并且学习指数对股票收益的负向预测能力不能被主流的多因子模型所解释。

表 4-4 的单变量组合排序分析的是学习指数 *LI* 低的股票和学习指数 *LI* 高的股票之间的差异，而不是某一个投资者进行学习之后得到的投资组合的预期收益。假设一个投资者学习更多高 *LI* 的股票、学习更少低 *LI* 的股票，信息选择并不意味着投资者会做多高 *LI* 股票，做空低 *LI* 股票。相反，投资者会根据自己学习的结果进行投资组合决策，这个投资者会做多他预期收益最高的股票，做空他预期收益最低的股票。由于对一种资产了解得越多，这些预期就越准确，投资者的投资组合预期收益随着他的学习能力而增加。因此，尽管高 *LI* 资产的预期均衡收益比低 *LI* 资产更低，但是了解这些资产的投资者比不知情的投资者拥有更高的投资组合预期收益。

表 4-4 单变量投资组合排序

组数	<i>Exret</i>	CAPM- α	FF3- α	FF5- α	CH3- α	CH4- α
Panel A: 等权重投资组合						
1(Low)	1.147*	0.367	0.067	0.277*	0.554***	0.689***
	(1.83)	(1.65)	(0.44)	(1.75)	(4.94)	(5.77)
2	0.966	0.202	-0.168	-0.051	0.091	0.193**
	(1.52)	(0.88)	(-1.51)	(-0.34)	(1.13)	(2.26)
3	0.890	0.131	-0.255**	-0.212	-0.068	0.003
	(1.38)	(0.58)	(-2.33)	(-1.41)	(-0.74)	(0.03)
4	0.856	0.119	-0.284**	-0.276*	-0.171*	-0.168*
	(1.32)	(0.51)	(-2.37)	(-1.88)	(-1.81)	(-1.79)
5(High)	0.754	0.058	-0.364**	-0.381**	-0.261**	-0.254**
	(1.18)	(0.24)	(-2.57)	(-2.23)	(-2.19)	(-2.10)
5-1(HML)	-0.393**	-0.309*	-0.431**	-0.658***	-0.815***	-0.943***
	(-2.45)	(-1.79)	(-2.21)	(-3.64)	(-4.66)	(-4.97)
Panel B: 流通市值加权投资组合						
1(Low)	1.069*	0.303*	0.281	0.560***	0.713***	0.837***
	(1.82)	(1.75)	(1.39)	(3.24)	(3.82)	(4.08)
2	0.596	-0.132	-0.245*	-0.084	-0.098	0.036
	(1.02)	(-0.95)	(-1.83)	(-0.53)	(-0.80)	(0.28)
3	0.698	-0.042	-0.198	-0.177	-0.141	-0.086
	(1.13)	(-0.30)	(-1.57)	(-1.21)	(-1.11)	(-0.67)
4	0.791	0.079	-0.072	-0.075	-0.191	-0.200
	(1.27)	(0.48)	(-0.52)	(-0.46)	(-1.48)	(-1.51)
5(High)	0.525	-0.126	-0.313**	-0.369**	-0.391**	-0.449***
	(0.88)	(-0.72)	(-2.01)	(-2.29)	(-2.51)	(-2.77)
5-1(HML)	-0.544**	-0.429*	-0.594**	-0.929***	-1.103***	-1.286***
	(-2.58)	(-1.85)	(-2.24)	(-4.02)	(-4.16)	(-4.41)

4.3.2 双变量组合排序

本节采用双变量组合排序的方法检验学习指数对股票横截面收益的预测能力。双变量组合排序的目的是检验在控制某一常见公司特征后，学习指数对股票横截面收益的负向预测能力是否依旧存在，确保学习指数对股票横截面收益的预测能力不能被其他变量所解释。双变量组合排序分为双变量独立排序和双变量条件排序两种方法，双变量独立排序的好处在于将两个变量置于同等地位进行分析，但是无法保证每个组合里面的股票个数接近，双变量条件排序的好处在于可以确保每个组合中股票数量接近，但是主要分析的变量是第二排序变量。

双变量独立排序的具体步骤如下：

(1) 在每个月末，首先根据控制变量的大小将股票分为 5 组，再根据学习指数 LI 的大小将股票分为 5 组，分别取每个控制变量组和 LI 组的交集，最后得到 5×5 一共 25 个组合。使用 25 个组合中的股票分别构建投资组合，构建组合的权重有等权重和流通市值两种，持有 25 个投资组合一个月，计算 25 个投资组合的超额收益。

(2) 计算同一 LI 排序组合中的 5 个不同控制变量排序组合的超额收益的均值，得到控制变量后的 LI 排序组合的超额收益，再计算最高 LI 和最低 LI 组合构建的差异，也就是多空组合 HML 的超额收益

(3) 在每个月末重复上述步骤，得到 LI 排序的 5 个组合和多空组合 HML 的月度超额收益时间序列。

(4) 计算组合超额收益的时间序列均值、风险调整后的超额收益以及 Newey-West (1987) 调整后的 t 统计量。

双变量条件排序的具体步骤如下：

(1) 在每个月末，先按照控制变量的大小将股票分为 5 组，得到 5 个控制变量组，再在控制变量的各组内将股票根据学习指数 LI 分为 5 组，最后得到 5×5 一共 25 个投资组合。使用 25 个组合中的股票分别构建投资组合，构建组合的权重有等权重和流通市值两种，持有 25 个投资组合一个月，计算 25 个投资组合的超额收益。

(2) 计算同一 LI 排序组合中的 5 个不同控制变量排序组合的超额收益的均值，得到控制变量后的 LI 排序组合的超额收益，再计算最高 LI 和最低 LI 组合构建的差异，也就是多空组合 HML 的超额收益

(3) 在每个月末重复上述步骤，得到 LI 排序的 5 个组合和多空组合 HML 的月度超额收益时间序列。

(4) 计算组合超额收益的时间序列均值、风险调整后的超额收益以及 Newey-West (1987) 调整后的 t 统计量。

表 4-5 报告了双变量组合排序的投资组合的超额收益均值，多空组合 HML 的

超额收益均值,多空组合HML的FF5(CH4)调整后超额收益以及相应的Newey-West(1987)调整后的 t 统计量。Panel A 报告的是双变量独立排序的等权重投资组合, Panel B 报告的是双变量独立排序的流通市值加权投资组合, Panel C 报告的是双变量条件排序的等权重投资组合, Panel D 报告的是双变量条件排序的流通市值加权投资组合。

表 4-5 的 Panel A 中,第 1 列表示了使用BETA作为控制变量与学习指数LI进行双变量组合排序,最低的LI组合的平均超额收益为每月 1.173%,最高LI组合的平均超额收益为每月 0.777%,最高LI组合和最低LI组合的月度超额收益的之间的差为-0.397%,在 5%的水平上显著(t 值为-2.57),对应的FF5(CH4)调整后超额收益是-0.652%(-0.897%), t 值为-3.70(-4.83),在 1%的水平上显著。这表明,在控制了BETA的影响之后,学习指数LI依旧对股票收益具有负向的预测能力,也就是说,学习指数LI的负向预测能力不能被BETA所解释。同样的,对于其他的控制变量,也都可以观察到类似的结果,在控制了BM、INV、IVOL和STR之后,多空组合HML的月度平均超额收益是-0.351%、-0.398%、-0.347%和-0.327%, t 值为-2.20、-2.53、-2.39 和-2.22,都在 5%的水平上显著,在控制其他变量后,多空组合的月度平均超额收益也都为负,且都在 1%的水平上显著。同时,所有控制常见公司特征的多空组合HML的CH4(FF5)调整后的超额收益都为负,且在 1%的水平上显著。这表明在控制了常见的公司特征后,学习指数对股票收益依然具有显著的负向预测能力,学习指数对股票收益的负向预测能力不能被常见的公司特征所解释。

在表 4-5 的 Panel B、Panel C、Panel D 中,也可以观察到类似的结果。在 Panel B 中,控制了BM和STR之后,多空组合HML的超额收益,在 10%的水平上显著为负,在控制了BETA、PROF、INV、IVOL、ILLIQ和LTR之后,多空组合HML的超额收益在 5%的水平上显著为负。对于其他控制变量,多空组合HML的超额收益在 1%的水平上显著为负。同时,所有控制常见公司特征的多空组合HML的CH4(FF5)调整后的超额收益都为负,且在 1%的水平上显著。在 Panel C 中,控制了BM、PROF、INV、IVOL和STR之后,多空组合HML的超额收益在 5%的水平上显著为负。对于其他的控制变量变量,多空组合HML的超额收益在 1%的水平上显著为负,同时,所有控制常见公司特征的多空组合HML的CH4(FF5)调整后的超额收益都为负,且在 1%的水平上显著。在 Panel D 中,在控制了BM和STR之后,多空组合HML的超额收益,在 10%的水平上显著为负,在控制了BETA、PROF、INV、IVOL和ILLIQ之后,多空组合HML的超额收益依旧在 5%的水平上显著为负。在控制了其他变量之后,多空组合HML的超额收益在 1%的水平上显著为负。同时,所有控制常见公司特征的多空组合HML的CH4(FF5)调整后的超额收益都为负,且在 1%的水平上显著。总的来说,Panel B、Panel C、

Panel D 的研究结果在定性上是相似的。

表 4-5 的双变量排序结果表明，在控制了其他公司特征控制变量后，学习指数 LI 对股票收益的负向预测能力依旧是显著的，这意味着 LI 对股票收益的负向预测能力不能被常见的公司特征所解释，同时，多空组合 HML 的多因子模型 $FF5$ 、 $CH4$ 调整后超额收益依旧是负的且都在 1% 的水平上显著，这说明了前文研究结果的稳健性。总而言之，双变量排序的结果与前文单变量排序分析得到的结论是一致的，表明了学习指数对未来股票收益具有显著的负向预测能力。

表 4-5 双变量组合排序

组数	<i>BETA</i>	<i>SIZE</i>	<i>BM</i>	<i>PROF</i>	<i>INV</i>	<i>MOM</i>	<i>IVOL</i>	<i>ILLIQ</i>	<i>STR</i>	<i>LTR</i>
Panel A: 双变量独立排序等权重投资组合										
1(Low)	1.173*	1.175*	1.138*	1.154*	1.153*	1.182*	1.011	1.176*	1.105*	1.160*
	(1.90)	(1.86)	(1.81)	(1.83)	(1.83)	(1.86)	(1.58)	(1.86)	(1.75)	(1.84)
2	0.970	1.018	0.948	0.963	0.967	0.999	0.891	0.983	0.956	0.991
	(1.54)	(1.60)	(1.49)	(1.50)	(1.52)	(1.59)	(1.36)	(1.55)	(1.50)	(1.55)
3	0.910	0.888	0.890	0.902	0.905	0.836	0.802	0.884	0.907	0.889
	(1.41)	(1.38)	(1.38)	(1.40)	(1.40)	(1.31)	(1.22)	(1.37)	(1.41)	(1.38)
4	0.860	0.833	0.880	0.847	0.866	0.721	0.725	0.830	0.884	0.867
	(1.32)	(1.28)	(1.36)	(1.31)	(1.33)	(1.11)	(1.11)	(1.28)	(1.36)	(1.34)
5(High)	0.777	0.675	0.787	0.743	0.755	0.491	0.664	0.712	0.778	0.738
	(1.19)	(1.07)	(1.23)	(1.18)	(1.19)	(0.74)	(1.03)	(1.12)	(1.24)	(1.16)
5-1(HML)	-0.397**	-0.500***	-0.351**	-0.411***	-0.398**	-0.691***	-0.347**	-0.464***	-0.327**	-0.422***
	(-2.57)	(-3.50)	(-2.20)	(-2.58)	(-2.53)	(-4.27)	(-2.39)	(-3.31)	(-2.22)	(-2.60)
FF5- α	-0.652***	-0.646***	-0.602***	-0.670***	-0.658***	-0.817***	-0.600***	-0.667***	-0.596***	-0.674***
	(-3.70)	(-4.02)	(-3.55)	(-3.88)	(-3.75)	(-4.47)	(-3.46)	(-4.14)	(-3.55)	(-3.94)
CH4- α	-0.897***	-0.963***	-0.823***	-0.922***	-0.926***	-0.851***	-0.849***	-0.929***	-0.880***	-0.920***
	(-4.83)	(-5.77)	(-4.33)	(-5.04)	(-4.92)	(-4.38)	(-4.55)	(-5.40)	(-4.97)	(-5.04)

表（续）4-5 双重排序法

组数	<i>BETA</i>	<i>SIZE</i>	<i>BM</i>	<i>PROF</i>	<i>INV</i>	<i>MOM</i>	<i>IVOL</i>	<i>ILLIQ</i>	<i>STR</i>	<i>LTR</i>
Panel B: 双变量独立排序流通市值加权投资组合										
1(<i>Low</i>)	1.054*	1.163*	1.053*	1.027*	1.066*	0.923	0.978	1.155*	1.063*	1.115*
	(1.79)	(1.87)	(1.77)	(1.69)	(1.78)	(1.49)	(1.60)	(1.94)	(1.74)	(1.85)
2	0.680	1.002	0.689	0.665	0.667	0.602	0.571	0.965	0.763	0.747
	(1.15)	(1.58)	(1.15)	(1.09)	(1.12)	(1.00)	(0.94)	(1.62)	(1.26)	(1.24)
3	0.745	0.867	0.697	0.668	0.711	0.577	0.586	0.881	0.777	0.690
	(1.18)	(1.36)	(1.13)	(1.07)	(1.15)	(0.94)	(0.93)	(1.43)	(1.25)	(1.13)
4	0.756	0.830	0.805	0.694	0.817	0.599	0.632	0.887	0.861	0.825
	(1.19)	(1.29)	(1.27)	(1.12)	(1.31)	(0.95)	(1.00)	(1.44)	(1.37)	(1.32)
5(<i>High</i>)	0.637	0.645	0.669	0.582	0.592	0.345	0.549	0.784	0.722	0.606
	(1.02)	(1.03)	(1.10)	(0.97)	(1.00)	(0.55)	(0.88)	(1.29)	(1.21)	(1.01)
5-1(<i>HML</i>)	-0.416**	-0.518***	-0.384*	-0.445**	-0.473**	-0.579***	-0.430**	-0.372**	-0.341*	-0.509**
	(-2.12)	(-3.58)	(-1.88)	(-2.32)	(-2.45)	(-3.15)	(-2.15)	(-2.30)	(-1.78)	(-2.55)
FF5- α	-0.762***	-0.683***	-0.711***	-0.849***	-0.816***	-0.756***	-0.795***	-0.588***	-0.748***	-0.892***
	(-3.50)	(-4.20)	(-3.32)	(-4.11)	(-3.80)	(-3.68)	(-3.37)	(-3.24)	(-3.56)	(-4.35)
CH4- α	-1.068***	-1.013***	-0.925***	-1.180***	-1.117***	-0.854***	-1.102***	-0.853***	-1.071***	-1.174***
	(-3.95)	(-6.03)	(-3.33)	(-4.82)	(-4.31)	(-3.71)	(-3.80)	(-4.41)	(-4.03)	(-4.95)

表(续) 4-5 双重排序法

组数	<i>BETA</i>	<i>SIZE</i>	<i>BM</i>	<i>PROF</i>	<i>INV</i>	<i>MOM</i>	<i>IVOL</i>	<i>ILLIQ</i>	<i>STR</i>	<i>LTR</i>
Panel C: 双变量条件排序等权重投资组合										
1(Low)	1.151*	1.185*	1.143*	1.147*	1.126*	1.184*	1.019	1.137*	1.107*	1.183*
	(1.86)	(1.86)	(1.80)	(1.81)	(1.80)	(1.90)	(1.59)	(1.81)	(1.76)	(1.88)
2	0.949	1.062*	0.940	0.955	0.996	0.986	0.907	1.039	0.945	0.961
	(1.50)	(1.66)	(1.49)	(1.51)	(1.55)	(1.56)	(1.38)	(1.61)	(1.48)	(1.51)
3	0.887	0.861	0.872	0.911	0.872	0.975	0.776	0.899	0.949	0.870
	(1.40)	(1.35)	(1.34)	(1.40)	(1.36)	(1.52)	(1.18)	(1.41)	(1.46)	(1.34)
4	0.889	0.832	0.864	0.847	0.879	0.862	0.731	0.829	0.840	0.832
	(1.36)	(1.28)	(1.36)	(1.32)	(1.36)	(1.33)	(1.11)	(1.28)	(1.31)	(1.29)
5(High)	0.752	0.668	0.799	0.744	0.753	0.604	0.650	0.711	0.790	0.767
	(1.15)	(1.06)	(1.24)	(1.18)	(1.18)	(0.94)	(1.01)	(1.12)	(1.24)	(1.21)
5-1(HML)	-0.399***	-0.517***	-0.344**	-0.403**	-0.373**	-0.581***	-0.369**	-0.426***	-0.317**	-0.416***
	(-2.81)	(-3.72)	(-2.23)	(-2.55)	(-2.43)	(-5.13)	(-2.57)	(-3.17)	(-2.19)	(-2.64)
FF5- α	-0.639***	-0.692***	-0.592***	-0.673***	-0.634***	-0.682***	-0.626***	-0.605***	-0.602***	-0.678***
	(-3.79)	(-4.31)	(-3.59)	(-3.84)	(-3.64)	(-5.12)	(-3.51)	(-3.77)	(-3.67)	(-3.97)
CH4- α	-0.892***	-0.968***	-0.823***	-0.933***	-0.923***	-0.737***	-0.865***	-0.862***	-0.876***	-0.932***
	(-4.91)	(-5.89)	(-4.37)	(-5.09)	(-4.99)	(-4.99)	(-4.52)	(-5.05)	(-4.97)	(-4.92)

表(续) 4-5 双重排序法

组数	<i>BETA</i>	<i>SIZE</i>	<i>BM</i>	<i>PROF</i>	<i>INV</i>	<i>MOM</i>	<i>IVOL</i>	<i>ILLIQ</i>	<i>STR</i>	<i>LTR</i>
Panel D: 双变量条件排序流通市值加权投资组合										
1(Low)	1.017*	1.172*	1.056*	1.017*	1.035*	0.918	0.980	1.111*	1.074*	1.157*
	(1.73)	(1.88)	(1.77)	(1.67)	(1.74)	(1.54)	(1.63)	(1.88)	(1.77)	(1.91)
2	0.676	1.041	0.666	0.693	0.648	0.686	0.560	1.042*	0.730	0.665
	(1.11)	(1.63)	(1.14)	(1.14)	(1.09)	(1.14)	(0.91)	(1.72)	(1.21)	(1.12)
3	0.713	0.840	0.653	0.624	0.681	0.706	0.567	0.852	0.802	0.720
	(1.17)	(1.33)	(1.03)	(1.00)	(1.10)	(1.17)	(0.88)	(1.40)	(1.29)	(1.17)
4	0.692	0.838	0.786	0.732	0.820	0.620	0.623	0.916	0.826	0.777
	(1.08)	(1.30)	(1.26)	(1.18)	(1.32)	(1.00)	(0.99)	(1.47)	(1.34)	(1.25)
5(High)	0.626	0.638	0.704	0.573	0.601	0.486	0.525	0.773	0.738	0.625
	(1.00)	(1.02)	(1.16)	(0.95)	(1.01)	(0.78)	(0.85)	(1.28)	(1.21)	(1.04)
5-1(HML)	-0.390**	-0.535***	-0.351*	-0.445**	-0.433**	-0.432***	-0.455**	-0.337**	-0.336*	-0.532***
	(-2.09)	(-3.78)	(-1.74)	(-2.32)	(-2.23)	(-3.05)	(-2.25)	(-2.15)	(-1.80)	(-2.72)
FF5- α	-0.711***	-0.726***	-0.679***	-0.871***	-0.790***	-0.577***	-0.826***	-0.539***	-0.713***	-0.912***
	(-3.27)	(-4.37)	(-3.24)	(-4.19)	(-3.68)	(-3.66)	(-3.37)	(-2.94)	(-3.52)	(-4.54)
CH4- α	-1.017***	-1.028***	-0.894***	-1.218***	-1.105***	-0.703***	-1.152***	-0.782***	-1.079***	-1.213***
	(-3.87)	(-6.06)	(-3.25)	(-5.15)	(-4.25)	(-3.42)	(-3.91)	(-3.96)	(-4.27)	(-5.09)

4.3.3 Fama-MacBeth 回归

本小节进一步使用 Fama-MacBeth (1973) 回归分析学习指数的对股票收益的负向预测能力。双变量排序法研究了控制单一公司特征后学习指数与股票预期收益的关系，但是双变量排序法无法同时控制多个公司特征的影响。Fama-MacBeth 回归适用于横截面分析，因为它解释了数据中的时间效应（给定月份的残差在不同公司之间是相关的）。Fama-MacBeth 回归的具体步骤如下：

(1) 在每个月末进行如下回归，

$$r_{i,t+1} - r_{f,t+1} = \alpha_{0,t} + \beta_{1,t} LI_{i,t} + \beta_{2,t} X_{i,t} + \varepsilon_{i,t+1} \quad (4-11)$$

其中， $r_{i,t+1} - r_{f,t+1}$ 是股票在 $t + 1$ 月的月度超额收益， $LI_{i,t}$ 是股票 i 在 t 时刻的学习指数， $X_{i,t}$ 是股票 i 在 t 时刻的控制变量。

(2) 在每个月末，重复进行上述横截面回归，可以得到各个变量系数估计值的时间序列。

(3) 计算变量估计值时间序列的均值得到回归系数以及对应的 Newey-West (1987) 调整的 t 统计量。

本节在 Fama-MacBeth 回归的基础范式上，逐步控制 4.3.2 节中使用的公司特征，分析在控制多个变量的情形下，学习指数对股票收益的负向预测能力是否存在。为了便于解释系数的估计值，本文将所有解释变量在当月标准化（均值为 0，标准差为 1）。同时，本节参考 Atilgan 等 (2020) 的做法，同时采用普通最小二乘 (OLS) 和加权最小二乘 (WLS) 两种方式进行回归分析，WLS 方法的权重是上一期的收益率，WLS 方法有助于缓解残差异方差的问题。

表 4-6 报告了不同模型设定下的 Fama-MacBeth 回归的系数估计值以及对应的 Newey-West (1987) 调整的 t 统计量。Panel A 报告了 OLS 回归的结果，Panel B 报告了 WLS 回归的结果。Panel A 的模型 (1) 报告了单变量回归的结果，学习指数 LI 的系数估计值为 -0.112， t 值为 -2.59，在 1% 的水平上显著，这表明在没有控制变量的情形下， LI 与预期收益之间存在显著的负相关关系。在经济意义上， LI 每增加一个标准差，股票的下一期收益下降 0.112%。模型 (2) 到模型 (11) 逐个增加了常见的公司特征作为控制变量，可以观察到 LI 的系数保持为负数，并且都在 1% 的水平上显著。在模型 (11) 中，考虑了所有的控制变量之后， LI 的系数估计值为 -0.088， t 值为 -3.18，在 1% 的水平上显著，这表明即使控制了所有常见的公司特征， LI 依旧对股票横截面收益具有显著的负向预测能力。在经济意义上，保持其他控制变量不变的情况下， LI 每增加一个标准差，股票下一期收益下降 0.088%。同时，在模型 (11) 中，盈利能力 $PROF$ 、动量 MOM 、非流动性 $ILLIQ$ 对股票横截面收益具有显著的正向预测能力，短期回报反转 STR 、特质波动率 $IVOL$ 对股票横截面收益具有显著负向预测能力，这与前人的研究发现一致。

在经济意义上, *IVOL*是所有变量中对股票横截面收益影响最大的变量, 在保持其他变量不变的情况下, *IVOL*每增加 1 个标准差, 下一期的股票收益将会下降 0.504%。

