

企业研发文本披露的同行溢出效应

符号亮 夏婷 孙凤娥

内容提要: 本文构建企业研发文本披露指标,分析同行企业研发文本是否对企业研发投入产生溢出效应。结果显示,中国上市企业研发文本披露存在溢出效应,同行企业披露的研发文本越多,目标企业的研发投入增加越多。机制分析结果表明,研发文本披露通过市场竞争机制和信息学习机制实现溢出效应;行业市场竞争会促使目标企业研发决策受到同行企业研发文本披露的影响;信息学习需求促使处于跟随地位的目标企业向行业领先企业学习。拓展研究结果显示,当企业融资约束程度低时,同行企业研发文本披露的溢出效应增强;当同行企业获得较多财政补贴时,研发文本披露的溢出效应增强;当同行企业信息可比度高时,研发文本披露的溢出效应增强。本文研究结论有助于相关部门完善企业研发文本披露规则和加快建立高效的企业创新生态。

关键词: 研发投入 信息披露 溢出效应 文本分析 信息可比度

中图分类号: F273

文献标识码: A

文章编号: 1000-7636(2023)11-0125-20

一、问题提出

长期以来,企业面临着如何平衡研发保护和研发共享的阿罗信息悖论:公开上市需要披露研发信息,而披露则意味着研发价值的流失^[1]。而有学者指出,即使缺乏产权保护,信息披露依然是改善信息悖论的有效途径^[2]。

现代企业产品迭代加快、研发周期缩短和技术不确定性增加,企业研发资源有限的弱点凸显,企业研发创新愈加强调合作开放。打破原有组织边界、内部研发与外部创新资源并重、与各类市场主体互动紧密,有助于企业找准市场焦点、加快技术研发和提高商业获利水平。企业作出研发决策需要进行市场调研、产品构思、比较论证等工作,参考或借助同行企业的研发信息对于企业决策者来说必不可少。有学者指出,企业的融资决策、股利决策等行为存在观察和模仿同行企业的现象^[3-5]。企业在获取研发决策信息时是否将同行企业研发信息纳入自然就成为本文的研究主题。本文从信息溢出视角切入,检验同行企业对外披露的研

收稿日期:2023-01-14;修回日期:2023-09-13

基金项目:广东省哲学社会科学规划项目“绿色金融、数字赋能与经济高质量发展研究”(GD23XYJ15)

作者简介:符号亮 广东财经大学金融学院副教授,广州,510320;

夏婷 广东财经大学金融学院讲师;

孙凤娥 北京林业大学经济管理学院副教授,通信作者,北京,100083。

作者感谢匿名审稿人的评审意见。

发文本是否被纳入目标企业研发决策的信息集合,以及何种机制促进了研发文本披露的外溢。通过探究这些问题,评估研发信息披露规则的有效性,有助于理解信息悖论问题。本文采用机器学习领域较为成熟的“种子词集+Word2Vec相似词扩充”词向量技术,通过文构财经文本数据平台(WinGo)统计上市企业年度报告中的相关变量展开分析。

本文的增量贡献在于:(1)参照已有研究^[6-7],通过文构财经文本数据平台分析中国上市公司年报,构建了研发信息指标,相关方法可以为后续度量中国企业研发文本披露提供借鉴。(2)相较于同行语调情绪指标而言^[8],本文深入分析同行研发活动信息指标,更为接近企业创新投资决策信息需求。本文的结果表明同行研发信息披露确实有助于目标企业研发投资,这为解释阿罗信息悖论问题增添更为直接的证据。(3)对中国上市企业研发信息的横向溢出效应进行研究。与纵向(供应链)溢出^[9]、区域溢出^[10]等相比较而言,本文对同一行业中的上市企业研发信息溢出效应的机制和特征进行了深入剖析,为企业研发信息的横向溢出效应提供了一种定性信息披露解释。

