

投资者预期协同的 资产定价研究*

——基于多特征机器学习视角

唐国豪 陈海玮 朱琳 姜富伟

摘要: 大数据时代下,随着人工智能技术的深度应用,投资者比以往更能有效关注资本市场的海量信息。投资者对多维信息的关注能否提升预期协同水平,这一问题对于深刻理解资本市场的预期形成具有重要意义。本文基于中国上市公司的多维度特征,引入机器学习框架,构造预期协同指标,探讨多特征学习下的投资者预期如何影响资产定价。实证结果表明,预期协同水平越低,股票未来收益越低,这说明提高投资者的预期协同水平有助于我国资本市场的稳定与可持续发展。本文进一步从卖空约束与有限套利两个角度,证明了错误定价是投资者预期协同与股票未来收益正相关关系的作用机制。此外,本文还发现会计信息质量与预期协同的定价效应显著负相关,而特征数量则对定价效应呈现先增强后减弱的非线性影响。本文针对不同交易风格的投资者进行了异质性分析,发现投资者对金融信号的偏好差异会导致预期协同的定价效应不同。最后,本文围绕我国的投资者信息学习、资本市场信息质量建设、人工智能技术应用与预期管理等多个方面提出了政策建议。

关键词: 特征学习 预期协同 预期管理 神经网络模型 资产定价

一、引言

长期以来,散户投资者能否有效学习上市公司的特征,从而使资产价格趋于理性并减少过度波动一直存在争议(薛健、汝毅,2020;席勒,1981)。中国股票市场多年来由散户交易主导,且散户驱动了80%以上的交易量(琼斯等,2025),不少散户的预期偏差对市场波动产生了显著影响(杨威等,2020)。散户投资者对瞬息万变的市场信息处理能力有限,在面对纷繁复杂的海量数据时,其预期形成的机制也愈发复杂。机构投资者虽具备较强的信息处理能力,但在面对同类信息时往往会进行同质交易,这种行为容易引发散户投资者的羊群效应,从而加剧市场波动。无论是2023年召开的中央金融工作会议,还是党的二十届四中全会都强调了投资者的预期管理以及资本市场“稳预期”的重要性。因此,探究基于公司多特征学习的投资者预期形成机制,以进一步理解投资者的主观预期如何形成及影响资产价格(内格尔、徐,2023),这对于我国金融市场的投资者预期管理至关重要。

近年来,大数据、机器学习与人工智能方法在经管领域得到广泛应用,对以往的研究范式带来了较大冲击(洪永森、汪寿阳,2021;刘景江等,2023)。在资产定价领域的研究中,不少学者认为上市公司的特征信息对于形成准确的价值预期应是“多多益善”(李斌等,2019;顾等,2020;凯利等,2019;莱特等,2017),而传统的研究则认为少量关键因素即可决定和解释资产价格的变动(张玉龙、李怡宗,2013;法马、弗伦奇,1992,2015)。在信息爆炸的时代,到底是应该“面面俱到”地关注金融市场的各类信息,还是应该“精益求精”地对少部分关键信息进行分析?预期的分歧和协同对市场又会产生何种影响?在中国市场以散户居多的环境中,不同类型公

收稿时间:2024-6-17;反馈外审意见时间:2024-10-9、2025-2-28、2025-5-29;拟录用时间:2025-10-21。

*本研究得到国家自然科学基金面上和专项项目(72072193、72342019、72573052)、国家社会科学基金重大项目(22&ZD063)、教育部人文社会科学研究青年基金项目(24YJC79059)、湖南省自然科学基金面上项目(2025JJ50425)、广东省哲学社会科学规划项目(GD25YYJ37)、广东省普通高校特色创新类项目(2025WTSCX029)的资助。朱琳为本文通讯作者。

司的资产价格变动如何受到投资者预期的影响?对上述问题的探究,对于深入理解金融市场如何通过信息与预期影响实体经济发展具有重要意义。

投资者间的预期差异在金融市场中扮演着关键角色,理论文献不仅探讨了预期差异的性质和来源,还分析了它如何影响资产的期望收益、价格波动性以及交易量等(阿特马兹、巴沙克,2018)。然而,目前的实证研究在准确度量投资者预期差异方面面临一系列挑战,尤其是如何量化投资者的主观心理和情感因素。一类较为有效的量化指标是分析师分歧度(安德森等,2005;迪特尔等,2002),然而,分析师的预期本身也被证明存在偏差(伍燕然等,2012;陈等,2007;莫科阿莱利-莫科泰利等,2009)。此外,无论是分析师还是散户投资者,其预期差异主要源于对金融市场信息的学习和接受程度的差异。因此,在数据充分的基础上,引入适当的机器学习方法模拟投资者信息学习的过程,为全面衡量投资者的预期差异提供了可行的新视角。其中,神经网络模型因其对非线性模式的高度敏感性和适应性,在模拟人类预期形成方面展现出独特优势,其可行性已在多个领域得到验证(拉比诺维茨等,2018;武,2019)。基于神经网络构建的投资者模型,不仅可以模拟单个投资者的信息处理和决策机制,还能再现当市场存在大量投资者时异质性信念的形成过程。

因此,本文采用神经网络模型,模拟投资者的信息处理和自适应学习过程,进而生成其对资产收益的不同预期。以此为基础,本文在中国股票市场上构建了一种新的投资者预期差异与协同的统计测度方法。具体而言,本文以105个公司特征作为信息集,利用神经网络模型,采用“先学习,再预测”的方式,模拟投资者对公司特征的处理和对股票收益的预测过程,最后通过计算所有预测值的标准差,构建了投资者预期协同的量化指标。本文从以下几个方面探究预期协同对资本市场的影响。首先,本文分析投资者的预期协同对资产未来收益的影响。其次,从行为金融学的错误定价视角出发,探讨卖空约束与有限套利如何影响预期协同的定价效应。再次,分析上市公司信息质量和数量如何影响投资者的预期协同定价效应,特别是在投资者信息偏好不同的情况下,预期协同的定价效应如何变化。最后,考察不同类型的上市公司如何增强或减弱投资者预期协同在横截面上的定价效应,在时间序列上不同市场周期对其定价效应又有何影响。

本文的研究得到以下几个方面的结论。第一,在中国股票市场中,投资者的预期协同与股票未来收益存在显著的正相关关系,投资者预期协同水平越低,则股票未来收益越低,且这一结论的稳健性不受神经网络模型超参数、滚动时间窗口长度等因素的影响。第二,本文从卖空约束和有限套利两个角度,检验了预期协同定价效应的错误定价机制。研究发现,当投资者缺乏预期协同时,卖空约束和有限套利这类市场摩擦会阻碍价格的有效发现,从而共同导致错误定价,这使得预期协同的定价效应在高摩擦环境中更为显著。第三,本文从特征学习过程中的信息质量和特征数量出发,发现会计信息质量与预期协同的定价效应显著负相关,而特征数量则对定价效应呈现先增强后减弱的非线性影响。第四,相较于国营企业,投资者预期协同的定价效应在民营企业中更加显著;相较于熊市,该效应在牛市中更加突出,这使得预期管理的价值在市场过热时尤为凸显。

本研究的创新之处体现在以下几个方面。首先,本文提供了基于公司特征数据形成投资者预期的新视角,通过神经网络模型模拟投资者行为,创新性地引入了投资者预期协同的测度指标。这一方法不仅可以解决投资者的预期协同程度难以量化的问题,还能深入分析信息如何影响投资者预期,进而引发资本市场的价格变动。其次,本文全面分析了投资者预期协同的定价效应。一方面,本文首次从多特征机器学习的视角,在中国市场上系统分析了投资者预期协同对股票未来收益的预测效果;另一方面,本文从卖空约束、有限套利等错误定价角度研究预期协同定价效应的来源,并深入探讨上市公司的信息质量和数量对预期协同定价效应影响的差异。最后,本文探讨了信息种类、市场环境和投资者风格的异质性对预期协同定价效应的影响,揭示了多特征学习下市场预期形成机制和经济效果,丰富了行为金融学中投资者预期的相关理论与实证证据,为我国市场的投资者预期管理提供了政策建议。

本文其余部分的结构安排如下。第二部分对已有文献进行梳理与评述,并阐明了本文的贡献;第三部分详细阐述了本文的研究设计,包括数据处理、模型构建以及指标计算;第四部分汇报了实证分析及稳健性检验

的结果;第五部分探讨了多特征学习下投资者预期协同定价效应的形成机制;第六部分进行了进一步分析,探究了不同股权性质和市场周期对预期协同定价效应的影响;第七部分则对全文进行了总结。

二、文献综述

(一)投资者预期差异与协同的测度与量化

投资者的预期差异作为金融市场的基本特征之一,是由于市场参与者对于未来资产价格、经济条件或投资表现存在不同看法(巴伯等,2009;洪等,2000)。预期差异使得金融交易的双方愿意买卖资产,它不仅是市场交易的根本推动力,更是市场流动性的重要来源,同时也是金融市场价格发现机制的核心组成部分(洪、斯坦,2007)。

对投资者预期的量化指标,以往研究主要集中在以下4类。第一类是基于调研的宏观预期指标。这类测度指标最初主要通过调研的方式获得,因为宏观经济预期相对易于通过调研获取,著名的指数包括密歇根大学消费者信心指数、MM经济预期指数等。然而,宏观预期往往存在滞后性,无法及时反映市场参与者的即时预期。同时,宏观调研信息也存在一定的局限性,其虽能揭示总体经济环境的趋势,却难以反映细分市场或个别资产的投资者预期。

第二类是基于金融市场交易的预期间接代理指标。这些间接指标包括收益波动率(高等,2006)、未预期交易量(刘燕、朱宏泉,2018;加芬克尔、索科宾,2006)、换手率(孟庆斌、黄清华,2018)、股票市场价格(尹海员、朱旭,2019)等。尽管间接代理指标在一定程度上有效,但其局限性也较为明显。例如,换手率、收益波动率等指标极易与市场其他信息交叉重叠,因而难以精准度量异质信念水平(加芬克尔,2009)。

第三类是分析师预期分歧指标。由于调查投资者对每只股票的真实预期非常困难,实证研究通常转而测度愿意披露预期信息的群体的预期差异,这类真实主体一般是分析师。诸多研究基于分析师在报告中披露的盈余预测构造了预期分歧指标,如分析师盈余预测分散度(迪特尔等,2002)和分析师盈余长期增长预测分散度(安德森等,2005)。然而,分析师预期分歧的测度同样存在诸多局限性。一方面,分析师群体不能代表市场上的所有投资者;另一方面,他们预测的准确性和客观性易受自身行为偏差和利益冲突的影响(伍燕然等,2012;陈等,2007;莫科阿莱利-莫科泰利等,2009)。

第四类是基于文本的预期指标。随着信息技术的发展,文本数据已成为测度投资者心理因素的重要数据来源。例如,不少研究使用公司年报文本、企业调研文本测度公司层面的预期(万海远等,2024;张成思等,2023),亦有研究使用社交平台上的发帖来衡量投资者预期(部慧等,2018;范小云等,2022;库克森、尼斯纳,2020)。然而,此类文本数据常面临噪音干扰,且存在刻意披露或稀疏性的问题,从而对指标准确性构成挑战。

