

doi: 10.19920/j.cnki.jmsc.2026.02.005

原油新闻语调与期货市场收益率： 基于大语言模型与垂直领域词典集成研究^①

陈荣达^{1,2}, 肖文昊¹, 金骋路^{1*}, 厉涵¹

(1. 浙江财经大学金融学院, 杭州 310018; 2. 浙江金融职业学院, 杭州 310018)

摘要: 大语言模型的垂直化发展已成为重要趋势,尤其在原油期货等高度专业化、跨市场场景下,通用大语言模型难以准确识别术语与复杂语境,且易产生前视偏差。本文提出集成大语言模型与垂直领域词典的原油新闻语调测度方法,通过融合领域词典的先验知识与大语言模型的语义理解能力,提升所测度语调的解释性与预测能力。具体地,基于 2018 年—2022 年来自 InfoBank 与 Factiva 的 93 004 篇中英文原油新闻,本文构建垂直领域词典并提纯高信噪比语料,在通用 BERT 模型上进行领域预训练,继而结合期货收益符号开展弱监督微调,集成测度原油期货新闻语调。实证结果表明,该方法得到的新闻语调在解释和预测上海原油期货收益率方面显著优于词典法与通用模型;各项稳健性检验结果一致,并在 2023 年—2024 年样本外预测中表现优越;新闻语调对原油期货收益率的影响主要经由投资者关注度这一中介渠道传导;在集成方法下,该机制的方向与显著性与经济逻辑保持一致。另外,进一步的风险溢出分析发现基于集成方法的新闻语调能够有效刻画国际新闻情绪对国内期货市场的尾部风险传导。本文贡献在于提出可复制的金融大模型垂直化应用路径,从收益与风险双视角揭示了新闻语调在市场信息传导中的作用机制,为原油期货市场的风险识别与政策制定提供了新的量化工具。

关键词: 新闻语调; 大语言模型; 垂直领域词典; 集成方法; 经领域适配的 BERT 模型

中图分类号: F832.5 **文献标识码:** A **文章编号:** 1007-9807(2026)02-0081-23

0 引言

习近平总书记在中共中央政治局第二十次集体学习时强调“人工智能作为引领新一轮科技革命和产业变革的战略性技术,深刻改变人类生产生活方式。”2024 年 7 月党的二十届三中全会明确将人工智能作为战略性新兴产业,旨在推动实现各行业的数智化转型;2025 年 3 月国务院总理李强在政府工作报告中首次提出“人工智能+”战略;2025 年 8 月国务院正式印发《关于深入实施“人工智能+”行动的意见》,宣布中国将在人工智能领域全面加速布局。近年来,基于 Transformer

架构的大语言模型(LLM)快速迭代,也已成为金融文本分析的重要工具^[1]。依托“通用语料预训练+下游微调”的范式,大语言模型在舆情监测、事件抽取与情绪量化等任务中表现出优于传统文本分析方法的上下文理解与迁移能力^[2]。然而,伴随金融业务向专业化、精细化演进,学界与业界逐渐认识到:通用大语言模型训练语料来源广泛,主题覆盖面大但专业聚焦度不足,难以精准解码衍生品市场中高频出现的专业术语、复杂交易逻辑与跨市场叙事,进而在情绪测度上产生“术语误判”“极性漂移”等失真问题^[1]。更为重要的是,现有方法往往面临“前视偏差”困境,即在事后评

① 收稿日期: 2025-04-16; 修订日期: 2025-11-03.

基金项目: 国家社会科学基金资助重大项目(22&ZD073)。

通讯作者: 金骋路(1990—),男,浙江金华人,博士,副教授。Email: chenglu.jin@zufe.edu.cn

价中看似有效,但在投资者当期信息环境下却难以获得同等的预测力,从而导致过拟合或虚假有效性。因此,在金融领域的人工智能应用中,如何平衡模型的上下文理解能力与领域知识约束,是当前亟待解决的关键问题。

原油期货市场对检验该问题提供了理想实验场景。自2018年上海原油期货上市以来,我国通过“人民币计价+国际平台”的机制设计,显著提升了在全球能源市场的话语权,交易规模已跃居全球第三。但是,市场参与主体的异质性(境内散户高频博弈与境外机构长期套保并存)与信息传导的双轨性(境内媒体报道的即时性与国际市场信号的滞后性交织),催生出独特的信息传导现象。例如,2025年1月境内原油期货涨停与国际油价波动脱钩的案例,揭示了境内投资者情绪对短期价格的过度反应,而夜盘交易量占比超60%的同步性则凸显国际情绪对长期价值中枢的锚定作用,正暴露出情绪信息传导链条的复杂性与风险外溢性。在此背景下,精准识别国际原油新闻产生的跨市场影响,有助于防范市场异常波动、优化风险管理工具的政策需求。

不同于以股票市场为主要对象的新闻文本研究^[3-5],期货市场的高杠杆性、合约到期限限制、双向交易机制、高频交易特征及机构投资者主导的结构^[6],决定了原油期货相关新闻在信息传导机制和市场反应特征上表现出显著差异。已有研究从不同角度讨论了影响原油期货市场的因素,主要包括汇率^[7]、通胀^[8]和相关资产价格^[9]等结构化变量。但是,媒体报道是信息被市场吸收并反映于价格的起点^[10],相较于仅对现状进行客观统计的低频数据,新闻文本不仅有更高的数据频率和时效性^[11],在内容上也包括了更多的关于未来趋势的预期。近年来基于非结构化文本的原油期货研究逐渐丰富,如从新闻内容中提取的经济不确定性^[12]、地缘政治风险^[13]等信息。然而,现有研究在强调新闻内容对资产价格影响的同时,往往忽视了新闻语调这一独立维度。与新闻内容所传递的客观事实不同,新闻语调反映的是信息的情感极性与叙事强度,其通过影响投资者的注意力分配与风险感知,改变订单流与风险溢价,从而对期货定价产生独立于内容的边际作用。在原油市场高度国际化与波动性强的背景下,投资者在有

限注意力条件下更倾向依赖语调的直观情绪信号做出快速决策;国际新闻语调也更容易通过情绪共振引发跨市场风险传染;在市场剧烈波动期,极端化的新闻语调往往更敏锐地预示潜在的尾部风险。因此,在原油期货市场这一高度复杂的场景中,如何更准确地测度其相关新闻语调,并识别其对市场收益率和风险的边际作用,具有重要意义。

值得注意的是,在该情境下,采用传统文本情绪测度方法来量化新闻语调易失效:其一,通用金融词典/规则计数存在术语覆盖不足与极性错置的问题(如“钻井平台增加/供给增加”在通用词典里易被误判为利好,“需求增加”与“供给增加”虽同含“增加”,但对价格方向相反),对“套期保值”“逐日盯市”等期货语汇体系的映射也不充分,且常因事后扩充词条而引入前视偏差^[11,14,15];其二,通用大语言模型虽具上下文理解与语义建模优势^[16],但缺乏垂直领域预训练与跨语种/跨市场语境对齐,同样容易对语义极性的方向与强度识别不稳健。基于此,在原油期货这样信息密集且跨市场耦合强的衍生品市场,仅依赖词典或通用模型,难以得到对收益与风险真正有解释力与可预测力的语调指标。

基于上述考虑,本文提出一种基于大语言模型与垂直领域词典的原油新闻语调集成测度方法,并利用2018年3月—2022年12月来自Info-Bank与Factiva的93 004篇中英新闻为原始语料,检验经垂直领域语料预训练、收益信号微调是否能够在语调测度的解释性与预测性能上显著优于文献中常见的词典法与通用模型所构建的新闻语调,并探究语调影响收益的传导机制来以验证集成方法的经济一致性。主要贡献与意义在三个方面:1)本文集成大语言模型与垂直领域词典的语调测度方法,为原油期货等高波动、跨语种、跨市场场景下复杂金融语料的量化研究提供了可解释的建模工具。该方法在保留模型语境理解能力的同时嵌入领域知识约束,实现语义方向的可控性与经济含义的透明化,解决了传统语调测度方法可解释性弱、极性漂移和经济一致性不足等问题;2)在经济理论上,本文发现原油期货市场中的新闻语调对收益率和尾部风险均具有显著影响,为理解原油等市场收益与风险的驱动机制提供了新的影响因素。具体而言,新闻语调无论在在样

本内还是样本外的收益率预测都显著,说明情绪并非对价格的被动反应,而是对市场预期形成的主动输入。另外,在风险溢出模型(CoVaR框架)中,国际新闻语调能够显著影响上海原油期货市场的尾部风险,体现出明显的跨市场传导特征,有助于深化对金融风险形成与跨市场传导机理的理解;3)在现实意义上,本文的集成方法不仅在实证中展现出优于传统模型的解释力与预测性能,更为人工智能在金融领域的可解释化应用提供了新的思路。通过将领域知识约束融入大语言模型的语义学习过程,该方法实现了技术逻辑与经济逻辑的结合,克服了通用人工智能模型在金融场景中普遍存在的黑箱化问题。这一框架为人工智能在其他高专业度领域的应用提供了可推广的范式,也为金融监管部门利用人工智能技术识别市场情绪极化、监测系统性风险提供了理论支持和可执行路径。

1 文献综述与研究假设

1.1 新闻语调及其市场影响

早期基于新闻文本的研究主要聚焦于新闻内容所承载的事实性信息。例如,Baker等^[18]提出的经济政策不确定性(EPU)指数,用新闻报道度量宏观政策不确定性;Caldara和Iacoviello^[19]构建地缘政治风险(GPR)指数,捕捉重大事件冲击;Brandt和Gao^[20]则比较了宏观基本面新闻与地缘政治事件新闻对原油期货的影响。这类基于新闻内容的研究强调新闻如何通过传递显性信息改变市场对未来供需、政策与宏观环境的预期,进而进入资产定价模型^[13,21];

随着行为金融学的发展,学者们认识到新闻的作用不仅在于传递事实,还在于引发投资者情绪。Tetlock^[22]的开创性研究发现,基于新闻报道中的负向词频统计构建的投资者情绪指数能够显著预测股票市场回报。此后,大量研究表明,媒体报道会影响投资者的心理预期与市场情绪,从而对股票、债券、外汇乃至加密货币市场产生广泛影响^[15,23-25];这一阶段的文献大多通过构建情绪指数(sentiment index),强调投资者心理偏差在资产定价中的作用。

在此基础上,研究进一步发现新闻报道除了情绪内容外,还存在影响市场的独立维度——新闻语调(news tone)。新闻语调指新闻在表述事件时所体现的情感极性(正/负)与叙事强度(强调/弱化)^[14,26-28]。与新闻内容强调事实不同,语调不改变事件本身,但会通过投资者的有限注意力分配与风险感知偏差直接作用于市场定价。与投资者情绪不同,语调并非投资者心理反应,而是信息传递方式的差异,它是情绪的外生触发器。例如,Huang等^[27]明确提出新闻语调不会改变新闻的事实性内容,但会影响投资者对信息的感知与反应,并证明机构交易对新闻语调高度敏感;在能源与商品市场领域,已有证据显示语调相较内容具有增量预测力,能够提升对油价和期货收益的预测表现^[28]。在中文语境下,汪昌云和武佳薇^[26]则在IPO定价研究中表明,“媒体语气”并非投资者情绪本身,而是影响和引致情绪的重要渠道。姚加权等^[14]通过区分正式文本(年报语调)与非正式文本(社交媒体情绪),展示了语调与情绪在文本体裁和信息载体上的差异。

为了量化新闻语调对市场的影响,现有文献发展了多种文本分析方法,主要包括从早期使用文章长度衡量新闻信息量^[29]发展到后来的人工标注和机器学习模型^[15]。例如,部分文献采用人工方式逐篇标记新闻情感,但由于人工标记成本高且口径难以统一,该方法难以规模化应用^[24,29]。相较之下,词典法因不依赖人工标注而被广泛采用,其中以Loughran和McDonald^[10]构建的LM金融情感词典最具代表性。大量研究证实了基于LM词典的情绪指标在不同市场情境下具有良好的解释效果^[30-32]。然而,通用情感词典难以准确匹配不同行业和市场的特定术语和语境,导致情绪度量偏差。唐国豪等^[24]指出在特定领域应对现有情感词典进行改进升级,以提高情绪识别的准确性。以原油市场为例,Lucey和Ren^[28]发现,用通用金融情感词典提取的原油新闻语调无法有效预测原油期货收益。因此,针对特定领域定制情感词典或语调指标,是提高新闻语调量化准确性的有效途径。

鉴于原油期货市场作为全球最重要的大宗商品市场之一,具有高度的国际联动性和信息敏感性^[33],其价格不仅受到宏观环境与地缘事件的直

接冲击,更依赖市场新闻信息的传导^[34].已有研究表明,上海原油期货市场与国际原油市场高度相关:上海原油期货作为“人民币计价+国际平台”的市场,信息接受程度较高,被视为国际能源市场的信息净接受者^[35],国际油价的波动可对国内原油期货产生显著冲击^[36].因此,新闻语调不仅在本土市场具有定价意义,其跨市场作用同样不容忽视.现有研究多聚焦对收益的影响,而对语调在极端市场条件下的风险传导与溢出效应关注不足,本文将进一步检验其对尾部风险的影响.