在表 4-6 的 Panel B 中, 可以观察到类似的结果。模型 (1) 是单变量回归的结果, 学习指数 *LI* 的系数估计值为 -0.108, *t* 值为 -2.10, 在 5% 的水平上显著, 这表明在没有控制变量的情形下, *LI* 与股票预期收益之间存在显著的负相关关系。在经济意义上, *LI* 每增加一个标准差, 股票的下一期收益下降 0.108%。模型 (2) 到模型 (11) 逐个增加了常见的公司特征作为控制变量, 可以观察到 *LI* 的系数都至少在 10% 的水平上显著。在模型 (11) 中, 控制了所有的控制变量之后, *LI* 的系数估计值为 -0.066, *t* 值为 -1.83, 在 10% 的水平上显著, 这表明即使控制了其他常见的公司特征, *LI* 依旧对股票横截面收益具有显著的负向预测能力。在模型 (11) 中, 盈利能力 *PROF*、动量 *MOM*、非流动性 *ILLIQ* 对股票横截面收益具有显著的正向预测能力, 短期回报反转 *STR*、特质波动率 *IVOL* 对股票横截面收益具有显著负向预测能力, 这与前人的研究发现一致。在经济意义上, 在保持其他控制变量不变的情况下, *LI* 每增加一个标准差, 股票下一期收益下降 0.066%。在经济意义上, *IVOL* 是所有变量中对股票横截面收益影响最大的变量, 在保持其他变量不变的情况下, *IVOL* 每增加 1 个标准差, 下一期的股票收益将会下降 0.512%。总体而言, Panel B 的结果和 Panel A 的结果在定性上是相似的, 在考虑了所有的常见公司特征控制变量后, 学习指数 *LI* 对股票收益的负向预测能力依旧显著存在。

综上所述, 通过 Fama-MacBeth 回归分析, 进一步说明了在控制了众多可以影响 *LI* 对预期收益的预测能力的公司特征后, 学习指数依旧对股票收益存在显著的负向预测能力, 这个结果与单变量组合排序分析、双变量组合排序分析的结果一致, Fama-MacBeth 回归的结果强化了前文研究的结论。

表 4-6 Fama-MacBeth 回归

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: OLS 回归											
<i>const</i>	0.784 (1.25)	0.795 (1.26)	0.793 (1.25)	0.797 (1.26)	0.798 (1.26)	0.806 (1.27)	0.811 (1.28)	0.814 (1.29)	0.826 (1.30)	0.830 (1.31)	0.841 (1.33)
<i>LI</i>	-0.112*** (-2.59)	-0.121*** (-3.02)	-0.143*** (-4.00)	-0.132*** (-3.70)	-0.127*** (-3.56)	-0.127*** (-3.51)	-0.138*** (-4.50)	-0.118*** (-3.91)	-0.106*** (-3.74)	-0.114*** (-4.02)	-0.088*** (-3.18)
<i>BETA</i>		-0.140* (-1.79)	-0.110 (-1.44)	-0.114 (-1.45)	-0.108 (-1.38)	-0.105 (-1.34)	-0.098 (-1.30)	-0.033 (-0.45)	-0.033 (-0.45)	-0.036 (-0.47)	-0.020 (-0.27)
<i>SIZE</i>			-0.104 (-1.10)	-0.116 (-1.26)	-0.153* (-1.72)	-0.156* (-1.77)	-0.208** (-2.44)	0.062 (0.72)	0.104 (1.18)	0.116 (1.40)	0.045 (0.55)
<i>BM</i>				0.083 (1.30)	0.094 (1.45)	0.097 (1.49)	0.079 (1.29)	0.075 (1.21)	0.095 (1.50)	0.078 (1.29)	0.036 (0.60)
<i>PROF</i>					0.136*** (4.08)	0.131*** (3.86)	0.142*** (4.44)	0.126*** (3.95)	0.109*** (3.26)	0.113*** (3.36)	0.092*** (2.76)
<i>INV</i>						0.034 (1.22)	0.035 (1.28)	0.037 (1.42)	0.019 (0.69)	0.025 (0.92)	0.020 (0.74)
<i>MOM</i>							0.155** (2.13)	0.193*** (2.65)	0.170** (2.33)	0.157** (2.10)	0.276*** (3.76)
<i>ILLIQ</i>								0.422*** (10.07)	0.436*** (9.73)	0.426*** (9.47)	0.385*** (8.62)
<i>STR</i>									-0.392*** (-5.64)	-0.400*** (-5.75)	-0.168*** (-2.34)
<i>LTR</i>										-0.040 (-0.89)	-0.019 (-0.41)
<i>IVOL</i>											-0.504*** (-11.47)
<i>avg. R_{adj}²</i>	0.006	0.028	0.051	0.060	0.063	0.065	0.078	0.083	0.099	0.100	0.107

表(续) 4-6 Fama-MacBeth 回归

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel B: WLS 回归											
<i>const</i>	0.803 (1.27)	0.818 (1.29)	0.818 (1.29)	0.819 (1.29)	0.820 (1.29)	0.823 (1.29)	0.824 (1.30)	0.833 (1.31)	0.937 (1.47)	0.937 (1.47)	0.829 (1.28)
<i>LI</i>	-0.108** (-2.10)	-0.109** (-2.27)	-0.137*** (-3.09)	-0.123*** (-2.74)	-0.121*** (-2.72)	-0.122*** (-2.74)	-0.132*** (-3.30)	-0.112*** (-2.88)	-0.080** (-2.10)	-0.087** (-2.36)	-0.066* (-1.83)
<i>BETA</i>		-0.103 (-1.17)	-0.053 (-0.62)	-0.060 (-0.69)	-0.059 (-0.67)	-0.060 (-0.69)	-0.046 (-0.54)	0.010 (0.12)	0.013 (0.15)	0.015 (0.18)	0.030 (0.35)
<i>SIZE</i>			-0.120 (-1.11)	-0.135 (-1.30)	-0.153 (-1.49)	-0.157 (-1.54)	-0.215** (-2.22)	-0.038 (-0.37)	0.009 (0.08)	0.033 (0.33)	-0.048 (-0.48)
<i>BM</i>				0.141** (2.08)	0.146** (2.15)	0.144** (2.12)	0.127* (1.93)	0.128* (1.94)	0.139** (2.10)	0.108* (1.70)	0.073 (1.12)
<i>PROF</i>					0.105*** (3.54)	0.100*** (3.36)	0.112*** (3.67)	0.102*** (3.40)	0.077** (2.54)	0.093*** (3.04)	0.097*** (3.36)
<i>INV</i>						0.026 (0.91)	0.023 (0.81)	0.027 (0.95)	0.016 (0.57)	0.026 (0.93)	0.017 (0.64)
<i>MOM</i>							0.129 (1.55)	0.153* (1.88)	0.125 (1.55)	0.110 (1.34)	0.240*** (3.04)
<i>ILLIQ</i>								0.302*** (5.95)	0.345*** (7.25)	0.345*** (7.22)	0.292*** (5.95)
<i>STR</i>									-0.418*** (-5.61)	-0.425*** (-5.75)	-0.166** (-2.02)
<i>LTR</i>										-0.083 (-1.55)	-0.068 (-1.32)
<i>IVOL</i>											-0.512*** (-8.67)
<i>avg. R_{adj}²</i>	0.006	0.029	0.053	0.062	0.064	0.066	0.078	0.082	0.098	0.101	0.108

4.3.4 长期预测能力

通过上述研究,本文发现学习指数 LI 对股票的预期收益具有显著的负向预测能力。本节继续检验学习指数对股票收益的负向预测能力是否长期存在,也就是检验 LI 对股票横截面收益的长期预测能力。具体的研究方法是单变量排序法,与前文不同的是,这里将投资组合的持有期限拉长,将投资组合持有1到12个月,观察组合在1到12个月的多空投资组合 HML 的月度平均超额收益以及 $CH4(FF5)$ 调整后的超额收益,从而研究 LI 对股票收益是否具有长期预测能力。

表4-7展示了投资组合持有期为1到12个月的各个组合的超额收益,多空投资组合 HML 的超额收益,以及 $CH4$ 和 $FF5$ 调整后的超额收益。Panel A展示的是等权重投资组合,Panel B展示的是流通市值加权投资组合。在Panel A中,可以看到随着学习指数 LI 的上升,组合的超额收益下降的趋势在短期内依旧存在。例如,在 $t+2$ 列,展示了持有投资组合2个月后,最低 LI 投资组合的月度平均超额收益是0.932%,最高 LI 投资组合的月度平均超额收益是0.501%,最高和最低组合的差异是-0.432%, t 值为-2.60%,在1%的水平上显著。这表明,持有多空投资组合 HML 的2个月内,每个月可以获得平均-0.432%的超额收益。同样的,在 $t+3$ 、 $t+4$ 和 $t+5$ 列,最高和最低组合的差异分别是-0.412%、-0.464%、-0.444%, t 值分别为-2.12、-1.94、-1.74,也都至少在1%的水平上显著。而在组合持有6到12个月后,多空组合 HML 的超额收益不再显著为负,逐渐向0靠近,但是也没有发生反转导致超额收益显著为正。同样的,多空组合 HML 经过 $FF5(CH4)$ 调整后的超额收益也有类似的表现,在持有组合的6(7)个月内,多空投资组合 HML 可以获得不能被模型解释的超额收益,显著性随后消失,超额收益也逐渐缩小,最终接近于0,最终也没有发生反转导致调整后的超额收益显著为正。这表明学习指数对股票横截面收益具有长期持续的预测能力,可以对未来至少5个月的横截面收益具有预测能力,并且这种预测能力会随着时间的推移逐渐消失。

表4-7的Panel B中可以观察到类似的结果,可以看到随着学习指数 LI 的上升,组合的超额收益下降的趋势在短期内依旧存在。在 $t+2$ 列,展示了持有投资组合2个月后,最低 LI 投资组合的月度平均超额收益是0.981%,最高 LI 投资组合的月度平均超额收益是0.373%,最高和最低组合的差异是-0.608%, t 值为-3.52,在1%的水平上显著。这表明,持有多空投资组合 HML 两个月,每个月可以获得平均-0.608%的超额收益。同样的,学习指数的预测能力一致持续到 $t+10$,多空组合 HML 的超额收益是-0.331%,在10%的水平上显著,也就是说在持有多空组合 HML 的10个月内,可以获得平均每个月-0.331%的超额收益,多空组合 HML 的收益不再显著为负,逐渐接近于0,但是也没有发生反转导致超额收益显著为正。同样的,多空组合 HML 经过 $FF5$ 和 $CH4$ 调整后的超额收益也有类似的表现,在持有组合的11个月内,多空投资组合 HML 可以获得不能被模型解释的超额收益,

显著性随后消失，超额收益也逐渐缩小，最终接近于 0，但是也没有发生反转导致调整后的超额收益显著为正。总的来说，Panel B 的结果与 Panel A 的结果在定性上是相似的，这表明学习指数对股票横截面收益具有长期持续的预测能力，可以对未来至少 10 个月的横截面收益具有预测能力，并且这种预测能力会随着时间的推移逐渐消失。

表 4-7 的分析结果表明学习指数 LI 对股票横截面收益的预测能力是长期存在的，并且随着时间的推移逐渐变小，这种预测能力在 1 年内也没有发生显著反转，这与 Cao 等（2020）在美国市场的发现一致。这个结果有力地支持了学习指数 LI 的理论模型（VNV 模型），学习指数 LI 的长期预测能力表明随着时间的推移，在股票市场中，有越来越多的投资者观察到足够多的信息，并根据信息进行投资决策，使得股票的价格逐渐回归到内在价值，逐渐修复了股票的错误定价。

表 4-7 学习指数的长期预测能力

组数	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$	$t+9$	$t+10$	$t+11$	$t+12$
Panel A: 等权重投资组合												
1(Low)	1.147*	0.932	0.808	0.761	0.718	0.675	0.636	0.615	0.551	0.514	0.477	0.461
	(1.83)	(1.64)	(1.53)	(1.52)	(1.51)	(1.49)	(1.49)	(1.51)	(1.42)	(1.38)	(1.33)	(1.35)
2	0.966	0.739	0.699	0.646	0.612	0.608	0.591	0.557	0.533	0.519	0.520	0.518
	(1.52)	(1.30)	(1.31)	(1.28)	(1.28)	(1.35)	(1.38)	(1.38)	(1.39)	(1.42)	(1.49)	(1.56)
3	0.890	0.731	0.646	0.614	0.604	0.581	0.527	0.512	0.507	0.507	0.510	0.504
	(1.38)	(1.24)	(1.21)	(1.23)	(1.29)	(1.33)	(1.27)	(1.31)	(1.35)	(1.42)	(1.49)	(1.54)
4	0.856	0.694	0.591	0.518	0.493	0.452	0.458	0.452	0.459	0.467	0.470	0.473
	(1.32)	(1.19)	(1.12)	(1.08)	(1.11)	(1.09)	(1.17)	(1.21)	(1.30)	(1.39)	(1.48)	(1.56)
5(High)	0.754	0.501	0.396	0.297	0.274	0.259	0.283	0.313	0.335	0.365	0.379	0.395
	(1.18)	(0.89)	(0.78)	(0.63)	(0.63)	(0.63)	(0.73)	(0.86)	(0.97)	(1.12)	(1.21)	(1.33)
5-1(HML)	-0.393**	-0.432***	-0.412**	-0.464*	-0.444*	-0.416	-0.353	-0.303	-0.216	-0.148	-0.098	-0.066
	(-2.45)	(-2.60)	(-2.12)	(-1.94)	(-1.74)	(-1.64)	(-1.48)	(-1.39)	(-1.11)	(-0.83)	(-0.57)	(-0.40)
FF5- α	-0.658***	-0.535***	-0.459**	-0.502**	-0.464*	-0.442*	-0.385	-0.319	-0.232	-0.181	-0.109	-0.064
	(-3.64)	(-3.09)	(-2.28)	(-2.11)	(-1.84)	(-1.74)	(-1.60)	(-1.47)	(-1.22)	(-1.02)	(-0.66)	(-0.40)
CH4- α	-0.943***	-0.604***	-0.436***	-0.440**	-0.401**	-0.387*	-0.344*	-0.279	-0.187	-0.105	-0.031	0.019
	(-4.97)	(-3.84)	(-2.59)	(-2.35)	(-2.15)	(-1.94)	(-1.72)	(-1.55)	(-1.21)	(-0.76)	(-0.23)	(0.13)

表(续) 4-7 学习指数的长期预测能力

组数	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$	$t+9$	$t+10$	$t+11$	$t+12$
Panel B: 流通市值加权投资组合												
1(Low)	1.069*	0.981*	0.888*	0.833*	0.780	0.738	0.691	0.648	0.582	0.545	0.504	0.481
	(1.82)	(1.82)	(1.72)	(1.67)	(1.62)	(1.58)	(1.54)	(1.50)	(1.41)	(1.38)	(1.33)	(1.34)
2	0.596	0.434	0.466	0.455	0.402	0.417	0.403	0.402	0.366	0.351	0.357	0.347
	(1.02)	(0.80)	(0.89)	(0.91)	(0.84)	(0.91)	(0.92)	(0.97)	(0.92)	(0.93)	(0.99)	(1.02)
3	0.698	0.533	0.483	0.444	0.466	0.408	0.387	0.382	0.367	0.372	0.377	0.378
	(1.13)	(0.92)	(0.90)	(0.87)	(0.96)	(0.91)	(0.90)	(0.94)	(0.94)	(1.01)	(1.08)	(1.14)
4	0.791	0.658	0.491	0.421	0.382	0.340	0.327	0.297	0.304	0.302	0.303	0.302
	(1.27)	(1.15)	(0.95)	(0.88)	(0.87)	(0.82)	(0.84)	(0.80)	(0.87)	(0.91)	(0.97)	(1.01)
5(High)	0.525	0.373	0.324	0.194	0.151	0.140	0.146	0.170	0.189	0.214	0.224	0.234
	(0.88)	(0.68)	(0.64)	(0.42)	(0.35)	(0.35)	(0.38)	(0.48)	(0.57)	(0.68)	(0.76)	(0.84)
5-1(HML)	-0.544**	-0.608***	-0.563***	-0.639***	-0.629***	-0.599**	-0.545**	-0.478**	-0.393*	-0.331*	-0.280	-0.247
	(-2.58)	(-3.52)	(-3.14)	(-2.97)	(-2.63)	(-2.39)	(-2.19)	(-2.05)	(-1.87)	(-1.71)	(-1.50)	(-1.39)
FF5- α	-0.929***	-0.797***	-0.677***	-0.757***	-0.701***	-0.657***	-0.607**	-0.513**	-0.415**	-0.365*	-0.298*	-0.261
	(-4.02)	(-4.29)	(-3.68)	(-3.64)	(-3.04)	(-2.65)	(-2.47)	(-2.28)	(-2.07)	(-1.93)	(-1.70)	(-1.57)
CH4- α	-1.286***	-0.935***	-0.696***	-0.725***	-0.648***	-0.610***	-0.579***	-0.482**	-0.396**	-0.328**	-0.267*	-0.215
	(-4.41)	(-4.81)	(-4.03)	(-3.96)	(-3.52)	(-3.02)	(-2.76)	(-2.44)	(-2.24)	(-2.05)	(-1.73)	(-1.35)

4.3.5 样本外检验

本小节检验了考虑学习指数投资组合的 CAPM 模型, 是否比其他模型具有更好的预测效果, 测试数据是学习指数单变量排序的市值加权投资组合的和等权重平均投资组合。使用 60 个月的滚动窗口, 比较 LI-CAPM 与 FF3、FF5、CH3、CH4 对 5 个学习指数 LI 组合收益率的预测差异。例如, 对投资组合 p 的预测误差为,

$$FELICAPM_{p,t} = r_{p,t+1} - \hat{b}_{M,p} E(r_{M,t+1}) - \hat{b}_{LI,p} E(r_{LI,t+1}) \quad (4-12)$$

随后, 将 LI-CAPM 和其他的模型进行比较的结果, 计算样本外 R_{OS}^2 的值, 和对应的 Clark 和 West (2007) 的 p 值。样本外 R_{OS}^2 衡量的是 LI-CAPM 相对于其他模型的均方误差 (MSFE)。

$$R_{OS}^2 = 1 - \frac{MSFE_{LI-CAPM}}{MSFE_{other\ model}} \quad (4-13)$$

若样本外 $R_{OS}^2 > 0$, 那么 $MSFE_{LI-CAPM} < MSFE_{other}$, 同时使用 Clark 和 West (2007) 的 $MSFE$ 调整的 p 值来评估样本外 R_{OS}^2 的显著性。

表 4-8 报告了样本外的 R_{OS}^2 和对应的 Clark 和 West (2007) 的 $MSFE$ 调整的 p 值。Panel A 报告了等权重投资组合的结果, Panel B 报告了流通市值加权投资组合的结果。本文比较的模型包括主流的 FF3、FF5、CH3 和 CH4 等多因子模型, 在 Panel A 中, 所有的样本外的 R_{OS}^2 都大于 0, 对应的 p 值也都小于 0.01, 也就是说 LI-CAPM 的模型的预测误差比其他模型显著更小。同样的, 在 Panel B 中, 可以观察到类似的结果, 所有的样本外的 R_{OS}^2 都大于 0, 对应的 p 值也都小于 0.01, 也就是说 LI-CAPM 的模型的预测误差比其他模型显著更小。Panel A 和 Panel B 的结果在定性上是相似的, 都表明了学习指数 LI 的预测能力难以被其他多因子模型所解释, 同时也表明了学习指数在资产定价模型中的重要性。

表 4-8 样本外检验

模型	1	2	3	4	5
Panel A: 等权重投资组合					
FF3	0.215 (0.00)	0.300 (0.00)	0.293 (0.00)	0.127 (0.00)	0.076 (0.00)
FF5	0.353 (0.00)	0.505 (0.00)	0.510 (0.00)	0.287 (0.00)	0.201 (0.00)
CH3	0.224 (0.00)	0.402 (0.00)	0.402 (0.00)	0.226 (0.00)	0.173 (0.00)
CH4	0.289 (0.00)	0.482 (0.00)	0.486 (0.00)	0.296 (0.00)	0.236 (0.00)
Panel B: 流通市值加权投资组合					
FF3	0.061 (0.00)	0.072 (0.00)	0.057 (0.00)	0.085 (0.00)	0.030 (0.00)
FF5	0.191 (0.00)	0.385 (0.00)	0.291 (0.00)	0.314 (0.00)	0.170 (0.00)
CH3	0.079 (0.00)	0.110 (0.00)	0.112 (0.00)	0.201 (0.00)	0.092 (0.00)
CH4	0.137 (0.00)	0.232 (0.00)	0.188 (0.00)	0.272 (0.00)	0.152 (0.00)

4. 4 学习指数对股票收益预测能力的来源

4. 4. 1 学习指数与信息流变量的关系

本文在这一小节中检验学习指数 LI 与信息流代理变量的之间的关系。投资者学习的过程需要获取信息，如果学习指数代表了投资者学习的过程，那么学习指数应该与市场上一些和信息流相关的变量具有相关关系。本小节使用的信息流相关变量包括：公司相关的新闻数量（ $News$ ）、公司公告的数量（ $Announce$ ）、分析师修正次数（ $nREV$ ）、机构投资者持有比例（ IO ）、投资者异常交易活动（ $ATURN$ ）。

公司相关的新闻数量（ $News$ ）和公司公告的数量（ $Announce$ ）代表了可能被投资者了解的公司相关的信息的数量。分析师修正次数（ $nREV$ ）计算为分析师对最近一个年度的收益预测的修正次数，分析师是金融市场中非常重要的信息中介，分析师的修正自己的预测代表了分析师更新自己的信念的过程，这个过程就是分析师在股票市场上获取信息，并根据对这些信息进行处理和分析，做出对股票收益的预测，这代表了分析师对股票的学习过程。机构投资者是金融市场上

的专业投资者,通常能够获取更多的信息、不容易受非理性行为的影响,那么机构投资者持有比例(*IO*)更高的股票应该具有更高的学习指数。*Barber*和*Odean* (2007)认为,随着投资者了解到公司的新信息,交易活动可能会增加,导致异常的交易活动。投资者交易活动计算为股票的换手率,股票异常交易活动(*ATURN*)可以计算为换手率除以前12个月的平均换手率,随后减去1,最后乘以100。

在现实的股票市场中,投资者更难获取小公司的相关信息,因为小公司信息透明度更低,*Hong*等(2000)发现公司规模 $SIZE$ 是决定分析师覆盖情况的最显著的影响因素。因此,本文基于 $SIZE$ 和 LI 使用双变量条件排序分析投资者学习与信息流代理变量的关系。这种方法可以在控制企业信息透明度(也就是控制 $SIZE$)的影响下,研究学习指数与信息流相关变量之间的关系,验证学习指数是否代表了投资者学习的大小。具体而言,在每个月末,本文将股票基于公司规模 $SIZE$ 分为5组,再在每个组内,使用学习指数 LI 将股票分为5组,得到25个投资组合,再求组合中各个信息流代理变量的等权重均值,最后计算每个 LI 组合中的5个 $SIZE$ 组合的信息流代理变量均值的均值,并计算最高 LI 组合和最低 LI 组合的差异 HML 。在每个月末重复上述步骤,得到5个 LI 组合与 HML 的时间序列,最后计算时间序列的均值和相应的Newey-West(1987)调整的 t 统计量。条件双重排序可以保证每个组合中的 $SIZE$ 变量有相似的分布,得到的结果不受公司规模 $SIZE$ 的影响。

表4-9展示了学习指数 LI 排序的五个信息流相关变量的组合均值。表4-9的第1列和第2列展示了 LI 和公司相关的新闻数量(*News*)和公司公告的数量(*Announce*)之间的关系。低(高) LI 组合的新闻数量是4.191(4.363),高 LI 组合和低 LI 组合的差异是0.172, t 值是3.17,在1%的水平上显著。同样的,低(高) LI 组合的公告数量是5.238(5.425),高 LI 组合和低 LI 组合的差异是0.187, t 值是1.95,在10%的水平上显著。这表明当月学习指数越高的公司在有更多的新闻和公告,也就是说学习指数越高的公司在有更多可能被投资者了解的信息。第3列展示了分析师修正次数(*nREV*)和 LI 的关系,低(高) LI 组合的分析师修正数量分别是2.433(2.645),低 LI 组合和高 LI 组合的差异是0.212, t 值是1.96,在5%的水平上显著。这表明当月学习指数越高的公司有更多的分析师修正,也就是说有更多的分析师对股票进行了学习。第4列展示了 LI 与机构投资者持有比例(*IO*)之间的关系,低(高) LI 组合的机构投资者持有比例分别是36.620%(38.292%),高 LI 组合与低 LI 组合的差异是1.672%, t 值是4.33,在1%的水平上显著。这表明学习指数越高的公司,机构投资者的持有比例越高,有可能有更多的机构投资者通过学习将该公司纳入投资组合。最后一列展示了学习指数 LI 与异常交易活动(*ATURN*)的关系。低(高) LI 股票的异常交易活动增加了7.910%