近年来,机器学习已成为金融研究的一个重要工具。机器学习能有效处理数据的高维特性以及提供更加丰富的预测函数形式,因而被广泛运用于基于特征的资产收益预测研究(李斌等,2019;张鹏等,2023;顾等,2020;凯利等,2019)。其中,以神经网络为代表的深度学习,因其能处理和提取复杂的非线性信息,从而展现出卓越的分析能力和适应性(马甜等,2022;陈等,2024;顾等,2021;卡尼尔等,2023)。随着研究的深入,机器学习的应用已从资产收益预测拓展至其他领域,如巴利等(2023)关注机器学习模型的预测分歧,凯利等(2024)探讨模型复杂度对预测的影响,约等(2025)研究机器学习模型的可解释性等。

与此同时,神经网络模型具备模拟投资者异质信念形成过程的潜力。一方面,投资者的信念或预期深受其信息获取与解读方式的影响。投资者信念的异质性主要源于以下两种因素。首先是信息获取的差异。西姆斯(2003)的研究表明,由于注意力有限,投资者可能接触到不同的信息,或对同一信息的关注程度存在差异。其次是解读方式的差异(丹尼尔等,1998,2002;赫舒拉发,2001;洪、斯坦,2007)。投资者在认知能力、教育背景、专业知识和经验等方面的差异会影响他们对信息的解读和利用。例如,更有经验的投资者可能更准确地解读复杂的经济报告或公司财报。另一方面,神经网络中的不同层和神经元可对应不同的信息处理层次和解读复杂性,其训练过程也类似于投资者基于个人经验的自适应学习过程。此外,神经网络作为一种“通用

近似器”,已被理论和实践证明具备逼近复杂函数的能力。即使是仅含单一隐藏层的前馈神经网络,只要包含足够多的神经元,也能以任意精度逼近任何连续函数(霍尼克等,1989)。

因此,神经网络模型凭借其灵活性及强大的数据处理能力,为模拟投资者预期形成过程提供了有效工具,尤其在理解和量化预期差异与协同方面展现出独特优势。这为借助前沿人工智能手段攻克投资者预期协同的度量难题提供了可能。然而,目前针对中国市场的相关研究中,还鲜有利用机器学习方法模拟投资者预期形成机制的探索。

(二)投资者预期差异与协同的资本市场定价效应

学术界关于投资者预期差异与协同在资本市场的定价效应主要存在两种观点:一种认为预期差异与股票期望收益负相关,另一种则认为两者呈正相关关系。

关于负向关系的研究,米勒(1977)提出在存在卖空约束的市场中,由于悲观投资者无法通过卖空表达观点,市场价格主要反映乐观投资者的预期,导致股票价格高估;坎德尔和皮尔逊(1995)则发现,过度自信的投资者对公共信息存在解读差异,特别是过度反应利好信号,这会进一步推动资产价格偏离基本面,加剧价格的高估现象。实证中,迪特尔等(2002)从分析师分歧角度、戈茨曼和马萨(2005)基于投资者持仓数据、约翰逊(2004)从期权定价视角、余(2011)从投资组合视角,均验证了预期差异较高的股票未来收益较低。

关于正向关系的研究,其核心观点是将预期差异视为一种未被定价的风险来源,从而要求更高的风险溢价。布卢姆菲尔德和费希尔(2011)指出,如果投资者认为其他人会过度重视噪音,则预期差异使投资者高估未来价格波动,从而要求更高的风险溢价。祖卡斯等(2006)提出使用多样性度量(diversity measure)作为更准确的指标,并发现股票收益与预期差异正相关。费希尔等(2022)强调需区分信息不对称与纯粹意见分歧,发现当不存在卖空约束时,纯粹的预期差异通过风险分担机制提升了股票的期望收益。

除了正向和负向关系的对立观点外,部分研究发现预期差异与股票期望收益的关系可能因时间跨度或其他条件而异。安德森等(2005)构建了一个将预期差异作为风险因子的一般均衡模型,发现短期预期差异与期望收益负相关,而长期预期差异与期望收益正相关。这一双向关系表明,预期差异的影响具有复杂性,取决于其持续时间和市场环境。

相较于机构投资者主导的美国股票市场,中国股票市场主要由散户交易驱动,如何衡量散户投资者预期差异与协同对资产收益的影响程度,对于提高市场的定价效率有着积极意义,但现有文献对此涉及较少。同时,衡量投资者预期协同的定价效果,有利于在中国股票市场进行更加有针对性的投资者预期管理,这也是本文的研究创新与意义所在。

(三)预期差异与收益之间关系的影响因素

对预期差异与股票收益之间关系的机制分析,主要基于错误定价和风险补偿两种理论。从错误定价的角度来看,卖空约束和有限套利引起的流动性约束是影响预期差异定价效应的重要因素。帕克(2005)提出,在卖空受限的市场中,预期差异较大的股票因投资者预期未来能以更高价格转售而被高估,这种投机性定价是负向关系的重要机制。类似地,贝姆等(2006)研究发现,只有当股票同时面临卖空约束和意见分歧时,才会出现显著的高估现象。若仅存在其中一种情况,股票价格不会系统性地偏离基本面。同时,由有限套利引致的流动性约束,是调节预期差异与股票收益关系的另一因素。萨德卡和谢尔宾娜(2007)发现,流动性较低会加剧预期差异较大的股票的高估现象,而流动性的提升能够降低高预期差异股票的错误定价程度。阿夫拉莫夫等(2009)则认为,分析师盈余预测分散度与未来股票收益的负相关性主要由信用风险驱动,并挑战了传统的错误定价解释。从风险补偿的角度来看,卡林等(2014)的实证研究表明,预期差异的增加与更高的期望收益、更高的收益波动率和更大的交易量相关,预期差异存在正的风险溢价。

总之,目前已有文献主要围绕美国股票市场中预期差异与收益的关系及其影响因素展开研究,而系统性地探索中国市场投资者预期差异或协同如何影响资产收益的研究则较为匮乏。本文尝试从信息种类、市场环境和投资者风格的异质性等角度出发,探索它们对预期协同定价效应的影响并进一步厘清机制,从而为我国

股票市场科学有效的投资者预期管理提供参考。

三、研究设计

(一)数据

首先,进行样本选取。本文以1997年1月至2021年12月为样本期间,以中国股票市场(含上证主板、深证主板和创业板)的全部股票为研究对象,所用数据均为月频数据。

其次,构建公司特征数据集。在机器学习过程中,本文构造了一个庞大的公司特征数据集,作为投资者所能获取的相同的公开信息。本文全面梳理了实证资产定价领域的近百篇文献,并考虑数据的可获得性,基于国泰安数据库中的财务报表数据和股票交易数据,整理出105个公司特征因子指标^①,其中包含25个估值与成长类指标、12个投资类指标、24个盈利类指标、7个惯性类指标、16个交易摩擦类指标和21个无形资产类指标。此外,在定价效应分析中,还使用了三因子和五因子模型指标、盈余披露质量指标、上市公司基本信息指标等,这些数据同样来源于国泰安数据库。

最后,进行数据处理。为保证样本的代表性与数据的可靠性,本文在数据处理方面借鉴了以往研究的经验和方法。具体处理过程如下。第一,剔除ST、PT股数据,这些股票对应的企业往往处于财务困境,且市场流动性不足,面临退市风险。第二,由于金融行业公司的部分财务指标计算方式与其他行业有所不同,为保持指标计算的一致性和可比性,本文剔除金融行业的股票数据。第三,鉴于新上市股票在上市初期往往会出现较大的价格波动,为缓解IPO抑价效应的影响,本文剔除上市首月的股票。第四,对于缺失值,参考李斌等(2019)的做法进行处理,具体过程分为两步:(1)若某只股票在第*t*月收益率数据缺失(通常由股票连续停牌造成),剔除该观测值;(2)若某只股票的公司特征数据缺失,则以0进行填充。

经过上述数据处理后,本文最终使用的样本涵盖4195只股票,时间跨度为300个月,共包含536514个观测值。如图1所示,随着时间的推移,每个月份的股票数量呈上升趋势,由1997年1月的470只上升至2021年12月的3830只。

由于不同公司特征的数据在数量级和分布上存在明显差异,可能引起预测偏差,并导致神经网络模型在训练过程中的迭代速度较慢。因此,在训练模型之前,本文先对公司特征数据进行标准化。标准化公式如下:

$$Z_{\text{normed}} = \frac{Z - \bar{Z}}{\sigma_Z} \quad (1)$$

其中, \bar{Z} 为均值, σ_Z 为标准差。

(二)模型构建

为了模拟投资者进行预测的过程,本文选择了机器学习模型中最接近人脑运作方式的神经网络模型,它模仿人脑神经元之间的连接和信息传递过程,能够处理和学习复杂的非线性关系。神经网络模型通常包括输入层、隐藏层和输出层。输入层接收数据,隐藏层执行一系列变换和计算,输出层则产生模型的输出值。

为确保研究更具一般性与代表性,本研究构建了一个基础的前馈神经网络模型,该模型包含3个主要部分:一个输入层、一个隐藏层和一个输出层,体现了神经

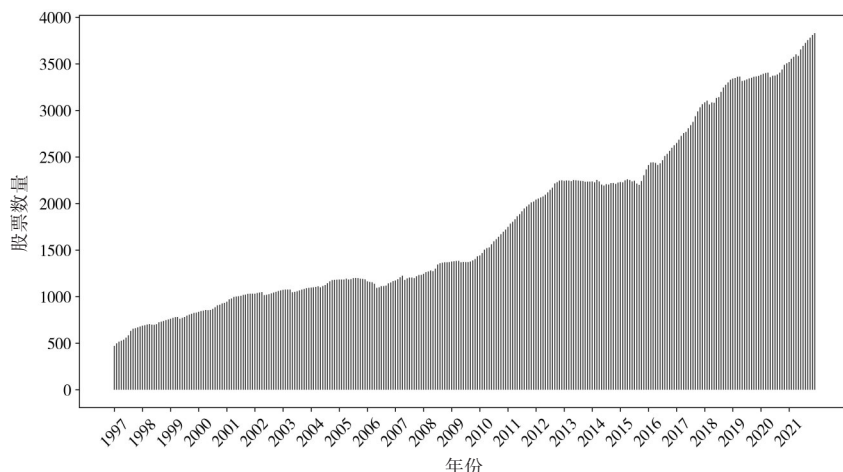


图1 样本期内月度股票有效数量

网络模型中最基本的架构。具体而言,输入层作为神经网络的入口,负责接收原始数据,其中每个输入特征对应一个神经元节点。隐藏层位于输入层和输出层之间,每个节点都与输入层的所有节点连接,承担着处理输入信息、抽象提取特征的核心任务。为实现非线性映射,隐藏层采用了ReLU激活函数,这使得网络能够学习数据中的复杂非线性关系。输出层是神经网络的最后一层,负责汇总处理隐藏层的信息并生成模型的最终输出结果。