二、制度背景、文献回顾和理论分析

(一) 制度背景

过去二十年,监管机构对上市企业研发活动的会计处理和信息披露要求不断完善,经历了全部费用化、有条件资本化、修订细化三个阶段。一是全部费用化阶段。2001年财政部出台《企业会计准则——无形资产》,规定企业发生的研发支出直接计入当期损益。研发支出全部费用化处理不能完全真实地反映公司的资产价值,导致企业存在大量的未确认的无形资产,企业相关文本的阅读者只知道企业正在进行研发活动,但无从知晓企业研发活动的进展、研发活动成功的可能性。二是有条件资本化阶段。2006年财政部出台无形资产会计准则规定,明确企业内部研究开发项目开发阶段的支出同时满足五项条件的,可以确认为无形资产。研发支出有条件资本化不仅能够向市场传递企业正在进行研发活动的信息,还能向市场传递关于企业研发进展、研发成果、研发商业化程度等方面的信息。研发支出资本化的五项条件本身意味着该无形资产能够为企业产生未来收益的极大可能性,企业的研发项目在技术上具有可行性、具备商业化条件。三是修订细化阶段。上述会计准则的2012年修订版明确了将“研发支出”单项列示,2015年修订版进一步要求对“研发投入”单项披露。研发费用的财务核算依照会计准则,需要设置一级科目“研发支出”。在其下设置二级科目“费用化支出”和“资本化支出”分别核算费用化和资本化的研发支出。

需要指出的是,从上市公司的实际运行情况看,研发战略、研发项目、研发进度以及研发关键因素等非数字化的定性信息对于投资者决策而言似乎更为关键。这些文本信息能够为众多投资者和决策者所使用并资本化,其阅读使用者数量远高出会计数字信息的阅读使用者的数量。企业会计准则修订使得企业创新行为信息披露初步具有强制性,不过企业在表述研发文本信息时依然存在强烈的选择性倾向。企业研发文本信息主要集中在产品研发创新上,较少涉及管理创新、销售创新等方面,造成企业创新信息披露以技术创新为主要内容。上市企业研发信息定量披露主要包括研发技术人员比例、研发支出、知识专利数量、新产品收入与预测等。而研发信息定性披露主要提及报告期内产品研发、研发项目进展、未来投资方向、研发重点、研发形式及组织机构、技术风险、市场风险与竞争对手,研发资本化与费用化的区分描述,以及创新项目的税收优惠与政府补助等。从企业研发决策者来讲,研发定性信息的及时性胜于研发定量信息,且更具有

前瞻性,潜在价值远高于定量信息。

(二) 文献回顾

从文献梳理看,学者关注企业研发信息披露的三类问题:披露目标、披露决定和披露后果。第一是披露目标的问题,指企业以减少与外部利益相关者之间的信息不对称为出发点,在资本市场吸引机构投资者、风险私募基金和潜在投资者的关注,达到股票上市、增发融资和增加创新投入等目的^[11-12]。一般情况下,企业信息披露有助于企业提高股票流动性、推动股票价格上涨,也有助于降低资本成本、减少信息不对称性、增加市场透明度^[13-14]。第二类是披露决定的问题,即研发信息如何披露以及披露多少,主要取决于披露收益与披露成本的权衡。影响研发信息披露的因素主要包括行业研发技术壁垒、产品替代目标、技术敏感度、技术纠纷与诉讼等^[15-16]。企业惯常策略是推迟披露,不过也有企业为了进行市值管理、占领产品市场会提早发布新产品信息、披露研发创新动态过程^[17-18],或者为了获得政府补贴、银行融资等特定目的披露研发详细信息^[19],以及存在研发信息的策略性披露^[20]。学者大多采用自己构建的信息披露指数或公开的披露指标来展开相关研究^[21]。第三类是披露后果的问题,主要涉及信息披露指标与信息披露的经济后果。信息披露的经济后果则是对股价和企业财务指标的影响,包括买卖价差、收益波动率、分析师预测分歧度^[22],或者资本成本、融资约束、企业绩效、企业投资等^[23-24],但鲜有关于研发信息披露对目标企业创新决策影响的研究。

随着使用机器语言度量的文本分析不断深化,少数学者开始探索同行研发语调或供应链信息对企业创新投资的影响。有学者研究发现,除了同行竞争对手投资增加会使得企业投资增加之外,更为重要的是同行竞争对手企业管理层分析与展望(MD&A)语调信息也对企业投资有促进作用,即具有溢出效应^[8]。底璐璐等(2020)则从供应链角度发现,处于行业供应链上下游的企业年报语调信息具有相互借鉴学习和传染效应,客户的年报语调越消极,供应商企业越会持有更多的现金^[9]。综合已有文献可以发现,企业对外披露的文本信息具有信息增量和有效性,已成为企业制定研发战略的重要参考。文本信息语调具有行业间或产业链间的溢出效应。然而,相较于同行语调这个情绪指标而言,构建一个更为接近企业研发决策信息的同行研发活动信息指标,并探讨该指标对企业研发决策的溢出效应,成为值得研究的问题。