1.2 基于大语言模型应用与发展的金融文本分析

随着人工智能的发展,基于 Transformer 架构的大语言模型(LLM)在金融文本分析中得到广泛应用^[1,16].大语言模型通过“通用预训练+任务微调”自动学习语言模式,在情感分析中具备更强的上下文理解与迁移能力.相比依赖静态词库的词典法,其可生成上下文相关词向量,更有效识别金融文本中的多义与隐喻^[14,16].例如,许雪晨和田侃^[2]采用预训练的 BERT 模型对中文财经新闻语调进行分析,发现提取的情绪指标对股指走势有良好的预测能力,且优于传统的 LSTM 模型预测效果.预训练语言模型 ERNIE 对上市公司年报文本的理解与分类效果明显优于基于情感词典的方法^[37].此外,大语言模型还能自动学习文本特征,避免了人工选择情绪关键词的主观偏差.这些进展表明,通用大语言模型为金融情绪量化提供了新的技术路径.

尽管具备上述优势,通用大语言模型(未经专业领域预训练的模型)在高度专业化的领域应用中仍面临挑战^[1].一方面,大语言模型的训练语料来源广泛、主题泛化,因而缺乏特定领域的专业聚焦.对于原油期货专业领域,通用大语言模型难以准确识别行业术语和复杂交易语境.例如,“backwardation(现货升水)”、“超额产能”等专业术语在通用模型中缺乏正确的语义权重,导致情绪判断失真.通用大语言模型直接用于金融场景时出现“术语误判”、“极性漂移”等问题,即模型对领域术语的情感极性判断错误或偏离真实含义.另一方面,原油期货市场具有跨市场特征和复杂逻辑,通用模型未经过领域语料训练,难以捕捉国际原油新闻与国内市场反应之间的细微关联.若大语言模型未针对原油领域进行适配,其对原

油新闻情绪的识别结果可能难以显著提升市场预测;通用模型的上下文感知能力若缺乏领域知识支撑,其情绪指标的市场解释力亦相对有限.因此,有必要检验通用大语言模型在原油新闻语调测度中的有效性.基于以上研究,本文提出:

假设 1 未经原油领域语料预训练的通用大语言模型难以准确衡量该领域新闻语调,因而不能显著提高对原油期货收益率的解释和预测能力.

针对上述局限,学术界与业界开始探索大语言模型的垂直化发展.垂直领域大语言模型是指在特定行业或领域的海量文本上进行二次预训练或训练,从而使模型更深入地掌握领域术语、语法风格和知识结构.将深度学习模型的语境理解优势与领域知识的专业优势相结合,是当前金融文本智能分析的关键课题之一.文献转向证明领域适配的有效性,例如在英文语境中,通过在金融新闻和报告上继续预训练 BERT 开发出的 FinBERT 等金融专用模型,大幅提高了金融情感分析的准确性^[38];在中文领域,有学者类似地使用财经新闻、研报等语料对通用模型进行再训练,以获取贴近金融语境的语言模型^[16].这类领域大语言模型能够更准确地识别专业词汇的语义极性,理解复杂表述,从而为下游预测提供更可靠的特征表示.

然而,仅靠深度学习模型仍然存在可解释性不足与前视偏差风险.其中,前视偏差是指当训练语料与回测样本在时间上重叠时,模型无意间“看到了未来”,从而在历史拟合中表现优异,但在真实预测中失效^[39].金融语言的动态演化(术语含义变化、叙事逻辑演进)加剧了这一问题^[40].已有研究尝试通过时间限制提示或遮蔽关键信息来缓解信息泄露,但效果有限^[41].在这一背景下,融入领域知识约束以降低前视偏差,成为金融人工智能应用亟待解决的核心问题.为兼顾情绪测度的准确性和可解释性,一种思路是将领域情感词典与大语言模型相结合,形成集成指标.词典提取的情绪得分直接、透明,能够体现重要的情感词汇和方向;而深度模型输出的情感概率融入了上下文和语义信息,具备更高的识别精度^[15].通过融合两者,有助于同步提升解释性以及预测能力.一方面,已有词典法研究为领域术语情绪极性提供了先验知识^[24],提供情绪分析的专业支撑;另

一方面,领域训练的BERT模型提供了对隐含情感的深度挖掘^[2]。这种“知识+数据驱动”融合的方法在包含大量行业术语和跨市场叙事的原油市场情绪分析中尤为适用。

本文在原油领域语料上进行预训练和微调,并与垂直领域情感词典相集成,在模型架构、算法优化和特征提取三个层面均做出针对性改进,构建新的原油新闻语调指标。与单一词典法相比,该方法通过知识约束提升了可解释性;与通用大语言模型相比,减少了前视偏差和噪声依赖,从而提高了预测稳健性。这种改进应当体现在两个方面:其一,解释力方面,领域适配模型识别的语调指标应当能够更充分地解释原油期货收益率的波动来源;其二,预测力方面,集成方法更准确捕捉新闻中的先导信号,因而在未来走势预测上表现更为稳健。因此本文提出:

假设2 经垂直领域语料预训练并以原油期货收益数据微调的大语言模型,相较于单一的词典法或未经领域适配的通用模型,能够更准确解释和预测原油期货收益率。

1.3 基于集成方法测度的新闻语调对市场收益率影响机制研究

上述分析表明,新闻语调作为独立于内容与情绪的维度,在金融市场中具有独特定价效应。然而,在新闻信息向市场定价的传导过程中,语调并非直接作用于价格,而是通过影响投资者的信息处理方式与注意力配置,进而改变交易行为和风险补偿要求。其中,投资者关注度是指投资者对于特定事件、信息或资产表现出来的兴趣和信息处理行为,是新闻情绪影响市场的重要渠道。

已有研究广泛使用搜索引擎数据、社交媒体热度等指标来捕捉投资者关注度,发现关注度对资产价格、交易量以及市场波动均有显著的预测和解释能力^[42,43]。在大宗商品市场中,Fiszeder等^[9]发现,投资者关注度的提升会明显增加原油及其关联市场(如黄金、股票)的价格波动和跨市场联动程度。新闻语调会影响投资者关注度进而传导至市场价格。在原油市场中,当国际新闻呈现明显负向语调时,受有限注意力约束下的投资者更倾向于聚焦此类高极性信息,增加搜索、讨论与交易行为,表现出更强的风险厌恶与对冲需求,从而引发原油期货收益率的显著波动。传统词典法

因缺乏语境理解,易导致极性误判或情绪稀释,使语调信号在传导过程中被削弱;而通用大语言模型虽具备上下文理解能力,但在原油等高专业性市场中易出现术语识别偏差和前视偏差,从而削弱机制检验的有效性。与之相比,本文提出的基于大语言模型与垂直领域词典的集成方法,并非单纯追求模型性能提升,而是通过引入领域知识约束,确保语调指标所反映的经济逻辑更加稳定和可解释。具体而言,集成方法在三个层面改进了语调指标的信号质量与经济一致性:其一,在模型架构上,引入垂直领域词典作为专业知识约束,确保语调识别贴合原油市场语境,减少无关信息干扰;其二,在算法优化上,通过交叉验证与样本外滚动预测有效缓解前视偏差,使语调—关注度—收益的因果链条更具时序严谨性;其三,在特征提取上,结合语义聚类与情绪密度筛选,提升了情绪信号的强度和区分度,从而持续有效地引导投资者注意力的集中与转移。

因此,本文验证集成语调指标下的经济机制是否与理论预期保持一致,提出:

假设3 新闻语调通过投资者关注度这一中介渠道影响原油期货收益率;在集成方法下,该机制的方向与显著性与经济逻辑保持一致。

1.4 新闻语调对尾部风险的影响及其跨市场溢出效应

上述研究主要关注新闻语调对收益或波动的平均效应,而在极端市场条件下,负面语调集中可能导致情绪同步化、风险偏好骤降、资产相关性上升,从而放大系统性尾部风险^[14,44]。这一现象在高杠杆、强联动的衍生品市场中尤为突出^[28,33]。由于投资者在极端情绪下往往表现出羊群行为与流动性挤兑倾向^[15,26],语调引发的情绪共振会加速价格下跌和风险外溢,使市场由个体性冲击演变为系统性波动。其中,姜富伟等^[16]引入市场换手率、平滑波动率等变量测度噪音交易者带来的非理性波动和市场风险的水平的影响,并分析了大模型情绪变量的极端事件预测能力,但是依然聚焦于是否助力于预测各种极端事件期间市场回报变化。针对大宗商品尤其是原油期货的跨市场尾部风险传导研究仍然有限,更没有基于大语言模型检验人工智能时代相关跨市场风险的表现与影响因素。

在此背景下,本文从新闻语调角度出发,关注其在极端市场情境下的风险传导效应,将 CoVaR 框架引入新闻语调研究,旨在从尾部风险视角识别国际新闻语调对上海原油期货市场的风险溢出效应。相较于传统的方差或波动指标,CoVaR 能够度量在极端负向情绪冲击下市场间的条件风险依赖关系,从而捕捉语调信号对系统性风险的边际贡献。与围绕收益率的检验不同,尾部风险分析更能揭示情绪极化下的系统性耦合机制,为理解金融市场尾部联动提供了新的文本化解释路径。

综上,本文的风险主线旨在回答两个关键问题:其一,国际新闻语调是否在极端市场条件下对上海原油期货的系统性风险具有显著溢出效应;其二,这种溢出效应是否表现出方向性与动态性特征。基于此,本文提出以下假设:

假设4 在极端市场条件下,国际新闻语调对上海原油期货具有显著的尾部风险溢出效应。

2 基于大语言模型与垂直领域词典的原油新闻语调集成方法

2.1 原油新闻样本集

对于新闻文本的获取通常选择具有多种新闻源的新闻文本数据库,以便获得更加丰富的语料,对于同一事件,不同媒体也可以提供更具有差异性语调的报道文本。本文采用中国资讯行数据库(InfoBank 数据库)和 Factiva 道琼斯新闻数据库获得原油新闻文本。具体地,选取2018年3月—2022年12月窗口期,英文语料全部来自 Factiva 道琼斯新闻数据库,选择英文“Crude Oil Market”新闻主题共计选择67 949条新闻。对于中文语料,在 InfoBank 数据库中选择“中国经济新闻库”,并搜索关键词“原油”来新闻链接、标题、标签以及发布日期等信息,共得到20 107条数据;其中,为了防止噪音,剔除标签中含有“石化”和“燃料”的新闻,再根据获取的新闻链接,得到新闻的正文内容,共包含了12 379条新闻文本。同时,在 Factiva 道琼斯新闻数据库选择“原油市场”这一新闻主题以同样获得新闻链接、标题、来源以及发布日期,共包含了15 406条新闻文本。考虑到不同数据库提供的新闻可能存在的重复问题,

本文剔除了标题完全相同的新闻,获得25 055条中文原油新闻文本。综上,本文样本共包括93 004篇中英文原油新闻文本。

2.2 集成式原油新闻语调构建

现有金融文本分析研究^[9,14,15]主要表现出两方面问题:其一,金融行业的高度专业化正在推动大语言模型由通用化向垂直化演进。通用大语言模型虽具备强大的语言理解能力,但面对原油市场特有的专业术语、交易逻辑与事件语境时,往往难以给出细粒度且一致性的情绪判别;其二,已有中文金融大语言模型^[15]主要依赖微博财经新闻等宽口径语料进行进一步预训练,虽然提升了对一般金融文本的适应性,但对原油期货这一更窄、更国际化的子领域仍显“口径过宽”。基于此,为充分挖掘原油新闻文本所蕴含的情绪信息,本文提出了一种基于大语言模型与垂直领域词典的原油新闻语调集成测度方法。