为定义投资者的预期差异,本文考虑一组投资者,每个个体用 n 表示, $n=1,2,\dots,N$ 。假设这些投资者面临一个公开、完整且相同的信息集 $\mathbf{Z}=(\mathbf{z}_{i,t})$,其中, $\mathbf{z}_{i,t}=(z_{i,t}^1, z_{i,t}^2, \dots, z_{i,t}^{K-1}, z_{i,t}^K)$ 是一个 K 维向量,该向量表示个股的 K 个公司特征。由于信念存在差异,不同投资者会对同一信息集产生不同的理解,具体表现为对其中的 K 个公司特征赋予不同的权重,因此,经过信息处理后,投资者 n 根据自己信念所形成的实际信息集 $\mathbf{x}_{n,i,t}$ 可表示为:

$$\mathbf{x}_{n,i,t} = g_n(\mathbf{z}_{i,t}) \quad (2)$$

$g_n(\cdot)$ 是投资者 n 对信息集 $\mathbf{z}_{i,t}$ 的处理函数,具体可表示为:

$$g_n(\mathbf{z}_{i,t}) = \mathbf{D}_n \mathbf{z}_{i,t} \quad (3)$$

其中, \mathbf{D}_n 为 K 阶对角矩阵且对角线上的元素 $d_{jj}(j=1,2,\dots,K)$ 满足 $d_{jj} \sim iidN(0,1)$, $\mathbf{x}_{n,i,t}=(x_{i,t}^1, x_{i,t}^2, \dots, x_{i,t}^{K-1}, x_{i,t}^K)$ 是一个 K 维向量,即投资者 n 面临的最终信息集。之后,投资者 n 在月份 t 对股票 i 的收益预测可表示为:

$$E_{n,t}[r_{i,t+1}] = f_{n,t}(\mathbf{x}_{n,i,t}) \quad (4)$$

函数 $f_{n,t}(\cdot)$ 是投资者 n 的预测函数,具体可表示为:

$$f_{n,t}(\mathbf{x}_{n,i,t}) = \mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{x}_{n,i,t} + \mathbf{b}_1) + b_2 \quad (5)$$

其中, \mathbf{W}_1 是 $m \times K$ 维的权重矩阵(m 为隐藏层神经元个数),用于连接输入层与隐藏层。 \mathbf{b}_1 是隐藏层的偏置向量,表示给隐藏层每个神经元添加的额外偏移值,维度为 $m \times 1$ 。ReLU是激活函数,用于将负值变为零而正值保持不变,以引入非线性变换。 \mathbf{W}_2 是 $1 \times m$ 维的权重矩阵(1为输出层神经元个数),用于连接隐藏层与输出层。 b_2 是输出层的偏置标量,表示给输出层神经元添加的额外偏移值。对于每一个投资者而言,相应的神经网络模型具有相同的结构,但是具体的参数即权重($\mathbf{W}_1, \mathbf{W}_2$)与偏置(\mathbf{b}_1, b_2)会被随机初始化为不同的初值,并在模型训练过程中不断更新,以提高模型的预测性能。

在确定用于训练的神经网络模型时,主要涉及2个需要预先设定的超参数,包括投资者数量 N 和隐藏层神经元数量 M ,本文将两者均设定为100。此外,还需确定滚动时间窗口的长度,如图2所示,本文采用了12个月的训练窗口。对于每个投资者 n ,其神经网络以训练窗口($t-12$ 至 $t-1$ 月)的公司特征数据作为输入值,以对应窗口内股票下一期的真实超额收益率(excess return)作为标签,以避免前视偏差。本文以最小化预测误差(均方误差, MSE)为目标进行模型训练,优化神经网络的参数即权重($\mathbf{W}_1, \mathbf{W}_2$)与偏置(\mathbf{b}_1, b_2)。该方法模拟了投资者基于历史特征—收益映射关系形成预期的过程,在 t 月投资者使用基于 $t-12$ 月至 $t-1$ 月数据训练好的神经网络模型,并输入 t 月的最新特征数据,以预测 $t+1$ 月的收益。这反映了现实中投资者对历史规律的有限理性学习^②。

为了确保结论的稳健性,在第六部分中,本文调整了上述参数的设定并进行了稳健性分析。同时,考虑到过拟合的可能性,在第六部分中,本文还引入了正则化技术,通过在模型训练过程中增加额外的约束条件,即引入正则化技术所需的 λ ^③,限制模型的参数空间,降低模型的复杂性,从而检验在进行正则化的情况下实证结果的稳健性。

(三)指标计算

在确定神经网络模型的具体结构后,本研究使用滚动窗口法来模拟各个投资者 n 在月份 t 对股票 i 的未来收益预测,



图2 滚动窗口法示例

预测值记为 $IF_{n,i,t}$, 进一步可得所有投资者预测值的均值 $MF_{i,t}$ 及标准差 $IED_{i,t}$, 见公式(6)~(8)。借鉴文献中以标准差衡量预期差异的做法(巴利等, 2023; 班纳吉, 2011; 洪、斯雷尔, 2016), 本文构建的 $IED_{i,t}$ 即为投资者预期分散度(Investor Expectation Dispersion, IED)指标。

$$IF_{n,i,t} = f_{n,t}(\mathbf{x}_{n,i,t}) \quad (6)$$

$$MF_{i,t} = \frac{1}{N} \sum_{n=1}^N IF_{n,i,t} \quad (7)$$

$$IED_{i,t} = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (IF_{n,i,t} - MF_{i,t})^2} \quad (8)$$

其中, IED 指标用来衡量投资者对未来收益预测的分散程度, IED 越小, 说明投资者的预测越一致, 分歧程度越低, 协同程度越高; 反之, IED 越大, 说明预测差异越大, 分歧程度越高, 协同程度越低。因此, IED 指标也反映了投资者预期协同的程度。本文构造的指标与其他相关研究存在明显差异。例如, 巴利等(2023)基于美国上市公司特征数据构建分歧指标, 与其相比, 本文的核心区别体现在以下几个方面。首先, 巴利等(2023)使用了随机森林模型, 该模型的可解释性更侧重于特征重要性排序和决策路径。而本文选择神经网络, 则是看重其模拟投资者非线性思维和自适应学习过程的潜力。此外, 顾等(2020)和李斌等(2019)分别在美国市场和中國市场上检验了不同机器学习模型对股票未来收益的预测能力, 均发现神经网络模型在众多机器学习模型中表现最佳。其次, 巴利等(2023)假设投资者注意力是有限的, 使用“随机选择”的方法定义不同的投资者, 而本文则假设投资者具有不同的先验信念, 通过“随机赋权”定义不同的投资者。再次, 相较于巴利等(2023)关注的机器预测分歧问题, 本文立足于我国金融市场投资者的预期管理问题, 构造了预期协同指标, 关注投资者基于公司特征的预期协同形成机制及其影响。最后, 本文进一步探讨了公司特征的异质性(信息质量与数量)及投资者的异质性(价值型投资者与技术型投资者)对预期协同定价效应的影响, 更深入地揭示公司特征、投资者信息偏好与资产定价之间的关系, 并为中国市场的预期管理和政策制定提供了参考依据。

四、实证分析

(一)描述性统计

表1展示了本文使用的主要指标的描述性统计结果, 样本区间为1998年1月至2021年12月。除了投资者预期分散度指标 IED , 表1还包含了在实证研究中用到的其他主要公司特征, 如市场Beta(β)、流通市值对数($Size$)、账面市值比(BM)、毛利率(GP)和投资资产比(IA)。所列出的描述性统计量包括均值、标准差、偏度、峰度、最小值、中位数、最大值以及月度观测值数量。为深入分析, 本文将样本分为民营企业 and 国营企业两个子样本, 并分别对这两类样本的投资者预期分散度进行了描述性统计。

表1 Panel A显示, 投资者预期分散度指标 IED 平均每月有1839个有效观测值, IED 的平均值为1.41, 这表明投资者对股票未来收益的预期存在一定程度的分散, 反映了市场参与者在预期形成过程中意见的异质性。 IED 的偏度为8.50, 峰度为192.41, 这表明 IED 存在一定的正偏与尖峰分布的特征。Panel B展示了民营企业与国营企业 IED 指标的描述性统计结果, 如其所示, 民营企业 IED 的均值为1.29, 与国营企业的1.30较为接近, 但其分布的正偏与尖峰特征更为明显。平均而言, 民营企业的数量略多于国营企业, 民营企业平均每月的观测值数量为1046个, 国营企业为792个。

(二)投资者预期协同的定价效应分析

1. 投资组合分析

组合分析是实证资产定价领域中广泛应用于检验一个或多个变量对股票未来收益预测能力的统计方法。组合分析的主要优点在于它是一种非参数分析方法, 无需对被检验变量在横截面上的关系提出线性假设, 且对个股的极端值也不敏感。

表1 描述性统计

Panel A: 全样本各指标的描述性统计								
	均值	标准差	偏度	峰度	最小值	中位数	最大值	数量
IED	1.41	0.76	8.50	192.41	0.52	1.28	16.68	1839
β	1.10	0.50	-0.14	1.66	-1.21	1.10	3.22	1817
$Size$	14.49	0.91	0.67	0.99	12.26	14.39	19.03	1839
BM	0.97	0.90	6.90	97.57	-0.31	0.78	15.40	1743
GP	0.14	0.09	1.81	6.32	-0.13	0.12	0.82	1743
IA	0.42	7.27	20.37	721.81	-0.69	0.11	270.09	1642
Panel B: 民营企业与国营企业 IED 的描述性统计								
	均值	标准差	偏度	峰度	最小值	中位数	最大值	数量
IED (民营企业)	1.29	0.64	5.81	92.93	0.51	1.17	10.26	1046
IED (国营企业)	1.30	0.61	5.36	71.73	0.52	1.19	9.07	792

本文使用投资组合分析法,在月度层面上,将个股按照 *IED* 指标从低到高排序并划分为 10 个投资组合。其中,第 1 组表示 *IED* 指标最低的 10% 的股票(即投资者预期协同度最高),第 10 组表示 *IED* 最高的 10% 的股票(即投资者预期协同度最低),其余组别按 *IED* 指标从低到高分别记为第 2~9 组,“对冲组合”为通过做多预期协同度最高的组合、做空预期协同度最低的组合所构建的多空对冲投资组合。所有这些投资组合的持有周期均为一个月,到了下一个月,重复上述过程构建和持有新的投资组合,各组合收益率均为流通市值加权计算。

如表 2 所示,投资者预期分散度最低(协同度最高)的组合收益率为 1.17%($t=1.86$),预期分散度最高(协同度最低)的组合收益率为 0.28%($t=0.44$),预期协同效应即多空对冲组合的收益率为月均 0.89%($t=3.58$)。随着 *IED* 的升高,投资组合的收益率整体呈下降趋势,从收益风险比的角度看,投资组合的夏普比率亦呈下降趋势,表明高 *IED* 组合承担单位风险所获收益更低。本文进一步采用资本资产定价模型(CAPM)、三因子模型(FF3)和五因子模型(FF5)计算各组合的风险调整后收益。由表 2 可知,对冲组合的 CAPM- α 、FF3- α 与 FF5- α 分别为 0.94%($t=3.77$)、0.74%($t=3.12$)和 0.61%($t=2.27$),均在 5% 水平上显著,表明其收益率无法被这些常见因子模型所解释。总体而言,投资者预期协同对股票横截面收益具有显著的预测能力,预期协同度越低,其未来收益也越低^④。