(三) 理论分析及研究假设

1. 研发文本披露与研发投入

与一般投资不同,企业研发投资存在着投入风险大、产出不确定、信息严重不对称等诸多风险特征。在开放创新环境下,同一行业其他企业的研发信息披露明显有利于减少企业研发投资的不确定性。从资源依赖角度看,作为资源约束和嵌入依赖关系中的组织,企业的发展能力取决于其外部资源获取能力。企业面对不确定性环境和资源约束,要提升发展能力就需要付出大量的时间成本来确定研发创新战略,而较为经济直接的方式就是搜集同行业其他企业的技术研发投资信息,包括私有秘密技术和公开披露信息,从而有效降低信息搜集和获取成本,以确定自身的研发投资范围、投入重点和投资强度^[25]。行业研发信息有助于企业获取市场信息和技术资源,弥补内部创新的资源基础不足,降低内部研发投入的不确定性。作为同行竞争对手,其他企业的研发信息披露可以帮助企业对比研发资金利用效率,改善利益相关者对企业的内外部评价。

从组织学习视角看,关注同行尤其是领先者的创新研发信息,了解竞争对手的产品研发情况以及行业的未来趋势,可帮助企业明确学习方向、制定组织愿景和调整发展战略。跨企业的资源搜寻让企业可以接

触到同行企业的最新动态,尤其是隐性知识和前沿技术,提高企业对已有技术知识的学习、复制和更新能力。为减少其资源约束和不确定性,企业与同行企业携手共进、深化联结进行价值共创,可大幅提升其生存能力和组织韧性。尽管存在研发专用成本的顾虑,企业管理层依然通过年报等方式提供创新研发活动的各类重要信息。企业越依赖创新产品,研发活动信息越详细^[26],并出现行业创新活动、行业信息披露“群体趋同”的特征,这有利于缓解企业创新的信息不对称问题和丰富企业组织愿景,吸引潜在投资者的注意,获取金融机构的融资支持。当技术方向、研发前景存在不确定性时,决策者会倾向于学习同行企业的技术路线,受到同行企业同类创新投资的启发,及时调整发展战略。

从外部网络视角看,开放性的研发信息获取过程能够助力企业拓展行业生存空间和提升获利水平,有助于企业构建外部创新同盟,尤其是与行业对手形成创新互动网络。企业仅仅局限于内部研发创新,缺乏开放的创新文化、灵活的知识战略、必要的吸收能力,则难以获得持续发展的空间^[27]。文本信息公开披露具有便捷性、高时效性等特点,压缩了时空距离,增强了地理邻近性,极大地缩短了各企业间的技术交流周期。在快速迭代的环境中,企业通过创新网络获得创新概念、前沿技术,技术发展周期缩短,课题攻关速度加快,促进跨企业和跨地域的研究成果可发现、可移植。有关信息披露的法律法规要求公司管理层向投资者、使用者提供真实可靠的高价值语义知识服务,而非提供无价值的庞杂信息。与同行对手的研发信息互动能够让企业具备广阔的技术知识搜寻范围,实现最直接的创新价值变现^[28]。

此外,基于非财务信息定价的重要性也因信息披露而得到增强,企业研发创新的行业认可度与行业应用前景能够影响企业股票市盈率和融资功能。已有研究发现,企业研发创新信息披露越详细,外部机构定价和分析师预测越准确,越有助于企业获得研发融资支持^[29]。以上这些因素无疑会增强同行企业研发信息在企业研发决策信息组合中的重要性,因此企业决策者具有较强的动机利用同行企业所披露的文本信息来改善研发投资决策,帮助企业获得长期稳定的发展。为此提出假设:

H₁:同行企业研发文本披露增加,企业的研发投入会相应增加,即研发文本披露存在溢出效应。

2. 同行企业研发文本披露促进企业研发投入的作用机制

已有研究表明,市场竞争和信息需求是促进企业投资决策模仿行为的重要因素^[30]。从市场竞争角度讲,同行企业研发信息披露有助于减轻创新的不确定性。首先,行业竞争一定程度促使同行企业研发文本披露成为行业创新战略互动的中介。在竞争激烈和技术快速变化的市场中,企业和同行竞争对手不得不同时披露翔实可靠的研发信息,以传递出企业高质量发展的信号。由于竞争激烈,即便是大企业也不得不对照业内披露的相关研发活动信息,领先于竞争对手提供更多的研发文本内容和发布更多创新产品以示先进^[31]。其次,除了行业战略互动之外,研发文本披露还与市场需求、产品成本有关。在相似经济条件下,同行企业研发文本披露对于预测市场需求和产品成本的意义重大。一方面,企业需要投资研发新产品满足社会新需求,开拓新领域;另一方面,企业在不断推出新性价比产品的同时也在不断淘汰老旧产品,形成创新迭代和棘轮驱动,带来创新升级的压力。再次,出于市场竞争压力,企业决策者倾向于在所在行业群体中进行比较,并通过工艺改良、产品研发、技术创新等手段赶超所感知到的行业对手。企业的赶超战略意味着企业可能使用同行竞争企业的信息披露改善创新投资决策,把握新产品机会,减少长期战略实施的不确定性。

从满足信息学习需求的角度看,同行企业研发文本披露同样有助于企业降低研发活动不确定性。首先,企业披露研发信息导致其他企业创新投资是一个帕累托改进,并不一定使信息披露企业处于竞争劣势。

以往研究表明,企业可以从同行企业披露的信息中学习,加强行业企业间以及产业链的信息比较和模仿互动^[32]。企业披露的专利技术、技术更新、技术布局等研发信息,有助于同行企业共同寻找行业增长机会,减少未来不确定性。其次,基于向上比较,大多数企业通常将行业中位企业或者行业中的最优企业作为投资决策的学习参照目标^[33]。除了减少行业增长路径选择的不确定性外,同行企业特别是领先者的信息披露报告可以帮助企业确定有前途的投资机会,比如新品研发、商业模式更新、产品扩张等可能性^[34]。最后,由于获取信息、处理信息的成本高昂,标准规范的信息披露报告为企业管理层提供了创新投资决策信息所需的低成本对比和参考渠道,减少企业创新投资决策的不确定性。为此提出假设:

H₂:市场竞争激烈程度在同行企业研发文本披露与企业研发投入的关系中具有中介促进作用。

H₃:企业信息学习需求在同行企业研发文本披露与企业研发投入的关系中具有中介促进作用。

3. 企业研发文本披露促进企业研发投入作用的异质性

第一,企业自身的融资约束。从企业自身角度看,能否获取充足稳定的资金支持是关乎创新成败的关键。企业良好的融资能力为研发创新提供更好的资金支持,促进企业创新水平的提升。研发创新外源融资主要依靠银行,银行市场结构会对企业研发投资的融资约束产生实质影响^[35]。研发投资不仅具有投入高风险、收益跨期性等突出问题,而且存在较为严重的信息不对称、道德风险问题,与固定资产投资相比,面临的融资约束更为严重。刘胜强等(2015)发现,融资约束使得企业研发投入数额较最优投资水平降低四成之多,导致研发投入不足^[36]。企业外部融资约束限制了企业内部研发支出,对企业技术创新投入具有显著的抑制作用。企业能否有效对标同行企业创新,很大程度需要自身融资能力的配合。

第二,企业获得的财政补贴。除了企业自身金融约束状况,企业面临的创新激励外部政策也会导致企业异质性。考虑到创新风险较高、技术研发的外部性等,各国对企业科研活动实施各项支持政策,企业能敏锐地捕捉到政府创新激励措施带来的潜在收益和资源,主动利用财政补助、税收优惠等激励资源抵消创新成本。政府的资金支持有利于激励企业创新,政府出台的产业政策和实施的财税激励会促使某一行业的企业增加研发投入,使得企业预期能获得更多的财政补贴和税收优惠^[37]。从示范效果看,政府会根据企业公开披露的研发信息材料给予企业相应的财政补贴或者税收优惠。