(10.740%)。高 LI 和低 LI 组合之间的异常交易活动的差异为2.830%, t 值为2.18, 在5%的水平上显著。这种差异表明, 在学习指数更高股票中, 股票的异常交易活动的程度更高, 也就是说, 有更多的投资者获取到了信息并根据信息进行交易。

总的来说, 表4-9关于学习指数与信息流相关变量的研究发现学习指数 LI 的确与一些信息流的代理变量正相关, 也就是说学习指数越高的股票中通常有更多信息和学习, 这与VNV模型的经济意义一致。

表4-9 学习指数和信息流相关的变量

组数	<i>News</i>	<i>Announce</i>	<i>nREV</i>	<i>IO</i>	<i>ATURN</i>
1(Low)	4.191*** (9.88)	5.238*** (7.50)	2.433*** (13.80)	36.620*** (17.18)	7.910 (1.64)
2	4.291*** (9.99)	5.364*** (7.59)	2.529*** (14.15)	37.078*** (17.21)	10.404** (2.22)
3	4.334*** (10.01)	5.317*** (7.53)	2.644*** (14.62)	37.498*** (17.23)	11.338** (2.44)
4	4.378*** (9.97)	5.425*** (7.57)	2.668*** (14.16)	37.795*** (17.11)	11.271** (2.49)
5(High)	4.363*** (9.86)	5.425*** (7.49)	2.645*** (13.90)	38.292*** (16.99)	10.740** (2.55)
5-1(HML)	0.172*** (3.17)	0.187* (1.95)	0.212** (1.96)	1.672*** (4.33)	2.830** (2.18)

4.4.2 两阶段分解

上一小节的研究发现学习指数与信息数量、机构投资者的持有比例、投资者异常交易活动呈正相关关系。这可能导致学习指数的预测能力有四个不同的假设: 第一, 学习指数只是一个信息数量的代理变量而不是代表投资者对股票的学习, 新闻的数量通常和股票收益率呈负相关关系 (Fang 和 Peres, 2019; Heston 和 Sinha, 2017); 第二, 学习指数代表了机构投资者的学习, 但是机构投资者没有能够通过降低股票的风险, 而是机构投资者持有比例更高的股票更容易导致股价崩盘 (许年行等, 2013; 曹丰等, 2015; 吴晓晖等, 2019), 导致股票未来收益降低。第三, 学习指数可能是一个投资者关注度的代理变量, Liu 等 (2019) 换手率高的股票往往意味着更多的非理性交易以及被舆情驱动的交易者过度关注的股票, 频繁交易往往造成这些股票近期价格的虚高, 而由于中国股票市场的做空限制, 又导致了空头投资者无法利用这种非理性行为修正股票价格, 最终导致股票未来收益率的下降。第四, 学习指数符合 VNV 模型的推断, 具有更高学习指

数的股票被投资者学习得越多, 股票未来的风险越小, 股票未来的收益越低。

本节参考 Boehmer 等 (2021) 使用的方法, 使用两阶段分解来区分方程中学习指数与对下期股票收益预测能力的四个不同的假设。本节使用股票相关新闻数量 (*News*), 投资者的异常交易活动 (*ATURN*), 机构投资者持有比例 (*IO*) 三个变量对学习指数 *LI* 的预测能力进行分解。同时, 保持和前文一致, 在进行两阶段分解前, 将所有变量标准化 (均值为 0, 方差为 1)。两阶段分解的具体步骤如下,

(1) 在第一阶段, 在 t 月, 将 t 月的学习指数 *LI*, 分解为四个分量。具体来说, 在 t 月, 进行一次横截面回归,

$$LI_{i,t} = \lambda_{0,t} + \lambda_{1,t}News_{i,t} + \lambda_{2,t}IO_{i,t} + \lambda_{3,t}ATURN_{i,t} + \varepsilon_{i,t} \quad (4-14)$$

在每个月末, 重复进行横截面回归, 可以得到横截面回归系数估计值的时间序列 $\{\lambda_{0,t}, \lambda_{1,t}, \lambda_{2,t}\}$, 接下来, 按照下面的公式计算学习指数 $LI_{i,t}$ 的四个成分,

$$\begin{aligned} LI_{i,t}^{News} &= \lambda_{1,t}News_{i,t}, \\ LI_{i,t}^{IO} &= \lambda_{2,t}IO_{i,t}, \\ LI_{i,t}^{ATURN} &= \lambda_{3,t}ATURN_{i,t}, \\ LI_{i,t}^{other} &= \lambda_{0,t} + \varepsilon_{i,t}. \end{aligned} \quad (4-15)$$

通过上面两式, 可以知道:

$$LI_{i,t} = LI_{i,t}^{News} + LI_{i,t}^{IO} + LI_{i,t}^{ATURN} + LI_{i,t}^{other} \quad (4-16)$$

定义与新闻相关部分与第一个假设相关, 定义机构投资者持有相关部分与第二个假设相关、异常交易活动相关的部分与第三个假设相关。在考虑到上述三个部分以后, 最后一个部分就是投资者学习对股票横截面收益的预测能力的部分。

(2) 第二阶段, 使用四个成分替代 *LI* 进行 Fama-MacBeth 回归,

$$r_{i,t+1} - r_f = \lambda_0 + \lambda_1 LI_{i,t}^{News} + \lambda_2 LI_{i,t}^{IO} + \lambda_3 LI_{i,t}^{ATURN} + \lambda_4 LI_{i,t}^{other} + \lambda_5 controls_{i,t} + \varepsilon_{i,t+1} \quad (4-17)$$

由于将原始的学习指数分解 *LI* 为新闻数量相关的部分、机构投资者相关的部分、投资者异常交易量相关的部分和残差的四个部分, 因此上式中的系数揭示了 *LI* 的每个组成部分对未来股票收益的贡献。两阶段分解的优势在于, 它在统一且内部一致的实证框架中包含了来自投资者学习预测能力的其他几个替代假设的学习指数 *LI* 的分量。

表 4-10 中报告了两阶段分解的回归系数和对应的 Newey-West (1987) 调整 t 统计量, 以及核心变量从 25% 分位数变化到 75% 的变化量, 以及这个变化量对应的收益率变化量。Panel A 展示了第一阶段的 Fama-Macbeth 回归的结果, Panel B 展示了第二阶段的 Fama-Macbeth 回归结果。在面板 A 中, 新闻数量、机构投资者持有比例、投资者异常交易活动都与学习指数正相关, 系数估计值分别为

0.026、0.028、0.016, t 统计量分别为 4.96、4.48、2.42, 分别在 1%、1%、5% 的水平上显著, 这与前文的研究结果相似。本文基于与前一节中相同的原因在第一阶段回归中控制了公司规模 $SIZE$, 同时, 由于很多学者发现动量效应可能与机构投资者参与相关 (Chui 等, 2022; Du 等, 2022; Hameed 等, 2023), 所以将股票的动量 MOM 也作为第一阶段横截面回归的控制变量。

表 4-10 的 Panel B 报告了使用学习指数的预测能力与信息数量相关的部分、与机构投资者相关的部分、与投资者异常交易活动相关的部分、其他部分放在一起预测未来股票收益的结果。 LI^{News} 的系数估计值为 71.354, t 统计量为 -1.10, 意味着不能拒绝原假设, 学习指数与信息数量相关的部分对学习指数的预测能力没有显著的贡献。 LI^{IO} 的系数估计值为 25.254, t 统计量为 1.24, 这也意味着不能拒绝原假设, 也就是说与机构投资者相关的预测能力对学习指数的预测能力没有显著贡献。同样的, LI^{ATURN} 的系数估计值是 0.646, t 统计量为 0.12, 这也意味着与异常活动相关的部分对学习指数的预测能力没有显著贡献。最后, 学习指数预测能力的其他部分 LI^{other} , 系数估计值为 -0.126, t 统计量为 -3.18, 这表明投资者学习对学习指数的预测能力有显著贡献。

表 4-10 的 Panel B 还展示了学习指数预测能力的所有四个组成部分在横截面上的从 25% 分位数增加到 75% 分位数的变化量和对应的收益率变化量。如果 LI^{other} 从 25% 分位数增加到 75% 分位数, 未来一个月股票超额收益的变化 -0.115%。对于其他三个部分, 符号是正的, 没有统计学意义。

总体而言, 两阶段分解结果展示了学习指数的预测能力与其他可能的假设无关, 学习指数的预测能力来源于投资者学习, 学习指数对股票收益具有显著的负向预测能力, 这符合 VNV 模型的推论。

表 4-10 两阶段分解

	<i>coef</i>	<i>t-stat</i>	<i>Int'Quartile change</i>	<i>return_diff</i>
Panel A: 第一阶段 FM 回归				
<i>const</i>	0.000	(0.22)		
<i>News</i>	0.026***	(4.96)		
<i>IO</i>	0.028***	(4.48)		
<i>ATURN</i>	0.016**	(2.42)		
<i>SIZE</i>	-0.077***	(-5.95)		
<i>MOM</i>	-0.043	(-1.09)		
<i>avg. R_{adj}²</i>	0.110			
Panel B: 第二阶段 FM 回归				
<i>const</i>	0.761	(1.11)		
<i>LI^{News}</i>	71.354	(1.10)	0.038	2.738
<i>LI^{IO}</i>	25.254	(1.24)	0.078	1.959
<i>LI^{ATURN}</i>	0.646	(0.12)	0.058	0.038
<i>LI^{Other}</i>	-0.126***	(-3.18)	0.912	-0.115
<i>SIZE</i>	-0.126	(-1.16)		
<i>MOM</i>	0.228***	(2.81)		
<i>LTR</i>	-0.067	(-1.34)		
<i>BM</i>	0.037	(0.42)		
<i>STR</i>	-0.152*	(-1.67)		
<i>IVOL</i>	-0.338***	(-4.96)		
<i>ILLIQ</i>	0.175**	(2.15)		
<i>INV</i>	0.049	(1.64)		
<i>PROF</i>	0.151***	(4.79)		
<i>BETA</i>	-0.104	(-0.93)		
<i>avg. R_{adj}²</i>	0.116			

4.5 稳健性检验

在前文研究中,本文对学习指数对股票横截面收益的负向预测能力进行了深入细致的研究,本节通过更换不同的多因子模型、更换排序组数、在回归中增加额外的控制变量、更换两阶段分解的控制变量等多种不同的方法进行稳健性检验,以保证研究结果稳健性。

4.5.1 更换多因子模型

这一节使用其他的多因子模型来计算单变量排序得到的投资组合的风险调整后超额收益,以确保学习指数多空组合的超额收益不是来源于其他已知的多因子模型。本节考虑的模型包括,Fama 和 French (2018) 提出的动量增强的六因子模型 FF5-MOM、Bali 等 (2017) 使用的六因子模型 FMAX6、Hou 等 (2015) 提出的 Q4 模型、Stambaugh 和 Yuan (2017) 的错误定价四因子模型 M4、Danieal 等 (2019) 的行为三因子模型 BF3 以及 Daniel 等 (1997) 提出的 DGTW 特征基准。

表 4-11 展示了使用其他的多因子模型计算风险调整后的超额收益, Panel A 展示了等权重投资组合的结果, Panel B 展示了流通市值加权投资组合的结果。例如,在 Panel A 中,第 1 列展示了学习指数 *LI* 投资组合在 FF5-MOM 调整后的超额收益,*LI* 多空组合 *HML* 的超额收益为 -0.588%, *t* 值为 -3.15, 在 1% 的水平上显著,这表明 *LI* 多空组合 *HML* 的超额收益不能被 FF5-MOM 所解释。同样的,剩余的列分别展示了 FMAX6、Q4、M4、BF3 和 DGTW 调整后的超额收益,在所有的列中,*LI* 多空组合 *HML* 的在其他的多因子模型调整后的超额收益都为负,且至少在 10% 的水平上显著,这表明 *LI* 多空组合 *HML* 的超额收益不能被这些多因子模型所解释。在 Panel B 中,可以观察到类似的结果,在考虑其他常见的多因子模型的情况之后,*LI* 多空组合 *HML* 的超额收益依旧为负,并且至少在 10% 的水平上显著。Panel A 和 Panel B 的结果在定性上与前文的研究结果是相似的,这些结果表明,学习指数 *LI* 多空组合 *HML* 的超额收益无法被多因子模型所解释,学习指数 *LI* 对股票收益具有显著的负向预测能力,这说明了前述研究的稳健性。

表 4-11 更换其他的多因子模型

组数	FF5-MOM- α	FMAX6- α	Q4- α	M4- α	BF3- α	DGTW- α
panel A: 等权重投资组合						
1(Low)	0.155 (0.93)	0.213 (1.23)	-0.029 (-0.17)	-0.008 (-0.04)	0.467** (2.17)	-0.875*** (-7.85)
2	-0.107 (-0.69)	-0.066 (-0.47)	-0.241* (-1.68)	-0.229 (-1.54)	0.312 (1.46)	-1.102*** (-10.92)
3	-0.270* (-1.86)	-0.178 (-1.47)	-0.303** (-2.27)	-0.330** (-2.52)	0.221 (1.06)	-1.260*** (-12.35)
4	-0.340** (-2.43)	-0.352*** (-2.89)	-0.327** (-2.22)	-0.362*** (-2.97)	0.196 (0.89)	-1.338*** (-12.67)
5(High)	-0.433*** (-2.64)	-0.488*** (-3.51)	-0.423*** (-2.60)	-0.449*** (-3.35)	0.161 (0.68)	-1.451*** (-12.05)
5-1(HML)	-0.588*** (-3.15)	-0.702*** (-3.19)	-0.394** (-2.13)	-0.441** (-2.13)	-0.306* (-1.74)	-0.576*** (-4.31)
panel B: 流通市值加权投资组合						
1(Low)	0.400** (2.38)	0.360* (1.82)	0.188 (1.01)	0.199 (0.96)	0.349** (2.11)	-0.789*** (-5.49)
2	-0.124 (-0.79)	-0.180 (-1.11)	-0.260* (-1.88)	-0.323** (-2.02)	-0.036 (-0.25)	-1.163*** (-8.40)
3	-0.229 (-1.56)	-0.128 (-0.92)	-0.199 (-1.50)	-0.323*** (-2.59)	0.024 (0.16)	-1.088*** (-10.34)
4	-0.114 (-0.72)	-0.157 (-1.08)	-0.088 (-0.60)	-0.218* (-1.85)	0.127 (0.72)	-1.081*** (-10.04)
5(High)	-0.424*** (-2.64)	-0.549*** (-3.83)	-0.364** (-2.31)	-0.456*** (-3.55)	-0.078 (-0.42)	-1.167*** (-8.98)
5-1(HML)	-0.824*** (-3.66)	-0.909*** (-3.55)	-0.553** (-2.35)	-0.655** (-2.52)	-0.427* (-1.94)	-0.378** (-1.98)

4.5.2 更换其他的排序组数

常见的单变量排序的分组数量有 5 组和 10 组，在前文的研究中本文根据学习指数 LI 的大小将股票分为 5 组进行研究。为了确保学习指数的预测能力不受分组数量的影响，在每个月末，使用单变量排序根据 LI 的大小将股票分为 10 组持有一个月，并计算多空组合的超额收益和风险调整后的超额收益。

表 4-12 展示了单变量排序组数为 10 组的结果, Panel A 展示了等权重投资组合, Panel B 展示了流通市值加权投资组合。在 Panel A 中, Panel A 的第 1 列的最后一行表示了多空组合 *HML* 的超额收益为 -0.520%, *t* 值为 -2.75, 在 1% 的水平上显著。除此之外, 在剩余几列中可以看到多空组合 *HML* 经过 CAPM、FF3、FF5、CH3 和 CH4 调整后的超额收益也都是显著为负的, 并且至少在 5% 的水平上显著。在 Panel B 中, 可以观察到类似的结果, 多空组合 *HML* 的超额收益和经过 CAPM、FF3、FF5、CH3 和 CH4 调整后的超额收益也都是显著为负的, 并且至少在 5% 的水平上显著。Panel A 和 Panel B 的结果与前文的研究结果在定性上是相似的, 也就是说学习指数多空组合 *HML* 的超额收益无法被主流的多因子模型所解释, 学习指数对股票横截面收益具有负向预测能力。被这表明了前述研究结果对单变量排序的分组的数量并不敏感, 表明了前述研究的稳健性。

表 4-12 更换排序的组数为 10 组

组数	<i>Exret</i>	CAPM- α	FF3- α	FF5- α	CH3- α	CH4- α
panel A: 等权重投资组合						
1(Low)	1.233** (1.96)	0.456** (1.97)	0.177 (0.98)	0.413** (2.26)	0.706*** (4.44)	0.869*** (5.48)
5	0.974 (1.48)	0.205 (0.86)	-0.166 (-1.33)	-0.117 (-0.68)	-0.011 (-0.10)	0.081 (0.74)
10(High)	0.713 (1.12)	0.037 (0.14)	-0.397** (-2.54)	-0.437** (-2.28)	-0.348** (-2.40)	-0.337** (-2.31)
10-1(<i>HML</i>)	-0.520*** (-2.75)	-0.419** (-2.05)	-0.574** (-2.40)	-0.850*** (-3.86)	-1.054*** (-4.76)	-1.205*** (-5.12)
panel B: 流通市值加权投资组合						
1(Low)	1.273** (2.12)	0.506** (2.56)	0.514** (2.21)	0.821*** (3.99)	0.886*** (3.83)	1.079*** (4.29)
5	0.722 (1.16)	-0.019 (-0.12)	-0.144 (-0.96)	-0.135 (-0.77)	-0.129 (-0.81)	-0.066 (-0.41)
10(High)	0.579 (0.98)	-0.059 (-0.31)	-0.286* (-1.70)	-0.324* (-1.84)	-0.343** (-2.06)	-0.397** (-2.17)
10-1(<i>HML</i>)	-0.694*** (-2.82)	-0.565** (-2.15)	-0.800*** (-2.69)	-1.144*** (-4.34)	-1.229*** (-3.98)	-1.475*** (-4.27)

4.5.3 增加一些额外的控制变量

本节在前文的 Fama-MacBeth 回归分析中, 增加了一些而额外的控制变量, 并保持前文的 Fama-Macbeth 回归使用的控制变量不变, 确保学习指数 *LI* 对股票

收益的预测能力不被其他的公司特征所解释,以保证前文 Fama-MacBeth 回归分析结果的稳健性。这些控制变量都在前人的研究中被发现可以预测股票横截面预期收益,本节增加的控制变量包括,George 和 Hwang (2004)提出的 52 周高点 $52MAX$, Bali 等(2016)提出的总偏度 $SKEW$ 、特质偏度 $ISKEW$, Harvey 和 Siddique (2000)提出的协偏度 $COSKEW$ 、Ang 等(2006)提出的下行贝塔 $DBETA$ 。保持和前文的一致性,本节的分析中也同时使用了 OLS 回归和 WLS 回归。

表 4-13 展示了增加了额外的控制变量之后的 Fama-MacBeth 回归分析,的 Panel A 展示了 OLS 回归的结果, Panel B 展示了 WLS 回归的结果。在 Panel A 中可以观察到,逐渐增加额外的控制变量之后,学习指数 LI 的系数估计值依旧为负,且至少在 1% 的水平上显著,模型(5)是考虑了所有额外控制变量的回归模型, LI 的系数估计值为 -0.093, t 值为 -2.81, 在 1% 的水平上显著,这些结果表明在增加了一系列的控制变量之后, LI 依旧对股票横截面收益具有显著的负向预测能力。在 Panel B 中可以观察到类似的结果,在控制了额外的控制变量之后, LI 回归系数依旧为负,并且至少在 10% 的水平上显著。Panel A 和 Panel B 的结果在定性上和前文的研究结果是相似的,这表明学习指数具有无法被常见公司特征解释的负向预测能力,这些结果表明的研究结果是稳健的。

表 4-13 增加额外的控制变量

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: OLS					Panel B: WLS					
<i>const</i>	0.836 (1.29)	0.837 (1.29)	0.841 (1.30)	0.842 (1.30)	0.842 (1.30)	0.832 (1.28)	0.833 (1.28)	0.837 (1.29)	0.837 (1.29)	0.838 (1.29)
<i>LI</i>	-0.091*** (-2.68)	-0.094*** (-2.76)	-0.090*** (-2.70)	-0.090*** (-2.75)	-0.093*** (-2.81)	-0.067* (-1.86)	-0.070* (-1.95)	-0.066* (-1.87)	-0.066* (-1.90)	-0.070** (-2.02)
<i>52MAX</i>	0.020 (0.96)	0.019 (0.93)	0.019 (0.89)	0.018 (0.87)	0.017 (0.82)	0.024 (1.08)	0.023 (1.02)	0.022 (0.95)	0.021 (0.94)	0.021 (0.91)
<i>SKEW</i>		-0.067* (-1.82)	-0.075** (-2.04)	-0.015 (-0.29)	-0.018 (-0.36)		-0.070* (-1.84)	-0.077** (-2.05)	-0.017 (-0.34)	-0.021 (-0.41)
<i>COSKEW</i>			0.280*** (4.45)	0.278*** (4.47)	0.288*** (4.56)			0.264*** (4.18)	0.261*** (4.19)	0.275*** (4.29)
<i>ISKEW</i>				-0.094** (-2.14)	-0.088** (-2.02)				-0.099** (-2.19)	-0.092** (-2.06)
<i>DBETA</i>					0.061 (0.96)					0.051 (0.78)
<i>Controls</i>	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
<i>avg. R_{adj}²</i>	0.105	0.107	0.112	0.114	0.115	0.108	0.110	0.115	0.116	0.118

4.5.4 更换两阶段分解中使用的信息变量

在两阶段分解中,信息数量的变量除了公司相关的新闻数量 $News$ 还有公司相关的公告数量 $Announce$ 。为了确保研究的稳健性,本节使用 $Announce$ 替代 $News$ 作为两阶段分解中代表的信息数量变量。

表 4-14 展示了使用 $Announce$ 替代 $News$ 的两阶段分解的结果。Panel A 展示了第一阶段 Fama-MacBeth 回归的结果,Panel B 展示了第二阶段 Fama-MacBeth 回归的结果。可以看到,在 Panel B 中, $Other$ 成分的系数估计值为-0.231, t 值为-2.12,在 5%的水平上显著。其他部分都不显著,这个结果和前文的研究结果在定性上是相似的,也就是说,只有投资者学习这个部分对学习指数的预测能力产生了贡献,这表明了前文研究的稳健性。

表 4-14 使用 $Announce$ 代替 $News$ 进行两阶段分解

	<i>coef</i>	<i>t-stat</i>	<i>Int'Quartile change</i>	<i>return_diff</i>
Panel A: 第一阶段 FM 回归				
<i>const</i>	-0.000	(-0.67)		
<i>Announce</i>	0.027***	(3.61)		
<i>IO</i>	0.036***	(3.71)		
<i>ATURN</i>	0.009	(1.08)		
<i>SIZE</i>	-0.076***	(-4.42)		
<i>MOM</i>	-0.074	(-1.64)		
<i>avg.R_{adj}²</i>	0.124			
Panel B: 第二阶段 FM 回归				
<i>const</i>	1.186	(1.64)		
<i>LI^{Announce}</i>	21.110	(0.99)	0.033	0.692
<i>LI^{IO}</i>	-73.797	(-0.86)	0.104	-7.653
<i>LI^{ATURN}</i>	-29.938	(-0.86)	0.071	-2.139
<i>LI^{Other}</i>	-0.231**	(-2.12)	0.937	-0.217
<i>SIZE</i>	0.128	(0.64)		
<i>MOM</i>	0.218**	(2.05)		
<i>LTR</i>	-0.266	(-1.26)		
<i>BM</i>	0.075	(0.70)		
<i>STR</i>	-0.189*	(-1.66)		
<i>IVOL</i>	-0.467***	(-3.43)		
<i>ILLIQ</i>	0.256**	(2.53)		
<i>INV</i>	0.040	(0.52)		
<i>PROF</i>	0.063	(0.78)		
<i>BETA</i>	0.050	(0.39)		
<i>avg.R_{adj}²</i>	0.112			