2.2.1 基于词典知识约束的原油垂直领域语料

初始新闻样本主要依据资讯供应商对“原油板块”或“石油能源”标签的粗粒度归档一次性抓取而来,这种“全量收集”策略虽然保证了覆盖面,却不可避免地引入了两类噪声:其一,大量宏观经济报道、企业公告乃至广告软文仅在标题或正文边角处提及“原油”字样,却并未包含任何与原油供需、价格预期相关的情绪信息;其二,相当部分快讯类稿件(如库存数据播报、航运日程更新)以纯客观陈述为主,缺乏可量化的情绪极性。若将这两类文本直接用于语调建模,不仅会稀释情绪信号、拉低极性分布的方差,导致模型在预训练阶段过度学习无情绪特征的中性模式,还会在微调阶段放大类别不平衡,最终削弱对收益率波动的解释力。因此,本文在检验国际新闻语调对上海原油期货市场的跨市场影响过程中,首先参照 Loughran 等^[17]的英文原油垂直领域词典构建办法,在第一轮词典标注后先行剔除“零情绪”或“弱情绪”文本,再借助语义聚类与人工校验去除主题游离于原油基本面的边缘新闻,并同步将残留高频新词迭代补充进词典。如此既压缩了无效样本规模,也确保语料在主题聚焦、情绪表达与极性平衡三方面均服务于后续大语言模型的领域学习需求,提高了预训练的信噪比与情绪判别精度。

在剔除无效聚类样本后,本文对保留语料进

一步执行词典反向扩展. 具体方法为: 统计当前语料中词典未覆盖的高频词与短语, 挖掘其与原油市场核心概念的情绪关联性. 在此基础上引入新词条并人工审核后, 将其纳入更新后的领域词典. 随后对保留语料执行第二轮情绪标注. 基于词典知识约束的原油垂直领域语料迭代优化不仅有效剔除了语义偏离领域主题或缺乏情绪表达的低质量文本, 同时增强了剩余语料在领域相关性与情绪极性分布方面的判别性.

2.2.2 使用垂直领域语料对 BERT 模型预训练与微调

鉴于词典法在语调分析中的可迁移性有限, 难以充分适应不同语境和复杂文本结构, 本文进一步引入大语言模型对原油新闻语调进行重构分析. 借助其在上下文理解与语义捕捉方面的优势, 大语言模型能够更准确地识别原油新闻中的潜在语调特征, 从而提升语调识别的广泛适应性与预测能力. 目前主流的 Transformer 模型的核心创新在于其自注意力机制, 该机制能够有效捕捉输入序列中任意两个位置之间的依赖关系, 突破了传统序列模型在处理长距离依赖时的限制. 此外, 注意力机制使模型能够并行处理输入序列的不同部分, 显著提升了计算效率^[1]. 本文使用基于 Transformer 的代表性模型 BERT 对原油新闻进行文本

语调分类.

原油新闻等高度专业化的领域文本通常包含大量行业术语、隐含逻辑以及特定的表达方式, 原始 BERT 模型难以有效捕捉其中隐含的语调信息与语义细节. 因此, 若缺乏针对性垂直领域预训练, 模型很难建立起对该类文本的深层理解, 进而影响其在语调识别与原油期货市场预测任务中的表现. 为了检验专业领域语料的预训练是否对模型识别的精确度产生显著的影响, 本文使用原油新闻语料对 BERT 模型进行预训练和微调. 具体做法如下: 首先使用新闻语调对模型进行预训练, 鉴于从头开始的预训练过程需要极高的资源和成本, 通常使用“进一步训练”的方式进行^[16]. 随机抽取已经获取的新闻样本作为语料执行遮掩语言模型 (masked language model, MLM) 和下文预测任务 (next sentence prediction, NSP) 两个预训练任务. MLM 通过遮蔽特定词汇, 使模型依据上下文推断被遮蔽内容. NSP 则通过判断两个句子的先后顺序, 考察模型对文本连贯性的理解. 这两个预训练任务使 BERT 模型能够有效捕捉文本的上下文关系和动态特征, 从而更准确地把握语义信息和句法结构. 相较于传统词典法模型, 这种基于上下文的动态学习机制显著提升了模型的语义理解能力. 图 1 展示了使用语料进行预训练的具体过程.

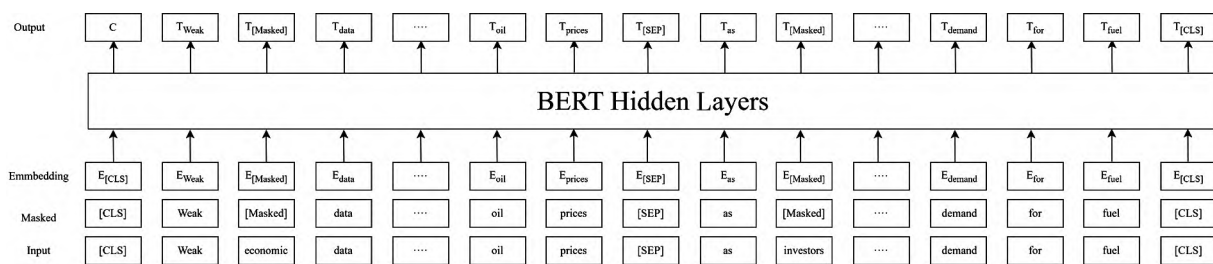


图 1 BERT 预训练过程

Fig. 1 BERT pre-training process

注: 在第一步输入语料, 以 [CLS] 作为句子 A 和句子 B 的开头, 以 [SEP] 为两句话的分隔, 最后以 [SEP] 作为两句话的结尾, 第二步将其中的部分词元 (economic、investors) 遮掩, 进行嵌入操作后对 BERT 模型进行预训练. 模型的输出即是预测 $T_{[Mask]}$. 而 C 的输出向量代表了 [SEP] 前后两句话的上下文关系.

其次, 参考姜富伟等^[16]的做法, 使用当日的市场回报对当日的新闻文本进行标记, 作为语调的标签, 当市场回报为正时标记为 1, 否则标记为 0. 并用带标签的语料对经过上述预训练的 BERT 模型进行微调. 通过使用带标签的语料对模型进行微调可以进一步优化上一步骤中产生的模型参数, 提高模型对于特定分类任务的适用性, 得到更

加准确的文本分类结果, 而在标记文本时使用市场回报可以规避人工标记造成的主观偏差. 具体过程如图 2 所示. 最后, 将前述步骤中未被使用过的语料代入微调过的 BERT 模型进行情绪判断, 使用新数据的目的是避免过拟合问题和可能造成的估计偏差^[16], 并且考虑到与词典法的整篇匹配不同, 由于 BERT 模型基于注意力机制, 避免句子

超过 BERT 模型最大输入长度(默认 512 Tokens)并确保模型能充分利用上下文,而不因截断导致信息丢失^②。

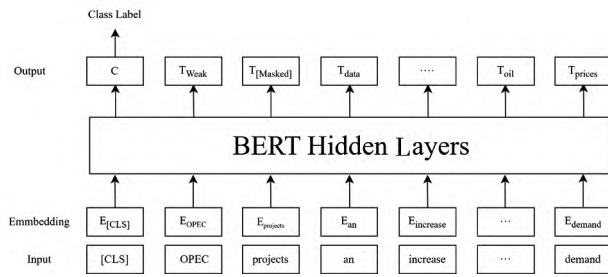


图2 BERT 模型微调过程
Fig. 2 BERT fine-tuning process

注:对于文本分类的微调任务,与预训练过程类似,但此时输入为单个文本,输出为分类结果(Class Label)

每个输入语调若为积极取 1,若为消极取 0.取均值后得到第 t 日中第 i 篇新闻文本语调 $tone_{it}$.进一步根据式(1)计算当日新闻语调

$$NewsToneBert_t = \frac{\sum_{i=1}^{N_t} tone_i}{N_t} \quad (1)$$

其中 N_t 为第 t 天的新闻总数.

因此,如图 2 所示,本文以原油期货市场为应用场景,通过“专用词典—语料精炼”形成垂直领域语料,并应用“领域语料预训练—收益率微调”共四步递进的流程,构建面向原油新闻的集成式语调指标,以期兼顾模型的上下文理解能力与词典的专业知识约束.

2.2.3 对比实验

总体上,与现有单一词典法或通用大语言模型相比,本研究提出的集成方法具有三方面创新:一是模型架构创新,即在深度学习框架中引入领域词典作为知识约束;二是算法优化创新,即在预训练—微调流程中设置交叉验证与稳健性检验,减少前视偏差;三是特征提取创新,即通过词典先验筛选与语义聚类,提高语料的信噪比与极性分布,确保集成方法在原油市场这一高专业性场景中的适用性.

为了检验本文提出的基于垂直领域词典与大

语言模型的原油新闻语调集成构建方法,对比另外三种原油新闻语调(基于垂直领域词典的新闻语调、基于通用语料的大语言模型构建的原油新闻语调、基于全部原油新闻语料的大语言模型构建的新闻语调)对原油期货收益率的影响.

对于采用了大语言模型训练的构建方法,为保证模型结果的稳健性,本文在训练过程中采用了 10 折交叉验证,并将数据划分为训练集、验证集和测试集:验证集用于调节超参数,测试集用于最终评估.以 F1 分数作为主要评估指标,同时参考准确率和精确率,以平衡模型的精度与召回表现.此外,本文引入了早停机制,当验证集的 F1 分数在连续两个训练轮次内无明显提升时,便停止训练,以避免过拟合和计算浪费^③.

另外,基于词典法的原油新闻语调构建参考姚加权等^[14]、Loughran 等^[17]的方法.具体地,在单篇中文新闻的语调分析中,本文采用以下处理流程:首先对文本进行分词处理,提取情绪(positive、negative 或 neutral)维度词汇的前后最多四个词语构成分析单元^④.对于 positive 和 negative 维度词语构成的单元,分别赋予初始语调值 1 和 -1.针对 neutral 维度词语构成的单元,若存在 positive modifier 或 negative modifier 维度词语,则根据 modifier 词语的方向与 neutral 词语的初始方向计算语调;若无 modifier 词语,则忽略该单元.这种分层处理方法有效提升了语调分析的精确性和可靠性.并且与 Lucey 和 Ren^[28]一致,本文不考虑存在多个修饰语的情况.接着计算单元中否定词的个数 n ,并按照 $(-1)^n$ 调整该单元的语调.最后,将每个单元的语调得到单个句子的语调,若语调大于 0,该句语调为正,计为 1,使得原油价格上升;反之,该句语调为负,计为 -1,使得原油价格下降.将一篇新闻中使得原油价格上升的句子数量减去原油价格下降的句子数量,并除以全文的词语数量以剔除极性,则为该篇新闻的语调.根据每条新闻的发布时间,加总每一天内新闻的语调,此外,同样考虑非交易日中新闻报道的影响,即可

② 在文本分类模型的训练中,采用 AdamW 优化器.初始学习率设置为 2×10^{-6} ,并配合逐步衰减策略,以避免过快收敛造成的不稳定.每批次输入 32 条样本,在整个数据集上反复迭代 5 轮训练.损失函数使用交叉熵,最终在验证集上模型性能在第 5 轮后趋于稳定.

③ 该验证过程省略,需要者可向作者索要.

④ 若前后四个词语中存在同类维度词语,则缩小提取范围直至排除同类词语.