2. Fama-MacBeth 回归分析

与组合分析法不同, Fama-MacBeth 回归分析(以下简称 FM 回归)允许在检验变量间关系时控制大量的其他变量。为了探究在控制一些常见的公司特征后 *IED* 对股票未来收益的预测能力,本文进行了 FM 回归,如公式(9)所示。因变量是未来一期的股票超额收益率,关键自变量为 *IED*,控制变量分别为 β 、Size、BM、GP、IA。

$$y_{i,t+1} = a_0 + b_1 IED_{i,t} + b_2 \beta_{i,t} + b_3 Size_{i,t} + b_4 BM_{i,t} + b_5 GP_{i,t} + b_6 IA_{i,t} + \varepsilon_{i,t+1} \quad (9)$$

表 3 展示了投资者预期分散度与股票未来收益之间关系的 FM 回归结果,列(1)~(3)分别为引入不同控制变量后进行 FM 回归的结果。由表 3 可知,在列(1)、(2)与(3)中, *IED* 的平均系数均显著为负^⑤,表明 *IED* 与股票未来收益之间的关系是负向的(即投资者预期协同度与股票未来收益之间的关系是正向的)。投资组合分析中,第 10 组与第 1 组之间的标准化 *IED* 差异约为 2.78(正文表未汇报),将这个差异乘以表 3 列(3)中 *IED* 平均系数的绝对值 0.19,得到的月度收益率差异为 0.53%。这表明在控制其他因素后,基于 *IED* 的对冲组合收益率为 0.53%,这在月度水平上已足够显著。综上,FM 回归的结果进一步表明,在中国股票市场中,投资者的高预期分散度(低预期协同度)对股票未来收益具有显著的负向影响,即使在控制多个变量后,该负向关系依然存在,表明较低的预期协同度往往预示着较低的期望收益率。

(三)稳健性检验

为验证结论的稳健性,本文进行了一系列检验,包括改变训练窗口的长度、改变投资者的数量、改变隐藏层神经元

表 2 投资者预期协同的单变量组合分析

投资组合	Ret	SR	CAPM- α	FF3- α	FF5- α
1	1.17* [1.86]	0.48	0.48** [2.42]	0.22 [1.24]	0.12 [0.61]
2	1.00* [1.66]	0.42	0.30* [1.80]	0.05 [0.31]	-0.01 [-0.09]
3	0.81 [1.41]	0.33	0.10 [0.67]	-0.17 [-1.27]	-0.18 [-1.25]
4	0.92 [1.58]	0.37	0.20 [1.41]	-0.07 [-0.50]	-0.04 [-0.34]
5	1.02* [1.69]	0.41	0.30* [1.93]	0.03 [0.23]	0.07 [0.51]
6	0.88 [1.49]	0.36	0.17 [1.22]	-0.08 [-0.66]	-0.05 [-0.36]
7	0.86 [1.44]	0.34	0.13 [0.91]	-0.15 [-1.11]	-0.14 [-0.96]
8	0.74 [1.29]	0.30	0.02 [0.13]	-0.19 [-1.58]	-0.10 [-0.80]
9	0.68 [1.18]	0.28	-0.04 [-0.25]	-0.19 [-1.20]	-0.08 [-0.51]
10	0.28 [0.44]	0.11	-0.46*** [-2.96]	-0.52*** [-3.33]	-0.49*** [-2.78]
对冲组合	0.89*** [3.58]	0.62	0.94*** [3.77]	0.74*** [3.12]	0.61** [2.27]

注: Ret 为持有某个投资组合的月度平均收益率(已减去无风险利率), SR 为投资组合的夏普比率。***、**和*分别代表在 1%、5% 和 10% 的水平上显著,括号中是使用经 Newey-West 调整后标准误所计算的 t 统计量,下同。

表 3 投资者预期协同的 FM 回归分析

	(1)	(2)	(3)	(4)	(5)	(6)
	全样本	全样本	全样本	剔除壳价值	民营企业	国有企业
<i>IED</i>	-0.15** [-2.51]	-0.20*** [-4.30]	-0.19*** [-4.17]	-0.21*** [-4.73]	-0.21*** [-3.90]	-0.18*** [-4.40]
β		0.31*** [4.48]	0.35*** [5.06]	0.33*** [4.41]	0.31*** [4.32]	0.38*** [5.55]
Size		-0.32*** [-2.66]	-0.36*** [-3.08]	-0.14 [-1.50]	-0.38*** [-3.38]	-0.29*** [-2.63]
BM		0.06 [1.08]	0.09* [1.79]	0.13** [2.17]	0.10* [1.88]	0.07 [1.38]
GP			0.15** [2.50]	0.18*** [2.73]	0.17** [2.52]	0.14** [2.35]
IA			-0.01 [-0.27]	0.03 [0.72]	0.00 [0.11]	-0.03 [-0.79]
Cons	1.20* [1.94]	1.09* [1.68]	1.10* [1.69]	0.88 [1.36]	1.15* [1.73]	1.00 [1.55]
N	1838	1721	1620	1232	881	739
Adj. R ²	0.95%	6.09%	7.05%	6.83%	6.61%	7.28%

注:各回归的自变量均在横截面上进行了标准化和 1% 水平的缩尾处理,下同。

的数量以及进行不同类型和不同强度的正则化。具体而言,本文分别将训练窗口长度设定为3、6、24和36个月;将投资者数量设定为50、200、500和1000;将隐藏层神经元数量设定为25、50和200。此外,在模型训练过程中,本文还引入了正则化技术,通过增加额外的约束条件,限制模型的参数空间,从而降低模型的复杂性和过拟合风险。为此,本文分别进行了L1和L2正则化,并在每种检验中,均考察了正则化系数 λ 设定为1、0.1、0.01和0.001时的结果稳健性。

表4展示了不同参数设定下的FM回归结果,限于篇幅,表4仅展示了 IED 的平均系数及 t 统计量、样本量(N)和调整后 R^2 ($Adj.R^2$),对控制变量(β 、 $Size$ 、 BM 、 GP 、 IA)的平均系数及 t 统计量予以省略。Panel A展示了不同训练窗口长度设定的结果,在不同的设定下, IED 的回归系数始终显著为负。Panel B展示了不同投资者数量设定的结果,在计算投资者预期分散度时,无论投资者数量设定为50、200、500还是1000, IED 与收益的负向关系均显著,验证了该指标对市场参与者规模的稳健性。Panel C展示了不同隐藏层神经元数量设定下的结果,当神经元的数量从25增至200时, IED 的回归系数始终在1%的水平上显著,显示其对神经元数量的设定不敏感。最后,Panel D和E的结果进一步表明,正则化技术的设定(L1或L2,以及不同的 λ 值)对预期协同的定价效应影响有限, IED 的回归系数始终保持显著为负。

总体而言,上述检验一致表明,在不同模型参数设定下,投资者预期分散度(IED)均对股票未来收益具有显著的负向预测能力,这进一步证明了低投资者预期协同度与低股票期望收益的显著关系,表明本文的研究结论不会因这些参数的设定而改变,具有很强的稳健性。

五、多特征学习下投资者预期协同定价效应的机制研究

(一)基于错误定价渠道的投资者预期协同定价效应研究

本节旨在探讨投资者预期分散度(IED)如何通过错误定价渠道对资产定价产生影响。具体而言,错误定价不仅是市场信息不完全或投资者认知偏差的结果,还可能由于投资者对未来现金流的不同预期导致的价格扭曲(黄等,2021)。本文认为,投资者的预期协同水平越高,市场的定价效率就越高。相反,当预期协同水平较低时,投资者分歧加剧,在存在市场摩擦的情况下,错误定价的程度更深。为了全面分析错误定价在投资者预期协同定价效应中的作用机制,在排除壳价值的影响后,本文从卖空约束和有限套利两个方面进行检验。

首先,探讨卖空约束对预期协同定价效应的影响。卖空约束作为影响市场错误定价的重要机制,在我国股票市场尤为突出。我国虽已实施融资融券制度,但对卖空交易仍持审慎态度。根据米勒(1977)的理论,当投资者对股票价格存在分歧且悲观投资者的卖空行为受限时,股票价格主要由乐观投资者的信念所决定,从而使得股票价格偏高,未来收益偏低。因此,本文假设,当卖空约束越严格时,投资者预期协同的定价效应越显著。

在卖空约束代理指标的选取上,本文采用融券可行性和机构投资者持股比例(内格尔,2005)两个变量来衡量股票是否受到卖空约束以及受到卖空约束的程度。股票可融券标识和机构投资者持股比例数据均来源于国泰安数据库。具体而言,本文根据股票是否可以融券卖出分为两组,在每组中按照 IED 从低到高分5个

表4 不同参数设定下的FM回归分析

Panel A: 不同训练窗口长度下的FM回归结果				
	(1)	(2)	(3)	(4)
	3个月	6个月	24个月	36个月
IED	-0.16*** [-3.90]	-0.18*** [-4.20]	-0.15*** [-4.31]	-0.20*** [-5.12]
Controls	Yes	Yes	Yes	Yes
N	1620	1620	1620	1643
$Adj.R^2$	6.98%	7.11%	6.98%	7.01%
Panel B: 不同投资者数量下的FM回归结果				
	(1)	(2)	(3)	(4)
	$N=50$	$N=200$	$N=500$	$N=1000$
IED	-0.19*** [-4.15]	-0.20*** [-4.39]	-0.21*** [-4.55]	-0.20*** [-4.57]
Controls	Yes	Yes	Yes	Yes
N	1620	1620	1620	1620
$Adj.R^2$	7.04%	7.05%	7.06%	7.06%
Panel C: 不同隐藏层神经元数量下的FM回归结果				
	(1)	(2)	(3)	
	$M=25$	$M=50$	$M=200$	
IED	-0.20*** [-4.90]	-0.20*** [-4.52]	-0.18*** [-3.99]	
Controls	Yes	Yes	Yes	
N	1620	1620	1620	
$Adj.R^2$	7.07%	7.07%	7.03%	
Panel D: L1正则化下不同参数的FM回归结果				
	(1)	(2)	(3)	(4)
	$L1; \lambda=1$	$L1; \lambda=0.1$	$L1; \lambda=0.01$	$L1; \lambda=0.001$
IED	-0.19*** [-4.45]	-0.19*** [-4.21]	-0.19*** [-4.18]	-0.19*** [-4.17]
Controls	Yes	Yes	Yes	Yes
N	1620	1620	1620	1620
$Adj.R^2$	7.06%	7.06%	7.06%	7.05%
Panel E: L2正则化下不同参数的FM回归结果				
	(1)	(2)	(3)	(4)
	$L2; \lambda=1$	$L2; \lambda=0.1$	$L2; \lambda=0.01$	$L2; \lambda=0.001$
IED	-0.20*** [-4.32]	-0.19*** [-4.18]	-0.19*** [-4.17]	-0.19*** [-4.17]
Controls	Yes	Yes	Yes	Yes
N	1620	1620	1620	1620
$Adj.R^2$	7.07%	7.06%	7.05%	7.05%