第五章 学习指数与错误定价的关系

上一章分析了学习指数对股票横截面收益的影响,研究发现学习指数越高的股票未来的收益越低。前文对学习指数长期预测能力的分析发现,随着时间推移,越来越多的投资者发现信息,最终使得股票价格回归内在价值,错误定价逐渐变小。因此,本章首先研究学习指数与错误定价之间的关系,分析学习指数对股票错误定价的影响,考察被低估和被高估股票之间学习指数对股票收益预测能力的差异,随后本章分析了学习指数与错误定价这种关系的成因,探讨了套利限制和投资者情绪在这种关系中的作用。

5.1 数据来源和变量说明

5.1.1 数据来源与样本选择

选取中国 A 股市场的所有上市公司作为实证研究的对象。所有使用的股票日度、月度交易数据,财务报表数据、Fama-French 多因子模型数据,无风险利率数据、投资者情绪数据、分析师数据、机构投资者持有数据等均来自国泰安数据库。Liu 等(2019)提出的中国市场三因子(CH3)、四因子(CH4)的数据来源于 Robert F. Stambaugh 的个人网站。

为了减少市场微观结构问题和小市值股票的影响,在每个月末,剔除公司规模小于沪深主板股票市值 20%分位数的股票和价格小于 5 元的股票。为了排除次新股的影响,剔除上市不到 1 年的股票。为了防止结果受到长时间停牌后的股票收益的影响,在每个月末,样本中包含的股票在过去一年至少有 120 天的交易记录,过去 20 个正常交易日至少有 15 天的交易记录。为了保持财务报表的一致性,剔除了金融行业股票。本章主要分析的时间区间从 2000 年 1 月 1 日到 2021 年 12 月 31 日。中国 A 股市场始于 1990 年上海和深圳交易所的建立,关注 2000 年之后的时期主要是为了确保财务报表数据的一致性,中国有关财务报告各个方面的规章制度的实施直到 1999 年才基本成型。

为了防止极端值对研究结果造成影响,本文对所有的研究变量在每个时间截面上进行截尾,剔除了小于 0.1%分位数和大于 99.9%分位数的数据,以保证研究结果的稳健性。

5.1.2 变量说明

本章的核心是研究学习指数与错误定价之间的关系。研究股票错误定价的难点在于错误定价本身是不可估计的,需要寻找合适的代理变量。Stambaugh 等(2015)使用 11 个不能被主流多因子模型解释的市场异象在美国市场构建了错误定价的代理指标。这 11 个异象包括:Fama 和 French(2008)发现股票净发行量(Net Stock Issues)与股票横截面预期收益负相关;Daniel 和 Titman(2006)

发现复合股权发行量 (Composite Equity Issues) 与股票横截面预期收益负相关; Sloan (1996) 发现应计利润 (Accruals) 与股票横截面预期收益负相关; Hirshleifer 等 (2004) 发现净营业资产 (Net Operating Assets) 与股票横截面预期收益负相关; Cooper 等 (2008) 发现总资产增长率 (Asset Growth) 与股票横截面预期收益负相关; Titman 等 (2004) 发现投资与总资产之比 (Investment to Asset) 与股票横截面预期收益负相关; Campbell 等 (2008) 构造了一个财务困境 (Distress) 指标, 与股票横截面预期收益负相关; Ohlson (1980) 构造了一个 O-分数 (O-score), 与股票横截面预期收益负相关; Jegadeesh 和 Titman (1993) 发现动量 (Momentum) 与股票横截面收益正相关; Novy 和 Marx (2013) 发现毛利润 (Gross Profitability) 与股票横截面预期收益正相关; Chen 等 (2010) 提出的总资产回报率 (Return of Assets) 与股票横截面预期收益正相关。由于中国市场的数据可得性与异象的显著性, Gu 等 (2018), 使用其中 6 个异象在中国股票市场构建了错误定价的代理变量 *MISP*, 这些异象包括应计利润 (ACC)、净营运资产 (NetOA)、毛利率 (GP)、股权回报率 ROE、动量 (MOM)、复合股权发行量 (CEI)。

本文参考 Stambaugh 等 (2015) 和 Gu 等 (2018) 使用 6 个异象构建错误定价指标, 用于计算错误定价的 6 个基础异象指标的计算方式如下: 应计利润 (ACC) 计算为利润的非现金部分除以上两年末的总资产的平均值。净营运资产 (NetOA) 计算为营运资产减营运负债的差除以上一年的总资产, 毛利率 (GP) 计算为总利润减去销售费用的差除以当年的总资产。股权回报率 (ROE) 计算为净利润除以上一期的账面价值。动量 (MOM) 计算为 $t-11$ 到 $t-1$ 的股票收益率。复合股权发行量 (CEI) 计算为过去 5 年的股票市值对数增长率减去股票收益率。

参考 Stambaugh 等 (2015) 的使用的方法, 在中国股票市场构建错误定价指标 (*MISP*), 在每个月末, 如果异象与收益负 (正) 相关, 则根据异象值从小 (大) 到大 (小) 对股票进行排序。最后将股票在所有异象变量上的排名根据求一个平均值, 这个值越大, 代表股票被高估得越严重, 这个值越小, 说明异象被低估越严重。

5.2 数据描述性统计特征

本章的核心是使用双变量组合排序研究学习指数 *LI* 与错误定价 *MISP* 之间的关系, 而不是类似前文将某一个变量当作控制变量研究另一个变量对股票横截面收益的预测能力, 所以, 确保双变量排序的有效性, 首先需要验证两个排序变量在很大程度上是不相关的, 随后才可以研究学习指数 *LI* 与错误定价 *MISP* 的排序的截面收益的情况, 两个相关性很高的变量进行排序会导致双重排序实际上是在一类变量中进一步排序。

本节使用双变量独立排序构造投资组合如下。在每个月末, 根据学习指数 *LI*

和错误定价MISP各自独立构建 5 个投资组合，从而产生 25 个双重排序的投资组合，计算各组的MISP和LI的均值。表 5-1 的 Panel A 和 Panel B 报告了每个投资组合的平均LI和平均MISP。在 Panel A 中，在每个LI组合中，不同MISP组合的平均 LI 几乎没有差异。在最高LI和最低MISP的underpriced组合中，LI的均值为 2.347，而在最高LI的最高MISP的overpriced组合中，LI的均值为 2.434。同样，在 Panel B 中，在每个MISP组合中，不同LI组合的MISP和没有太大差异。在 Panel C 中，两个排序变量LI和MISP之间的相关性仅为 0.03。总之，表 5-1 中展示的投资组合特征表明，对LI和MISP的相关性很低，对LI和MISP进行双重排序并不是对某一个变量的进一步排序。

表 5-1 投资组合特征					
Panel A: LI 的均值					
组数	1(Low LI)	2	3	4	5(High LI)
1(underpriced)	0.318	0.595	0.874	1.296	2.347
2	0.319	0.592	0.882	1.301	2.410
3	0.319	0.595	0.884	1.297	2.456
4	0.323	0.595	0.883	1.302	2.418
5(overpriced)	0.328	0.594	0.883	1.303	2.434
Panel B: MISP 的均值					
组数	1(Low LI)	2	3	4	5(High LI)
1(underpriced)	0.305	0.307	0.310	0.314	0.310
2	0.413	0.414	0.414	0.414	0.414
3	0.479	0.479	0.479	0.479	0.480
4	0.543	0.544	0.544	0.545	0.544
5(overpriced)	0.642	0.639	0.639	0.638	0.637
Panel C: 相关系数矩阵					
	LI	MISP			
LI	1				
MISP	0.03	1			

5. 3 学习指数与错误定价

5. 3. 1 双变量排序

使用前文构造的错误定价变量MISP作为错误定价的代理变量。错误定价的高估和低估的程度是相对的，例如，在极端的牛市行情中，股票可能会出现系统

性的高估,这就导致即使分在被低估的股票组合中的股票可能也是被高估的股票,同样的,在极端的熊市行情中,股票也可能会出现系统性的低估,导致即使被分在被高估的股票组合中的股票也是被低估的。相对被低估的股票拥有相对更高的未来收益,相对被高估的股票拥有相对更低的未来收益,导致被低估的股票和被高估的股票未来收益率存在显著差异。因此,参考 Cao 和 Han (2016) 以及 Han 等 (2022),使用被高估股票组合 (Overpriced) 和被低估股票组合 (Underpriced) 的收益率差异 (OMU) 来评估错误定价程度。

在股票市场中,投资者会寻找具有错误定价股票,通过做多被低估的股票,做空被高估的股票进行套利。投资者可以通过学习对加深对股票的了解,然后做多被低估的股票的同时去做空高估的股票,以缩小错误定价。因此,在学习指数 LI 比较高的股票中,错误定价会被投资者所缩小,也就是说被低估的股票和被高估的股票的组之间的未来异常收益差异 OMU 要更小一些。同时,在被低估的股票中, LI 的负向预测能力可能比在被高估的股票中更强,这是因为股票市场中可能存在的套利限制,特别是 A 股市场存在的做空限制,部分股票无法做空,投资者可能不愿意学习被高估的资产,或者即使投资者学习了被高估的资产也没有办法根据自己对资产的学习构建相应的投资组合,这些原因可能导致了 LI 的对股票收益的负向预测能力在被高估的股票中失效。

本节使用双变量独立排序法研究学习指数 LI 和错误定价 $MISP$ 对股票收益的影响。具体而言,在每个月末,根据 $MISP$ 和学习指数 LI 将股票从低到高分别分为 5 组,得到 25 个组合,在 25 个组合内使用流通市值作为权重构建投资组合,并持有一个月,计算各个组合的超额收益与风险调整后超额收益,以及对应的 Newey-West (1987) 调整后 t 统计量。

双变量独立排序的结果展示在表 5-2 中,Panel A 展示了各组合的超额收益,Panel B 展示了 FF5 调整后的超额收益,Panel C 展示了 CH4 调整后的超额收益。在 Panel A 中可以看到,从被低估的股票组到被高估的股票组中,股票的收益呈现下降的趋势,被低估股票未来的收益更高,被高估的股票未来收益更低,这和错误定价产生异常收益率的性质一致。在最低 LI 组合中,被高估和被低估的股票组合的未来收益率的差异 OMU 为 -1.028%, t 值为 -2.67,在 1% 的水平上显著。但是在高 LI 组合中,被高估和被低估的股票组合的收益率差异 OMU 为 -0.188%, t 值为 -0.55,与 0 没有显著差异。同时, OMU 随着 LI 的增加而减少, OMU 在低 LI 和高 LI 组合之间的差异是 0.840%, t 值是 2.22。也就是说,在学习指数 LI 更高的股票组合中,错误定价 OMU 的程度更小。同时,在最后一列,被低估股票组合中的 LI 的差异 HML 是 -0.726%, t 值为 -2.38,在 5% 的水平上显著,被高估的股票组合中, LI 的差异 HML 是 0.114%, t 值是 0.114,同时,被高估股票组合和被低估股票组合中的 HML 的差异是 0.840%, t 值是 2.22。这就说明,在被低估的股票中的学

习指数对股票收益的负向预测能力比被高估的股票中显著更强。在 Panel B (Panel C) 中可以观察到类似的结果。最低 *LI* 组合中的 *OMU* 的 FF5 (CH4) 调整后超额收益是 -1.209% (-0.922%), *t* 值是 -3.27 (-2.07), 在 1% (5%) 的水平上显著, 最高 *LI* 组合中的 *OMU* 的调整后超额收益是 -0.420% (-0.067%), *t* 值是 -1.51 (0.22), 与 0 没有显著差异。同时, 随着 *LI* 的增加 *OMU* 不断下降, 最高 *LI* 组合和最低 *LI* 组合的 *OMU* 差异 *HML* 是 0.789% (0.989%), *t* 值是 1.83 (2.00), 在 10% (5%) 的水平上显著, 这表明最高 *LI* 组合中比最低 *LI* 组合中的错误定价程度 *OMU* 显著更小。同时, 在被低估的股票组合中, 学习指数 *LI* 多空组合 *HML* 的超额收益是 -0.881% (-1.248%), *t* 值是 -2.40 (-2.98), 在 5% (1%) 的水平上显著, 在被高估的股票组合中, 学习指数 *LI* 多空组合 *HML* 的超额收益是 -0.092% (-0.259%), *t* 值是 -2.40 (-2.98), 与 0 没有显著差异, 在被高估和被低估的股票中, *HML* 组合的差异 *OMU* 是 0.789% (0.989%), *t* 值是 1.83 (2.00), 在 10% (5%) 的水平上显著, 这表明学习指数的预测能力在被低估的股票组合中更强。

总的来说, Panel A、Panel B、Panel C 的结果在定性上是相似的, 学习指数在被低估的股票中对股票收益的负向预测能力显著更强, 在学习指数更高的股票组合中错误定价显著更小。

表 5-2 学习指数与错误定价

组数	1 (Low LI)	2	3	4	5 (High LI)	5-1 (HML)
Panel A: <i>Exret</i>						
1(<i>underpriced</i>)	1.336** (2.35)	0.918 (1.56)	0.973 (1.54)	1.017 (1.56)	0.610 (1.03)	-0.726** (-2.38)
3	0.952 (1.49)	0.617 (1.02)	0.775 (1.33)	0.756 (1.17)	0.780 (1.31)	-0.172 (-0.58)
5(<i>overpriced</i>)	0.308 (0.46)	0.226 (0.35)	0.339 (0.49)	0.400 (0.59)	0.422 (0.62)	0.114 (0.37)
5-1(<i>OMU</i>)	-1.028*** (-2.67)	-0.691* (-1.83)	-0.634** (-2.07)	-0.618* (-1.94)	-0.188 (-0.55)	0.840** (2.22)
Panel B: <i>FF5-α</i>						
1(<i>underpriced</i>)	0.737*** (3.07)	0.245 (1.03)	0.139 (0.59)	0.309 (1.21)	-0.144 (-0.54)	-0.881** (-2.40)
3	0.371 (1.45)	-0.216 (-1.12)	-0.000 (-0.00)	-0.171 (-0.87)	-0.220 (-0.94)	-0.591* (-1.82)
5(<i>overpriced</i>)	-0.472* (-1.68)	-0.548** (-1.99)	-0.590*** (-2.80)	-0.555*** (-3.11)	-0.565** (-2.46)	-0.092 (-0.26)
5-1(<i>OMU</i>)	-1.209*** (-3.27)	-0.793** (-2.40)	-0.729*** (-2.64)	-0.864*** (-2.96)	-0.420 (-1.51)	0.789* (1.83)
Panel C: <i>CH4-α</i>						
1(<i>underpriced</i>)	0.807*** (2.65)	0.165 (0.69)	0.015 (0.07)	-0.172 (-0.68)	-0.441 (-1.54)	-1.248*** (-2.98)
3	0.741*** (2.79)	-0.083 (-0.42)	0.119 (0.70)	-0.145 (-0.90)	-0.096 (-0.44)	-0.837** (-2.34)
5(<i>overpriced</i>)	-0.115 (-0.41)	-0.022 (-0.08)	-0.313* (-1.72)	-0.282* (-1.72)	-0.374* (-1.69)	-0.259 (-0.76)
5-1(<i>OMU</i>)	-0.922** (-2.07)	-0.187 (-0.47)	-0.329 (-1.32)	-0.110 (-0.35)	0.067 (0.22)	0.989** (2.00)

5.3.2 长期持续分析

通过上述研究,发现在学习指数较高的股票组合中股票的错误定价显著更小。本节继续检验这种关系是否长期存在,也就是检验 LI 对股票错误定价 $MISP$ 的长期影响。同样使用双变量独立排序进行研究,在每个月末,根据错误定价 $MISP$

和学习指数 LI 和大小对股票进行双变量独立排序,构建流通市值加权的投资组合,并将这些投资组合分别持有 1、12、24、36、48、60 个月,观察所有 LI 排序组合中的 UMO 组合的超额收益以及 $CH4$ 、 $FF5$ 调整后的超额收益,从而研究学习指数 LI 对错误定价的长期影响。

表 5-3 展示了双变量独立排序的结果, Panel A 报告了双变量独立排序的超额收益率, Panel B 报告了双变量独立排序的 $FF5$ 调整后的超额收益率, Panel C 报告了双变量独立排序的 $CH4$ 调整后的超额收益率。在 Panel A 中,可以看到,学习指数对错误定价的影响,也就是最高 LI 组合和最低 LI 组合的错误定价程度 OMU 的差异 HML 在 $t + 12$ 的时候是 0.499% ($t = 2.64$),在 $t + 24$ 期的时候是 0.315% ($t = 1.99$),并且随着时间的推移,学习指数对错误定价的影响逐渐减弱,最终在 $t + 48$ 期的时候缩小到 0.282% ($t = 2.05$),随后消失。同样的,在

表 5-3 表 5-3 的 Panel B 和 Panel C 中,经过 $FF5$ 和 $CH4$ 调整后,依然可以观察到类似的现象,学习指数对错误定价的影响会随着时间的消失逐渐缩小,并且随着时间的推移而逐渐消失。Panel A、Panel B 和 Panel C 的结果在定性上是相似的,这些结果表明学习指数对错误定价的影响具有长期的持续性,但是这种影响会随着时间的推移不断缩小并最终消失,这表明被学习更多的股票在长期来看都会出现更小的错误定价,因为投资者对该股票的信息了解更多,学习更多。

表 5-3 学习指数的对错误定价的长期影响

	1 (Low LI)	2	3	4	5 (High LI)	5-1 (HML)
Panel A: OMU组合的超额收益						
$t + 1$	-1.028*** (-2.67)	-0.691* (-1.83)	-0.634** (-2.07)	-0.618* (-1.94)	-0.188 (-0.55)	0.840** (2.22)
$t + 12$	-3.012*** (-16.90)	-2.731*** (-15.29)	-2.702*** (-15.37)	-2.696*** (-14.57)	-2.513*** (-12.43)	0.499*** (2.64)
$t + 24$	-1.727*** (-11.47)	-1.512*** (-11.07)	-1.489*** (-9.64)	-1.494*** (-10.76)	-1.412*** (-7.05)	0.315** (1.99)
$t + 36$	-1.220*** (-8.45)	-1.036*** (-9.24)	-1.043*** (-8.18)	-0.968*** (-8.39)	-0.845*** (-6.40)	0.374*** (2.71)
$t + 48$	-0.921*** (-6.89)	-0.676*** (-5.88)	-0.680*** (-6.14)	-0.754*** (-6.28)	-0.639*** (-4.55)	0.282** (2.05)
$t + 60$	-0.618*** (-5.53)	-0.402*** (-4.08)	-0.503*** (-4.30)	-0.586*** (-4.63)	-0.572*** (-5.02)	0.045 (0.36)
Panel B: OMU组合的 FF5- α						
$t + 1$	-1.209*** (-3.27)	-0.793** (-2.40)	-0.729*** (-2.64)	-0.864*** (-2.96)	-0.420 (-1.51)	0.789* (1.83)
$t + 12$	-3.001*** (-16.88)	-2.742*** (-16.39)	-2.691*** (-15.69)	-2.659*** (-15.25)	-2.546*** (-13.63)	0.455** (2.52)
$t + 24$	-1.748*** (-11.50)	-1.568*** (-11.51)	-1.538*** (-10.54)	-1.441*** (-10.92)	-1.455*** (-8.13)	0.293** (1.98)
$t + 36$	-1.224*** (-8.40)	-1.027*** (-9.37)	-1.076*** (-8.74)	-0.989*** (-8.70)	-0.865*** (-6.35)	0.359** (2.54)
$t + 48$	-0.932*** (-6.82)	-0.654*** (-5.38)	-0.714*** (-6.39)	-0.773*** (-6.70)	-0.652*** (-4.74)	0.280* (1.90)
$t + 60$	-0.596*** (-5.46)	-0.398*** (-4.39)	-0.508*** (-4.01)	-0.608*** (-4.68)	-0.607*** (-4.97)	-0.012 (-0.09)
Panel C: OMU组合的 CH4- α						
$t + 1$	-0.922** (-2.07)	-0.187 (-0.47)	-0.329 (-1.32)	-0.110 (-0.35)	0.067 (0.22)	0.989** (2.00)
$t + 12$	-2.942*** (-16.21)	-2.755*** (-15.17)	-2.628*** (-15.27)	-2.613*** (-13.65)	-2.503*** (-13.94)	0.438** (2.52)
$t + 24$	-1.711*** (-10.45)	-1.525*** (-10.56)	-1.480*** (-9.24)	-1.450*** (-10.60)	-1.449*** (-7.97)	0.262* (1.77)
$t + 36$	-1.259*** (-7.85)	-0.979*** (-8.34)	-1.026*** (-7.55)	-1.025*** (-8.54)	-0.872*** (-6.33)	0.387*** (2.67)
$t + 48$	-0.940*** (-6.65)	-0.631*** (-5.08)	-0.759*** (-6.05)	-0.782*** (-6.69)	-0.648*** (-4.45)	0.291** (2.00)
$t + 60$	-0.599*** (-5.68)	-0.375*** (-3.96)	-0.523*** (-4.11)	-0.651*** (-4.95)	-0.620*** (-4.93)	-0.022 (-0.16)

5.3.3 子样本分析

很多学者认为股票的高估主要在投资者情绪比较高的时期出现（Stambaugh 等，2015；Han 等，2022）。本文使用易志高和茅宁（2009）参考 Baker 和 Wurgler（2006）结合中国市场实际构建的投资者情绪指数作为投资者情绪的代理变量，将样本划分为低情绪期和高情绪期，研究学习指数 LI 和错误定价之间的关系是否在不同的情绪期存在差异。具体的做法是：根据与 $MISP$ 和 LI 同期的情绪指数相对于中位数的大小划分为高情绪期和低情绪期的子样本，使用 $MISP$ 和 LI 进行双变量独立排序，构建流通市值加权投资组合，并持有 1 个月，计算组合的 $CH4$ 调整后的超额收益。

表 5-4 展示了双变量独立排序的 $CH4$ 调整后超额收益，Panel A 展示的低情绪期的子样本，Panel B 展示了高情绪期的子样本。在 Panel A 中，低情绪期的低估组，学习指数 LI 的多空组合 HML 组合的 $CH4$ 调整后的超额收益是 -1.180%， t 值是 -2.56，在 5% 的水平上显著，在 Panel B 中，高情绪期的低估组， LI 的多空组合 HML 的 $CH4$ 调整后超额收益是 -1.150%。 t 值是 -2.15，在 5% 的水平上显著，这说明不管是在低情绪期还是高情绪期，说明 LI 对股票收益的负向预测能力在股票被低估的组合中总是存在的。同时，在 Panel A 中可以观察到，在低情绪期， LI 的多空组合 HML 在被低估和被高估的组合中的差异 OMU 的值是 0.643%， t 值是 0.93，这表明在低情绪期被高估的股票和被低估的股票中学习指数的负效应没有显著差异。但是，在 Panel B 中可以观察到，在高情绪期， HML 组合中的被低估组合和被高估组合 HML 的差异 OMU 的值是 1.620%， t 值是 2.43，在 5% 的水平上显著。

这表明，这表明在低情绪期被高估的股票和被低估的股票中学习指数对股票收益的负向预测能力存在显著差异，这说明股票被高估的情况经常出现在高情绪期，这与 Stambaugh 等（2015）和 Han 等（2022）的发现一致。