得到月度原油新闻语调指标。

3 新闻语调与原油期货收益

3.1 原油期货数据样本与变量构建

原油期货较于现货具有更好的流动性^[17], 投资者和金融机构普遍将原油期货价格视作对原油现货价格的最佳预测。本文对应原油新闻样本时间窗口选择 2018 年 3 月 26 日—2022 年 12 月 31 日期间共 1 161 个交易日, 选择各个原油期货

的主力连续合约作为研究对象, 使用收盘价的自然对数差分作为原油期货收益率的代理变量。数据来源为 Wind 数据库。

根据现有文献对原油期货其他影响因素的梳理^[45], 本文选取贸易加权人民币指数作为汇率的代理变量; 使用上海金交所黄金现货收盘价作为黄金市场情况的代理变量; 采用十年期国债收益率描述市场宏观经济因素。对上述变量, 均取指数或者价格序列的自然对数差分。所有变量说明及其描述性统计如表 1 及表 2 所示。

表 1 变量说明

Table 1 Descriptions of variables

变量含义	变量符号	数据来源
原油期货收益率	<i>Return</i>	基于 SC 连续合约收盘价计算的原油期货收益率
原油新闻语调 NewsTone	<i>Dict-NT</i>	使用词典法构建的新闻语调
	<i>BERT-NT</i>	使用原始 BERT 模型识别的新闻语调
	<i>FT-BERT-NT</i>	全部原油新闻适配的 BERT 模型识别的新闻语调
	<i>VT-BERT-NT</i>	垂直领域语料适配的 BERT 模型识别的新闻语调
汇率指数变化率	<i>Exchange</i>	根据贸易加权人民币指数计算
黄金收益率	<i>Gold</i>	根据上海金交所黄金现货收盘价计算
国债收益率差分	<i>Treasury</i>	十年期国债收益率取对数差分

表 2 描述性统计

Table 2 Statistical summary

变量	观测数	Mean	SD	Min	Max	Skewness	Kurtosis
<i>Return</i>	1 160	0.000 2	0.025	-0.141 3	0.101 1	-0.210 4	5.665 1
<i>BERT-NT</i>	1 160	-0.095 5	0.123	-0.863 6	0.342 2	-0.792 4	5.888 1
<i>FT-BERT-NT</i>	1 160	0.026 5	0.025	0.000 0	0.200 0	1.344 8	6.701 9
<i>VT-BERT-NT</i>	1 160	0.469 2	0.172	0.062 5	0.952 4	0.067 7	2.448 1
<i>Dict-NT</i>	1 160	0.602 0	0.206	0.034 0	1.000 0	-0.403 8	2.369 2
<i>Gold</i>	1 160	0.000 3	0.008	-0.061 9	0.050 3	-0.377 8	9.386 1
<i>Treasury</i>	1 160	-0.000 8	0.025	-0.144 9	0.113 2	-0.398 0	6.894 6
<i>Exchange</i>	1 160	-0.000 0	0.004	-0.079 2	0.074 7	-1.053 7	198.046 5

3.2 国际原油新闻语调对国内原油期货收益的跨市场影响

在人民币国际化的战略背景下, 上海原油期货作为我国首个上市交易的原油期货合约, 采用了与国际市场接轨的日夜盘交易模式。而通过交易模式的设定, 国内原油期货市场与国际能源交易市场产生了更为密切的关联^[33]。因此, 本文重点分析国际市场中的新闻语调是否能够影响国内市场的收益, 讨论跨市场的新闻语调、投资者关注度与原油期货收益率之间的关系。

不同模型度量的新闻语调对收益率的影响结

果如表 3 和表 4 所示。表 3 报告了基于词典法 (Dict-NT) 与通用大语言模型 (BERT-NT) 的结果。可以看到, 词典法在滞后一阶的系数显著为正 (0.022, 均在 1% 水平显著), 而通用大语言模型在各期均未呈现显著性。这一发现表明, 国际原油新闻的语调确实能够通过跨市场传导影响国内原油期货的次期收益, 从而验证了假设 1。然而, 通用大语言模型未能有效捕捉这一效应, 其原因在于通用模型在能源领域术语识别与极性判别方面存在偏差。例如, “钻井平台增加” “减产协议” 等高频术语在缺乏领域适配时容易被误判, 导致模

型提取的信号被削弱.此外,跨市场传导往往通过夜盘交易窗口才能在次期收益中显现,词典法依赖极性词计数在这一时点表现更为稳健,而通用

大语言模型因过度依赖语境解释而稀释了滞后效应.因此,在缺乏领域约束的情境下,词典法在捕捉国际新闻滞后效应方面反而更具解释力.

表 3 不同模型度量的跨市场新闻语调对收益率的影响

Table 3 The impact of cross-market news tone on returns measured by different models

变量	(1)	(2)	(3)	(4)
	基于 <i>Dict-NT</i>	基于 <i>Dict-NT</i>	基于 <i>BERT-NT</i>	基于 <i>BERT-NT</i>
<i>NewsTone</i>	0.016 *** (3.352)	0.015 *** (3.451)	0.018 (0.936)	0.017 (0.885)
<i>L. NewsTone</i>	0.022 ** (2.842)	0.022 ** (2.819)	-0.034 (-1.182)	-0.035 (-1.254)
<i>L2. NewsTone</i>	-0.007 (-1.039)	-0.008 (-1.048)	0.026 (1.261)	0.028 (1.330)
<i>L. Gold</i>		0.049 (0.657)		0.050 (0.710)
<i>L. Treasury</i>		0.041 (1.374)		0.042 (1.348)
<i>L. Exchange</i>		0.223 (1.172)		0.250 (1.386)
<i>_cons</i>	0.003 *** (4.007)	0.003 *** (3.998)	0.000 (0.018)	0.000 (0.034)
<i>Year FE</i>	Yes	Yes	Yes	Yes
<i>Month FE</i>	Yes	Yes	Yes	Yes
<i>Week FE</i>	Yes	Yes	Yes	Yes
<i>N</i>	1 158	1 158	1 158	1 158
<i>R²</i>	0.026	0.030	0.011	0.015

注: 变量前 *L* 表示滞后, 数字表示滞后期; ***、**、* 分别代表 1%、5% 和 10% 的显著性水平. 在表格中所有列的被解释变量均为原油期货收益率 *Return*, 对于解释变量 *NewsTone*, 列(1)与列(2)使用由词典法构建的新闻语调 *Dict-NT* 对收益率回归, 列(3)与列(4)代表使用通用 BERT 模型构建的新闻语调 *BERT-NT* 对收益率回归.

表 4 全量原油新闻语料训练的大语言模型与垂直领域语料训练大语言模型语调的回归结果

Table 4 Regression results of large language models trained on full-scale oil news corpora and those trained on domain-specific corpora

变量	(1)	(2)	(3)	(4)
	基于 <i>FT-BERT-NT</i>	基于 <i>FT-BERT-NT</i>	基于 <i>VT-BERT-NT</i>	基于 <i>VT-BERT-NT</i>
<i>NewsTone</i>	0.030 *** (5.463)	0.030 *** (5.374)	0.061 *** (10.987)	0.061 *** (11.132)
<i>L. NewsTone</i>	0.050 *** (9.229)	0.050 *** (9.243)	0.018 *** (4.857)	0.018 *** (4.826)
<i>L2. NewsTone</i>	-0.014 ** (-2.505)	-0.014 ** (-2.496)	-0.023 *** (-6.462)	-0.023 *** (-6.546)
<i>L. Gold</i>		0.065 (1.121)		-0.026 (-0.390)
<i>L. Treasury</i>		0.032 (1.160)		-0.001 (-0.026)
<i>L. Exchange</i>		0.116 (0.987)		0.015 (0.112)

续表 4

Table 4 Continues

变量	(1)	(2)	(3)	(4)
	基于 <i>FT-BERT-NT</i>	基于 <i>FT-BERT-NT</i>	基于 <i>VT-BERT-NT</i>	基于 <i>VT-BERT-NT</i>
<i>_cons</i>	-0.031 *** (-7.149)	-0.031 *** (-7.071)	-0.034 *** (-11.212)	-0.034 *** (-10.895)
<i>Year FE</i>	Yes	Yes	Yes	Yes
<i>Month FE</i>	Yes	Yes	Yes	Yes
<i>Week FE</i>	Yes	Yes	Yes	Yes
<i>N</i>	1 158	1 158	1 158	1 158
<i>R²</i>	0.186	0.187	0.284	0.284

注: 变量前 *L* 表示滞后, 数字表示滞后期; ***, **, * 分别代表 1%、5% 和 10% 的显著性水平。表格中所有列的被解释变量均为原油期货收益率 *Return*, 对于解释变量 *NewsTone*, 列(1) 与列(2) 使用由全量原油新闻预训练的 BERT 构建的新闻语调 *FT-BERT-NT* 对收益率回归, 列(3) 与列(4) 代表使用垂直领域语料预训练的 BERT 构建的的新闻语调 *VT-BERT-NT* 对收益率回归。

表 4 进一步比较了全量新闻语料预训练模型 (*FT-BERT-NT*) 与垂直语料预训练模型 (*VT-BERT-NT*) 的差异。结果显示, 两类模型识别的语调变量在当期和滞后一阶均显著为正, 表明大语言模型在提取跨市场情绪信号时具有一致性。然而, 两者在解释力上存在显著差异: *FT-BERT-NT* 的 *R²* 仅为 0.186, 而 *VT-BERT-NT* 的 *R²* 提升至 0.284。这一结果验证了假设 2, 即经垂直领域语料预训练并结合收益符号微调的大语言模型, 能够更为准确和稳健地解释及预测原油期货收益率。机制上看, 垂直语料训练显著提升了模型的信噪比, 剔除了与原油市场关联度较低的泛金融新闻, 使模型学习到的特征更集中于价格驱动因素。同时, 词典知识约束与收益微调提供了方向性校准, 使模型在语义表征与市场反应之间建立了更紧密的映射关系, 降低了极性漂移的风险。更重要

的是, 领域特化语料增强了模型对供需冲击类叙事的识别能力, 从而提高了跨市场信息对齐的效率。

为进一步验证实证结果的稳健性, 本文从交易时段与模型参数设定两个角度开展扩展检验。一方面, 考虑到国际新闻的发布时间与传播通常集中在欧美交易时段, 其影响更可能通过夜盘交易传导至国内原油期货市场, 因此选取夜盘样本 (*NightSession*) 有助于评估新闻语调在不同交易时段下的适用性与一致性。另一方面, 在方法层面, 为避免结论受到特定技术设定的质疑, 本文将大语言模型的最大输入长度由默认的 512 Tokens 调整为 256 Tokens (*ShortTokens*), 以检验模型在文本截断情境下的表现。该检验能够揭示模型输出是否依赖于特定的参数选择, 从而提高结论的可信度与普适性。检验结果如表 5 所示。

表 5 更换样本和训练参数的稳健性检验

Table 5 Robustness test of replacement samples and training parameters

变量	(1)	(2)	(3)	(4)
	<i>NightSession</i>	<i>NightSession</i>	<i>ShortTokens</i> BERT	<i>ShortTokens</i> BERT
<i>VT-BERT-NT</i>	0.022 *** (7.838)	0.023 *** (7.849)	0.074 *** (10.823)	0.074 *** (17.738)
<i>L. VT-BERT-NT</i>	0.013 *** (2.774)	0.013 *** (2.788)	0.010 ** (2.409)	0.010 ** (1.958)
<i>L2. VT-BERT-NT</i>	-0.013 *** (-3.573)	-0.013 *** (-3.537)	-0.025 *** (-5.735)	-0.025 *** (-5.323)
<i>L. Gold</i>		0.070 (1.128)		0.010 (0.123)

续表 5

Table 5 Continues

变量	(1)	(2)	(3)	(4)
	NightSession	NightSession	ShortTokens BERT	ShortTokens BERT
<i>L. Treasury</i>		-0.027 (-1.446)		0.000 (0.003)
<i>L. Exchange</i>		-0.107 (-0.747)		0.023 (0.185)
<i>_cons</i>	-0.014 *** (-5.443)	-0.015 *** (-5.725)	-0.037 *** (-10.426)	-0.037 *** (-10.061)
<i>Year FE</i>	Yes	Yes	Yes	Yes
<i>Month FE</i>	Yes	Yes	Yes	Yes
<i>Week FE</i>	Yes	Yes	Yes	Yes
<i>N</i>	863	863	1 158	1 158
<i>R²</i>	0.125	0.129	0.277	0.277