投资组合,共计10(2×5)个投资组合,结果如表5 Panel A所示^⑥。本文进一步根据机构投资者持股比例,按照30%和70%分位数从低到高分3个投资组合,同样,在每组中将股票根据IED从低到高分5个投资组合,共计15(3×5)个投资组合,结果如表5 Panel B所示。

表5显示,在不可融券的股票组合中,IED的对冲组合收益率为1.04%($t=4.39$),投资者预期协同的定价效应显著,而在可融券的股票组合中,定价效应不显著。这表明只有在存在卖空约束的情况下,投资者的异质信念才会导致股价高估。同样地,在机构投资者持股比例低(卖空约束强)的股票组合中,IED的对冲组合收益率为0.82%($t=2.94$),投资者预期协同的定价效应显著,而在机构投资者持股比例高(卖空约束弱)的股票组合中,定价效应不显著,这进一步验证了卖空约束是驱动投资者预期协同定价效应的重要作用机制。

其次,探讨有限套利对预期协同定价效应的影响。如果投资者预期协同的定价效应确实源于错误定价,那么这种效应在更难套利的股票组合中应更加显著。本文选择非流动性作为有限套利的代理指标(阿米胡德,2002)。具体而言,根据非流动性指标的30%和70%分位数,将股票从低到高分3个投资组合,然后在每组中将股票根据IED从低到高分5个投资组合,共计15(3×5)个投资组合,结果如表5 Panel C所示。结果表明,在非流动性水平高(更难套利)的股票组合中,IED的对冲组合收益率为0.69%($t=3.01$),投资者预期协同的定价效应显著,而在非流动性水平低(更易套利)的股票组合中,定价效应不显著。

以上分析表明,IED的定价效应主要是通过错误定价渠道实现的^⑦。当投资者缺乏预期协同时,卖空约束、有限套利等因素共同作用,阻碍价格的有效发现,从而导致错误定价,这使得预期协同的定价效应在高摩擦环境中更为显著。

(二)不同信息质量和数量下的投资者预期协同定价效应研究

投资者预期差异源于信息获取和解读方式的差异。为了深入探索和理解投资者预期形成的经济机制,本节将从上市公司的信息质量与数量两个角度进行分析,探究其对投资者预期协同定价效应的影响。

1. 信息质量对定价效应的影响

由于投资者在市场上观察到相同的信息,因而信息质量会对投资者的预期形成产生至关重要的影响。因此,本文通过考察不同会计信息质量条件下投资者预期协同与股票未来收益关系的变化,来探究预期形成的机制。

参考徐经长等(2022)和刘贯春等(2023)的研究,本文选取上交所和深交所发布的信息披露考评结果作为评估上市公司会计信息质量的代理变量,数据来源于国泰安数据库,评估结果包括“优秀”、“良好”、“合格”与“不合格”4个等级。对于每个信息披露等级,将投资者预期分散度(IED)按照30%和70%分位数从低到高分3个投资组合,共计12(4×3)个投资组合,结果如表6所示。

表6表明,会计信息质量越差的公司,即信息质量考评结果越低的公司,对冲组合的收益越显著。具体而言,信息质量考评结果为“不合格”公司的对冲组合收益率为1.74%($t=3.19$),收益的绝对值和显著性明显高于考评结果为“优秀”公司的0.48%($t=1.79$)。由此可见,在会计信息质量相对较差的公司,预期协同的定价效应更强。

2. 信息数量对定价效应的影响

除了信息质量,信息数量也对投资者预期的形成有着重要影响。一方面,投资者因其金融专业知识的差异,在信息获取能力上存在显著不同。例如,散户投资者通常被认为在注意力总量和分配效率上不如机构投资者。此外,由于缺乏专业知识,散户投资者往往无法获取所有相关信息。这可能是因为他们未能意识到

表5 基于错误定价的机制检验

投资组合	Panel A: 融券可行性		Panel B: 机构持股比例		Panel C: 非流动性	
	不可融券	可融券	低	高	低	高
低 IED	1.25* [1.90]	0.76 [1.30]	1.01 [1.39]	0.60 [1.05]	0.68 [1.09]	1.49** [2.53]
高 IED	0.21 [0.28]	0.53 [0.85]	0.19 [0.23]	0.42 [0.66]	0.27 [0.45]	0.80 [1.22]
对冲组合	1.04*** [4.39]	0.24 [0.80]	0.82*** [2.94]	0.18 [0.68]	0.40 [1.63]	0.69*** [3.01]

表6 会计信息质量与投资者预期协同的双变量组合分析

投资组合	信息质量考评结果			
	优秀	良好	合格	不合格
低 IED	1.40** [2.09]	1.23* [1.68]	0.95 [1.32]	1.03 [1.19]
高 IED	0.92 [1.39]	0.72 [0.93]	0.35 [0.47]	-0.71 [-0.81]
对冲组合	0.48* [1.79]	0.51** [2.12]	0.60** [2.22]	1.74*** [3.19]

某些信息的存在,或缺乏获取这些信息的必要工具与知识(赫舒拉发、张,2003;洪、斯坦,1999)。另一方面,对于投资者而言,注意力是一种稀缺的认知资源(卡尼曼,1973),因此,即使信息可得,投资者受有限注意力约束,也无法全面关注所有信息,只能进行选择性地处理。基于此,本文通过考察不同信息可得性下投资者预期协同与股票未来收益关系的变化,来探究其背后的预期形成机制。

本节基于文献综述中“信息获取”这一分歧来源进行模型设定,假设投资者只能获得完整数据集的一个子集。具体而言,本节假设每个投资者 n 只能随机获得 K 个公司特征中的一定比例,记为 P 。然后,本文为每个投资者 n 构建一个特征索引的有序序列 $k_n=(k_{n,1},k_{n,2},\cdots,k_{n,p-1},k_{n,p})$,其中 $p=K\times P$,该序列从 $\{1,2,\cdots,K-1,K\}$ 中无放回随机抽取,从而确保索引的唯一性。在此设定下,公式(3)中投资者对信息集的处理函数被重新定义为:

$$g_n(z_{i,t})=\left(z_{i,t}^{k_{n,1}},z_{i,t}^{k_{n,2}},\cdots,z_{i,t}^{k_{n,p-1}},z_{i,t}^{k_{n,p}}\right) \quad (10)$$

本节令 P 分别取20%、40%、60%和80%,其他条件不变,重复机器学习与定价检验的过程。

表7展示了在 P 的不同取值下,单变量组合分析的结果。由结果可知,当投资者可获得公司特征的比例适中时,对冲组合的收益较为显著,而当投资者可获得公司特征的比例太高或太低时,对冲组合的收益有所下降。这种倒U型的结果表明,当学习的特征数量太少时,投资者难以全面把握公司特征与股票收益之间的关系,导致其无法形成稳定有效的预期,从而削弱了预期协同的定价效应;而当学习的特征数量太多时,投资者“信息获取”的差异越来越小,对股票收益的预期趋于一致,也会削弱预期协同的定价效应。

(三)投资者不同信息偏好下的预期协同定价效应研究

在我国的金融市场中,存在价值型与技术型两种投资者,其关注的信息存在较大差异。为了更深入地理解这两种投资风格对市场的具体影响,本文提出了一个理想化的假设:设想一个完全由价值型投资者构成的市场,与之相对的是一个完全由技术型投资者构成的市场。这种设想虽然在现实中难以完全实现,但通过神经网络模型分别模拟这两种极端的市場类型,有助于清晰揭示不同风格投资者的预期协同对金融市场的差异化影响。

价值型投资者通常基于公司的基本面进行投资决策,如财务状况、行业地位、市场竞争力等,追求长期价值的实现。相反,技术型投资者主要依赖历史价格走势和市场交易量等技术信息来预测市场的短期波动,寻求从中获利。为了模拟这两种经典的投资风格,本文构建了两组独立的投资者群体。具体而言,价值型投资者仅关注与公司基本面高度相关的“估值与成长类”和“盈利类”指标,所有其他类型特征的权重均设为0;相应地,技术型投资者则仅关注与市场交易行为相关的“交易摩擦类”和“惯性类”指标。通过这种严格的信息隔离设定,本文旨在干净地分离并检验完全依赖于不同信息类型的预期协同,其定价效应是否存在显著差异。

在设定好两类投资者的基础上,本文遵循前述方法计算不同风格投资者的预期分散度指标,并采用投资组合分析法,检验不同投资风格下的预期协同所产生的差异化定价效应。表8报告了技术型和价值型投资者预期分散度的单变量组合分析结果。Panel A结果显示,两类投资者的预期协同定价效应存在显著差异。具体而言,在完全由技术型投资者构成的市场中,预期协同的定价效应显著,对冲组合收益达0.73%,且在1%的水平上显著($t=2.71$),而在完全由价值型投资者构成的市场中,该效应则不显著。

然而,上述结果可能源于不同类型特征更新频率的差异。技术型投资者所依赖的技术类指标(如换手率)是高频更新的,而价值型投资者依赖的基本面指标(如账面市值比)则是低频更新的。在12个月的训练窗口中,高频变化的技术类指标更易被模型捕捉并赋予较高权重,这可能引发对高频信号的过拟合。因此,为控制特征更新频率的影响,本文对训练集中的所有变量进行时序均值化处理,将12个月的面板数据压缩

表7 不同数据可得性下的单变量组合分析

投资组合	投资者可获得公司特征的比例			
	20%	40%	60%	80%
低 IED	0.95 [1.56]	1.06* [1.73]	1.08* [1.74]	1.04* [1.71]
高 IED	0.37 [0.62]	0.34 [0.54]	0.24 [0.39]	0.40 [0.63]
对冲组合	0.58** [2.28]	0.72*** [2.94]	0.84*** [3.37]	0.63*** [2.59]

表8 不同投资者风格下的单变量组合分析

投资组合	Panel A: 原始特征		Panel B: 特征同频	
	技术型	价值型	技术型	价值型
低 IED	0.92 [1.40]	0.84 [1.25]	0.99 [1.48]	0.93 [1.35]
高 IED	0.20 [0.29]	0.64 [1.05]	0.39 [0.61]	0.65 [1.08]
对冲组合	0.73*** [2.71]	0.20 [0.62]	0.61** [2.07]	0.28 [0.90]