表 5-4 子样本分析

组数	1 (Low LI)	2	3	4	5 (High LI)	5-1 (HML)
Panel A: 低情绪期						
1(Underpriced)	0.320 (0.84)	0.011 (0.03)	0.180 (0.54)	-0.273 (-0.76)	-0.860* (-1.86)	-1.180** (-2.56)
3	0.064 (0.14)	-0.213 (-0.51)	0.339 (1.08)	-0.438* (-1.86)	-0.093 (-0.26)	-0.157 (-0.29)
5(Overpriced)	0.205 (0.42)	0.185 (0.33)	-0.133 (-0.42)	-0.349 (-1.26)	-0.332 (-0.70)	-0.537 (-0.81)
5-1(OMU)	-0.116 (-0.19)	0.174 (0.25)	-0.313 (-0.85)	-0.076 (-0.14)	0.527 (0.88)	0.643 (0.93)
Panel B: 高情绪期						
1(Underpriced)	1.106*** (2.68)	0.151 (0.37)	-0.047 (-0.14)	-0.335 (-0.89)	-0.044 (-0.11)	-1.150** (-2.15)
3	1.156*** (3.23)	0.028 (0.10)	0.152 (0.60)	0.011 (0.04)	-0.028 (-0.08)	-1.184** (-2.16)
5(Overpriced)	-0.670** (-1.99)	-0.214 (-0.86)	-0.349 (-1.48)	-0.203 (-0.70)	-0.200 (-0.85)	0.470 (1.31)
5-1(OMU)	-1.776*** (-2.85)	-0.365 (-0.73)	-0.302 (-0.74)	0.132 (0.27)	-0.156 (-0.37)	1.620** (2.43)

5.3.4 套利限制的影响

这一小节研究学习指数在被高估股票和被低估股票中对股票收益的负向预测能力的差异是否与套利限制相关。学习指数对股票收益的负向预测能力主要集中在被低估的股票组合中，并且在被低估组合与被高估的股票中存在显著差异。投资者学习了被高估的股票后，可能由于套利限制的缘故，无法将其加入到自己的投资组合中，所以导致学习指数对股票收益的负向预测能力失效。由于本节主要研究的是被高估股票组合与被低估股票中的套利限制的大小，因此本节基于 *MISP* 和 *LI* 使用双变量条件排序构建一个等权重投资组合。

本节使用四个常见衡量套利限制程度的变量。首先，本文使用 Amihud (2002) 的非流动性测度 (*ILLIQ*) 来衡量交易成本的大小，非流动性越高的股票交易成本越高，套利限制程度越大。其次，参考 Nagel (2005) 使用机构投资者持有比例 (*IO*) 来捕捉卖空限制的程度，因为机构持有比例较高的股票更容易被空头投资者借入。第三，Zhang (2006) 提出使用分析师覆盖率 (*NOA*) 作为捕捉套利限制的变量，因为分析师覆盖率更高的股票有更低的信息不对称，更容易被套利。第四，参考 Barberis 等 (2018) 使用的公司规模 (*SIZE*) 作为套利限制的代理变

量，公司规模越大的股票套利限制更小。

表 5-5 展示了错误定价 *MISP* 与学习指数 *LI* 各组中套利限制的情况。Panel A 展示了错误定价 *MISP* 与学习指数 *LI* 的双变量排序组合中的机构投资者持有比例均值，Panel B 展示了错误定价 *MISP* 与学习指数 *LI* 的双变量排序组合中的非流动性均值，Panel C 展示了错误定价 *MISP* 与学习指数 *LI* 的双变量排序组合中的公司规模均值，Panel D 展示了错误定价 *MISP* 与学习指数 *LI* 的双变量排序组合中的机构投资者持有分析师覆盖率均值。

在 Panel A 的第 1 列中可以观察到，从被低估股票组到被高估组，机构投资者持有比例逐渐变小，低估 (*underpriced*) 组和高估 (*overpriced*) 组的差异 *OMU* 组合的差值是 -7.852%，这表明被高估的股票组合中机构投资者持有比例比被低估股票组合平均少 7.852%，*t* 值是 -8.16，在 1% 的水平上显著，对于剩余的其他列，也可以观察到类似的现象，所有被高估的股票组合中，机构投资者持有比例更小，套利限制更大。同样的，可以在 Panel B、Panel C 和 Panel D 中观察到类似的结果，与被低估的股票相比被高估的股票的流动性更大、公司规模更小、分析师覆盖率更低，同时被告组和被低估的差异 *OMU* 都在 1% 的水平上显著。Panel A、Panel B、Panel C 和 Panel D 中的结果在定性上是相似的，这些结果都表明相对于被低估的股票而言，被高估的股票的套利限制更大。

总的来说，表 5-5 的结果表明，投资者学习对股票横截面收益的预测能力更差可能是由于被高估的股票套利限制更大，投资者进行了学习但是因为套利的限制没有把被高估的股票纳入自己的投资组合当中。

表 5-5 套利限制

组数	1 (Low LI)	2	3	4	5 (High LI)
Panel A: 机构投资者持有比例					
1(<i>underpriced</i>)	42.548*** (16.80)	42.167*** (17.14)	41.743*** (17.30)	42.001*** (17.09)	42.705*** (16.77)
3	36.874*** (16.96)	36.320*** (17.03)	36.393*** (17.13)	36.736*** (17.19)	36.497*** (17.09)
5(<i>overpriced</i>)	34.696*** (16.59)	34.366*** (16.66)	34.586*** (16.71)	34.557*** (16.75)	35.697*** (16.86)
5-1(<i>OMU</i>)	-7.852*** (-8.16)	-7.800*** (-9.30)	-7.157*** (-10.43)	-7.443*** (-11.88)	-7.008*** (-10.12)
Panel B: 非流动性					
1(<i>underpriced</i>)	0.116*** (5.80)	0.118*** (5.86)	0.121*** (5.93)	0.110*** (6.23)	0.112*** (6.40)
3	0.165*** (5.07)	0.144*** (5.76)	0.133*** (5.97)	0.137*** (5.96)	0.138*** (6.33)
5(<i>overpriced</i>)	0.181*** (5.06)	0.167*** (5.15)	0.166*** (5.31)	0.155*** (5.69)	0.146*** (5.98)
5-1(<i>OMU</i>)	0.066*** (3.26)	0.050*** (3.02)	0.045*** (3.51)	0.045*** (3.95)	0.034*** (3.36)
Panel C: 公司规模					
1(<i>underpriced</i>)	8.404*** (58.96)	8.258*** (63.94)	8.228*** (66.02)	8.235*** (68.22)	8.211*** (68.01)
3	8.076*** (62.33)	8.017*** (66.47)	7.972*** (68.69)	7.971*** (71.91)	7.914*** (69.99)
5(<i>overpriced</i>)	7.986*** (64.05)	7.928*** (65.38)	7.892*** (68.31)	7.860*** (69.26)	7.850*** (74.84)
5-1(<i>OMU</i>)	-0.418*** (-6.81)	-0.330*** (-8.52)	-0.336*** (-9.29)	-0.374*** (-12.46)	-0.361*** (-9.60)
Panel D: 分析师覆盖率					
1(<i>underpriced</i>)	0.673*** (10.52)	0.650*** (11.36)	0.630*** (11.47)	0.608*** (11.42)	0.598*** (11.32)
3	0.461*** (10.81)	0.443*** (11.46)	0.445*** (11.33)	0.452*** (11.40)	0.428*** (10.60)
5(<i>overpriced</i>)	0.280*** (9.91)	0.301*** (10.86)	0.303*** (10.24)	0.297*** (10.00)	0.293*** (10.03)
5-1(<i>OMU</i>)	-0.393*** (-8.74)	-0.349*** (-9.87)	-0.327*** (-10.30)	-0.310*** (-10.53)	-0.305*** (-9.85)

5.4 稳健性检验

5.4.1 更换错误定价的衡量指标

由于错误定价 *MISP* 和学习指数 *LI* 的双变量独立排序结果在很大程度上取决于使用什么变量来衡量错误定价 *MISP*，因此研究实证结果对 *MISP* 的敏感性是很重要的。本节使用以下两个指标来替代错误定价，这两个指标比 *MISP* 更容易计算。第一个错误定价的代理变量是 CAPM 的 Alpha，它是投资者在做出投资组合配置决策时广泛使用的绩效指标 (Barber, 2016; Berk 和 van Binsbergen, 2016)，所以 Han (2022) 使用 CAPM 的 Alpha 作为错误定价的代理变量。除此之外，本节还参考 Han 等 (2022) 构建了一个复合的 Alpha (Composite Alpha) 作为本节使用的错误定价 *MISP* 的第二个代理变量，复合 Alpha 计算为 CAPM 的 AlphaFF3 的 Alpha、FF5 的 Alpha、CH3 的 Alpha 和 CH4 的 Alpha 的均值。具体来说，使用过去两年的股票收益对相应的因子收益进行回归，计算得到的对应的多因子模型的 Alpha。

表 5-6 报告了使用 CAPM 的 Alpha 作为错误定价 *MISP* 的代理变量的研究结果，Panel A 报告了双变量独立排序的超额收益，Panel B 报告了双变量独立排序的 FF5 调整后的超额收益，Panel C 报告了双变量独立排序的 CH4 调整后的超额收益。在 Panel A 中，从被低估的股票组到被高估的股票组中，股票的收益呈现下降的趋势，被低估股票未来的收益更高，被高估的股票未来收益更低。在低 *LI* 组合中，被高估和被低估的股票组合的未来收益率的差异 *OMU* 为 -0.844%，*t* 值为 -1.94，在 1% 的水平上显著。但是在高 *LI* 组合中，被高估和被低估的股票组合的收益率差异 *OMU* 为 0.012%，*t* 值为 -0.03。同时，*OMU* 随着 *LI* 的增加而减少，*OMU* 在低 *LI* 和高 *LI* 组合之间的差异 *HML* 是 0.856%，*t* 值是 2.08，在 5% 的水平上显著。也就是说，在学习指数 *LI* 更高的股票组合中，错误定价 *OMU* 的程度更小。同时，在最后一列，被低估股票组合中的 *LI* 的差异 *HML* 是 -1.234%，*t* 值为 -3.78，在 1% 的水平上显著，被高估的股票组合中，*LI* 的差异 *HML* 是 -0.378%，*t* 值是 -1.36，与 0 没有显著差异同时，被高估股票组合和被低估股票组合中的 *HML* 的差异 *OMU* 是 0.856%，*t* 值是 2.08，在 5% 的水平上显著。这就说明，在被低估的股票中的学习指数对股票收益的负向预测能力比被高估的股票中显著更强。类似的，在 Panel B (Panel C) 中，*OMU* 组合的 *HML* 或者说 *HML* 组合的 *OMU* 的值是 1.063% (1.096%)，*t* 值是 2.48 (2.22)，在 5% (5%) 的水平上显著，这表明学习指数对股票横截面收益的预测能力在被低估的股票组合中更强，同时，学习指数高的组合中，股票的错误定价更小。Panel B 和 Panel C 的结果在定性上是相似的。

表 5-7 报告了使用 CAPM 的 Alpha 作为错误定价 *MISP* 的代理变量的研究结

果, Panel A 报告了双变量独立排序的超额收益, Panel B 报告了双变量独立排序的 FF5 调整后的超额收益, Panel C 报告了双重排序的 CH4 调整后的超额收益。在 Panel A 中, 可以看到, 从被低估的股票组到被高估的股票组中, 股票的收益呈现下降的趋势, 被低估股票未来的收益更高, 被高估的股票未来收益更低。在最低 *LI* 组合中, 被高估和被低估的股票组合超额收益的差异 *OMU* 为 -0.628%, *t* 值为 -1.65, 在 10% 的水平上显著。但是在高 *LI* 组合中, 被高估和被低估的股票组合的收益率差异 *OMU* 为 0.202%, *t* 值为 0.55, 与 0 没有差异。同时, *OMU* 随着 *LI* 的增加而减少, *OMU* 在低 *LI* 和高 *LI* 组合之间的差异是 0.831%, *t* 值是 1.82, 在 10% 的水平上显著。也就是说, 在学习指数 *LI* 更高的股票组合中, 错误定价 *OMU* 的程度更小。同时, 在最后一列, 被低估股票组合中的 *LI* 的差异 *HML* 是 -1.047%, *t* 值为 -3.10, 在 1% 的水平上显著, 被高估的股票组合中, *LI* 的差异 *HML* 是 -0.216%, *t* 值是 -0.74, 与 0 没有显著差异, 同时, 被高估股票组合和被低估股票组合中的 *HML* 的差异是 0.831%, *t* 值是 1.82, 在 10% 的水平上显著。这就说明, 在被低估的股票中的学习指数对股票收益的负向预测能力比被高估的股票中显著更强。类似的, 在 Panel B (Panel C) 中, *OMU* 组合的 *HML* 或者说 *HML* 组合的 *OMU* 的值是 0.823% (0.986%), *t* 值是 1.71 (1.67), 在 10% (10%) 的水平上显著, 这表明学习指数对股票横截面收益的预测能力在被低估的股票组合中更强, 同时, 学习指数高的组合中, 股票的错误定价更小。Panel B 和 Panel C 的与 Panel A 的结果在定性上是相似的。

总的来说, 表 5-6 和表 5-7 的结果与前文的研究在定性上是相似的, 这表明学习指数对股票收益的负向预测能力在被高估组合和被低估组合之间存在显著差异, 在被低估的股票组合中要更强。此外, *UMO* 组合的超额收益随着 *LI* 的增加而减小, 最高 *LI* 组合的 *UMO* 组合的超额收益与最低 *LI* 组合的 *UMO* 组合的超额收益之间存在显著差异, 这表明前文发现学习指数与错误定价的关系是稳健的, 不受错误定价代理变量的选择的影响。

表 5-6 使用 CAPM Alpha 作为错误定价的代理变量

组数	1 (Low LI)	2	3	4	5 (High LI)	5-1 (HML)
Panel A: 超额收益						
1(<i>underpriced</i>)	1.558** (2.18)	0.682 (1.00)	0.693 (1.03)	0.837 (1.21)	0.324 (0.52)	-1.234*** (-3.78)
3	1.018 (1.55)	0.664 (1.11)	0.538 (0.92)	0.631 (1.00)	0.644 (1.13)	-0.374 (-1.23)
5(<i>overpriced</i>)	0.713 (1.16)	0.594 (0.96)	0.635 (1.04)	0.721 (1.11)	0.335 (0.54)	-0.378 (-1.36)
5-1(<i>OMU</i>)	-0.844* (-1.94)	-0.088 (-0.21)	-0.058 (-0.16)	-0.117 (-0.29)	0.012 (0.03)	0.856** (2.08)
Panel B: FF5 Alpha						
1(<i>underpriced</i>)	1.088*** (3.35)	0.078 (0.25)	-0.138 (-0.56)	0.001 (0.00)	-0.387 (-1.54)	-1.475*** (-4.38)
3	0.383 (1.49)	-0.104 (-0.51)	-0.296 (-1.51)	-0.167 (-0.79)	-0.405* (-1.88)	-0.788** (-2.49)
5(<i>overpriced</i>)	-0.100 (-0.42)	-0.275 (-1.13)	-0.285 (-1.55)	-0.225 (-1.01)	-0.513** (-2.19)	-0.413 (-1.28)
5-1(<i>OMU</i>)	-1.189*** (-2.81)	-0.353 (-0.96)	-0.148 (-0.50)	-0.226 (-0.59)	-0.126 (-0.40)	1.063** (2.48)
Panel C: CH4 Alpha						
1(<i>underpriced</i>)	1.403*** (3.95)	0.387 (1.15)	0.271 (0.87)	0.163 (0.49)	-0.180 (-0.64)	-1.583*** (-4.09)
3	0.496** (1.99)	-0.101 (-0.48)	-0.309* (-1.88)	-0.117 (-0.63)	-0.397* (-1.86)	-0.893** (-2.54)
5(<i>overpriced</i>)	-0.059 (-0.20)	-0.274 (-0.93)	-0.291 (-1.51)	-0.512** (-2.43)	-0.546** (-2.35)	-0.487 (-1.35)
5-1(<i>OMU</i>)	-1.462*** (-2.73)	-0.661 (-1.23)	-0.561 (-1.42)	-0.674 (-1.46)	-0.366 (-1.02)	1.096** (2.22)

表 5-7 使用 Composite Alpha 作为错误定价的代理变量

组数	1 (Low LI)	2	3	4	5 (High LI)	5-1 (HML)
Panel A: 超额收益						
1(<i>underpriced</i>)	1.545** (2.22)	0.609 (0.88)	0.552 (0.82)	0.809 (1.12)	0.499 (0.75)	-1.047*** (-3.10)
3	0.902 (1.50)	0.570 (0.97)	0.525 (0.88)	0.766 (1.19)	0.684 (1.19)	-0.219 (-0.75)
5(<i>overpriced</i>)	0.917 (1.41)	0.831 (1.38)	0.804 (1.31)	0.983 (1.53)	0.701 (1.13)	-0.216 (-0.74)
5-1(<i>OMU</i>)	-0.628* (-1.65)	0.222 (0.64)	0.252 (0.75)	0.174 (0.49)	0.202 (0.55)	0.831* (1.82)
Panel B: FF5 Alpha						
1(<i>underpriced</i>)	1.132*** (3.51)	0.120 (0.44)	-0.147 (-0.53)	0.184 (0.66)	-0.198 (-0.69)	-1.330*** (-3.49)
3	0.157 (0.76)	-0.180 (-0.96)	-0.439** (-2.00)	-0.104 (-0.44)	-0.199 (-0.85)	-0.356 (-1.16)
5(<i>overpriced</i>)	0.090 (0.38)	-0.164 (-0.76)	-0.181 (-0.97)	-0.037 (-0.16)	-0.417** (-1.97)	-0.507* (-1.67)
5-1(<i>OMU</i>)	-1.043*** (-2.66)	-0.284 (-0.90)	-0.035 (-0.13)	-0.221 (-0.68)	-0.220 (-0.61)	0.823* (1.71)
Panel C: CH4 Alpha						
1(<i>underpriced</i>)	1.583*** (4.97)	0.400 (1.49)	0.196 (0.69)	0.235 (0.71)	-0.073 (-0.23)	-1.656*** (-3.84)
3	0.308 (1.35)	-0.168 (-0.83)	-0.355* (-1.82)	-0.181 (-0.86)	-0.378 (-1.52)	-0.686* (-1.80)
5(<i>overpriced</i>)	0.172 (0.57)	-0.195 (-0.84)	-0.153 (-0.78)	-0.304* (-1.68)	-0.498** (-2.42)	-0.670* (-1.80)
5-1(<i>OMU</i>)	-1.410*** (-2.88)	-0.595 (-1.44)	-0.349 (-1.06)	-0.539 (-1.38)	-0.424 (-1.07)	0.986* (1.67)

5.4.2 其他公司特征正交化的学习指数

为了确保学习指数与错误定价之间的关系不受其他常见公司特征的影响,在每一个时间截面上,使用学习指数对所有常见公司特征做横截面回归,进行正交化处理,使用学习指数的残差作为正交化的学习指数 LI_{orth} ,随后使用正交化后的学习指数与错误定价进行双重排序,考察前文研究的稳健性。

表 5-8 报告了使用正交化学习指数 LI_{orth} 与错误定价 $MISP$ 进行双变量排序的研究结果, Panel A 报告了双变量独立排序的超额收益, Panel B 报告了双变量独立排序的 FF5 调整后的超额收益, Panel C 报告了双重排序的 CH4 调整后的超额收益。在 Panel A 中,可以看到,从被低估的股票组到被高估的股票组中,股票的收益呈现下降的趋势,被低估股票未来的收益更高,被高估的股票未来收益更低。在最低 LI 组合中,被高估和被低估的股票组合的收益率的差异 OMU 的值是-1.429%, t 值为-4.25,在 1%的水平上显著。但是在高 LI 组合中,被高估和被低估的股票组合的收益率差异 OMU 为-0.315%, t 值为-0.84,与 0 没有差异。同时, OMU 随着 LI 的增加而减少, OMU 在低 LI 和高 LI 组合之间的差异是 1.114%, t 值是 3.16,在 1%的水平上显著。也就是说,在学习指数 LI 更高的股票组合中,错误定价 OMU 的程度更小。同时,在最后一列,被低估股票组合中的 LI 的多空组合 HML 是-0.955%, t 值为-3.22,在 1%的水平上显著,被高估的股票组合中, LI 的多空组合 HML 是 0.160%, t 值是 0.75,与 0 没有显著差异,同时,被高估股票组合和被低估股票组合中的 HML 的差异 OMU 是 1.114%, t 值是 3.16,在 1%的水平上显著。这就说明,在被低估的股票中的学习指数对股票收益的负向预测能力比被高估的股票中显著更强。类似的,在 Panel B (Panel C) 中, OMU 组合的 HML 或者说 HML 组合的 OMU 的值是 0.831% (0.880%), t 值是 2.08 (2.18),在 5% (5%)的水平上显著,这表明学习指数对股票横截面收益的预测能力在被低估的股票组合中更强,同时,学习指数高的组合中,股票的错误定价更小。Panel B 和 Panel C 的与 Panel A 的结果在定性上是相似的。

总的来说,表 5-8 的结果与前文的研究在定性上是相似的,这表明学习指数对股票收益的负向预测能力在被高估组合和被低估组合之间存在显著差异,在被低估的股票组合中要更强。此外, UMO 组合的超额收益随着 LI 的增加而减小,最高 LI 组合的 UMO 组合的超额收益与最低 LI 组合的 UMO 组合的超额收益之间存在显著差异。同时,这种关系不能被其他常见的公司特征所解释,这表明前文发现学习指数与错误定价的关系是稳健的。

表 5-8 学习指数与错误定价

组数	1 (Low LI)	2	3	4	5 (High LI)	5-1 (HML)
Panel A: <i>Exret</i>						
1(<i>underpriced</i>)	1.504** (2.54)	1.167** (2.07)	0.898 (1.43)	0.660 (1.00)	0.549 (0.87)	-0.955*** (-3.22)
3	0.709 (1.20)	0.902 (1.35)	0.464 (0.79)	0.628 (0.92)	0.600 (0.97)	-0.109 (-0.48)
5(<i>overpriced</i>)	0.075 (0.11)	0.372 (0.51)	0.337 (0.51)	0.380 (0.54)	0.234 (0.33)	0.160 (0.75)
5-1(<i>OMU</i>)	-1.429*** (-4.25)	-0.795** (-2.02)	-0.561* (-1.73)	-0.280 (-0.80)	-0.315 (-0.84)	1.114*** (3.16)
Panel B: <i>FF5-α</i>						
1(<i>underpriced</i>)	0.745** (2.56)	0.549** (2.49)	0.438 (1.43)	0.188 (0.79)	0.001 (0.00)	-0.744** (-2.43)
3	-0.020 (-0.09)	0.150 (0.64)	-0.183 (-0.97)	0.008 (0.04)	-0.144 (-0.65)	-0.123 (-0.49)
5(<i>overpriced</i>)	-0.662*** (-2.86)	-0.373 (-1.40)	-0.543** (-2.19)	-0.300 (-1.48)	-0.575** (-2.53)	0.088 (0.35)
5-1(<i>OMU</i>)	-1.407*** (-3.73)	-0.922*** (-2.73)	-0.981*** (-2.62)	-0.488* (-1.67)	-0.576 (-1.51)	0.831** (2.08)
Panel C: <i>CH4-α</i>						
1(<i>underpriced</i>)	0.517** (1.97)	0.570** (2.36)	0.339 (1.05)	0.089 (0.33)	-0.270 (-0.84)	-0.787** (-2.55)
3	-0.018 (-0.09)	0.312 (1.41)	0.127 (0.65)	0.336 (1.33)	-0.011 (-0.05)	0.007 (0.03)
5(<i>overpriced</i>)	-0.490* (-1.75)	-0.047 (-0.21)	-0.137 (-0.67)	0.152 (0.71)	-0.397* (-1.68)	0.094 (0.30)
5-1(<i>OMU</i>)	-1.007** (-2.48)	-0.618* (-1.70)	-0.477 (-1.17)	0.063 (0.18)	-0.127 (-0.32)	0.880** (2.18)

第六章 学习因子及其定价能力检验

通过前面章节的研究,本文明确了中国股票市场的投资者学习对股票收益的负向预测能力无法被主流的多因子模型解释,并且学习指数来源于一个 VNV 模型,这为使用学习指数构建因子提供了理论基础。本章首先在中国股票市场构建了一个学习因子 LI,将其加入到 CH4 模型中,得到一个 CH4-LI 五因子模型。随后本章将 CH4-LI 和其他多因子模型进行定价能力的比较,从而对学习因子 LI 在中国市场的定价能力进行检验。