注：变量前 *L* 表示滞后，数字表示滞后期；***、**、* 分别代表 1%、5% 和 10% 的显著性水平。

表 5 报告了在不同样本区间与训练参数设定下的稳健性检验结果。首先，基于表 5 列(1)和列(2)的夜盘样本，新闻语调的滞后一期变量在所有模型中均在 1% 的显著性水平下为正，显示出国际新闻语调在境外投资者主导的夜盘交易中对价格波动的持续驱动作用。其次，如表 5 列(3)和列(4)所示，在将大语言模型的最大输入长度由 512 Tokens 缩短至 256 Tokens 的情况下，当期语调变量的显著性略有下降，但滞后一期的语调依然高度显著，且系数方向与规模保持稳定，表明新闻语调对市场回报的预测效应并非由样本选择或技术设定所驱动。值得注意的是，Token 长度缩短导致回归模型整体 R^2 较基准设定有所下降，这说明输入文本长度直接影响模型对复杂语义关系的捕捉能力，从而间接影响语调指标的解释力。总体来看，集成方法在不同市场环境与技术设定下均表现出良好的稳定性与可迁移性。

该稳健性检验从实证上强化了集成方法的可靠性，并揭示出两个层面的理论与实践意义。一方面，夜盘样本的结果显示，国际新闻语调在境内外投资者信息交互与交易时段差异中，仍然能够通过跨市场风险传导显著作用于国内原油期货收益，进一步印证了新闻语调对跨市场原油期货定价的重要影响。另一方面，大语言模型在处理金融新闻时需要兼顾文本覆盖度与建模效率，较长的 Token 长度能够更全面地捕捉跨市场叙事与专业

术语，从而提升语调指标的解释力。这不仅证明了集成方法在不同实验设定下的稳健性，也提示未来研究与实践应重视模型参数选择对金融语义挖掘和风险预测的影响。

3.3 中文新闻语境下新闻语调对上海原油期货收益的影响

为考察实证结果的稳健性与广泛适用性，本文基于中文原油期货的垂直领域词典检验集成方法在中文新闻语境下对上海原油期货收益的影响。

参考 Loughran 等^[17] 构建英文原油垂直领域词典的方法，首先构建中文原油垂直领域词典，进而通过“专用词典—语料精炼”形成垂直领域语料，并应用“领域语料预训练—收益率微调”流程，最终获得获得中文原油新闻的集成语调指标，以此检验该方法的可复制性。具体如下：

首先，翻译 Loughran 和 McDonald^[10]、Loughran 等^[17] 提出的 LM 英文金融情感词典为母库，并在此基础上结合中文词典，包括姚加权等^[14] 和姜富伟等^[15] 构建的金融情绪词典、清华大学李军中文褒贬义词典、知网 HowNet 情感词典和台湾大学 NTUSD 简体中文情感词等^[44]，初步获得适用于中文语境的情感词条，借助词频统计与共现网络分析，对国际及国内原油新闻语料中的高频词、共现词进行筛选，并结合专家复核，迭代生成涵盖正向、负向、中性及其修饰词的“原油新闻语调词典”。该词典不仅保留了“钻井平台”“套期保值”

等行业专属词汇,还对“需求增加”“库存下降”等在原油定价中方向效应相反的同义表达进行了极性重定向,显著提高了词典的领域贴合度。其次,选取余弦相似度在 80% 以上的词语,并人工筛选

确定其是否属于原油语调词典以及相应的维度;最后,使用 Word2Vec 模型,以原油新闻样本集作为语料库,使用 Skip-gram 模式进行训练,得到一个词向量矩阵。具体如图 3 所示^⑤。

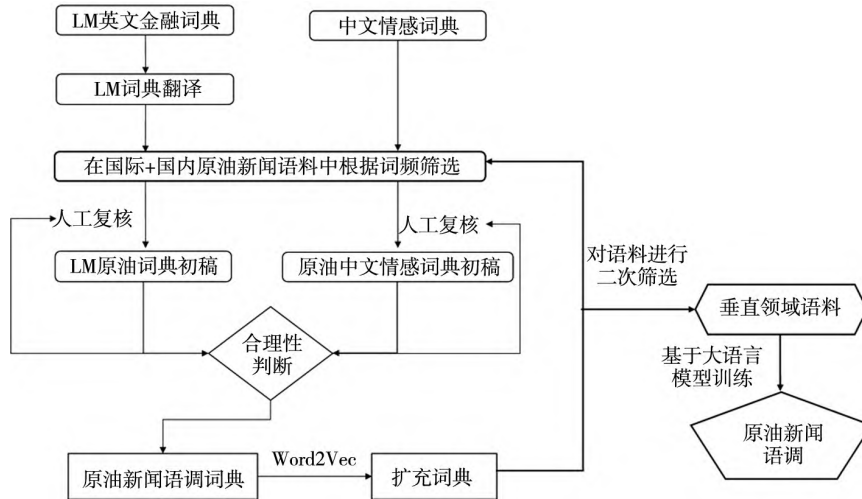


图 3 集成大语言模型与中文垂直领域词典的原油新闻语调构建

Fig. 3 Flow chart of Chinese crude oil tone construction based on the combination approach

如表 6 所示,中文新闻语调的回归结果与前文基于英文新闻的结论保持高度一致,进一步验证了本文方法的跨语种稳健性。具体地,在词典法(表 6 的列(1)与列(2))下,滞后两期的新闻语调

表 6 不同方法测度的国内新闻语调对收益率的影响

Table 6 The impact of Chinese news tone on returns measured by different models

变量	(1)	(2)	(3)	(4)
	基于 Dict-NT	基于 Dict-NT	基于 BERT-NT	基于 BERT-NT
NewsTone	-0.013 (-0.328)	-0.011 (-0.286)	-0.005 (-1.523)	-0.005 (-1.363)
L. NewsTone	-0.017 (-0.404)	-0.021 (-0.485)	-0.011** (-2.485)	-0.011** (-2.520)
L2. NewsTone	-0.090** (-2.124)	-0.086** (-1.995)	0.001 (0.164)	0.001 (0.204)
L. Gold		0.056 (0.565)		0.040 (0.544)
L. Treasury		0.038 (1.045)		0.043 (1.365)
L. Exchange		0.224 (1.410)		0.241 (1.331)
_cons	0.001 (1.274)	0.001 (1.275)	0.006** (2.485)	0.006** (2.572)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
N	1 158	1 158	1 156	1 156
R ²	0.018	0.021	0.016	0.019

注: 变量前 L 表示滞后, 数字表示滞后期; **、* 分别代表 1%、5% 和 10% 的显著性水平。在表格中所有列的被解释变量均为原油期货收益率 Return。对于解释变量 NewsTone 列(1)与列(2)使用由词典法构建的新闻语调 Dict-NT 对收益率回归, 列(3)与列(4)代表使用通用 BERT 模型构建的新闻语调 BERT-NT 对收益率回归。

⑤ 中文词典及相关统计省略, 需要者可向作者索要。

系数显著为负(-0.090 与 -0.086 ,分别在 5% 和 10% 水平显著) ,表明投资者对新闻报道存在一定的过度反应: 前期乐观语调的累积效应在随后两个交易日转化为价格下跌 符合行为金融学中“过度自信—反转效应”的解释. 而通用大语言模型(表 6 的列(3) 与列(4)) 在当期未能显著解释收益率, 但在滞后一阶出现了稳定的负向作用(系数约为 -0.011 ,且在 5% 水平显著) ,显示出其在捕捉短期延迟反应时具备一定能力. 这与英文结果中“通用模型相对滞后、解释力有限”的表现相吻合.

值得注意的是, 无论是基于词典法还是通用大语言模型, 其构建的新闻语调对收益率的解释

力整体较低(R^2 在 0.016 ~ 0.021 之间) ,明显低于表 7 中基于领域预训练的大语言模型结果. 这说明, 仅依靠词典计数或未经领域适配的通用模型, 难以充分揭示国内新闻语调与期货收益之间的系统性关系. 结合前文的跨市场检验可以得出结论: 尽管词典法在方向性识别上表现稳健, 通用大语言模型在局部滞后效应上具备一定解释力, 但只有引入领域预训练与收益微调的集成方法, 才能显著提升语调指标的解释力与预测力. 因此, 该部分稳健性检验不仅验证了集成方法的跨语种可迁移性, 也进一步印证了假设 1 与假设 2 的结论.

表 7 中文全量原油新闻语料训练与垂直领域语料训练大语言模型语调的回归结果

Table 7 Regression results of large language models trained on Chinese full-scale oil news corpora and those trained on domain-specific corpora

变量	(1)	(2)	(3)	(4)
	基于 FT-BERT-NT	基于 FT-BERT-NT	基于 VT-BERT-NT	基于 VT-BERT-NT
<i>NewsTone</i>	0.059 *** (10.100)	0.059 *** (10.061)	0.059 *** (11.310)	0.059 *** (11.269)
<i>L. NewsTone</i>	-0.005 (-1.109)	-0.005 (-1.179)	-0.007 *** (-3.555)	-0.007 *** (-3.625)
<i>L2. NewsTone</i>	-0.006* (-2.176)	-0.006* (-2.186)	-0.001 (-0.408)	-0.001 (-0.421)
<i>L. Gold</i>		0.056 (0.923)		0.077 (1.293)
<i>L. Treasury</i>		0.033 (1.398)		0.036 (1.618)
<i>L. Exchange</i>		0.116 (0.678)		0.136 (0.758)
<i>_cons</i>	-0.017 *** (-6.679)	-0.017 *** (-6.656)	-0.022 *** (-6.966)	-0.021 *** (-7.118)
<i>Year FE</i>	Yes	Yes	Yes	Yes
<i>Month FE</i>	Yes	Yes	Yes	Yes
<i>Week FE</i>	Yes	Yes	Yes	Yes
<i>N</i>	1 155	1 154	1 156	1 156
R^2	0.230	0.232	0.331	0.333

注: 变量前 *L* 表示滞后, 数字表示滞后期; ***, **, * 分别代表 1%、5% 和 10% 的显著性水平. 表格中所有列的被解释变量均为原油期货收益率 *Return*, 对于解释变量 *NewsTone*, 列(1) 与列(2) 使用由全量原油新闻预训练的 BERT 构建的新闻语调 FT-BERT-NT 对收益率回归, 列(3) 与列(4) 代表使用垂直领域语料预训练的 BERT 构建的新闻语调 VT-BERT-NT 对收益率回归.

3.4 样本外预测检验

本文进一步进行样本外预测分析. 具体而言, 使用 2018 年—2022 年新闻文本训练得到的大语言模型, 对 2023 年—2024 年的原油新闻进行预测检验. 其中, 模型(2) 与模型(4) 使用新样本单独进行回归, 但由于控制变量(如人民币实际有效汇率) 仅更新至 2024 年 3 月, 为避免样本量不

足引致的偏误, 模型(1) 与模型(3) 将全部样本合并进行回归.

在实证设计上, 本文不仅进行常规的样本外回归, 还参考姜富伟等^[15]的做法, 构建样本外预测 R 方(R_{out}^2). 具体地, 将样本外的样本逐次加入回归, 并产生新的拟合值. 设 P 为设定的初始训练区间大小, 则本文需要对时刻 $t = P + 1, P + 2, \dots$,

T 的回报进行样本外预测, 从而共有 $q = T - P$ 个样本外预测值. 初始训练区间内数据用于估计第一个预测性回归方程参数, 并得到第一个样本外预测值. 此后, 随时间推移, 不断增添新的训练样本, 以滚动回归的方式得到新的回归方程参数, 并以此计算对应的样本外预测值. R_{out}^2 表示如下

$$R_{out}^2 = 1 - \frac{\sum_{t=P+1}^T (R_t - \widehat{R}_t)^2}{\sum_{t=P+1}^T (R_t - \bar{R}_t)^2} \quad (2)$$

其中 \widehat{R}_t 为预测性回归方程得到的未来回报预测, \bar{R}_t 代表历史均值预测基准, 即将股票回报的历史

均值当作回报预测值, 计算方法如下

$$\bar{R}_t = \frac{1}{t-1} \sum_{s=1}^{t-1} R_s \quad (3)$$

表 8 的结果显示, 无论是基于全量语料还是垂直领域语料训练的大语言模型, 其识别的新闻语调在样本外回归中均对未来一期收益率保持显著预测力, 表明预测效果并非样本内拟合的虚假结果. 进一步比较发现, 垂直语料训练的大语言模型在解释力 (R^2) 与样本外预测 R_{out}^2 统计量上均优于全量语料训练的模型. 根据姜富伟等^[15] 的标准, 只要 R_{out}^2 大于 0, 即可说明预测指标超越了历史均值基准, 具备现实应用价值. 显然, 集成方法在样本外预测中表现更优越.