第三,同行企业的信息可比性。可比性是信息披露的重要质量属性,可帮助信息使用者识别和比较各类信息,优化其投资决策^[38]。首先,可比性作为投资者选择参照对象的重要标准,可降低信息加工成本。当目标企业的财务报告与同行企业相比表现出高可比性时,收购者能够做出更好的收购投资决策,协同效应和运营绩效更好,促进更高效的资本配置^[39]。当可比性较高时,主要股东或高管能通过同行业竞争者信息来推断目标公司的重要信息,对投资项目做充分比较、辨别和论证^[40]。其次,高可比性的同行企业对外报告能够提供有助于投资者准确预测的信息。通过对比目标企业与可比企业的财务状况、经营成果和现金流量,投资者能在事前清晰地认识企业创新的必要性,在事中有有效监督企业高管对创新项目的投入及估计创新性投资对企业短期业绩的不利影响,并在事后对企业创新做出更为客观的综合评价^[41]。同行企业会计信息可比性越高,同行企业对目标公司盈余公告的反应越强,可比性促进了盈余信息的跨公司传递^[42]。为此提出假设:

H₄:同行企业研发文本披露对融资约束低的目标企业的研发投入正向影响更为显著。

H₅:获得财政补贴较多行业的同行企业研发文本披露对目标企业的研发投入正向影响更为显著。

H₆:信息可比性强的同行企业研发文本披露对目标企业的研发投入正向影响更为显著。

三、模型和描述统计

(一) 数据

本文选取2009—2020年沪深A股上市企业年度报告和相关数据展开分析,其中,研发信息披露指标在文构财经文本数据平台(WinGo)完成文本数据挖掘,其他财务特征数据来自深圳希施玛数据科技有限公司中国经济金融研究(CSMAR)数据库。行业按照证监会《上市企业行业分类指引(2012年修订)》分类,制造业按照大类、其他行业按照门类进行行业分类,共70个行业。

按如下步骤剔除部分样本:(1)剔除ST公司;(2)剔除金融类公司;(3)剔除资产总额、固定资产净值小于0以及资产负债率大于1的样本;(4)剔除销售收入增长率或资产增长率超过100%的观测样本;(5)剔除行业内公司少于五家的样本。最终确定29816家公司-年度观测值。为避免异常值干扰,对连续变量在1%和99%水平上进行缩尾处理。

(二) 模型设定

为检验基准假设H₁,借鉴已有研究^[16],建立基准模型:

$$RD_{ijt+1} = \alpha_0 + \alpha_1 PeerRDTEXT_{kjt} + \sum ControlVariables_{ijt} + \sum Year + \sum Industry + \varepsilon_{ijt} \quad (1)$$

式中, RD_{ijt+1} 为被解释变量,表示j行业目标企业i在t+1年的研发投入。 $PeerRDTEXT_{kjt}$ 是解释变量,表示j行业的同行企业k在t年的研发文本披露。 $ControlVariables_{ijt}$ 是控制变量,表示j行业同行企业、目标企业在t年的基本面特征指标,包括企业规模(SIZE)、现金持有(CASH)、资产负债率(DEBT)、总资产收益率(ROA)、企业年龄(AGE)、资本人力比(KL)等。

本文的变量定义见表1。

表1 变量定义及说明

变量符号	变量名称	变量说明
RD	企业研发投入	研发投入/营业收入
PeerRDTEXT	同行企业的研发文本披露	除目标企业的同行其他企业的研发文本信息披露
RDTEXT	企业的研发文本披露	同行所有企业的研发文本信息披露
SIZE	公司规模	总资产的自然对数
AGE	企业年龄	企业上市年份至统计年份年数长短赋值
CASH	现金持有	现金等价物/总资产
ROA	资产收益率	净利润/总资产
LEV	资产负债率	总负债/总资产
SUBSIDY	财政补贴	政府补助/销售收入
BOARD	董事会人数	企业董事会人数的对数

表1(续)

变量符号	变量名称	变量说明
<i>MNGHOLD</i>	管理层持股比例	管理层持股数量占企业发行股份总数的比例
<i>KL</i>	资本员工比	每一员工人力占用固定资产
<i>TOBINQ</i>	托宾 <i>Q</i>	企业股票市值/总资产