6.1 数据来源和变量说明

选取中国 A 股市场的所有上市公司作为实证研究的对象。所有使用的股票日度、月度交易数据,财务报表数据、Fama-French 多因子模型数据,无风险利率数据等均来自国泰安数据库。Liu 等(2019)提出的中国市场三因子(CH3)、四因子(CH4)的数据来源于 Robert F. Stambaugh 的个人网站。Hou 等(2015)的 q 因子模型、Stambaugh 和 Yuan(2017)的错误定价四因子模型 M4 以及 Daniel 等(2020)的行为三因子模型 BF3,参考原始文献对因子进行构造,在中国市场构造。至于 Bali 等(2017)的六因子模型中因子有三个是 FF3 是中的因子,另外两个则分别是 Carhart(1997)的动量因子 UMD 与 Pastor 和 Stambaugh(2003)的流动性因子 PS ,这些因子的数据来源于国泰安数据库,而博彩偏好因子 MAX 则参考原始文献对因子进行构造。

为了减少市场微观结构问题和小市值股票的影响,在每个月末,剔除公司规模小于沪深主板股票市值 20%分位数的股票和价格小于 5 元的股票。为了排除次新股的影响,剔除上市不到 1 年的股票。为了防止结果受到长时间停牌后的股票收益的影响,在每个月末,样本中包含的股票在过去一年至少有 120 天的交易记录,过去 20 个正常交易日至少有 15 天的交易记录。为了保持财务报表的一致性,剔除了金融行业股票。本章主要分析的时间区间从 2000 年 1 月 1 日到 2021 年 12 月 31 日。中国 A 股市场始于 1990 年上海和深圳交易所的建立,关注 2000 年之后的时期主要是为了确保财务报表数据的一致性,中国有关财务报告各个方面的规章制度的实施直到 1999 年才基本成型。

为了防止极端值对研究结果造成影响,本文对所有的研究变量在每个时间截面上进行截尾,剔除了小于 0.1%分位数和大于 99.9%分位数的数据,以保证研究结果的稳健性。

6.2 CH4-LI 五因子模型

6.2.1 学习因子的构建

本节参考 Fama 和 French (1993, 2015)、Stambaugh 和 Yuan (2017) 和 Daniel 等 (2020) 等人的研究方法, 使用实证资产定价中构建因子常用的双变量独立排序构建中国股票市场的学习因子。具体步骤为如下,

在每个月末, 对所有的股票使用公司规模 *SIZE* 和学习指数 *LI* 两个变量进行 2×3 的双变量独立排序。对于公司规模 *SIZE*, 以股票流通市值的中位数将股票划分为小市值 (*S*) 和大市值 (*B*) 两个组合。对于学习指数 *LI*, 以 30% 和 70% 作为分界线, 将股票根据学习指数 *LI* 的大小划分为低学习 (*L*)、中学习 (*M*)、高学习 (*H*) 三个组合。随后, 三个学习指数 *LI* 组合和两个公司规模 *SIZE* 组合取交集, 一共得到 6 个股票投资组合 (*L/S*、*L/B*、*M/S*、*M/B*、*H/S*、*H/B*)。在 6 个组合内构造流通市值加权投资组合, 持有组合一个月, 组合到下个月末的超额收益率, 使用上述 6 个组合的超额收益率构造学习因子, 学习因子 (记作 *LI*) 的收益率定义为:

$$LI = \frac{1}{2}(L/S + L/B) - \frac{1}{2}(H/S + H/B) \quad (6-1)$$

6.2.2 因子描述性统计

本节首先对前述提到的所有多因子模型中的因子进行描述性统计, 包括本节构造的 CH4-LI 五因子模型。

具体的描述性统计结果展示在表 6-1 中。第 1 列展示了因子收益率均值, 第 2 列展示了均值对应的 *t* 统计量, 第 3 列展示了因子收益率的标准差, 第 4 列展示了因子收益率的年化夏普比率, 第 5 列展示了因子收益率的偏度, 第 6 列展示了因子收益率的峰度。学习因子 *LI* 的月度收益率均值是 0.368%, *t* 值是 2.58, 在 1% 的水平上显著。在所有的模型中规模因子都有显著的因子溢价。FMAX6 中的 *UMD*、*PS* 和 *MAX* 因子的月度显著溢价为 0.635%、0.252% 和 0.929%, M4 模型的 *PERF* 因子的月度显著溢价为 0.482%, BF3 模型中的 *PEAD* 因子的月度显著溢价为 0.323%, CH4 中的 *VMG* 和 *PMO* 因子的月度显著溢价为 1.097% 和 0.804%, 在所有具有显著的因子溢价的因子中, 可见 CH4 中 *VMG* 因子的溢价最高。*LI* 因子的标准差在所有因子中处于低位, 在标准差比 *LI* 低的因子中, *LI* 也有更高的夏普比率。具体来说, *LI* 因子的夏普比率达到 0.50, 夏普比率高于 *LI* 因子的因子包括: CH3、CH4 中的 *VMG* 因子 (1.01) 和 CH4 中的 *PMO* 因子 (0.78), FF5、Q4 中的规模因子 (分别为 0.65 和 0.52), FMAX6 中的 *MAX* 因子 (0.86)。

表 6-2 提供了学习因子 *LI* 与其他因子的相关系数矩阵。可以观察到, *LI* 与 FMAX6 中的 *MAX* 因子相关系数最大, 相关系数值为 -0.47, 其次, BF3 中的 *FIN*、CH4 中的 *PMO*, FF5 中的 *MKT*、*CMA*, FMAX6 中的 *UMD*, CH4 中的 *VMG*、*PMO* 相关性较高。但是总体来看, *LI* 因子和其他模型中的因子相关性不强。

表 6-1 常见多因子模型的描述性统计

模型	因子	均值	<i>t-stat</i>	标准差	夏普比率	偏度	峰度
CH4-LI	<i>MKT</i>	0.662	1.21	7.43	0.31	-0.24	1.32
	<i>SMB</i>	0.515*	1.78	4.50	0.40	0.06	1.76
	<i>VMG</i>	1.097***	5.52	3.76	1.01	0.11	1.19
	<i>PMO</i>	0.804***	4.24	3.55	0.78	-0.73	6.20
	<i>LI</i>	0.368***	2.58	2.53	0.50	0.65	3.69
FF3	<i>MKT</i>	0.701	1.24	7.49	0.32	-0.19	1.40
	<i>SMB</i>	0.659**	2.06	5.10	0.45	-0.07	2.95
	<i>HML</i>	0.112	0.59	3.24	0.12	-0.15	3.73
FF5	<i>MKT</i>	0.701	1.24	7.49	0.32	-0.19	1.40
	<i>SMB</i>	0.705***	2.97	3.78	0.65	0.20	3.92
	<i>HML</i>	0.141	1.18	2.12	0.23	-0.02	3.88
	<i>RMW</i>	0.051	0.39	2.07	0.09	-0.28	3.08
	<i>CMA</i>	-0.006	-0.09	1.24	-0.02	-0.07	1.03
FMAX6	<i>MKT</i>	0.701	1.24	7.49	0.32	-0.19	1.40
	<i>SMB</i>	0.659**	2.06	5.10	0.45	-0.07	2.95
	<i>HML</i>	0.112	0.59	3.24	0.12	-0.15	3.73
	<i>UMD</i>	0.635**	1.97	5.23	0.42	-0.05	1.38
	<i>PS</i>	0.252*	1.86	2.29	0.38	-0.61	4.35
	<i>MAX</i>	0.929***	4.74	3.72	0.86	0.13	1.60
Q4	<i>MKT</i>	0.701	1.24	7.49	0.32	-0.19	1.40
	<i>SIZE</i>	0.615**	2.46	4.11	0.52	-0.15	6.85
	<i>ROE</i>	0.182	0.65	3.82	0.17	-0.98	14.68
	<i>I/A</i>	-0.045	-0.23	3.62	-0.04	0.04	7.60
M4	<i>MKT</i>	0.701	1.24	7.49	0.32	-0.19	1.40
	<i>MGMT</i>	0.235	1.57	2.48	0.33	-0.17	0.92
	<i>PERF</i>	0.482*	1.85	4.29	0.39	0.11	0.99
	<i>SIZE</i>	0.313**	2.07	2.51	0.43	-0.16	5.04
BF3	<i>MKT</i>	0.701	1.24	7.49	0.32	-0.19	1.40
	<i>PEAD</i>	0.323**	2.01	2.74	0.41	-0.24	1.84
	<i>FIN</i>	0.246	1.23	3.38	0.25	-0.04	0.56
CH3	<i>MKT</i>	0.662	1.21	7.43	0.31	-0.24	1.32
	<i>SMB</i>	0.515*	1.78	4.50	0.40	0.06	1.76
	<i>VMG</i>	1.097***	5.52	3.76	1.01	0.11	1.19
CH4	<i>MKT</i>	0.662	1.21	7.43	0.31	-0.24	1.32
	<i>SMB</i>	0.515*	1.78	4.50	0.40	0.06	1.76
	<i>VMG</i>	1.097***	5.52	3.76	1.01	0.11	1.19
	<i>PMO</i>	0.804***	4.24	3.55	0.78	-0.73	6.20

表 6-2 因子相关系数矩阵

	<i>LI</i>	<i>PS</i>	<i>MAX</i>	<i>ROE</i>	<i>I/A</i>	<i>MGMT</i>	<i>PERF</i>	<i>PEAD</i>	<i>FIN</i>	<i>MKT</i>	<i>SMB</i>	<i>HML</i>	<i>RMW</i>	<i>CMA</i>	<i>UMD</i>	<i>VMG</i>	<i>PMO</i>
<i>LI</i>	1.00																
<i>PS</i>	0.05	1.00															
<i>MAX</i>	-0.47	0.08	1.00														
<i>ROE</i>	0.10	-0.03	0.03	1.00													
<i>I/A</i>	-0.00	0.10	0.08	-0.07	1.00												
<i>MGMT</i>	-0.09	0.16	0.43	0.08	-0.04	1.00											
<i>PERF</i>	0.02	-0.28	-0.07	0.13	-0.31	0.37	1.00										
<i>PEAD</i>	0.07	-0.20	-0.15	0.15	-0.39	0.19	0.44	1.00									
<i>FIN</i>	-0.27	0.07	0.22	-0.14	0.20	0.05	-0.29	-0.24	1.00								
<i>MKT</i>	0.29	0.12	-0.29	0.17	0.28	-0.25	-0.25	-0.16	-0.06	1.00							
<i>SMB</i>	-0.18	0.05	-0.06	-0.25	0.22	-0.48	-0.44	-0.38	0.36	0.07	1.00						
<i>HML</i>	-0.09	0.19	0.35	-0.00	0.15	0.62	0.16	0.02	-0.09	-0.03	-0.48	1.00					
<i>RMW</i>	-0.10	-0.15	0.21	0.32	-0.31	0.50	0.55	0.47	-0.28	-0.31	-0.70	0.35	1.00				
<i>CMA</i>	-0.28	0.23	0.33	-0.18	0.44	0.22	-0.28	-0.26	0.43	-0.02	0.20	0.31	-0.23	1.00			
<i>UMD</i>	0.20	-0.26	-0.24	-0.03	-0.26	0.10	0.64	0.56	-0.20	-0.07	-0.25	-0.09	0.30	-0.26	1.00		
<i>VMG</i>	-0.21	0.07	0.44	0.21	-0.26	0.64	0.38	0.36	-0.16	-0.26	-0.62	0.52	0.78	-0.07	0.10	1.00	
<i>PMO</i>	-0.37	-0.25	0.54	-0.01	0.11	0.04	-0.04	-0.13	0.33	-0.26	0.17	-0.12	0.077	0.11	-0.06	0.06	1.00

6.2.3 因子的冗余性检验

Barillas 和 Shanken (2017) 认为, 检验一个因子的定价能力的前提是这个因子无法被其他多因子模型解释, 如果模型中的因子能够被其他多因子模型所解释, 那么这个因子就是冗余因子, 没有包含增量信息。因此, 本小节对学习因子 *LI* 进行冗余性检验, 考察学习因子 *LI* 的是否能够被其他的多因子模型所解释。学习因子 *LI* 是 CH4-LI 五因子模型的核心新因子, 因此, 只需要对学习因子进行考察, 使用学习因子 *LI* 的收益率对其他多因子模型进行时间序列回归, 查看回归结果常数项 (*Const*), 如果常数项 (*Const*) 与 0 有显著差异, 那就表明学习因子 *LI* 无法被其他的多因子模型所解释, 学习因子 *LI* 就不是一个冗余的因子。

本小节使用学习因子 *LI* 收益率依次对其他的多因子模型进行时间序列回归, 考察 *LI* 的冗余性。时间序列回归中考虑的多因子模型有型 FF3、FF5、CH3、CH4、FMAX6、Q4、M4 和 BF3。

表 6-3 展示了因子冗余性检验的时间序列回归的结果。每一列展示了一个模型的回归结果, 学习因子 *LI* 对所有模型的时间序列回归截距项都是正的, 并且在所有因子的时间序列回归中, 常数项的 *t* 值都大于 1.96, 也就是说, 学习因子的溢价都至少在 5% 的水平上显著, 且为正数。因此, 学习因子不是一个冗余的因子, 他不能被主流的多因子模型所解释。具体而言, FF3 的市场因子 *MKT* 表现出了对 *LI* 因子的正向解释能力, FF5 的市场因子 *MKT* 表现出了正向的解释能力, 而 FF5 的规模因子 *SMB*、盈利因子 *RMW*、投资因子 *CMA* 均对 *LI* 表现出负向的解释能力。了学习指数因子 *LI* 的解释能力, CH3、CH4 中的各个因子都表现出了对 *LI* 的一定解释能力, CH3、CH4 中除 *MKT* 表现出正向解释能力外都表现出负向解释能力, FMAX6 中的市场因子 *MKT* 对 *LI* 表现出了正向的解释能力、彩票偏好因子 *MAX* 对 *LI* 表现出了负向解释能力, Q4 中的市场因子 *MKT* 对 *LI* 表现出了正向解释能力、规模因子 *SMB* 对学习因子 *LI* 表现出了负向解释能力, M4 中市场因子 *MKT* 对 *LI* 表现出了正向解释能力、规模因子 *SMB* 表现出了负向解释能力, BF3 中市场因子 *MKT* 对 *LI* 表现出了正向解释能力和长时间尺度 *FIN* 因子对 *LI* 表现出了负向解释能力。在所有的多因子模型中, BF3 对因子 *LI* 的解释能力最强, 使得截距项最小, 达到了 0.326%, *t* 值也最小, 达到了 2.04。

总的来说, 因子冗余性检验的结果表明, 没有一个多因子模型可以完全解释学习因子 *LI* 的超额收益, 也就是说, 学习因子 *LI* 含有其他因子没有的增量信息, 并不是一个冗余性因子, 这为后续将学习因子与其他模型进行比较奠定了基础。

表 6-3 因子冗余性检验

	FF3	FF5	CH3	CH4	FMAX6	Q4	M4	BF3
<i>const</i>	0.367** (2.05)	0.564*** (3.66)	0.748*** (4.76)	0.866*** (5.09)	0.581*** (3.05)	0.359** (2.17)	0.401** (2.17)	0.326** (2.04)
<i>MKT</i>	0.104*** (5.05)	0.068*** (3.96)	0.085*** (3.80)	0.060*** (3.00)	0.064*** (3.23)	0.107*** (4.26)	0.095*** (4.19)	0.097*** (4.42)
<i>SMB</i>	-0.097 (-1.48)	-0.297*** (-4.24)	-0.243*** (-4.11)	-0.216*** (-3.97)	-0.069 (-1.35)	-0.112* (-1.92)	-0.246** (-2.08)	
<i>HML</i>	-0.080 (-0.70)	-0.098 (-0.79)			0.044 (0.52)			
<i>RMW</i>		-0.465*** (-3.88)						
<i>CMA</i>		-0.512*** (-3.32)						
<i>UMD</i>					0.049 (0.91)			
<i>VMG</i>			-0.247*** (-4.22)	-0.254*** (-4.39)				
<i>PMO</i>				-0.184*** (-2.86)				
<i>PS</i>					0.095 (1.27)			
<i>MAX</i>					-0.294*** (-5.52)			
<i>ROE</i>						0.005 (0.11)		
<i>IA</i>						-0.038 (-0.74)		
<i>MGMT</i>							-0.160 (-1.38)	
<i>PERF</i>							0.031 (0.55)	
<i>PEAD</i>								0.056 (0.57)
<i>FIN</i>								-0.180*** (-3.36)
R^2_{adj}	0.104	0.237	0.222	0.280	0.277	0.112	0.126	0.144

6.3 模型定价能力比较

学术界通常使用两种方法比较不同的多因子模型的定价能力。第一种方法 Barillas 和 Shanken (2017) 提出的夏普比率测试, 这种方法通过比较不同多因子模型中的因子资产构造能够构造的所有组合中的夏普比率最大组合的夏普比率, 和相应因子在组合中的权重大小, 对模型进行比较。第二种方法是比较不同多因子模型对市场异象的解释能力, 如果一个多因子模型对市场异象集的解释能力强于另一个多因子模型对同样市场异象集的解释能力, 就可以认为前一个多因子模型的定价能力强于后一个多因子模型。本节同时使用这两种方法比较 CH4-LI 与其他多因子模型的在中国股票市场的定价能力, 以确保研究的稳健性。

6.3.1 夏普比率测试

Barillas 和 Shanken (2017) 认为应该使用多因子模型中的因子作为测试资产, 对模型进行比较, 具体而言, 对不同的多因子模型中的因子资产按照不同的权重进行配置, 可以得到很多个不同的组合, 对于每一个给定的预期收益率 μ , 都能够找到唯一的配置权重 w , 使得该组合在所有预期收益率 μ 的组合中方差最小, 使用所有的最小方差组合可以构建一个均值-方差的最小方差前沿, 考虑无风险利率的情况下, 可以计算切点组合的夏普比率。如果一个多因子模型的切点组合夏普比率高于另一个模型的切点组合的夏普比率, 就可以认为这个多因子模型的定价能力更强。

本小节根据这个思路, 计算依次计算 FF3、FF5、CH3、CH4、FMAX6、Q4、M4、BF3 和 CH4-LI 等多因子模型的最大夏普比率。具体步骤如下,

(1) 计算每一个模型中每一个因子的收益率均值和标准差, 以及因子之间的协方差矩阵, 作为基础因子资产的收益和风险特征。

(2) 随机地给模型的基础因子资产赋予不同的权重 (每个基础因子资产的权重均在范围都在 0 和 1 之间, 并且所有权重的和为 1), 构造 5000 组资产组合, 根据权重计算每个资产组合收益率均值, 以及资产组合的方差。

(3) 使用者 5000 个资产组合构造均值-方差的最小方差前沿, 在考虑无风险资产前提下, 寻找切点投资组合, 并记录切点投资组合的权重。

(4) 使用切点投资组合的权重计算切点组合的风险收益特征以及年化夏普比率。

表 6-4 的 Panel A 给出了各个因子模型切点组合对应的模型因子资产权重, 可以观察到在切点组合中各个因子对夏普比率的贡献。FF5 的投资因子 *CMA* 权重接近于 0, FMAX6 的价值因子 *HML*、Q4 中的投资因子 *I/A* 的权重也只有 0.03、0.02。在 CH4-LI 模型中, 学习因子 *LI* 的权重仅低于 *VMG* 因子, 达到 0.27, 这说明 CH4-LI 模型中夏普比率有很大一部分来源于学习因子 *LI* 的贡献, CH4 和 CH3 中权重

最大的因子也是VMG因子，权重达到 0.45、0.56。Panel B 展示了不同多因子模型切点组合的超额收益率、标准差、夏普比率。可以将这些模型根据年化夏普比率是否大于 1 大致分为两档，在年化夏普比率低于 1 的模型中，FF3、HXZ-Q4 和 BF3 夏普比率较低的原因可能是切点组合的风险较大，而 FF5 和 M4 夏普比率较低的原因则是切点组合的组合收益率较低。在年化夏普比率大于 1 的模型中，CH4-LI 模型的夏普比率最高，达到 1.716。因此，从夏普比率测试的结果来看，学习因子 LI 具有其他多因子模型无法解释的收益率，并且对切点组合的夏普比率贡献很高。

表 6-4 夏普比率测试

	FF3	FF5	FMAX6	Q4	M4	BF3	CH3	CH4	CH4-LI
Pannel A: 模型张成的切点组合各因子资产的权重									
<i>MKT</i>	0.20	0.09	0.09	0.20	0.13	0.23	0.09	0.11	0.05
<i>SMB</i>	0.52	0.46	0.17				0.35	0.23	0.19
<i>HML</i>	0.28	0.19	0.03						
<i>RMW</i>		0.25							
<i>CMA</i>		0.00							
<i>UMD</i>			0.22						
<i>VMG</i>							0.56	0.45	0.34
<i>PS</i>			0.16						
<i>MAX</i>			0.32						
<i>SIZE</i>				0.68	0.43				
<i>ROE</i>				0.11					
<i>I/A</i>				0.02					
<i>MGMT</i>					0.18				
<i>PERF</i>					0.26				
<i>PEAD</i>						0.54			
<i>FIN</i>						0.24			
<i>PMO</i>								0.21	0.16
<i>LI</i>									0.27
Panel B: 模型张成的切点组合的特征									
<i>Exret</i>	0.328	0.242	0.470	0.385	0.203	0.202	0.724	0.665	0.537
<i>std</i>	2.945	1.448	1.481	3.224	1.363	2.039	1.796	1.586	1.084
<i>SR</i>	0.386	0.579	1.099	0.414	0.515	0.343	1.396	1.452	1.716

6.3.2 解释市场异象

Fama 和 French (2018) 使用异象解释能力比较多因子模型之间的差异很容易出现数据窥探的问题，研究者很有可能会选择容易被自己模型的多解释的市场

异象。Liu 等（2019）在检验 CH4 模型在中国市场的定价能力的时候，只选择了中国 A 股市场中的 14 个异象作为异象集，并且没有测试所检验的异象在中国股票市场的显著性。一个更合理的做法是尽可能将所有前人研究发现的异象资产都纳入异象集中，同时首先考察异象变量的显著性，使用显著的异象进行研究。

Qiao（2019）参照 Hou 等（2020）对美股市场的 447 个异象的过程在中国 A 股市场构建了 105 个市场异象。本小节参考 Qiao（2019）和 Hou 等（2020）在中国市场构建了包含交易摩擦、动量、价值成长、投资、盈利、无形资产类等所有类别的市场异象，总计有 105 个市场异象。在进行多因子模型的定价能力比较之前，本文首先检验这些异象在中国股票市场的显著性，随后只保留所有显著的异象进行因子定价能力的比较。

使用单变量排序计算异象变量的收益率：依据上月末的异象变量值将所有股票分为 5 组，然后计算每个组合的等权重平均收益率，将异象收益率计算为买入组合 5、卖出组合 1 的得到的多空组合的收益率，在每月末重复上述步骤，就可以得到异象收益率的时间序列。

首先，在比较多因子模型的定价能力之前，需要检验异象的显著性，通过计算异象的 Newey-West（1987） t 统计量，检验 105 个异象变量在 A 股市场的显著性。若 t 值的绝对值大于 1.65，则认为该异象在 10% 的水平上显著，那么认为该异象在 A 股市场是一个显著的异象。通过这种方法，一共筛选出了 45 个在 A 股市场具有显著超额收益的市场异象，这些异象来自于 6 个不同的分类。表 6-5 提供了 45 个异象变量的分类以及月度超额收益率的描述性统计特征。45 个显著异象分散在 6 个不同的类别，交易摩擦类的异象最多，有 19 个；其次是价值成长类的异象，有 9 个；随后是动量类和投资类异象，都各自有 6 个；紧接着是无形资产类异象，一共有 4 个，最后是盈利类异象只有 1 个。

然后，使用上述 45 个显著的市场异象对多因子模型的定价能力进行比较，依次将这些异象的收益率对上述提到的主流多因子模型以及 CH4-LI 模型进行时间序列回归，观察时间序列回归的结果，包括观察回归常数项（ α ）、常数项（ α ）对应的 Newey-West（1987）调整的 t 统计量、GRS 检验（Gibbons 等，1989）的 F 统计量。本节使用以下 4 个指标比较不同多因子模型对市场异象的解释能力，进而对因子模型的定价能力进行比较。具体如下：