表 8 全量原油新闻语料训练的大语言模型与垂直领域语料训练大语言模型语调的样本外预测结果

Table 8 Out-of-sample prediction results of large language models trained on full-scale oil news corpora and those trained on domain-specific Corpora

变量	(1)	(2)	(3)	(4)
	基于 FT-BERT-NT	基于 FT-BERT-NT	基于 VT-BERT-NT	基于 VT-BERT-NT
NewsTone	0.030 *** (7.921)	0.049 *** (3.956)	0.059 *** (19.094)	0.059 *** (6.221)
L. NewsTone	0.047 *** (11.526)	0.022 * (1.711)	0.018 *** (4.761)	0.018 ** (2.020)
L2. NewsTone	-0.015 *** (-3.797)	-0.018 (-1.570)	-0.024 *** (-6.667)	-0.032 *** (-3.656)
L. Gold	0.103 (1.174)	0.316 (1.201)	0.043 (0.502)	0.390 (1.577)
L. Treasury	0.036 (1.094)	0.130 (0.585)	0.006 (0.185)	0.179 (0.786)
L. Exchange	0.135 (1.001)	0.330 (0.672)	0.024 (0.183)	0.090 (0.185)
_cons	-0.032 *** (-9.344)	-0.037 *** (-3.224)	-0.033 *** (-11.090)	-0.033 *** (-3.466)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
N	1 447	289	1 447	289
R^2	0.168	0.125	0.266	0.236
R_{out}^2	0.030		0.096	

注: 变量前 L 表示滞后, 数字表示滞后期; ***, **, * 分别代表 1%、5% 和 10% 的显著性水平. 表格中所有列的被解释变量均为原油期货收益率 Return. 对于解释变量 NewsTone, 列(1) 与列(2) 使用由全量原油新闻预训练的 BERT 构建的新闻语调 FT-BERT-NT 对收益率回归, 列(3) 与列(4) 代表使用垂直领域语料预训练的 BERT 构建的新闻语调 VT-BERT-NT 对收益率回归.

上述结果表明, 集成方法不仅能够在样本内解释市场收益, 更能在样本外环境中保持有效性, 避免前视偏差带来的虚假拟合. 相比通用大语言

模型, 经过领域预训练与收益微调的垂直模型在捕捉新闻语调与收益率的动态关系方面更具泛化能力和解释力. 这一证据说明, 领域化大模型能够

更贴近投资者当期可获得的信息集,从而提升预测结果的现实适用性.对于实践而言,该方法为投资者风险管理和监管部门跨期市场预警提供了一个可操作、稳健的量化工具.

3.5 影响机制分析

作为全球最重要的战略性能源商品之一,原油市场的动态始终受到媒体的高度关注.媒体对原油供需变化、地缘政治冲突等重大事件的报道具有即时性和持续性特征.研究表明,当原油市场发生重大事件时,媒体报道的频次与强度呈现显著正相关^[46]:事件重要性越高,媒体报道数量越多,用词强度越大,新闻语调也越强烈.这种强烈的新闻语调能够有效吸引投资者关注,引导其将注意力资源投向原油期货市场,进而影响其交易决策,最终作用于市场价格形成机制.基于上述信息传递机制,本文聚焦我国原油期货市场,重点检验原油新闻语调、投资者关注度与原油期货收益率三者之间的传导路径.

具体地,参考江艇^[47],本文建立如下回归模型以分析中介效应.其中,以集成语调(VT_BERT_NT)为主要检验变量,控制变量包括汇率、黄金市场以及宏观经济变量的代理变量,并且同样控制年份、月份以及周内等固定效应

$$Return_t = \alpha_0 + \alpha_1 VT_BERT_NT_{t-1} + \alpha_2 Controls_{t-1} + \gamma + \eta + \rho + \varepsilon \quad (4)$$

$$Attention_t = \lambda_0 + \lambda_1 VT_BERT_NT_{t-1} + \lambda_2 Controls_{t-1} + \gamma + \eta + \rho + \varepsilon \quad (5)$$

其中式(4)中回归系数 α_1 表示新闻语调 $VT_BERT_NT_{t-1}$ 对原油期货收益率 $Return$ 影响的总效应.式(5)中回归系数 λ_1 表示新闻语调 $VT_BERT_NT_{t-1}$ 对投资者关注度 $Attention$ 的影响效应.通过依次估计式(4)和式(5),若 α_1 、 λ_1 均显著不为0,则说明投资者关注度在新闻语调影响原油期货收益率的过程中起到中介作用,从而证明假设3.

表9的回归结果表明,滞后一期的集成语调对原油期货收益率呈显著正向影响(表9列(1)),这一结论与前文的跨市场回归结果保持一致,说明语调能够在较短时滞内影响价格波动.同时,集成语调对投资者关注度亦呈显著负向作用(表9列(2)),表明新闻报道中更为强烈的情绪极性和叙事强度能够有效吸引投资者的注意力,

但这些新闻报道降低了主动讨论和搜索行为.在引入关注度变量后,新闻语调对收益率的直接作用系数下降,而关注度保持显著,反映出投资者关注度在二者之间发挥了部分中介效应.这一结果说明,新闻语调不仅直接作用于市场价格,还通过塑造投资者的注意力分布和交易预期,间接影响市场收益.

为检验该结果的稳健性,本文采用东方财富“原油期货吧”的浏览量(Click)、发帖量(Post)与评论量(Comment)对数值等社交媒体行为变量作为投资者关注度的替代性指标,结果同样显示滞后一期的集成语调能够显著提升投资者关注度(表9列(3)~列(5)),说明语调不仅存在于媒体文本之中,还会在投资者的实际行为中得到体现,并通过关注度这一渠道对市场产生放大效应.

从机制上看,集成语调与投资者关注度之间的负向关系揭示了一个重要特征:语调越强烈,投资者越容易被新闻吸引而转移注意力,导致降低进行主动的搜索和讨论行为.这与有限注意力理论相契合,即当市场面临信息过载时,投资者更倾向于依赖直观的新闻语调或者情绪信号来分配注意力和做出决策.部分中介效应的结果进一步表明,新闻语调对收益率的影响并非完全由关注度驱动,语调仍保留了直接作用路径.这意味着市场价格既对新闻的情绪极性敏感,也受到投资者注意力分布的调节,二者共同决定了价格的短期波动.这种“直接作用+间接作用”的双路径机制,也解释了原油期货市场中情绪冲击往往表现为放大且持续的特征,上述结果验证了假设3.

在政策层面,这一发现具有多重启示.首先,对于投资者而言,关注度的中介效应意味着新闻语调不仅反映在价格中,更体现在行为活跃度上,因此,投资者在进行交易决策时应警惕情绪化报道引发的“羊群效应”.其次,对于监管者而言,集成语调指标与关注度指标的结合可成为监测市场情绪与跨境风险传染的重要工具,特别是在极端市场情境下,可以通过对新闻语调和投资者关注的联动监测实现更及时的预警.最后,对于市场制度建设而言,新闻语调通过关注度影响价格的链条提示监管部门应重视媒体在市场预期管理中的作用,适度引导新闻报道的平衡性和专业性,以降低极端语调放大市场波动的风险.

表 9 中介效应检验结果
Table 9 Results of mediating effect test

变量	Attention				
	(1)	(2)	(3)	(4)	(5)
	Return	BaiduIndex	Click	Post	Comment
<i>L. VT-BERT-NT</i>	0.036 *** (8.851)	-2.236 *** (-3.958)	-0.406 *** (-3.614)	-0.393 *** (-3.635)	-0.524 *** (-3.104)
<i>L. Gold</i>	0.022 (0.231)	10.807 (0.934)	0.532 (0.192)	0.928 (0.375)	2.431 (0.626)
<i>L. Treasury</i>	0.019 (0.541)	2.696 (0.683)	1.447 (1.417)	0.754 (0.805)	0.583 (0.388)
<i>L. Exchange</i>	0.166 (1.115)	30.388 ** (2.041)	7.569* (1.880)	9.943 ** (2.415)	14.524 *** (2.702)
<i>_cons</i>	-0.021 *** (-7.893)	1.400 *** (3.652)	9.897 *** (138.727)	3.135 *** (45.732)	3.410 *** (31.840)
<i>Year FE</i>	Yes	Yes	Yes	Yes	Yes
<i>Month FE</i>	Yes	Yes	Yes	Yes	Yes
<i>Week FE</i>	Yes	Yes	Yes	Yes	Yes
<i>N</i>	1 159	1 159	1 151	1 151	1 151
<i>R²</i>	0.090	0.366	0.564	0.369	0.426

注: 变量前 *L* 表示滞后, 数字表示滞后期; ***, **, * 分别代表 1%、5% 和 10% 的显著性水平。

4 新闻语调与风险溢出

上述分析考察了新闻语调与原油期货收益率间的线性关系,但是在诸多资产市场中,情绪具有显著的非对称影响^[48]。为了进一步从信息和风险传导的角度验证新闻语调的跨市场影响,本文使用 CoVaR 对原油新闻语调的风险溢出效应进行分析。该模型最初被用于分析银行资产所面临的极端风险,以便考察金融体系可能存在的系统性风险^[49-51],后续研究将这一模型进一步扩展为变量的极端变动对其他变量的影响^[52]。相较于线性的回归分析,CoVaR 方法不仅能够刻画变量之间的相关性,还能够进一步揭示在极端市场情境下变量间的非线性影响关系。它通过条件风险溢出效应衡量一个变量在极端情况下对另一个变量风险水平的影响,因此更适用于考察新闻语调等外生冲击因素在尾部风险传染、市场压力集中释放等情形下对原油期货收益率的潜在影响。这一方法有助于捕捉在常规分析中可能被忽视的尾部依赖结构,提升风险识别的敏感性与准确性。本文进一步通过考察尾部风险依赖关系的方式检验原油新

闻语调对原油期货收益率的影响。首先进行分位数回归,回归模型如下

$$Q_{\tau}(NewsTone_t | F_{t-1}) = \alpha_0(\tau) + \alpha_1(\tau) Controls_{t-1} + \varepsilon_t \quad (6)$$

$$Q_{\tau}(Return_t | F_{t-1}) = \beta_0(\tau) + \beta_1(\tau) \times NewsTone_{t-1} + \gamma(\tau) \times Controls_{k,t-1} + \varepsilon_t \quad (7)$$

其中 $Q_{\tau}(NewsTone_t | F_{t-1})$ 为信息集 (F_{t-1}) 条件下,原油新闻语调在分位数 $\tau \in (0, 1)$ 处的条件分位数。在 CoVaR 方法中代表了新闻语调的极端情况。而 $Q_{\tau}(Return_t | F_{t-1})$ 为信息集 (F_{t-1}) 条件下,原油期货收益率在分位数 $\tau \in (0, 1)$ 处的条件分位数。 $NewsTone_{t-1}$ 为滞后一期的新闻语调指标, $Controls_{t-1}$ 为滞后一期的控制变量。本文使用 5% 的分位数进行估计

$$VaR_t(q) = \hat{\alpha}_0 + \hat{\alpha}_1 Controls_{t-1} \quad (8)$$

$$CoVaR_t(q) = \hat{\beta}_0 + \hat{\beta}_1 VaR_t(q) + \hat{\beta}_3 Controls_{t-1} \quad (9)$$

最终得到 $\Delta CoVaR$

$$\Delta CoVaR_t(q) = CoVaR_t(q) - CoVaR_t(50\%) \quad (10)$$

通过分位数回归计算得到的原油期货市场价格收益率在四种语调(使用词典法构建的国际新闻语

调、使用原始模型或全部原油新闻语料训练的大语言模型构建的国际新闻语调、使用垂直领域语

料预训练的大语言模型构建的国际新闻语调) 下的 $\Delta CoVaR$ 描述性统计如表 10 所示.