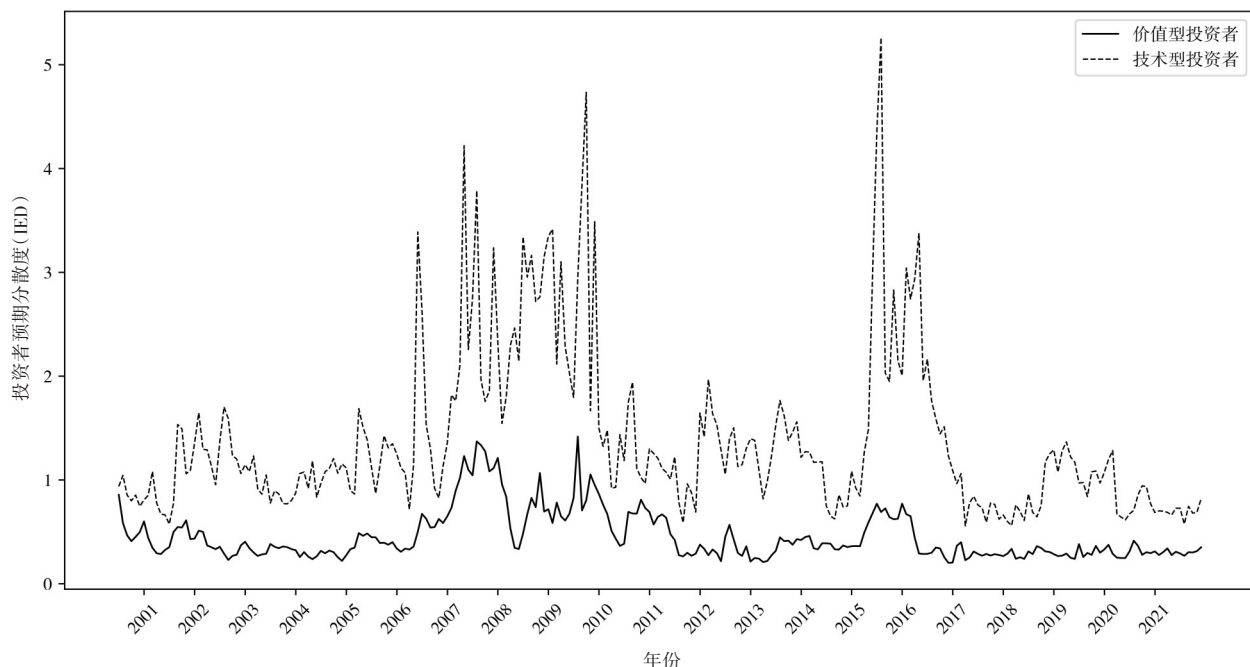


图3 不同风格投资者预期分散度的时间序列图

为单一横截面,从而统一不同特征的更新频率。Panel B展示了在“特征同频”设定下的结果,技术型投资者的对冲组合收益依然达到了0.61%,且在5%的水平上显著($t=2.07$),而价值型投资者的定价效应依旧不显著。综上,在技术型投资者主导的市场中,错误定价的程度更深,预期协同的定价效应更强;而在价值型投资者主导的市场中,预期相对理性,预期协同的定价效应也就不明显。

为了更加直观地展示两类投资者在预期协同程度上的差异,本文将个股的投资者预期分散度在横截面上进行平均,图3展示了价值型和技术型两类投资者预期分散度的时间序列。结果显示,技术型投资者的预期分散度明显高于价值型投资者,且波动率更大。图3表明,投资者基于技术信号产生的预期更分散,预期协同水平更低,而基于价值指标产生的预期更一致,预期协同水平更高,这说明在散户居多且大多关注技术信号的中国市场,预期管理尤其重要。

六、进一步分析

（一）异质性分析

本节从公司市值和股权性质两个维度,对投资者预期协同的定价效应进行异质性分析。特别地,针对市值效应的分析,本文还专门考察并排除“壳价值”这一特殊因素的潜在影响。

1. 市值效应分析

本文采用双变量组合分析法,考察投资者预期协同的定价效应与公司市值的关系。具体而言,在月度层面上,分别将个股按公司市值和投资者预期分散度进行排序并划分投资组合。其中,按照市值从低到高分为小市值组合(市值位于30%分位数以下)、中市值组合(市值位于30%~70%分位数之间)和大市值组合(市值位于70%分位数以上)。随后,在各个市值组合内,将股票按照投资者预期分散度从低到高分为10个投资组合。

表9展示了各个市值组合内IED对冲组合的收益表现。由表9可知,大市值组合和中市值组合内的IED对冲组合收益率分别为1.09%和0.54%,均在1%水平上显著,且经风险调整后依然显著。然而,小市值组合内的IED对冲组合的收益率接近于零,且统计上不显著。上述结果表明,投资者预期协同的定价效应存在明显的市值异质性,在大市值和中市值公司中表现较强,而在小市值公司中则不显著。

小市值公司中投资者预期协同定价效应的消失可能是受到“壳价值”的影响。“壳价值”指在中国股票市场

中,一家已上市的小市值公司因其可能在借壳上市中被用作“壳公司”而产生的价值,这主要源于上市资格的稀缺性。一方面,“壳价值”难以在公司特征中得到充分体现,因而会对基于公司特征生成的投资者预期分散度指标产生影响。另一方面,对小市值股票,“壳价值”可能成为投资者关注的主要因素,从而超越其他的公司特征,并影响投资者的预期协同效应,最终使得IED和股票未来收益之间的关系变得不显著。因此,为更清晰地识别市值效应,本文在每个月份内排除了市值最小的30%的股票(刘等,2019),以排除“壳价值”的影响,然后重新训练神经网络模型并计算投资者预期分散度指标。

在此基础上,本文基于公式(9)的设定进行FM回归分析,重新估计排除“壳价值”后投资者预期协同的定价效应。结果如表3所示,列(4)表明,在排除“壳价值”股票后,IED系数的统计显著性较全样本及其他异质性分组情形有所提高。由此可见,在排除“壳价值”影响后,投资者预期协同的定价效应显著增强,这进一步支持了本文的核心结论。

2. 股权性质分析

由于民营企业[®]与国有企业[®]在所有权结构和经营模式上存在本质区别,投资者对这两类企业的信息解读和预期形成过程也存在显著差异。为深入探讨股权性质对投资者预期协同定价效应的影响,本研究分别对民营企业与国有企业子样本重新计算投资者预期分散度并进行定价效应分析。

表10 Panel A报告了民营企业和国有企业的单变量组合分析结果,结果显示,民营企业的IED对冲组合收益率为0.65%($t=2.59$),国有企业的IED对冲组合收益率为0.49%($t=1.77$)。在民营企业中,IED对股票未来收益的预测能力显著更强。根据公式(9)进行FM回归,结果如表3所示,列(5)和列(6)分别报告了在民营企业和国有企业中的FM回归结果。其中,民营企业IED的平均系数为-0.21,国有企业为-0.18,两者均在1%水平上显著,但在民营企业中,预期协同的定价效应更强。这可能源于民营企业面临的信息环境更为复杂,从而放大了错误定价的程度。相较于信息披露更规范的国有企业,民营企业更高的信息不对称性,使得卖空约束等市场摩擦对价格的扭曲作用更强。因此,即使是同等程度的投资者分歧,也更可能在民营企业中引发显著的价格高估。最终,这种被放大的错误定价需要未来更大幅度的价格修正,从而使得预期协同的定价效应在民营企业中表现得更为显著。

(二) 市场周期对定价效应的影响

本节考察投资者预期协同的定价效应在不同股市周期内的差异。参考马休和麦柯迪(2000)的分析框架,本文划分中国市场的牛市和熊市两个周期,结果如图4所示,图中阴影部分为熊市周期,而非阴影部分为牛市周期。

为研究定价效应的差异,本节分别在两个市场周期内进行了单变量组合分析。表10 Pan-

表9 市值与投资者预期协同的双变量组合分析

对冲组合	Ret	SR	CAPM- α	FF3- α	FF5- α
小市值	0.11 [0.40]	0.09	0.16 [0.61]	0.26 [1.05]	0.04 [0.16]
中市值	0.54*** [2.65]	0.47	0.59*** [3.05]	0.70*** [3.60]	0.53** [2.38]
大市值	1.09*** [3.92]	0.69	1.12*** [3.98]	0.90*** [3.45]	0.79*** [2.66]

表10 不同股权性质和市场周期下的单变量组合分析

对冲组合	Ret	SR	CAPM- α	FF3- α	FF5- α
Panel A: 不同股权性质下的组合分析结果					
民营企业	0.65*** [2.59]	0.44	0.69*** [2.76]	0.45* [1.81]	0.30 [1.05]
国有企业	0.49* [1.77]	0.32	0.58** [2.08]	0.38 [1.39]	0.24 [0.83]
Panel B: 不同市场周期下的组合分析结果					
熊市	0.88** [2.40]	0.65	0.48 [1.37]	0.36 [1.09]	-0.13 [-0.38]
牛市	0.89*** [2.97]	0.59	1.10*** [2.99]	0.70** [1.96]	0.79* [1.91]

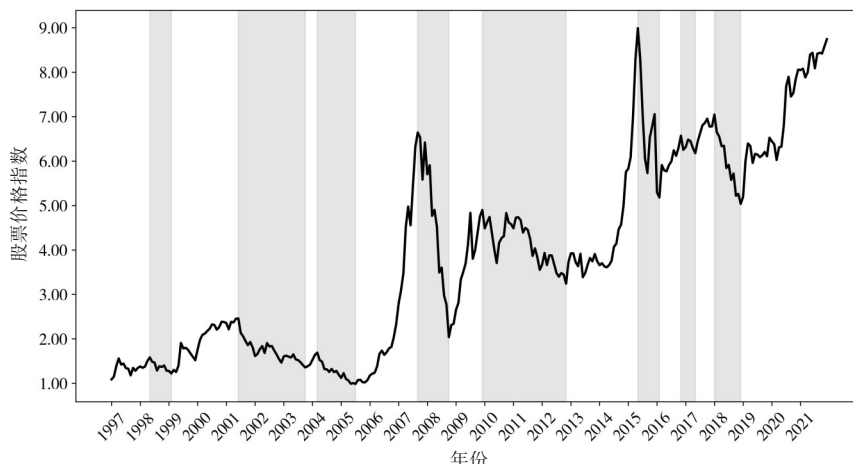


图4 股市牛熊市周期划分时序图

注:股票价格指数为上证主板、深证主板和创业板的所有股票按照流通市值加权计算的平均价格指数,以1996年12月31日为基日,基点为1。

el B展示了在熊市和牛市周期内,IED对冲组合的月度平均收益率、夏普比率以及经因子模型调整后的CAPM- α 、FF3- α 和FF5- α 。由表10 Panel B的结果可知,尽管对冲组合的月度平均收益率在牛市、熊市中较为接近,但风险调整后收益存在显著差异。在牛市周期内,对冲组合的CAPM- α 高达1.10%($t=2.99$),且FF3- α 和FF5- α 依然显著。然而,在熊市周期内,对冲组合的风险调整后收益均已不再显著。由此可知,投资者预期协同的定价效应在牛市更加显著且稳健,这一发现凸显了预期管理对于中国资本市场的重要性,尤其是在市场过热时期,其对维护长期稳定的意义不容忽视。