（1）模型可以解释的市场异象个数 N_{exp} ：如果回归的截距项无法在 10% 或者更小的水平上拒绝原假设，也就是 t 统计量小于 1.64，就说明回归截距项与 0 没有显著差异，那么认为该异象可以被模型所解释。一个模型可以解释的异象越多，那么这个模型在中国股票市场中的定价能力就越强。

（2）回归常数项（ α ）绝对值的均值 $|\alpha|_{avg}$ ：一个异象的常数项（ α ）越大，代表这个异象不能被多因子模型解释的超额收益越大，代表这个异象具有越多不

能被模型解释的部分,也就是模型对这个异象解释的能力越小。因此,模型的对所有异象的常数项平均值 $|\alpha|_{avg}$ 越大,代表模型无法解释的异象成分越多,模型对市场异象的解释能力越弱。

(3) 回归常数项(α)对应的 t 值的绝对值的均值 $|t|_{avg}$:这个指标是对 $|\alpha|_{avg}$ 的一个补充,一个异象的常数项(α)对应的 t 值的绝对值越大,代表这个异象不能被多因子模型解释的超额收益越显著,也就是模型对这个异象解释的能力越小。两个指标是相似的指标,一般情况下 $|\alpha|_{avg}$ 越小,对应的 $|t|_{avg}$ 也会越小。也就是说, $|t|_{avg}$ 越小,模型对异象的解释能力强。

(4) GRS 检验的 F 统计量($GRS-F$): GRS 检验考察的是模型对多个异象的联合解释能力,检验模型对多个异象的截距项(α)是否联合为零(原假设)。GRS 检验的前提是被检验的异象个数小于样本期的长度,本节使用的异象有 45 个,样本期的长度为 264 个月。GRS 检验对应的 F 统计量越小,表明原假设被拒绝的概率更小,也就是回归截距项(α)联合靠近 0 的程度更大,模型对股票市场异象的解释能力越强,模型的定价能力越强。

表 6-6 展示了所有的多因子模型对市场异象的解释情况。在解释异象个数方面,CH4-LI 的表现最好,可以解释 31 个是藏异象,略好 CH4 和 CH3 的 28 个,在剩余的模型中,FMAX6 的表现最好,可以解释 20 个异象,其他的模型只能解释 10 个以下的异象,表现最差的是 BF3,只能解释 4 个异象。在回归截距项(α)和对应的 t 值方面,CH4-LI 都是最小的,CH3 和 CH4 的取值比较接近,仅次于 CH4-LI,相比较于其他模型,CH4-LI、CH4、CH3 的(α)和, t 统计量的值更小。同样,在 GRS 检验方面,GRS 检验的 F 统计量,CH4-LI 明显最小,CH4 和 CH3 之间差异不大,仅次于 CH4-LI,也明显优于其他模型。总的来说,CH4-LI 对异象的解释能力最强,CH4、CH3 对异象的解释能力仅次于 CH4-LI。

为了对 CH4 和 CH4-LI 进行进一步比较,观察 LI 对 CH4 的增强作用,本节在表 6-7 中列出不能被 CH4-LI 或 CH4 所解释的 17 个市场异象,这些异象中有 3 个可以被 CH4-LI 模型解释,而 CH4 则无法解释其中的任何一个。能被解释的异象分别是总波动率 TV 、季度更新的负债增长率 LGQ 、最大日收益率 MDR 。从 t 值的绝对值的角度看,除了 BVG 、 DTV 、 $VDTV$ 、 $RDMQ$ 以外,CH4-LI 的 t 值的绝对值都比 CH4 的更低。

综合以上对异象的解释能力,可知 CH4-LI 的定价能力优于 CH4 和 CH3 模型,且明显优于其他模型这就表明学习指数因子 LI 能够提供对一部分市场异象的解释能力。

表 6-5 显著的 45 个异象的描述性统计

异象代码	异象名称	均值	<i>t-stat</i>	标准差	偏度	峰度
交易摩擦类						
<i>RA</i>	季节性 (Heston 和 Sadka, 2008)	0.27	1.71	2.55	-0.10	2.66
<i>SIZE</i>	公司规模 (Banz, 1981)	-1.07	-3.62	4.74	-0.08	2.32
<i>IDSC</i>	CAPM 特质偏度	-0.62	-4.91	2.35	0.54	12.53
<i>IDSFF</i>	FF3 特质偏度 (Boyer 等, 2009)	-0.35	-2.91	1.97	0.31	8.20
<i>IDVC</i>	年度更新 CAPM 特质波动率 (Ali 等, 2003)	-0.42	-1.92	4.09	0.21	3.24
<i>IDVCM</i>	月度更新 CAPM 特质波动率	-0.81	-4.03	4.05	-0.02	2.70
<i>IDVFF</i>	FF3 特质波动率 (Ang 等, 2006)	-1.07	-5.38	3.64	-0.11	3.37
<i>TS</i>	总偏度 (Amaya 等, 2015)	-0.49	-3.61	2.34	0.10	1.35
<i>CS</i>	协偏度 (Harvey 和 Siddique, 2000)	0.67	2.61	4.27	0.70	2.12
<i>TV</i>	总波动率 (Ang 等, 2006)	-0.47	-2.00	4.67	0.36	0.94
<i>TURN</i>	换手率 (Datar 等, 1998)	-0.45	-1.78	4.72	0.04	1.61
<i>VDTV</i>	成交额变化 (Chordia 等, 2001)	-1.549	-6.49	4.13	-0.25	2.17
<i>VTURN</i>	换手率变化率 (Chordia 等, 2001)	-0.61	-2.41	4.55	-0.08	2.35
<i>DTV</i>	六个月日均成交额 (Brennan 等, 1998)	-1.37	-5.45	4.21	-0.21	2.06
<i>BETADM</i>	市场 BETA (Dimson, 1979)	0.58	2.90	3.33	0.71	3.21
<i>BETAFP</i>	市场 BETA (Frazzini 和 Pedersen, 2014)	-0.43	-1.73	4.74	0.39	0.53
<i>MDR</i>	最大日收益率 (Bali 等, 2011)	-0.69	-3.28	4.26	-0.22	1.83
<i>PR</i>	股票价格 (Miller 和 Scholes, 1982)	-0.88	-4.34	3.55	-0.12	1.73
<i>SREV</i>	短期回报反转 (Jegadeesh, 1990)	-0.88	-3.50	4.38	-0.35	2.18
动量类						
<i>M24</i>	24 个月动量 (Jegadeesh 和 Titman, 1993)	-0.69	-3.53	3.65	0.31	3.16
<i>MAX</i>	52 周高点 (George 和 Hwang, 2004)	-0.84	-4.33	3.44	-0.14	1.63
<i>MCHG</i>	动量变化 (Gettleman 和 Marks, 2006)	-0.55	-2.88	3.47	-0.76	2.08
<i>SUE</i>	标准化盈利惊喜 (Foster 等, 1984)	0.38	2.96	2.31	1.11	4.56
<i>TES</i>	税费支出惊喜 (Thomas 和 Zhang, 2011)	0.30	2.17	2.55	0.65	6.25
<i>RS</i>	利润惊喜 (Jegadeesh 和 Livnat, 2006)	0.26	1.79	2.67	-0.32	2.12

表 (续) 6-5 显著的 45 个异象的描述性统计

异象代码	异象名称	均值	<i>t-stat</i>	标准差	偏度	峰度
价值和成长类						
<i>BM</i>	账面市值比 (Basu, 1983)	0.56	2.54	4.03	0.11	3.04
<i>BMJ</i>	账面市值比 (Asness 和 Frazzini, 2013)	0.63	2.61	4.58	-0.08	3.57
<i>BMQ</i>	季度更新的账面市值比	0.64	2.36	5.05	-0.05	2.75
<i>AM</i>	年度更新的资产市值比	0.43	1.79	4.29	0.10	3.14
<i>AMQ</i>	季度更新的资产市值比	0.57	1.97	5.20	-0.05	3.11
<i>LREV</i>	长期反转 (De Bondt 和 Thaler, 1985)	-0.68	-3.38	3.63	0.17	3.20
<i>EM</i>	企业乘数 (Loughran 和 Wellman, 2011)	-0.20	-1.68	2.01	-0.31	1.37
<i>SP</i>	销售市值比 (Barbee 等, 1996)	0.35	1.73	3.64	-0.24	4.72
<i>SPQ</i>	季度更新的销售市值比	0.51	2.29	3.97	-0.18	4.52
投资类						
<i>CEI</i>	复合股权发行 (Daniel 和 Titman, 2006)	-0.33	-2.66	2.35	-0.09	0.14
<i>ACC</i>	应计资产比 (Richardson 等, 2005)	-0.31	-2.39	2.15	-1.57	10.22
<i>ACCP</i>	应计利润百分比 (Hafzalla 等, 2011)	-0.21	-1.99	1.71	-0.55	1.63
<i>AGE</i>	公司年龄 (Jiang 等, 2005)	-0.43	-2.52	3.20	-0.18	2.87
<i>LGQ</i>	季度更新的负债增长率	0.21	1.79	1.94	0.26	1.41
<i>BVG</i>	账面价值增长 (Richardson 等, 2005)	-0.27	-1.84	2.43	-0.11	1.34
盈利类						
<i>SGQ</i>	季度更新的销售增长	0.27	1.97	2.42	-0.09	0.82
无形资产类						
<i>RDM</i>	研发市值比 (Chan 等, 2011)	0.65	4.29	2.57	0.26	2.55
<i>RDMQ</i>	季度更新的研发市值比	0.87	5.01	3.24	0.11	1.64
<i>RDS</i>	研发销售比 (Chan 等, 2011)	0.29	1.78	3.09	0.34	4.35
<i>RDSQ</i>	季度更新的研发销售比	0.30	1.79	3.09	0.14	3.45

表 6-6 不同模型对测试异象集的定价表现

	FF3	FF5	FMAX6	Q4	M4	BF3	CH3	CH4	CH4-LI
N_{exp}	7	6	20	6	9	4	28	28	31
$ \alpha _{avg}$	0.479	0.430	0.344	0.615	0.516	0.570	0.300	0.321	0.295
$ t _{avg}$	2.897	2.618	1.974	3.536	2.817	3.002	1.636	1.612	1.403
<i>GRS-F</i>	5.448	4.409	4.457	5.342	5.191	5.547	4.001	3.934	3.122

表 6-7 CH4 或 CH4-LI 不能解释的异象

异象代码	超额收益	Std dev	CH4	CH4-LI
<i>SIZE</i>	-1.068 (-3.60)	4.743	-0.915 (-5.69)	-0.946 (-5.51)
<i>ACC</i>	-0.306 (-2.38)	2.151	-0.361 (-2.39)	-0.382 (-2.29)
<i>BETADM</i>	0.577 (2.89)	3.327	0.745 (3.50)	0.416 (2.15)
<i>BVG</i>	-0.273 (-1.82)	2.426	-0.271 (-1.90)	-0.303 (-2.03)
<i>CS</i>	0.669 (2.59)	4.271	1.032 (4.20)	0.656 (2.97)
<i>DTV</i>	-1.369 (-5.56)	4.207	-0.584 (-3.18)	-0.872 (-4.47)
<i>IDSC</i>	-0.621 (-4.96)	2.345	-0.758 (-4.70)	-0.758 (-3.85)
<i>IDSFF</i>	-0.353 (-2.95)	1.967	-0.327 (-2.53)	-0.374 (-2.35)
<i>LGQ</i>	0.208 (1.79)	1.942	0.242 (1.68)	0.211 (1.31)
<i>LREV</i>	-0.675 (-3.40)	3.625	-0.497 (-2.12)	-0.471 (-1.92)
<i>M24</i>	-0.690 (-3.53)	3.646	-0.540 (-2.39)	-0.503 (-2.03)
<i>MDR</i>	-0.690 (-3.31)	4.256	0.427 (1.96)	0.208 (0.72)
<i>RDS</i>	0.293 (1.77)	3.086	0.510 (2.98)	0.554 (2.84)
<i>RDMQ</i>	0.299 (1.78)	3.092	0.581 (3.34)	0.653 (3.44)
<i>TS</i>	-0.494 (-3.62)	2.337	-0.391 (-2.37)	-0.321 (-1.76)
<i>TV</i>	-0.467 (-1.99)	4.666	0.572 (2.31)	0.181 (0.61)
<i>VDTV</i>	-1.549 (-6.59)	4.128	-0.691 (-3.73)	-0.939 (-4.78)

6.4 稳健性检验

6.4.1 Bootstrap 方法

为了进一步研究学习因子在中国市场定价能力的稳健性, 本小节采用 Fama 和 French (2016) 的 Bootstrap 检验进行稳健性检验。Bootstrap 的本质是通过模拟抽样进行夏普比率测试, 比较 CH4-LI 模型和其他模型在样本外的最大月度夏普比率, 如果 CH4-LI 模型的比其他模型的夏普比率更高, 就可以说明 CH4-LI 模型具有更好的定价能力, 也就是说 LI 因子对 CH4 模型进行了很好的补充。

Bootstrap 检验的具体步骤如下:

(1) 本文的研究数据包括 264 个月, 将多因子模型的因子收益按时间顺序排序好, 随后从第一期开始, 每 2 个月地因子数据放在一组, 即 (1, 2)、(3, 4) ... (263, 264), 括号中的数字代表对应的月数, 最终可以得到 132 组数据, 这些数据构成了 Bootstrap 抽样的基础数据。

(2) 从 132 组数据中有放回地抽取一组数据, 被抽取出来的这组数据中有两个数据。再从这两个数据中随机抽取一个数据, 这个数据放入样本内的数据集,

另一个数据放入样本外的数据集，总共抽取 132 次，最后得到 132 个样本内数据和 132 个样本外数据。

(3) 使用样本内数据集进行夏普比率测试，得到多因子模型的样本内的最大夏普比率和对应的因子权重。

(4) 使用样本内数据中计算出的因子权重，在样本外数据中构建因子投资组合，这个组合的夏普比率也就是样本外的数据集的夏普比率。

(5) 依次重复步骤 (1) (2) (3) (4) 一共 9999 次。

(6) 计算总共 10000 次地模拟结果的夏普比率地均值以及中位数，比较不同多因子模型的最大夏普比率的差异。

表 6-8 给出了不同多因子模型的 Bootstrap 检验结果。Panel A 展示了样本内、样本外、全样本的均值和中位数，其中 CH4-LI 的夏普比率全面领先于其他的多因子模型，样本内夏普比率的均值达到 0.310，样本外夏普比率均值达到 0.200，全样本夏普比率均值达到 0.243，样本内、样本外、全样本的中位数也由于其他的多因子模型，达到 0.299、0.190、0.238，除了 CH4-LI 之外，表现最好的依次是 CH4、CH3、FMAX6。Panel B 展示了 CH4-LI 与其他多因子模型的夏普比率均值之差和对应的 t 统计量，与 Panel A 的结果一致，与 CH4-LI 差异最小的三个模型依次是 CH4、CH3、FMAX6，CH4-LI 与 CH4 的在样本内夏普比率之差为 0.084， t 统计量的值为 126.34，在 1% 的水平上显著，样本外的夏普比率之差为 0.054， t 统计量的值为 85.36，在 1% 的水平上显著，全样本的夏普比率之差是 0.066， t 统计量的值为 159.04，这些结果表明 CH4-LI 显著优于 CH4 模型，同样的，CH4-LI 也优于其他的多因子模型。表 6-8 的结果表明，学习因子 LI 显著增强了 CH4 模型，提高了 CH4 模型的夏普比率，这说明学习因子 LI 对 CH4-LI 模型在中国股票市场的定价能力具有显著的边际贡献。

表 6-8 Bootstrap 方法

Panel A: Bootstrap 抽样夏普比率						
模型	IS-mean	OS-mean	ALL-mean	IS-median	OS-median	ALL-median
FF3	0.046	0.015	0.018	0.033	0.006	0.012
FF5	0.063	0.020	0.032	0.055	0.010	0.028
FMAX6	0.177	0.060	0.101	0.167	0.046	0.094
HXZ	0.043	0.019	0.020	0.032	0.008	0.015
M4	0.066	0.014	0.027	0.058	0.007	0.022
BF3	0.035	0.012	0.016	0.028	0.006	0.012
CH3	0.204	0.131	0.159	0.193	0.122	0.154
CH4	0.226	0.146	0.177	0.216	0.137	0.173
CH4-LI	0.310	0.200	0.243	0.299	0.190	0.238
Panel B: CH4-LI 和其他模型的夏普比均值的差						
	样本内	<i>t-stat</i>	样本外	<i>t-stat</i>	全样本	<i>t-stat</i>
FF3	0.264	271.52	0.185	206.01	0.225	333.64
FF5	0.247	256.78	0.180	203.09	0.211	322.65
FMAX6	0.133	119.33	0.141	149.73	0.142	202.86
HXZ	0.267	272.35	0.181	204.11	0.223	332.99
M4	0.244	227.24	0.186	202.57	0.216	325.07
BF3	0.275	257.20	0.188	203.46	0.227	333.68
CH3	0.106	132.60	0.069	86.18	0.084	155.12
CH4	0.084	126.34	0.054	85.36	0.066	159.04

6.4.2 剔除“壳污染”后的股票

Liu 等（2019）指出中国严格的 IPO 制度会致使一些市值较小的公司存在被借壳上市或反向并购的可能，而“壳公司”的价值波动会显著削弱因子的定价能力，因子对异象的解释能力下降。Liu 等（2019）在构造 CH4 模型的时候剔除了 30% 的小市值股票，以解决“壳价值污染”的问题。为了防止“壳价值污染”对学习因子 *LI* 在中国股票市场的定价能力造成影响，本文参考 Liu 等（2019）剔除了 30% 小市值股票，再按照 6.2.1 的步骤构造学习因子 *LI*，将学习因子 *LI* 与 CH4 结合得到 CH4-LI 五因子模型，并 CH4-LI 对市场异象的解释能力。

表 6-9 展示了剔除了“壳污染”的股票之后的学习指数对股票市场异象的解释情况。CH4-LI 依旧能够解释 31 个市场异象，在所有模型中表现最佳。在异象解释的 Alpha 绝对值的平均值、*t* 统计量绝对值的平均值、GRS 检验的 *F* 统计量方面，CH4-LI 也是表现最优的。表 6-9 的结果在定性上和前文的研究结果是相似的，在剔除了“壳污染”的股票后，学习因子的定价能力依旧是稳健的。

表 6-9 剔除“壳污染”后的学习因子对异象的解释能力

	FF3	FF5	FMAX6	Q4	M4	BF3	CH3	CH4	CH4-LI
N_{exp}	7	6	20	6	9	4	28	28	31
$ \alpha _{avg}$	0.479	0.430	0.344	0.615	0.516	0.570	0.300	0.321	0.300
$ t _{avg}$	2.897	2.618	1.974	3.536	2.817	3.002	1.636	1.612	1.419
$GRS-F$	5.448	4.409	4.457	5.342	5.191	5.547	4.001	3.934	3.117

第七章 结论、启示与展望

7.1 研究结论

本文以中国 A 股市场的所有上市公司作为主要研究对象, 基于个股的周数据, 估计了个股的学习指数, 研究了学习指数对股票横截面收益的预测能力, 对学习指数的预测能力进行了分解分析, 考察了学习指数对错误定价的影响, 还使用学习指数构建了一个学习因子, 考察了学习因子在中国股票市场的定价能力, 本文的研究丰富了已有的关于投资者学习的相关文献。本文的主要结论如下:

第一, 个股的学习指数与股票的横截面收益呈现出负相关关系, 也就是说一个资产被投资者学习的程度越高, 其未来的回报就相对越低, 这种负相关关系不能被常见的公司特征变量和主流的多因子模型所解释。本文对学习指数与股票长期收益的分析表明, 学习指数对股票横截面收益的预测能力可以持续到第 5 个月, 随后消失, 最终接近于 0 且不会发生反转导致超额收益显著为正, 这说明随着越来越多的投资者观察到与资产相关的信息并且进行学习, 股票的价格逐渐回归到股票的基本价值。样本外检验的结果也表明考虑了学习指数的 CAPM 模型优于其他主流的多因子模型。本文还进一步对学习指数的预测能力进行了分解分析, 由于学习指数与信息流相关的变量存在正相关关系, 本文使用一个两阶段分解的方法对学习指数的预测能力进行分解, 证实了学习指数的预测能力主要来源于投资者学习而不是其他方面, 支持了 VNV 模型的推论。

第二, 本文还分析了学习指数对错误定价的影响, 当投资者获取信息并构建投资组合的时候应当可以降低股票的错误定价。本文研究发现股票的错误定价程度在学习指数较高的股票组合中显著更小, 同时学习指数对股票横截面收益的预测能力在被低估的股票组中显著更强, 长期分析表明这种关系具有长期的持续性, 并且随后逐渐缩小最终消失, 这说明随着投资者对信息的获取, 可以对股票的错误定价程度具有长期的影响。本文的研究表明学习指数和股票预期收益的负相关关系在被低估的股票中更强, 这种现象的原因可能是被高估的股票具有更多的套利限制, 投资者无法根据学习的结果构建投资组合, 同时这种现象在高情绪期要更为明显, 这可能是因为股票高估的情况主要出现在高情绪期。

第三, 学习指数背后的具有坚实的理论基础——VNV 模型, 学习指数对股票横截面收益的负向预测能力, 不能被主流的多因子模型和常见的公司特征所解释, 因此本文使用学习指数构建了一个学习因子。因子冗余性检验的结果表明, 学习因子无法被其他主流多因子模型所解释, 包含了其他因子无法解释的增量信息, 排除了学习因子是冗余性因子的可能, 这为使用学习因子与其他多因子模型结合奠定了基础。本文随后将学习因子 LI 与 Liu 等 (2019) 提出的中国市场的

四因子模型 CH4 结合构建了一个新的五因子模型 CH4-LI。相比于其他主流的多因子模型, CH4-LI 在中国股票市场具有较强的市场定价能力。首先, 由 CH4-LI 五因子模型中的因子张成的切点组合的具有最高的夏普比率, 达到 1.716, 优于其他主流的多因子模型。其次, 学习因子 LI 在切点组合中的权重达到 0.27, 为 CH4-LI 的切点组合贡献了较高的权重。最后, 本文使用在中国股票市场显著的 45 个异象考察 CH4-LI 和其他主流多因子模型对市场异象的解释能力, CH4-LI 五因子模型可以解释的异象最多, 达到了 31 个, 优于其他主流的多因子模型。

7.2 研究启示

本文研究的发现可以为在中国市场进行交易的投资者提供投资建议, 为对中国市场感兴趣的研究者提供研究参照和启示, 为金融市场监管机构提供政策建议。鉴于投资者学习在中国股票市场的重要影响, 不管是在中国市场进行交易的投资者、对中国市场感兴趣的研究者还是金融市场监管机构, 都应该给予充分的重视。

(1) 对于在中国市场进行交易的投资者而言, 本文的研究结果有助于指导投资者在中国股票市场的投资实践。

第一, 学习指数较高的股票未来收益更低, 这与 VNV 模型的推论保持一致, 这个结果表明投资者应该在市场上获取一些独特的信息, 深入研究一些其他投资者没有深入研究的资产, 根据信息获取和学习的结果来构建自己的投资组合, 才能获得超额收益。如果投资者选择研究很多人都能获得的信息、与其他投资者学习相同的资产, 最终构建与其他投资者类似的投资组合, 导致资产组合的同质化, 进而导致投资收益率的下降, 这也是学习指数对股票横截面预期收益存在预测能力的原因。

第二, 本文的研究结果可以帮助股票市场的量化投资者构建相关的投资策略, 由于学习指数可以显著预测股票的横截面收益, 学习指数高的股票未来显著收益更低, 量化投资者可以将学习指数作为股票收益的预测指标之一, 在每个月末估计股票的学习指数, 根据本文的研究结果, 做多学习指数较低的股票, 做空学习指数较高的股票, 构建零成本的多空组合, 获取超额收益。同时, 学习指数无法被其他常见的多因子模型所解释, 与常见的公司特征相关性较低, 将学习指数多空组合添加到策略组合当中, 可以帮助量化投资者改善最终策略的表现。