表 10 $\Delta CoVaR$ 描述性统计

Table 10 Statistics summary of $\Delta CoVaR$

变量	观测数	Mean	SD	Min	Median	Max
$\Delta CoVaR1$	1 159	-0.005 2 [-0.013 6 ρ . 003 1]	0.001	-0.009 5	-0.005 2	0.000 0
$\Delta CoVaR2$	1 159	0.000 4 [-0.004 7 ρ . 005 3]	0.000	0.000 2	0.000 4	0.000 5
$\Delta CoVaR3$	1 159	-0.023 8 [-0.028 5, -0.019 0]	0.000	-0.027 8	-0.023 8	-0.019 8
$\Delta CoVaR4$	1 159	-0.022 0 [-0.029 4, -0.014 5]	0.001	-0.026 5	-0.022 0	-0.017 0

注: 表中 $\Delta CoVaR1$ 代表词典法构建的国际市场语调对期货收益率的跨市场风险溢出; $\Delta CoVaR2$ 代表使用原始大语言模型构建的国际新闻语调对期货收益率的跨市场风险溢出; $\Delta CoVaR3$ 代表使用全部原油新闻语料预训练的大语言构建的国际新闻语调对期货收益率的跨市场风险溢出; $\Delta CoVaR4$ 代表使用垂直领域语料训练的大语言模型构建的国际新闻语调对期货收益率的跨市场风险溢出. 在第 3 列 Mean 下列了 95% 置信区间.

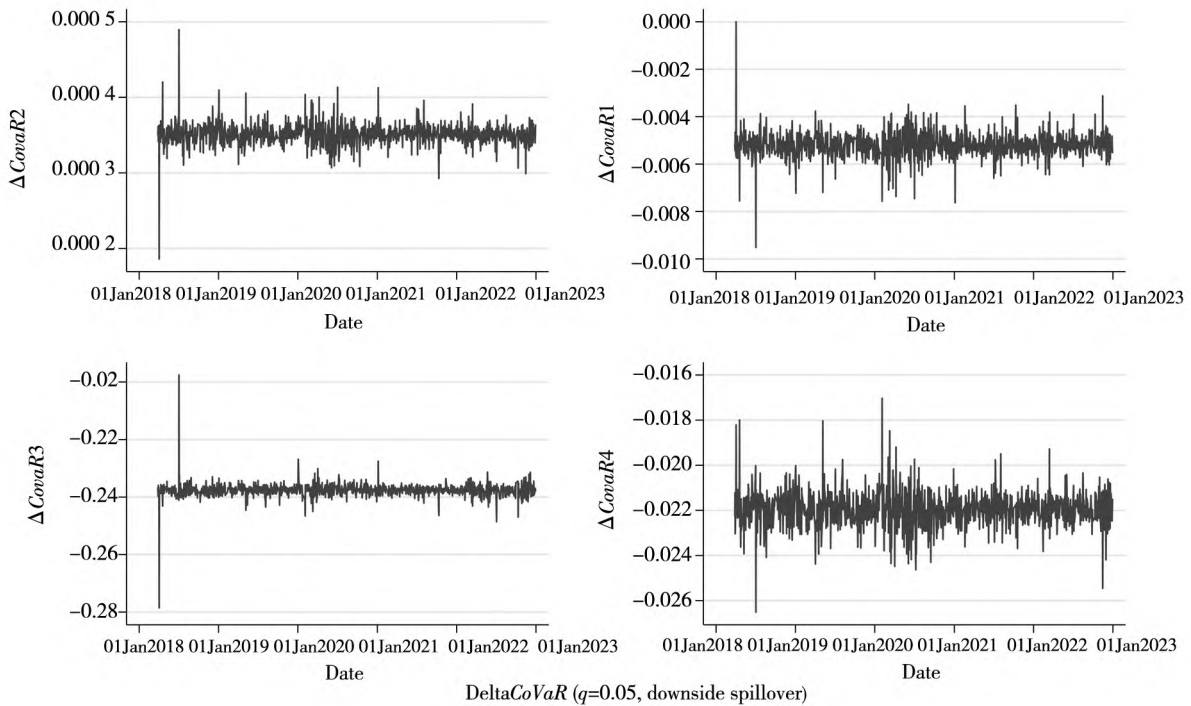


图 4 原油期货市场收益在四种语调条件下的 $\Delta CoVaR$

Fig. 4 Crude oil futures market returns in four tones of $\Delta CoVaR$

表 10 与图 4 均报告了不同语调测度下的 $\Delta CoVaR$ 结果,词典法 ($\Delta CoVaR1$) 与原始大语言模型 ($\Delta CoVaR2$) 的置信区间均跨越零值,说明其在极端市场条件下缺乏稳健的风险溢出解释力,甚至出现与经济直觉相悖的负向结果. 基于全量

语料预训练的大语言模型 ($\Delta CoVaR3$) 在统计上虽显著,但其识别出的负向溢出效应与市场逻辑并不完全一致^[53,54],可能源于模型在非筛选语料下对语调特征的模糊提取. 而基于垂直领域语料训练并结合收益信号微调的大语言模型

($\Delta CoVaR4$) 不仅在统计上显著不为零,且所反映的溢出方向与市场经验相符,能够准确刻画国际新闻情绪在极端情况下对国内原油期货收益率的正向风险传导。

上述结果表明:传统词典方法因过度依赖既有词条扩展而忽视极端语境,难以在尾部风险条件下捕捉真实的情绪冲击;通用大语言模型虽然具备语境理解能力,但因缺乏领域适配而在极端市场环境下表现失效;相比之下,集成方法通过结合领域词典的知识约束与大模型的语境建模能力,能够更好地识别新闻语调的尾部依赖结构,验证了假设4,同样支撑了假设1与假设2。这一发现不仅验证了新闻语调在跨市场风险传导中的作用机制,也说明大语言模型的领域化与弱监督微调对于提升金融市场极端风险识别的有效性具有关键价值,为完善风险预警工具和制定跨市场监管政策提供了新的实证支持。

5 结束语

近年来,大语言模型(LLM)在金融文本分析领域的应用迅速扩展,其基于海量语料的语义捕捉能力为市场情绪测度提供了新工具。然而,现有研究多聚焦于通用场景,对原油期货等专业领域复杂语义的适配性不足——通用模型易因行业术语歧义与跨语言信息损耗等问题容易导致新闻语调误判。鉴于此,本文围绕原油市场新闻语调,提出了基于大语言模型与垂直领域词典的原油新闻语调集成方法,为金融大语言模型的垂直化应用提供了可复制路径。

首先,现有文献多强调新闻内容的事实性信息,而忽视了语调在情绪传递与注意力分配中的独立作用。本文通过验证新闻语调对原油期货收益率的显著影响,揭示了语调在跨市场信息传导与尾部风险形成中的重要角色,并进一步明确了其通过投资者关注度中介影响价格的机制。这一发现为投资者情绪与市场定价关系提供了新的理论证据。其次,本文提出的集成方法有效结合了领域词典的专业知识约束与大语言模型的上下文建

模能力,并通过收益符号的弱监督微调实现了与市场反应的一致性。与词典法或未经适配的通用大语言模型相比,该方法不仅在收益预测和风险刻画上表现出更高的显著性和稳健性,而且在样本外预测和极端市场事件的风险溢出检验中均展现出优势。这一结果说明,集成方法能够克服传统方法的前视偏差与语义漂移问题,为金融大语言模型的垂直化应用提供了可复制的路径,也为未来文本分析方法在金融市场中的发展提供了借鉴。

本文为市场参与者提供了新的风险识别与管理工具。一方面,对于投资者而言,基于集成方法的语调指标能够在高频环境下提供更及时、可靠的情绪信号,有助于优化交易决策和风险对冲策略。另一方面,对于监管部门而言,采用集成方法测度新闻语调有助于识别对国际新闻冲击下市场异常波动,从而提高对跨境风险传染的防范与应对能力。更为重要的是,实证结果表明国际新闻语调对上海原油期货收益具有显著的跨市场影响,并在极端市场情境下加剧尾部风险溢出,这揭示了我国原油期货市场在全球能源金融体系中的深度嵌入。在这一背景下,本文的研究不仅有助于监管层更好地把握国际情绪冲击下的风险传导机制,也为健全风险防控机制和深化金融开放提供了政策参考,凸显其在我国金融战略中的现实价值。

需要指出的是,本文提出的集成方法更适用于在原油期货这类复杂、跨市场耦合度极高的场景,在语料有限或缺乏稳定收益信号的市场中,其依然可行但是相较于传统词典法活通用大语言模型的优势可能受到一定制约。另外,在股票或债券等信息结构与交易机制差异较大的市场中,直接迁移该方法会面临语境适配与语料特征差异的挑战。因此,在这些市场中需在领域词典构建和模型微调环节进行进一步改进,以确保方法的稳健性与解释力。未来研究可在跨市场、跨语种、多模态数据的维度上进一步拓展集成方法的适用性,从而提升金融情绪测度与风险管理的普适性。

参考文献:

- [1]林建浩,孙乐轩. 大语言模型与经济金融文本分析: 基本原理、应用场景与研究展望[J]. 计量经济学报, 2025, 5(1): 1-34.
Lin Jianhao, Sun Lexuan. Text analysis in economics and finance with large language models: Fundamentals, applications, and future prospects[J]. China Journal of Econometrics, 2025, 5(1): 1-34. (in Chinese)
- [2]许雪晨,田侃. 一种基于金融文本情感分析的股票指数预测新方法[J]. 数量经济技术经济研究, 2021, 38(12): 124-145.
Xu Xuechen, Tian Kan. A novel financial text sentiment analysis-based approach for stock index prediction[J]. The Journal of Quantitative & Technical Economics, 2021, 38(12): 124-145. (in Chinese)
- [3]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(7): 141-152.
You Jiaying, Wu Jing. Spiral of silence: Media sentiment and the asset mispricing[J]. Economic Research Journal, 2012, 47(7): 141-152. (in Chinese)
- [4]黄俊,郭照蕊. 新闻媒体报道与资本市场定价效率——基于股价同步性的分析[J]. 管理世界, 2014, (5): 121-130.
Huang Jun, Guo Zhaorui. News media coverage and capital market pricing efficiency: An analysis based on stock price synchronicity[J]. Management World, 2014, (5): 121-130. (in Chinese)
- [5]何勇,焦丽,杨艺,等. AI大语言模型赋能金融市场量化投资? 基于另类数据与传统金融数据的研究[J]. 计量经济学报, 2024, 4(3): 761-783.
He Yong, Jiao Li, Yang Yi, et al. AI large language model empowers quantitative investment in financial markets? Research based on alternative data and traditional financial data Chinese full text[J]. China Journal of Econometrics, 2024, 4(3): 761-783. (in Chinese)
- [6]陶利斌,潘婉彬,黄筠哲. 沪深300股指期货价格发现能力的变化及其决定因素[J]. 金融研究, 2014, (4): 128-142.
Tao Libin, Pan Wanbin, Huang Yunzhe. The determinants of price discovery in CSI300 Stock Index futures[J]. Journal of Financial Research, 2014, (4): 128-142. (in Chinese)
- [7]Beckmann J, Czudaj R L, Arora V. The relationship between oil prices and exchange rates: Revisiting theory and evidence[J]. Energy Economics, 2020, (88): 104772.
- [8]Zhao L, Zhang X, Wang S, et al. The effects of oil price shocks on output and inflation in China[J]. Energy Economics, 2016, (53): 101-110.
- [9]Fiszeder P, Faldziński M, Molnár P. Attention to oil prices and its impact on the oil, gold and stock markets and their covariance[J]. Energy Economics, 2023, (120): 106643.
- [10]Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks[J]. The Journal of Finance, 2011, 66(1): 35-65.
- [11]沈艳,陈赟,黄卓. 文本大数据分析在经济学和金融学中的应用: 一个文献综述[J]. 经济学(季刊), 2019, 18(4): 1153-1186.
Shen Yan, Chen Yun, Huang Zhuo. A literature review of textual analysis in economic and financial research[J]. China Economic Quarterly, 2019, 18(4): 1153-1186. (in Chinese)
- [12]Yang K, Wei Y, Li S, et al. Global financial uncertainties and China's crude oil futures market: Evidence from interday and intraday price dynamics[J]. Energy Economics, 2021, (96): 105149.
- [13]Albulescu C T, Demirel R, Raheem I D, et al. Does the U. S. economic policy uncertainty connect financial markets? Evidence from oil and commodity currencies[J]. Energy Economics, 2019, (83): 375-388.
- [14]姚加权,冯绪,王赞钧,等. 语调、情绪及市场影响: 基于金融情绪词典[J]. 管理科学学报, 2021, 24(5): 26-46.