七、结论与建议

本文引入机器学习框架,采用神经网络模型学习上市公司的多维度特征,以模拟中国股票市场中投资者的预期形成过程,进而构建投资者预期分散度指标。该指标用以度量投资者预期协同的程度,数值越低,预期协同程度越高。研究发现,投资者的预期协同程度越低,股票的未来收益越低,这表明提升预期协同程度有助于促进股票市场的合理定价与稳定运行。机制分析表明,卖空约束与有限套利引起的错误定价是投资者预期协同定价效应的关键作用机制。具体而言,当投资者缺乏预期协同时,卖空约束、有限套利等因素共同作用,阻碍价格的有效发现,从而导致了错误定价,这使得预期协同的定价效应在高摩擦环境中更为显著。此外,本文发现会计信息质量与预期协同的定价效应显著负相关,而特征数量则对定价效应呈现先增强后减弱的非线性影响。在投资者存在信息偏好的情况下,技术型投资者和价值型投资者的预期形成和预期协同的定价效应均表现出明显的差异。这表明,在中国这样一个以技术信息偏好的散户投资者为主的市场中,加强投资者预期管理具有尤为突出的现实意义。进一步分析发现,在不同类型的上市公司中,民营企业的投资者预期协同定价效应更为明显;在不同的市场周期中,牛市周期下预期协同的定价效应更为显著。上述发现都揭示了多特征学习下投资者预期的形成机制和经济效果,对于理解金融市场如何通过信息与预期影响实体经济发展具有重要意义。本文具有如下政策与研究启示。

首先,对投资者而言,预期的低协同度可能意味着股票被高估,从而预示着较低的未来收益。因此,在进行投资决策时,投资者应充分考虑其他投资者的预期,并适当分散化投资,以降低由高预期差异引致的潜在风险,规避买入高估资产造成的未来损失。此外,投资者未来可借助人工智能技术,对市场的多维度高质量信息进行学习,从而提升人类投资者有限的信息处理能力,以更科学地进行资产配置和管理,争做中国市场价值投资的耐心资本。

其次,对市场监管机构而言,应持续重视资本市场信息质量的建设与提升。通过促进上市公司披露更多维度的信息,包括ESG、人力资本、技术创新等,引导投资者对优质上市公司形成更高程度的预期协同,进而提升市场的定价效率,助力金融市场更好地服务于实体经济发展。此外,应健全我国资本市场的交易制度,降低资本市场的错误定价程度,从而提升预期管理的效果。特别是当外部市场存在丰富的做空机制时,国内市场如何减少不对称交易所导致的错误定价,是未来学术研究和制度设计的关键课题。

最后,对政策制定者而言,应进一步推动人工智能技术在金融管理中的应用。为此,需积极制定相应的激励措施,鼓励相关机构在资产定价、风险管理和监管科技等方面开展“人工智能+金融市场”的应用研究,切实保证人工智能技术更广泛地服务于资本市场的“数智化”转型,推动形成智能向善理论体系,助力金融市场的健康与可持续发展。此外,应完善市场沟通机制,通过精准测度并有效引导市场预期,平抑市场波动,增强投资者对中国金融市场的发展信心,进一步夯实金融稳定的微观基础^①。

(作者单位:唐国豪,湖南大学金融与统计学院、高性能分布式账本与数字金融教育部重点实验室;陈海玮,上海财经大学金融学院;朱琳,广东财经大学金融学院;姜富伟,厦门大学经济学院、厦门大学王亚南经济研究院)

注释

①由于篇幅限制,公司特征因子指标名称与文献出处详见《管理世界》网络发行版附录1附表1。

②与现实情况相符,单个模拟投资者的预测效果并不十分准确,而所有模拟投资者作为一个整体,预测效果会随着投资者数量的增加而提升,详见《管理世界》网络发行版附录2附图1~附图4。

③正则化系数,在正则化技术中用来控制正则化项的强度。在L1和L2正则化中, λ 的取值越大,模型的复杂度越小,过拟合的风险越低。相反, λ 的取值越小,模型的复杂度越大,过拟合的风险越高。

④进一步的分析表明,无论市场整体预期是乐观还是悲观,高预期协同的股票组合总是表现出最温和、反极端的预期,而低协同组合则总是由最极端的预期所驱动,且始终预示着更低的未来收益,这佐证了高协同是理性定价的体现,而低协同与错误定价高度相关。详见《管理世界》网络发行版附录3附表2。

⑤在加入了多达15个控制变量后,IED的系数仍在1%的水平上显著,详见《管理世界》网络发行版附录4附表3。

⑥证监会于2010年3月启动A股融资融券业务试点,之后多次对融资融券标的进行了大规模扩容,根据俞红海等(2018)和郭彪等(2020),本文选择2011年12月(第一次扩容)后的数据进行分析。

⑦使用Fama-French五因子模型对各组合收益率进行风险调整,可以得到相同的结论,详见《管理世界》网络发行版附录5附表4。

⑧包括中外合作企业、中外合资企业、外商独资企业、民营企业与其他企业。

⑨包括国营或国有控股企业、集体企业、事业单位与政府机关。

⑩中外文人名(机构名)对照:席勒(Shiller);琼斯(Jones);内格尔(Nagel);徐(Xu);顾(Gu);凯利(Kelly);莱特(Light);法马(Fama);弗伦奇(French);阿特马兹(Atmaz);巴沙克(Basak);安德森(Anderson);迪特(Deether);陈(Chan);莫科阿莱利-莫科泰利(Mokoaleli-Mokoteli);拉比诺维茨(Rabinowitz);武(Wu);巴伯(Barber);洪(Hong);斯坦(Stein);高(Gao);加芬克尔(Garfinkel);索科宾(Sokobin);库克森(Cookson);尼斯纳(Niessner);陈(Chen);卡尼尔(Kaniel);巴利(Bali);约(Yeo);西姆斯(Sims);丹尼尔(Daniel);赫舒拉发(Hirschleifer);霍尼克(Hornik);米勒(Miller);坎德尔(Kandel);皮尔逊(Pearson);戈茨曼(Goetzmann);马萨(Massa);约翰逊(Johnson);余(Yu);布卢姆菲尔德(Bloomfield);费希尔(Fischer);祖卡斯(Doukas);帕克(Park);贝姆(Boehme);萨德卡(Sadka);谢尔宾娜(Scherbina);阿夫拉莫夫(Avramov);卡林(Carlin);班纳吉(Banerjee);斯雷尔(Sraer);黄(Huang);阿米胡德(Amihud);卡尼曼(Kahneman);张(Teoh);刘(Liu);马休(Maheu);麦柯迪(McCurdy)。

参考文献

- (1) 邵慧、解峥、李佳鸿、吴俊杰:《基于股评的投资者情绪对股票市场的影响》,《管理科学学报》,2018年第4期。
- (2) 范小云、王业东、王道平、郭文璇、胡煊翊:《不同来源金融文本信息含量的异质性分析——基于混合式文本情绪测度方法》,《管理世界》,2022年第10期。
- (3) 郭彪、刘普阳、姜圆:《卖空限制与收益可预测性——A股融资融券制度的证据》,《金融研究》,2020年第8期。
- (4) 洪永森、汪寿阳:《大数据如何改变经济学研究范式?》,《管理世界》,2021年第10期。
- (5) 李斌、邵新月、李玥阳:《机器学习驱动的基本面量化投资研究》,《中国工业经济》,2019年第8期。
- (6) 刘贯春、叶永卫、陈肖雄、张军:《固定资产折旧、税收筹划与异常投资》,《经济研究》,2023年第4期。
- (7) 刘景江、郑畅然、洪永森:《机器学习如何赋能管理研究?——国内外前沿综述和未来展望》,《管理世界》,2023年第9期。
- (8) 刘燕、朱宏泉:《个体与机构投资者,谁左右A股股价变化?——基于投资者异质信念的视角》,《中国管理科学》,2018年第4期。
- (9) 马甜、姜富伟、唐国豪:《深度学习与中国股票市场因子投资——基于生成式对抗网络方法》,《经济学(季刊)》,2022年第3期。
- (10) 孟庆斌、黄清华:《卖空机制是否降低了股价高估?——基于投资者异质信念的视角》,《管理科学学报》,2018年第4期。
- (11) 万海远、张尉、陈基平、辛柯均:《税收政策支持与企业预期转变》,《经济研究》,2024年第4期。
- (12) 伍燕然、潘可、胡松明、江婕:《行业分析师盈利预测偏差的新解释》,《经济研究》,2012年第4期。
- (13) 徐经长、柯劭婧、何乐伟:《新收入准则能否提高会计信息质量?——基于分析师预测视角的研究》,《会计研究》,2022年第9期。
- (14) 薛健、汝毅:《信息披露业务关系与新闻报道质量》,《管理世界》,2020年第10期。
- (15) 杨威、冯璐、宋敏、李春涛:《锚定比率可以衡量股价高估吗?——基于崩盘风险视角的经验证据》,《管理世界》,2020年第1期。
- (16) 尹海员、朱旭:《投资者异质信念、预期演化与股票市场流动性》,《中国管理科学》,2019年第10期。
- (17) 俞红海、陈百助、蒋振凯、钱仪倬:《融资融券交易行为及其收益可预测性研究》,《管理科学学报》,2018年第1期。
- (18) 张成思、孙宇辰、阮睿:《经济政策不确定性、企业货币政策感知与实业投资》,《财贸经济》,2023年第7期。
- (19) 张鹏、党世力、黄梅雨、李璟欣:《基于机器学习预测股票收益率的两步骤M-SV投资组合优化》,《中国管理科学》,2023年第12期。
- (20) 张玉龙、李怡宗:《基于随机折现因子方法的流动性定价机制研究》,《管理世界》,2013年第10期。
- (21) Amihud, Y., 2002, "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects", *Journal of Financial Markets*, vol.5 (1), pp.31~56.
- (22) Anderson, E. W., Ghysels, E. and Juergens, J. L., 2005, "Do Heterogeneous Beliefs Matter for Asset Pricing?", *The Review of Financial Studies*, vol.18(3), pp.875~924.
- (23) Atmaz, A. and Basak, S., 2018, "Belief Dispersion in the Stock Market", *The Journal of Finance*, vol.73(3), pp.1225~1279.
- (24) Avramov, D., Chordia, T., Jostova, G. and Philipov, A., 2009, "Dispersion in Analysts' Earnings Forecasts and Credit Rating", *Journal of Financial Economics*, vol.91(1), pp.83~101.
- (25) Bali, T. G., Kelly, B. T., Mörke, M. and Rahman, J., 2023, "Machine Forecast Disagreement", NBER Working Paper, No.31583.
- (26) Banerjee, S., 2011, "Learning from Prices and the Dispersion in Beliefs", *The Review of Financial Studies*, vol.24(9), pp.3025~3068.