第三, 本文发现学习指数与错误定价的关系与投资者情绪密切相关。所以学习指数对股票收益的预测能力可能在不同的市场状态变量下存在差异, 这意味着依据学习指数构建的股票多空组合超额收益具有时变性, 投资者需要在学习指数多空组合进行择时, 寻找投资者情绪、市场流动性等变量尝试预测学习指数多空组合的超额收益, 在预期学习指数多空组合表现好的时候配置更高的权重, 在预期学习指数多空组合表现差的时候配置更低的权重。

(2) 对于对中国市场感兴趣的研究者而言, 本文提供了学习指数与股票横截面预期收益的实证证据。

第一, 中国股票市场在投资者结构和制度安排等方面与国外成熟的市场不同, 因此具有不同的市场特征, 在中国市场进行研究时, 要充分考虑中国市场的市场特征如投资者结构、投资者行为、做空限制等对研究问题的影响。中国市场具有不同的投资者结构, 相比较于国外成熟的股票市场, 中国股票市场的机构投资者比例相对较低, 而散户投资者比例相对较高, 散户投资者更容易受到非理性行为的影响, 因此市场的投机氛围较强, 投资者也更短视, 所以在构建指标的时候根据中国市场的实际情况做出调整。同时, 中国市场的制度相对于国外成熟的股票市场而言不够完善, 存在做空限制等制度缺陷, 这就导致中国股票市场股票更容易被高估。这些原因都可能导致国外成熟市场的研究结果与中国市场存在差异, 很多研究结论不能简单照搬到中国市场。

第二, 学习指数对股票横截面预期收益预测能力在中国股票市场显著存在, 并且无法被其他主流的多因子模型所解释, 这表明学习指数多空组合是一个市场异象。其他学者在构造多因子模型的时候, 可以将学习指数多空组合作为一个市场异象, 纳入市场异象的资产测试集, 考察学者提出的多因子模型是否具有足够强的异象解释能力, 是否优于其他的主流多因子模型。相比较于美国市场, 中国市场的异象测试集较小, 本文实证检验的学习指数多空组合为中国股票市场的因子定价研究提供了一个新的异象资产。

第三, 本文提出的 CH4-LI 五因子模型可以解释更多的市场异象, 具有更强的市场异象解释能力, 其他学者在研究中国股票市场中的市场异象时, 可以将 CH4-LI 当作一个新的多因子模型, 计算异象的 CH4-LI 调整后超额收益, 如果超额收益不显著, 说明这个变量不是一个真正的市场异象, 在股票横截面预期收益的预测方面不能提供增量信息, 只是一个冗余的股票收益预测变量。

(3) 对于市场金融监管机构而言, 本研究可以帮助市场金融监管机构全面了解投资者学习在中国股票市场中的影响。这有助于金融监管机构制定相关的政策, 缓解市场中的信息不对称问题, 为市场提供一个良好的信息环境, 促进金融市场的健康发展。针对这个方面, 本文提出以下政策建议。

第一, 推进上市公司信息披露制度的完善, 增加企业的信息透明度。对于公告信息造假、财务造假的上市公司, 给予严厉地惩罚, 给投资者提供更多真实的信息。一个良好的信息环境, 一方面可以让投资者可以获取信息进行学习, 全方位了解上市公司的实际情况, 减少投资风险, 也减少股票的错误定价, 另一方面可以让投资者获得真实的信息, 不需要担心虚假信息, 才能转变投资理念, 改善当前中国股票市场的投机氛围。

第二, 扩大融资融券业务覆盖的股票范围, 降低融资融券的准入门槛。长期

以来中国市场都存在严重的做空限制，投资者无法通过空头获取收益，导致投资者不愿意学习被高估的股票，即使学习了高估的股票也没有办法纳入投资组合，导致了被高估的股票持续高估，导致错误定价的长期存在，最终导致股票的崩盘风险。融资融券业务是当前卖空制度中最重要的一环，然而融资融券制度的股票范围较小，很多股票无法进行融资融券，同时融资融券的准入门槛较高，一般的投资者无法达到，这就导致在中国股票市场中难以通过全市场的力量修正股票的错误定价。

第三，进一步发展和规范化分析师市场。证券分析师是金融市场中重要的信息中介，证券分析师有更多的渠道与上市公司沟通，获取独特的信息，同时，证券分析师也是专业的投资研究人员，专业能力更强。证券分析师提供的研究报告能够帮助普通投资者了解上市公司的信息，减少市场中的信息不对称。进一步发展和规范化分析师市场，可以改善市场中的信息环境，让投资者有更多更深入的信息进行学习，降低投资者的风险。

7.3 研究展望

本文针对投资者学习在中国市场的影响研究取得了一定的成果，对学习指数对股票横截面收益的负向预测能力进行了实证检验，然而，本文还是存在一些不足之处，未来可以从以下几个方面展开更深入的研究和讨论：

第一，进一步研究投资者学习与股票横截面收益之间的关系在不同的数据频率上是否存在。可以在保持学习指数指标构建方法不变的基础上，考虑使用日度数据构建学习指数，或者考察学习指数在日度、周度、季度、年度的频率上对股票横截面收益的预测能力是否存在，并比较现有的月度频率的学习指数预测能力来源是否与其他频率相似，也可以使用更高频的日内数据进行研究，不同频率学习指数对股票收益的预测效果可能不同，值得进行更深入的挖掘。

第二，研究学习因子的择时问题。本文构造一个学习因子并研究学习因子对异象的解释能力，构造了一个新的多因子模型，但是没有对学习因子的风险收益特征进行更进一步的研究。因子的收益率实际上表现的是全样本期收益率的平均值，但是从时间序列的角度上看，因子的收益率在不同的时期表现存在很大的差异，每个因子都有表现更好的时期，也有表现更差的时期。因子择时问题一直都是业界和学术界最关注的问题之一，主要目的是在时序上对因子的权重进行配置，通过因子择时的手段提高量化投资策略的收益。但是本文对学习因子的在时序上的变化情况没有进行深入的讨论，后续可以在这个方向进行更深入的研究。

第三，本文构建的学习指数源于理性预期一般均衡模型，投资者对于资产的选择仍然是理性的，但是非理性的现象在市场上广泛存在，后续的研究可以针对理性预期一般均衡模型进行修改，加入投资者非理性行为的部分，构建一个全新

的行为金融模型,并在此基础上使用相同方法估计一个考虑了非理性行为的学习指数,这个全新的学习指数可能会有完全不一样的特征和股票收益的预测能力。

参考文献

- [1] Amihud Y. Illiquidity and stock returns: cross-section and time-series effects[J]. Journal of financial markets, 2002, 5(1): 31-56.
- [2] Ang A, Chen J, Xing Y. Downside risk[J]. The review of financial studies, 2006, 19(4): 1191-1239.
- [3] Ang A, Hodrick R J, Xing Y, et al. The cross-section of volatility and expected returns[J]. The journal of finance, 2006, 61(1): 259-299.
- [4] Atilgan Y, Bali T G, Demirtas K O, et al. Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns[J]. Journal of Financial Economics, 2020, 135(3): 725-753.
- [5] Baker M, Wurgler J. Investor sentiment in the stock market[J]. Journal of economic perspectives, 2007, 21(2): 129-151.
- [6] Bali T G, Brown S J, Murray S, et al. A lottery-demand-based explanation of the beta anomaly[J]. Journal of Financial and Quantitative Analysis, 2017, 52(6): 2369-2397.
- [7] Bali T G, Engle R F, Murray S. Empirical asset pricing: The cross section of stock returns[M]. John Wiley & Sons, 2016.
- [8] Banerjee S. Learning from prices and the dispersion in beliefs[J]. The Review of Financial Studies, 2011, 24(9): 3025-3068.
- [9] Banz R W. The relationship between return and market value of common stocks[J]. Journal of financial economics, 1981, 9(1): 3-18.
- [10] Barber B M, Huang X, Odean T. Which factors matter to investors? Evidence from mutual fund flows[J]. The Review of Financial Studies, 2016, 29(10): 2600-2642.
- [11] Barber B M, Lee Y T, Liu Y J, et al. Is the aggregate investor reluctant to realise losses? Evidence from Taiwan[J]. European Financial Management, 2007, 13(3): 423-447.
- [12] Barberis N, Greenwood R, Jin L, et al. Extrapolation and bubbles[J]. Journal of Financial Economics, 2018, 129(2): 203-227.
- [13] Barberis N, Shleifer A, Vishny R. A model of investor sentiment[J]. Journal of financial economics, 1998, 49(3): 307-343.
- [14] Barillas F, Shanken J. Which alpha?[J]. The Review of Financial Studies, 2017, 30(4): 1316-1338.
- [15] Berger P G, Ofek E. Diversification's effect on firm value[J]. Journal of financial economics, 1995, 37(1): 39-65.
- [16] Berk J B, Van Binsbergen J H. Assessing asset pricing models using revealed preference[J]. Journal of Financial Economics, 2016, 119(1): 1-23.
- [17] Biais B, Bossaerts P, Spatt C. Equilibrium asset pricing and portfolio choice under asymmetric information[J]. The Review of Financial Studies, 2010, 23(4): 1503-1543.
- [18] Boehmer E, Jones C M, Zhang X, et al. Tracking retail investor activity[J]. The Journal of Finance, 2021, 76(5): 2249-2305.
- [19] Burlacu R, Fontaine P, Jimenez-Garcés S, et al. Risk and the cross section of stock returns[J]. Journal of Financial Economics, 2012, 105(3): 511-522.
- [20] Campbell J Y, Hilscher J, Szilagyi J. In search of distress risk[J]. The Journal of finance, 2008, 63(6): 2899-2939.
- [21] Cao C, Gempesaw D, Simin T T. Information choice, uncertainty, and expected returns[J]. The Review of Financial Studies, 2021, 34(12): 5977-6031.

- [22] Cao J, Han B. Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns[J]. *Journal of Banking & Finance*, 2016, 73: 1-15.
- [23] Carhart M M. On persistence in mutual fund performance[J]. *The Journal of finance*, 1997, 52(1): 57-82.
- [24] Chen L, Novy-Marx R, Zhang L. An alternative three-factor model[J]. Available at SSRN 1418117, 2011.
- [25] Chui A C W, Subrahmanyam A, Titman S. Momentum, reversals, and investor clientele[J]. *Review of Finance*, 2022, 26(2): 217-255.
- [26] Clark T E, West K D. Approximately normal tests for equal predictive accuracy in nested models[J]. *Journal of econometrics*, 2007, 138(1): 291-311.
- [27] Cooper M J, Gulen H, Schill M J. Asset growth and the cross-section of stock returns[J]. *the Journal of Finance*, 2008, 63(4): 1609-1651.
- [28] Daniel K, Grinblatt M, Titman S, et al. Measuring mutual fund performance with characteristic-based benchmarks[J]. *The Journal of finance*, 1997, 52(3): 1035-1058.
- [29] Daniel K, Hirshleifer D, Sun L. Short-and long-horizon behavioral factors[J]. *The review of financial studies*, 2020, 33(4): 1673-1736.
- [30] Daniel K, Titman S. Market reactions to tangible and intangible information[J]. *The Journal of Finance*, 2006, 61(4): 1605-1643.
- [31] De Bondt W F M, Thaler R. Does the stock market overreact?[J]. *The Journal of finance*, 1985, 40(3): 793-805.
- [32] De Long J B, Shleifer A, Summers L H, et al. Noise trader risk in financial markets[J]. *Journal of political Economy*, 1990, 98(4): 703-738.
- [33] Du J, Huang D, Liu Y J, et al. Retail investors and momentum[J]. Available at SSRN 4163257, 2022.
- [34] Eugene F, French K. The cross-section of expected stock returns[J]. *Journal of Finance*, 1992, 47(2): 427-465.
- [35] Fairfield P M, Whisenant J S, Yohn T L. Accrued earnings and growth: Implications for future profitability and market mispricing[J]. *The accounting review*, 2003, 78(1): 353-371.
- [36] Fama E F, French K R, Booth D G, et al. Differences in the risks and returns of NYSE and NASD stocks[J]. *Financial Analysts Journal*, 1993, 49(1): 37-41.
- [37] Fama E F, French K R. A five-factor asset pricing model[J]. *Journal of financial economics*, 2015, 116(1): 1-22.
- [38] Fama E F, French K R. Choosing factors[J]. *Journal of financial economics*, 2018, 128(2): 234-252.
- [39] Fama E F, French K R. Dissecting anomalies with a five-factor model[J]. *The Review of Financial Studies*, 2016, 29(1): 69-103.
- [40] Fama E F, French K R. Profitability, investment and average returns[J]. *Journal of financial economics*, 2006, 82(3): 491-518.
- [41] Fama E F, MacBeth J D. Risk, return, and equilibrium: Empirical tests[J]. *Journal of political economy*, 1973, 81(3): 607-636.
- [42] Fang L, Peress J. Media coverage and the cross-section of stock returns[J]. *The journal of finance*, 2009, 64(5): 2023-2052.
- [43] Feltham G A, Ohlson J A. Valuation and clean surplus accounting for operating and financial activities[J]. *Contemporary accounting research*, 1995, 11(2): 689-731.
- [44] Frankel R, Lee C M C. Accounting valuation, market expectation, and cross-sectional stock

- returns[J]. *Journal of Accounting and economics*, 1998, 25(3): 283-319.
- [45] Gargano A, Rossi A G. Does it pay to pay attention?[J]. *The Review of Financial Studies*, 2018, 31(12): 4595-4649.
- [46] George T J, Hwang C Y. The 52-week high and momentum investing[J]. *The Journal of Finance*, 2004, 59(5): 2145-2176.
- [47] Grossman S J, Stiglitz J E. On the impossibility of informationally efficient markets[J]. *The American economic review*, 1980, 70(3): 393-408.
- [48] Gu M, Kang W, Xu B. Limits of arbitrage and idiosyncratic volatility: Evidence from China stock market[J]. *Journal of Banking & Finance*, 2018, 86: 240-258.
- [49] Hameed A, Ni Z, Tan C A. Momentum and individual investor trades: Evidence from Singapore[J]. *Pacific-Basin Finance Journal*, 2023, 82: 102186.
- [50] Han Y, Huang D, Huang D, et al. Expected return, volume, and mispricing[J]. *Journal of financial economics*, 2022, 143(3): 1295-1315.
- [51] Harvey C R, Siddique A. Conditional skewness in asset pricing tests[J]. *The Journal of finance*, 2000, 55(3): 1263-1295.
- [52] Heston S L, Sinha N R. News vs. sentiment: Predicting stock returns from news stories[J]. *Financial Analysts Journal*, 2017, 73(3): 67-83.
- [53] Hirshleifer D, Hou K, Teoh S H, et al. Do investors overvalue firms with bloated balance sheets?[J]. *Journal of accounting and economics*, 2004, 38: 297-331.
- [54] Hong H, Lim T, Stein J C. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies[J]. *The Journal of finance*, 2000, 55(1): 265-295.
- [55] Hong H, Stein J C. A unified theory of underreaction, momentum trading, and overreaction in asset markets[J]. *The Journal of finance*, 1999, 54(6): 2143-2184.
- [56] Hou K, Xue C, Zhang L. Digesting anomalies: An investment approach[J]. *The Review of Financial Studies*, 2015, 28(3): 650-705.
- [57] Hou K, Xue C, Zhang L. Digesting anomalies: An investment approach[J]. *The Review of Financial Studies*, 2015, 28(3): 650-705.
- [58] Hu G X, Chen C, Shao Y, et al. Fama–French in China: size and value factors in Chinese stock returns[J]. *International Review of Finance*, 2019, 19(1): 3-44.
- [59] Jegadeesh N, Titman S. Returns to buying winners and selling losers: Implications for stock market efficiency[J]. *The Journal of finance*, 1993, 48(1): 65-91.
- [60] Jegadeesh N. Evidence of predictable behavior of security returns[J]. *The Journal of finance*, 1990, 45(3): 881-898.
- [61] Kacperczyk M, Van Nieuwerburgh S, Veldkamp L. A rational theory of mutual funds' attention allocation[J]. *Econometrica*, 2016, 84(2): 571-626.
- [62] Kacperczyk M, Van Nieuwerburgh S, Veldkamp L. Rational attention allocation over the business cycle[R]. *National Bureau of Economic Research*, 2009.
- [63] Lin Q. Noisy prices and the Fama–French five-factor asset pricing model in China[J]. *Emerging Markets Review*, 2017, 31: 141-163.
- [64] Lintner J. Security prices, risk, and maximal gains from diversification[J]. *The journal of finance*, 1965, 20(4): 587-615.
- [65] Liu J, Stambaugh R F, Yuan Y. Size and value in China[J]. *Journal of financial economics*, 2019, 134(1): 48-69.
- [66] Mondria J. Portfolio choice, attention allocation, and price comovement[J]. *Journal of Economic Theory*, 2010, 145(5): 1837-1864.

- [67] Mossin J. Equilibrium in a capital asset market[J]. *Econometrica: Journal of the econometric society*, 1966: 768-783.
- [68] Nagel S. Short sales, institutional investors and the cross-section of stock returns[J]. *Journal of financial economics*, 2005, 78(2): 277-309.
- [69] Newey W K, West K D. Hypothesis testing with efficient method of moments estimation[J]. *International Economic Review*, 1987: 777-787.
- [70] Novy-Marx R. The other side of value: The gross profitability premium[J]. *Journal of financial economics*, 2013, 108(1): 1-28.
- [71] Ohlson J A. Financial ratios and the probabilistic prediction of bankruptcy[J]. *Journal of accounting research*, 1980: 109-131.
- [72] Pástor L, Stambaugh R F. Liquidity risk and expected stock returns[J]. *Journal of Political economy*, 2003, 111(3): 642-685.
- [73] Peress J. The tradeoff between risk sharing and information production in financial markets[J]. *Journal of Economic Theory*, 2010, 145(1): 124-155.
- [74] Peress J. Wealth, information acquisition, and portfolio choice[J]. *The Review of Financial Studies*, 2004, 17(3): 879-914.
- [75] Qiao F. Replicating anomalies in China [J]. Available at SSRN 3263990, 2019.
- [76] Rhodes-Kropf M, Robinson D T, Viswanathan S. Valuation waves and merger activity: The empirical evidence[J]. *Journal of financial Economics*, 2005, 77(3): 561-603.
- [77] Rosenberg B, Reid K, Lanstein R. Persuasive evidence of market inefficiency[J]. *The Journal of Portfolio Management*, 1985, 11(3): 9-16.
- [78] Sadka R, Scherbina A. Analyst disagreement, mispricing, and liquidity[J]. *The Journal of Finance*, 2007, 62(5): 2367-2403.
- [79] Sharpe W F. Capital asset prices: A theory of market equilibrium under conditions of risk[J]. *The journal of finance*, 1964, 19(3): 425-442.
- [80] Sloan R G. Do stock prices fully reflect information in accruals and cash flows about future earnings?[J]. *Accounting review*, 1996: 289-315.
- [81] Stambaugh R F, Yu J, Yuan Y. Arbitrage asymmetry and the idiosyncratic volatility puzzle[J]. *The Journal of Finance*, 2015, 70(5): 1903-1948.
- [82] Stambaugh R F, Yuan Y. Mispricing factors[J]. *The review of financial studies*, 2017, 30(4): 1270-1315.
- [83] Titman S, Wei K C J, Xie F. Capital investments and stock returns[J]. *Journal of financial and Quantitative Analysis*, 2004, 39(4): 677-700.
- [84] Van Hai H, Park J W, Tsai P C, et al. Lottery mindset, mispricing and idiosyncratic volatility puzzle: Evidence from the Chinese stock market[J]. *The North American Journal of Economics and Finance*, 2020, 54: 101266.
- [85] Van Nieuwerburgh S, Veldkamp L. Information acquisition and under-diversification[J]. *The Review of Economic Studies*, 2010, 77(2): 779-805.
- [86] Verrecchia R E. Information acquisition in a noisy rational expectations economy[J]. *Econometrica: Journal of the Econometric Society*, 1982: 1415-1430.
- [87] Xie H. The mispricing of abnormal accruals[J]. *The accounting review*, 2001, 76(3): 357-373.
- [88] Zhang X F. Information uncertainty and stock returns[J]. *The journal of Finance*, 2006, 61(1): 105-137.
- [89] Zhong A, Gray P. The MAX effect: An exploration of risk and mispricing explanations[J].

- Journal of Banking & Finance, 2016, 65: 76-90.
- [90] 曹丰,鲁冰,李争光,等.机构投资者降低了股价崩盘风险吗?[J].会计研究,2015,(11):55-61+97.
- [91] 陈莹.中国股市账面-市值比效应的检验及解释[J].证券市场导报,2008,(02):35-41.
- [92] 高春亭,周孝华.公司盈利、投资与资产定价:基于中国股市的实证[J].管理工程学报,2016,30(04):25-33.DOI:10.13587/j.cnki.jieem.2016.04.004.
- [93] 李志冰,杨光艺,冯永昌,等.Fama-French 五因子模型在中国股票市场的实证检验[J].金融研究,2017,(06):191-206.
- [94] 林虎,孙博,刘力.换手率波动、转售期权与股票横截面收益率[J].金融研究,2013,(12):181-193.
- [95] 刘博,皮天雷.惯性策略和反转策略:来自中国沪深 A 股市场的新证据[J].金融研究,2007,(08):154-166.
- [96] 鲁臻,邹恒甫.中国股市的惯性与反转效应研究[J].经济研究,2007,(09):145-155.
- [97] 潘莉,徐建国.A 股个股回报率的惯性与反转[J].金融研究,2011,(01):149-166.
- [98] 田利辉,王冠英,张伟.三因素模型定价:中国与美国有何不同?[J].国际金融研究,2014,(07):37-45.
- [99] 汪炜,周宇.中国股市“规模效应”和“时间效应”的实证分析——以上海股票市场为例[J].经济研究,2002,(10):16-21+30-94.
- [100] 王永宏,赵学军.中国股市“惯性策略”和“反转策略”的实证分析[J].经济研究,2001,(06):56-61+89.
- [101] 吴世农,许年行.资产的理性定价模型和非理性定价模型的比较研究——基于中国股市的实证分析[J].经济研究,2004,(06):105-116.
- [102] 吴文锋,芮萌,陈工孟.中国股票收益的非流动性补偿[J].世界经济,2003,(07):54-60.
- [103] 吴晓晖,郭晓冬,乔政.机构投资者抱团与股价崩盘风险[J].中国工业经济,2019,(02):117-135.DOI:10.19581/j.cnki.ciejournal.20190131.007.
- [104] 谢谦,唐国豪,罗倩琳.上市公司综合盈利水平与股票收益[J].金融研究,2019,(03):189-206.
- [105] 许年行,于上尧,伊志宏.机构投资者羊群行为与股价崩盘风险[J].管理世界,2013,(07):31-43.DOI:10.19744/j.cnki.11-1235/f.2013.07.004.
- [106] 杨华蔚,韩立岩.中国股票市场特质波动率与横截面收益研究[J].北京航空航天大学学报(社会科学版),2009,22(01):6-10.DOI:10.13766/j.bhsk.1008-2204.2009.01.018.
- [107] 杨旸,陈展辉.中国股市三因子资产定价模型实证研究[J].数量经济技术经济研究,2003,(12):137-141.
- [108] 易志高,茅宁.中国股市投资者情绪测量研究:CICSI 的构建[J].金融研究,2009,(11):174-184.
- [109] 尹玉刚,谭滨,陈威.套利非对称性、误定价与股票特质波动[J].经济学(季刊),2018,17(03):1235-1258.DOI:10.13821/j.cnki.ceq.2018.04.16.
- [110] 游家兴,吴静.沉默的螺旋:媒体情绪与资产误定价[J].经济研究,2012,47(07):141-152.
- [111] 张静,王生年.资产误定价对过度投资的影响路径[J].财经科学,2016,(03):69-78.
- [112] 张然,平帆,汪荣飞.线上销售与未来股票收益[J].金融研究,2022,(06):189-206.

-
- [113] 张然,汪荣飞,王胜华.分析师修正信息、基本面分析与未来股票收益[J].金融研究,2017,(07):156-174.
- [114] 左浩苗,郑鸣,张翼.股票特质波动率与横截面收益:对中国股市“特质波动率之谜”的解释[J].世界经济,2011,34(05):117-135.DOI:10.19985/j.cnki.cassjwe.2011.05.00