- Yao Jiaquan, Feng Xu, Wang Zanjun, et al. Tone, sentiment and market impacts: The construction of Chinese sentiment dictionary in finance [J]. *Journal of Management Sciences in China*, 2021, 24(5): 26–46. (in Chinese)
- [15] 姜富伟, 孟令超, 唐国豪. 媒体文本情绪与股票回报预测 [J]. *经济学(季刊)*, 2021, 21(4): 1323–1344.
Jiang Fuwei, Meng Lingchao, Tang Guohao. Media textual sentiment and Chinese stock return predictability [J]. *China Economic Quarterly*, 2021, 21(4): 1323–1344. (in Chinese)
- [16] 姜富伟, 刘雨旻, 孟令超. 大语言模型、文本情绪与金融市场 [J]. *管理世界*, 2024, 40(8): 42–64.
Jiang Fuwei, Liu Yumin, Meng Lingchao. Large language model and textual sentiment analysis in Chinese stock markets [J]. *Management World*, 2024, 40(8): 42–64. (in Chinese)
- [17] Loughran T, McDonald B, Pragidis I. Assimilation of oil news into prices [J]. *International Review of Financial Analysis*, 2019, (63): 105–118.
- [18] Baker S R, Bloom N, Davis S J. Measuring economic policy uncertainty [J]. *The Quarterly Journal of Economics*, 2016, 131(4): 1593–1636.
- [19] Caldara D, Iacoviello M. Measuring geopolitical risk [J]. *American Economic Review*, 2022, 112(4): 1194–1225.
- [20] Brandt M W, Gao L. Macro fundamentals or geopolitical events? A textual analysis of news events for crude oil [J]. *Journal of Empirical Finance*, 2019, (51): 64–94.
- [21] Balciilar M, Bekiros S, Gupta R. The role of news-based uncertainty indices in predicting oil markets: A hybrid nonparametric quantile causality method [J]. *Empirical Economics*, 2017, 53(3): 879–889.
- [22] Tetlock P C. Giving content to investor sentiment: The role of media in the stock market [J]. *The Journal of Finance*, 2007, 62(3): 1139–1168.
- [23] Löffler G, Norden L, Rieber A. Negative news and the stock market impact of tone in rating reports [J]. *Journal of Banking & Finance*, 2021, (133): 106256.
- [24] 唐国豪, 姜富伟, 张定胜. 金融市场文本情绪研究进展 [J]. *经济学动态*, 2016, (11): 137–147.
Tang Guohao, Jiang Fuwei, Zhang Dingsheng. Research progress on text sentiment in financial markets [J]. *Economic Perspectives*, 2016, (11): 137–147. (in Chinese)
- [25] Liu S, Han J. Media tone and expected stock returns [J]. *International Review of Financial Analysis*, 2020, (70): 101522.
- [26] 汪昌云, 武佳薇. 媒体语气、投资者情绪与 IPO 定价 [J]. *金融研究*, 2015, (9): 174–189.
Wang Changyun, Wu Jiawei. Media tone, investor sentiment, and IPO pricing [J]. *Journal of Financial Research*, 2015, (9): 174–189. (in Chinese)
- [27] Huang A G, Tan H, Wermers R. Institutional trading around corporate news: Evidence from textual analysis [J]. *The Review of Financial Studies*, 2020, 33(10): 4627–4675.
- [28] Lucey B, Ren B. Does news tone help forecast oil? [J]. *Economic Modelling*, 2021, (104): 105635.
- [29] You J, Zhang B, Zhang L. Who captures the power of the pen? [J]. *The Review of Financial Studies*, 2018, 31(1): 43–96.
- [30] Basu S, Ma X, Briscoe-Tran H. Measuring multidimensional investment opportunity sets with 10-K Text [J]. *The Accounting Review*, 2022, 97(1): 51–73.
- [31] Li X, Wu P, Wang W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong [J]. *Information Processing & Management*, 2020, 57(5): 102212.
- [32] Barakat A, Ashby S, Fenn P, et al. Operational risk and reputation in financial institutions: Does media tone make a difference? [J]. *Journal of Banking & Finance*, 2019, (98): 1–24.
- [33] 龚旭, 刘堂勇, 文凤华. 国际与中国原油市场间的波动率动态溢出分析 [J]. *管理科学学报*, 2023, 26(11): 125–141.
Gong Xu, Liu Tangyong, Wen Fenghua. Dynamic volatility spillovers between international crude oil futures and China's crude oil spot [J]. *Journal of Management Sciences in China*, 2023, 26(11): 125–141. (in Chinese)

- [34] Su T, Lin B. Reassessing the information transmission and pricing influence of Shanghai crude oil futures: A time-varying perspective [J]. *Energy Economics*, 2024, (140): 107977.
- [35] Chen X, Tongurai J. Informational linkage and price discovery between China's futures and spot markets: Evidence from the US-China trade dispute [J]. *Global Finance Journal*, 2023, (55): 100750.
- [36] 张大永, 姬强. 中国原油期货动态风险溢出研究 [J]. *中国管理科学*, 2018, 26(11): 42–49.
Zhang Dayong, Ji Qiang. Studies on the dynamic risk spillovers for China's crude oil futures [J]. *Chinese Journal of Management Science*, 2018, 26(11): 42–49. (in Chinese)
- [37] 金星晔, 左从江, 方明月, 等. 企业数字化转型的测度难题: 基于大语言模型的新方法与新发现 [J]. *经济研究*, 2024, 59(3): 34–53.
Jin Xingye, Zuo Congjiang, Fang Mingyue, et al. Measurement problem of enterprise digital transformation: New methods and findings based on large language models [J]. *Economic Research Journal*, 2024, 59(3): 34–53. (in Chinese)
- [38] Huang A H, Wang H, Yang Y. FinBERT: A large language model for extracting information from financial text [J]. *Contemporary Accounting Research*, 2023, 40(2): 806–841.
- [39] Breitung C, Müller S. Global business networks [J]. *Journal of Financial Economics*, 2025, (166): 104007.
- [40] Navigli R, Conia S, Ross B. Biases in large language models: Origins, inventory, and discussion [J]. *Journal of Data and Information Quality*, 2023, 15(2): 1–21.
- [41] Tomalin M, Byrne B, Concannon S, et al. The practical ethics of bias reduction in machine translation: Why domain adaptation is better than data debiasing [J]. *Ethics and Information Technology*, 2021, 23(3): 419–433.
- [42] 杨涛, 郭萌萌. 投资者关注度与股票市场——以 PM2.5 概念股为例 [J]. *金融研究*, 2019, (5): 190–206.
Yang Tao, Guo Mengmeng. Investor attention and the stock market: A new perspective on PM2.5 concept stocks [J]. *Journal of Financial Research*, 2019, (5): 190–206. (in Chinese)
- [43] Da Z, Engelberg J, Gao P. In search of attention [J]. *The Journal of Finance*, 2011, 66(5): 1461–1499.
- [44] 刘春林, 石睿. 交易所问询是否影响管理层语调? ——基于上市公司年报的文本分析 [J]. *管理科学学报*, 2024, 27(11): 119–135.
Liu Chunlin, Shi Rui. Do exchange inquiries affect managerial tone?: A textual analysis based on listed companies' annual reports [J]. *Journal of Management Sciences in China*, 2024, 27(11): 119–135. (in Chinese)
- [45] 马嫣然, 吴菲, 张大永, 等. 谁是驱动中国原油期货价格波动的关键信息? [J]. *管理科学学报*, 2024, 27(1): 113–125.
Ma Yanran, Wu Fei, Zhang Dayong, et al. What is the key factor driving price volatility of China's crude oil futures? [J]. *Journal of Management Sciences in China*, 2024, 27(1): 113–125. (in Chinese)
- [46] Plante M. OPEC in the news [J]. *Energy Economics*, 2019, (80): 163–172.
- [47] 江艇. 因果推断经验研究中的中介效应与调节效应 [J]. *中国工业经济*, 2022, (5): 100–120.
Jiang Ting. Mediating effects and moderating effects in causal inference [J]. *China Industrial Economics*, 2022, (5): 100–120. (in Chinese)
- [48] 张大永, 姬强. 中国原油期货动态风险溢出研究 [J]. *中国管理科学*, 2018, 26(11): 42–49.
Zhang Dayong, Ji Qiang. Studies on the dynamic risk spillovers for China's crude oil futures [J]. *Chinese Journal of Management Science*, 2018, 26(11): 42–49. (in Chinese)
- [49] Yan W L, Cheung A. Quantile connectedness among climate policy uncertainty, news sentiment, oil and renewables in China [J]. *Research in International Business and Finance*, 2025, (76): 102814.
- [50] Adrian T, Brunnermeier M K. Covar [J]. *American Economic Review*, 2016, 106(7): 1705–1741.
- [51] Ma Y, Zhang Y, Ji Q. Do oil shocks affect Chinese bank risk? [J]. *Energy Economics*, 2021, (96): 105166.
- [52] 汪宁丽, 汤珂. 保证金调整与期货价格波动 [J]. *管理科学学报*, 2025, 28(1): 160–176.
Wang Ningli, Tang Ke. Margin adjustment and futures price fluctuation [J]. *Journal of Management Sciences in China*, 2025, 28(1): 160–176. (in Chinese)

[53] 龚旭, 刘堂勇, 文凤华. 国际与中国原油市场间的波动率动态溢出分析[J]. 管理科学学报, 2023, 26(11): 125–141.

Gong Xu, Liu Tangyong, Wen Fenghua. Dynamic volatility spillovers between international crude oil futures and China crude oil spot[J]. Journal of Management Sciences in China, 2023, 26(11): 125–141. (in Chinese)

[54] 吴国兵, 陈平, 李帆, 等. 全球金融市场风险溢出效应研究——基于时频视角和网络关联的分析[J]. 管理科学学报, 2025, 28(1): 119–144.

Wu Guobing, Chen Ping, Li Fan, et al. Risk spillover effects in global financial markets: Analysis based on time-frequency perspective and network connection[J]. Journal of Management Sciences in China, 2025, 28(1): 119–144. (in Chinese)

Crude oil news tone and futures market returns: An integrated approach based on large language models and domain-specific lexicons

CHEN Rong-da^{1,2}, XIAO Wen-hao¹, JIN Cheng-lu^{1*}, LI Han¹

1. School of Finance, Zhejiang University of Finance and Economics, Hangzhou 310018, China;

2. Zhejiang Financial College, Hangzhou 310018, China

Abstract: The verticalization of large language models (LLMs) has become an important trend. In highly specialized and cross-market contexts, such as crude oil futures, general-purpose LLMs struggle to accurately interpret domain-specific terminology and complex semantics, and they are prone to look-ahead bias. This paper proposes an integrated approach that combines LLMs with a domain-specific sentiment lexicon to measure the tone of crude oil news. By integrating the prior knowledge of a domain lexicon with the semantic understanding of LLMs, the proposed method enhances both the interpretability and predictive power of tone measures. Specifically, based on 93 004 Chinese and English crude oil news articles from InfoBank and Factiva between 2018 and 2022, this study constructs a vertical domain lexicon, refines high signal-to-noise corpora, performs domain-specific pretraining on a general BERT model, and applies weakly supervised fine-tuning guided by futures return signs to generate an integrated tone index. Empirical results show that the tone measure derived from the proposed method significantly outperforms dictionary-based and general LLM methods in explaining and predicting Shanghai crude oil futures returns. Robustness tests confirm consistent results, and the method also exhibits superior out-of-sample predictive performance for 2023–2024. Further analysis reveals that the impact of news tone on crude oil futures returns is transmitted through investor attention as a mediating channel, and under the integrated framework, this mechanism aligns closely with economic logic in both direction and significance. Moreover, the risk spillover analysis indicates that the integrated tone measure effectively captures the tail-risk transmission from international news sentiment to China's crude oil futures market. This study contributes by proposing a reproducible vertical application framework for financial large language models, revealing the dual role of news tone in market return formation and risk transmission, and providing new quantitative tools for risk identification and policy formulation in the crude oil futures market.

Key words: news tone; large language model; domain-specific lexicon; integrated method; domain-adapted BERT model