(三) 研发文本披露 (*RDTEXT*) 构建

描述性研发文本披露是企业披露的与真实研发活动有关的文本信息。在文本分析领域,词频是目前度量文本词汇重视程度的常见方式。不同于传统词典法,本文通过 WinGo 数据平台采用“种子词+相似词”词向量方式对企业描述性研发文本信息进行统计以衡量研发文本披露水平,采集范围限定在新企业会计准则全面实施之后的企业年度报告。具体构建过程如下:

第一步,确定种子词集。通过翻译现有英文文献中与描述性研发相关的词汇来确定种子词集。现有文献主要分为两类:一是与研发活动相关的研究^[43],二是广义创新的相关研究^[44]。考虑到描述性研发以技术创新为主,以第一类研究为主要参考确定种子词集,同时参照财报样本进行校验。为减少第一类统计错误的发生概率,在人工筛选的过程中剔除了诸如“突破”“创造”等词汇,原因在于:(1)“突破”等词汇在财报文本中除了与技术创新相关以外,还和盈余水平紧密相关,这类词汇的存在会让描述性研发指标因携带过多盈余信息而失去准确性;(2)“创造”等词汇的外延远远超出了技术创新的范畴,无法用来准确揭示相关变量的特征。

第二步,扩充相似词。现有文献使用的相似词扩充方法多为人工扩充法和通用同近义词扩充法。人工扩充法主要基于研究人员的主观经验判断,可靠性和可复制性较弱;通用同近义词词典扩充法普通语料较为适用,但对年报等财经专业语料的适用性较弱^[6]。Word2Vec 神经网络模型是深度学习领域的里程碑式成果^[7]。Word2Vec 模型根据上下文内容将词汇表征为多维向量,通过计算向量的相似度得到词汇间的语义相似性。该模型不仅能克服稀疏向量表征的维数灾难,而且能克服中文词汇的语境鸿沟。本文依照相似度大小对扩充词进行排序,得到种子词汇的前 200 个相似词汇。经过去除重复词汇和部分低频词汇之后,总共得到了 376 个描述性研发关键词。将该组词汇词频占该企业年度报告总词频的比例乘以 100 后得到企业研发文本披露的总指标 *RDTEXT*,指标值越大,代表企业披露的研发信息越丰富。出于稳健性考虑,单独测量了管理层分析和展望(MD&A)部分的研发文本披露指标 *RDTEXT_MDA*,将其作为稳健性检验的替代变量。

第三步,指标有效性验证。一是相关性验证。将研发文本披露指标和定量研发指标进行相关性验证,选取研发强度和申请专利作为衡量企业创新水平的传统指标,进行相关性分析,结果显示研发文本披露指标与传统创新指标呈现出较高且显著的正相关关系,且时间趋势和行业分布较为一致。二是邀请专家对研发文本披露指标进行核验,按行业抽取描述性研发词频比例最高的前 5 家公司,认真研读其年度报告所披露的企业研发相关信息,将企业在公司官网、媒体报道等其他渠道发布的研发信息披露指标进行比较,认为研发文本披露指标能够较为准确地衡量企业的研发水平。

第四步,指标基本面情况说明。表 2 报告了研究样本的 *RDTEXT* 指标的时间趋势以及行业差异性。如

表2所示,研发文本词频占比的均值逐年稳步上升,从2008年的0.0052增加到2020年的0.0088,尤其是2014年后,研发文本词频占比均值增长迅速,这与中国强调创新驱动发展战略下上市企业研发创新的事实特征符合。同时,尽管企业研发文本词频呈增长态势,但是标准差从2008年的0.0026变成2020年的0.0083,呈现逐年扩大的趋势,说明企业之间研发创新行为的差距越来越大。