- (27) Barber, B. M., Odean, T. and Zhu, N., 2009, "Do Retail Trades Move Markets?", *The Review of Financial Studies*, vol.22(1), pp.151~186.
- (28) Bloomfield, R. and Fischer, P. E., 2011, "Disagreement and the Cost of Capital", *Journal of Accounting Research*, vol.49(1), pp.41~68.
- (29) Boehme, R. D., Danielsen, B. R. and Sorescu, S. M., 2006, "Short-Sale Constraints, Differences of Opinion, and Overvaluation", *Journal of Financial and Quantitative Analysis*, vol.41(2), pp.455~487.
- (30) Carlin, B. I., Longstaff, F. A. and Matoba, K., 2014, "Disagreement and Asset Prices", *Journal of Financial Economics*, vol.114(2), pp.226~238.
- (31) Chan, L. K. C., Karceski, J. and Lakonishok, J., 2007, "Analysts' Conflicts of Interest and Biases in Earnings Forecasts", *Journal of Financial and Quantitative Analysis*, vol.42(4), pp.893~913.
- (32) Chen, L., Pelger, M. and Zhu, J., 2024, "Deep Learning in Asset Pricing", *Management Science*, vol.70(2), pp.714~750.
- (33) Cookson, J. A. and Niessner, M., 2020, "Why Don't We Agree? Evidence from a Social Network of Investors", *The Journal of Finance*, vol.75(1), pp.173~228.
- (34) Daniel, K., Hirshleifer, D. and Subrahmanyam, A., 1998, "Investor Psychology and Security Market Under- and Overreactions", *The Journal of Finance*, vol.53(6), pp.1839~1885.
- (35) Daniel, K., Hirshleifer, D. and Teoh, S. H., 2002, "Investor Psychology in Capital Markets: Evidence and Policy Implications", *Journal of Monetary Economics*, vol.49(1), pp.139~209.
- (36) Diether, K. B., Malloy, C. J. and Scherbina, A., 2002, "Differences of Opinion and the Cross Section of Stock Returns", *The Journal of Finance*, vol.57(5), pp.2113~2141.
- (37) Doukas, J. A., Kim, C. F. and Pantzalis, C., 2006, "Divergence of Opinion and Equity Returns", *Journal of Financial and Quantitative Analysis*, vol.41(3), pp.573~606.
- (38) Fama, E. F. and French, K. R., 1992, "The Cross-section of Expected Stock Returns", *The Journal of Finance*, vol.47(2), pp.427~465.
- (39) Fama, E. F. and French, K. R., 2015, "A Five-Factor Asset Pricing Model", *Journal of Financial Economics*, vol.116(1), pp.1~22.
- (40) Fischer, P., Kim, C. and Zhou, F., 2022, "Disagreement about Fundamentals: Measurement and Consequences", *Review of Accounting Studies*, vol.27(4), pp.1423~1456.
- (41) Gao, Y., Mao, C. X. and Zhong, R., 2006, "Divergence of Opinion and Long-term Performance of Initial Public Offerings", *Journal of Financial Research*, vol.29(1), pp.113~129.
- (42) Garfinkel, J. A., 2009, "Measuring Investors' Opinion Divergence", *Journal of Accounting Research*, vol.47(5), pp.1317~1348.
- (43) Garfinkel, J. A. and Sokobin, J., 2006, "Volume, Opinion Divergence, and Returns: A Study of Post-Earnings Announcement Drift", *Journal of Accounting Research*, vol.44(1), pp.85~112.
- (44) Goetzmann, W. N. and Massa, M., 2005, "Dispersion of Opinion and Stock Returns", *Journal of Financial Markets*, vol.8(3), pp.324~349.
- (45) Gu, S., Kelly, B. T. and Xiu, D., 2020, "Empirical Asset Pricing via Machine Learning", *The Review of Financial Studies*, vol.33(5), pp.2223~2273.
- (46) Gu, S., Kelly, B. T. and Xiu, D., 2021, "Autoencoder Asset Pricing Models", *Journal of Econometrics*, vol.222(1), pp.429~450.
- (47) Hirshleifer, D., 2001, "Investor Psychology and Asset Pricing", *The Journal of Finance*, vol.56(4), pp.1533~1597.
- (48) Hirshleifer, D. and Teoh, S. H., 2003, "Limited Attention, Information Disclosure, and Financial Reporting", *Journal of Accounting and Economics*, vol.36(1~3), pp.337~386.
- (49) Hong, H., Lim, T. and Stein, J. C., 2000, "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies", *The Journal of Finance*, vol.55(1), pp.265~295.
- (50) Hong, H. and Sraer, D. A., 2016, "Speculative Betas", *The Journal of Finance*, vol.71(5), pp.2095~2144.
- (51) Hong, H. and Stein, J. C., 1999, "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets", *The Journal of Finance*, vol.54(6), pp.2143~2184.
- (52) Hong, H. and Stein, J. C., 2007, "Disagreement and the Stock Market", *Journal of Economic Perspectives*, vol.21(2), pp.109~128.
- (53) Hornik, K., Stinchcombe, M. and White, H., 1989, "Multilayer Feedforward Networks Are Universal Approximators", *Neural Networks*, vol.2(5), pp.359~366.
- (54) Huang, D., Li, J. and Wang, L., 2021, "Are Disagreements Agreeable? Evidence from Information Aggregation", *Journal of Financial Economics*, vol.141(1), pp.83~101.
- (55) Johnson, T. C., 2004, "Forecast Dispersion and the Cross Section of Expected Returns", *The Journal of Finance*, vol.59(5), pp.1957~1978.
- (56) Jones, C. M., Shi, D., Zhang, X. and Zhang, X., 2025, "Retail Trading and Return Predictability in China", *Journal of Financial and Quantitative Analysis*, vol.60(1), pp.68~104.
- (57) Kahneman, D., 1973, *Attention and Effort*, Englewood Cliffs, NJ: Prentice-Hall.
- (58) Kandel, E. and Pearson, N. D., 1995, "Differential Interpretation of Public Signals and Trade in Speculative Markets", *Journal of Political Economy*, vol.103(4), pp.831~872.

- (59) Kaniel, R., Lin, Z., Pelger, M. and Van Nieuwerburgh, S., 2023, "Machine-Learning the Skill of Mutual Fund Managers", *Journal of Financial Economics*, vol.150(1), pp.94~138.
- (60) Kelly, B. T., Malamud, S. and Zhou, K., 2024, "The Virtue of Complexity in Return Prediction", *The Journal of Finance*, vol.79(1), pp.459~503.
- (61) Kelly, B. T., Pruitt, S. and Su, Y., 2019, "Characteristics Are Covariances: A Unified Model of Risk and Return", *Journal of Financial Economics*, vol.134(3), pp.501~524.
- (62) Light, N., Maslov, D. and Rytchkov, O., 2017, "Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach", *The Review of Financial Studies*, vol.30(4), pp.1339~1381.
- (63) Liu, J., Stambaugh, R. F. and Yuan, Y., 2019, "Size and Value in China", *Journal of Financial Economics*, vol.134(1), pp.48~69.
- (64) Maheu, J. M. and McCurdy, T. H., 2000, "Identifying Bull and Bear Markets in Stock Returns", *Journal of Business & Economic Statistics*, vol.18(1), pp.100~112.
- (65) Miller, E. M., 1977, "Risk, Uncertainty, and Divergence of Opinion", *The Journal of Finance*, vol.32(4), pp.1151~1168.
- (66) Mokoaleli-Mokoteli, T., Taffler, R. J. and Agarwal, V., 2009, "Behavioural Bias and Conflicts of Interest in Analyst Stock Recommendations", *Journal of Business Finance & Accounting*, vol.36(3~4), pp.384~418.
- (67) Nagel, S., 2005, "Short Sales, Institutional Investors and the Cross-Section of Stock Returns", *Journal of Financial Economics*, vol.78(2), pp.277~309.
- (68) Nagel, S. and Xu, Z., 2023, "Dynamics of Subjective Risk Premia", *Journal of Financial Economics*, vol.150(2), No.103713.
- (69) Park, C., 2005, "Stock Return Predictability and the Dispersion in Earnings Forecasts", *The Journal of Business*, vol.78(6), pp.2351~2376.
- (70) Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. M. A. and Botvinick, M., 2018, "Machine Theory of Mind", in *Proceedings of the 35th International Conference on Machine Learning*, vol.80, pp.4218~4227.
- (71) Sadka, R. and Scherbina, A., 2007, "Analyst Disagreement, Mispricing, and Liquidity", *The Journal of Finance*, vol.62(5), pp.2367~2403.
- (72) Shiller, R. J., 1981, "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?", *American Economic Review*, vol.71, pp.421~436.
- (73) Sims, C. A., 2003, "Implications of Rational Inattention", *Journal of Monetary Economics*, vol.50(3), pp.665~690.
- (74) Vo, N. N., He, X., Liu, S. and Xu, G., 2019, "Deep Learning for Decision Making and the Optimization of Socially Responsible Investments and Portfolio", *Decision Support Systems*, vol.124, No.113097.
- (75) Yeo, W. J., Van Der Heever, W., Mao, R., Cambria, E., Satapathy, R. and Mengaldo, G., 2025, "A Comprehensive Review on Financial Explainable AI", *Artificial Intelligence Review*, vol.58, No.189.
- (76) Yu, J., 2011, "Disagreement and Return Predictability of Stock Portfolios", *Journal of Financial Economics*, vol.99(1), pp.162~183.

Asset Pricing Based on Investor Expectation Synergy: A Machine Learning Perspective with Multiple Characteristics

Tang Guohao^{a,b}, Chen Haiwei^c, Zhu Lin^d and Jiang Fuwei^{e,f}

(a. College of Finance and Statistics, Hunan University; b. Ministry of Education Key Laboratory of High-Performance Distributed Ledger and Digital Finance; c. School of Finance, Shanghai University of Finance and Economics; d. School of Finance, Guangdong University of Finance and Economics; e. School of Economics, Xiamen University; f. The Wang Yanan Institute for Studies in Economics, Xiamen University)

Abstract: In the era of big data, with the deep application of artificial intelligence technology, investors are more capable than ever of effectively focusing on the vast amount of information in the capital markets. Whether investors' attention to multidimensional information can enhance expectation synergy is of great significance for a deeper understanding of the formation of expectations in capital markets. Based on the multidimensional characteristics of Chinese listed companies, this paper introduces a machine learning framework to construct a measure of expectation synergy, exploring how investor expectations under multiple-characteristic learning affect asset pricing. Empirical results show that in the Chinese stock market, the lower the level of expectation synergy, the lower the future returns of stocks, indicating that enhancing investor expectation synergy contributes to the stability and sustainable development of the capital market in China. This paper further demonstrates, from the perspectives of short-sale constraints and limited arbitrage, that mispricing is the underlying mechanism for the positive correlation between investor expectation synergy and future stock returns. Additionally, this paper finds that the quality of accounting information has a significant negative correlation with the pricing effect of expectation synergy, while the number of learning characteristics exhibits an inverted U-shaped non-linear impact on the pricing effect. A heterogeneity analysis is conducted for investors with different trading styles, revealing that differences in investors' preferences for financial signals lead to different pricing effects of expectation synergy. Finally, this paper proposes policy recommendations from the perspectives of investors' information learning, capital market information disclosure quality, AI technology application, and expectation management.

Keywords: feature learning; expectation synergy; expectation management; neural network model; asset pricing