表2 研发文本披露(RDTEXT)年度词频分布

年份	观测数	均值	标准差	最小值	25%分位数	50%分位数	75%分位数	最大值
2008	1 162	0.005 2	0.002 6	0.000 9	0.003 4	0.004 7	0.006 4	0.024 5
2009	1 379	0.005 2	0.002 6	0.000 8	0.003 5	0.004 8	0.006 4	0.024 9
2010	1 494	0.005 6	0.002 8	0.000 9	0.003 6	0.005 1	0.007 1	0.023 6
2011	1 587	0.006 1	0.003 3	0.000 6	0.003 8	0.005 4	0.007 7	0.028 7
2012	1 937	0.006 5	0.003 7	0.000 7	0.004 0	0.005 7	0.008 1	0.032 0
2013	2 238	0.006 8	0.003 9	0.000 2	0.004 0	0.005 8	0.008 6	0.031 3
2014	2 401	0.008 5	0.004 3	0.001 1	0.005 2	0.007 6	0.010 9	0.033 9
2015	2 435	0.008 2	0.004 0	0.000 7	0.005 2	0.007 5	0.010 5	0.030 1
2016	2 557	0.007 5	0.003 6	0.000 5	0.004 8	0.006 9	0.009 5	0.024 8
2017	2 782	0.008 3	0.004 1	0.000 6	0.005 2	0.007 8	0.010 8	0.033 9
2018	2 997	0.008 7	0.004 1	0.001 4	0.005 5	0.008 2	0.011 3	0.030 9
2019	3 336	0.008 9	0.004 1	0.001 0	0.005 7	0.008 4	0.011 4	0.029 1
2020	3 511	0.008 8	0.004 0	0.001 4	0.005 8	0.008 3	0.011 2	0.028 3
全部样本	29 816	0.007 8	0.004 0	0.000 2	0.004 7	0.007 0	0.010 1	0.033 9

为呈现企业研发文本披露程度的行业差异,将上市企业研发文本披露指标进行分行业统计,统计结果如表3所示,平均研发文本披露程度较高的行业主要集中在创新、研究属性较强的行业,比如科学研究与技术服务业研发词频均值达到0.0119,信息设备仪表制造业研发词频均值为0.0102,信息软件和信息信息技术服务业的研发词频均值达到0.0100,这些行业明显属于技术含量较高的板块。同时,因竞争较激烈,其行业内研发词频的标准差也最高,分别为0.0052、0.0042、0.0044,表明行业内披露意愿波动较大。相比而言,批发零售、交通仓储和邮政、住宿餐饮业的研发词频较低,行业内标准差也较小。研发文本披露的行业分布特征基本符合当前的现实情况,即科创行业的企业会更加注重其技术研发的信息披露。

表3 研发文本RDTEXT行业词频分布

行业代码	行业名称	观测数	均值	标准差	最小值	50%分位数	最大值
A	农林牧渔	371	0.007 5	0.004 7	0.001 5	0.006 5	0.025 9
B	采矿业	822	0.005 5	0.003 2	0.001 3	0.004 7	0.025 9
C1	纺织服装毛皮	1 235	0.006 2	0.002 4	0.001 4	0.005 9	0.017 7
C2	木材家具印刷化工	5 594	0.008 6	0.003 9	0.000 9	0.007 9	0.026 5
C3	金属冶炼设备制造	9 881	0.009 1	0.004 0	0.000 9	0.008 6	0.033 9
C4	信息设备仪表制造	1 355	0.010 2	0.004 2	0.000 7	0.010 2	0.021 9
D	电力热力燃气及水生产供应业	1 197	0.004 5	0.001 9	0.000 9	0.004 2	0.014 0

表3(续)

行业代码	行业名称	观测数	均值	标准差	最小值	50%分位数	最大值
E	建筑业	752	0.007 6	0.003 1	0.000 9	0.007 2	0.020 5
F	批发零售	1983	0.004 7	0.002 3	0.000 2	0.004 3	0.019 3
G	交通仓储和邮政	978	0.003 9	0.001 9	0.000 6	0.003 6	0.014 0
H	住宿餐饮业	124	0.004 4	0.002 6	0.000 5	0.004 2	0.010 4
I	信息软件和信息技术业	1987	0.010 0	0.004 4	0.000 7	0.009 6	0.026 0
K	房地产业	1 481	0.005 6	0.001 9	0.001 5	0.005 3	0.014 9
L	租赁和商务服务	354	0.004 9	0.002 2	0.001 1	0.004 6	0.013 6
M	科学研究和技术服务	295	0.011 9	0.005 2	0.001 7	0.011 7	0.030 9
N	水利环境和公共设施管理	382	0.007 1	0.003 9	0.001 3	0.006 1	0.019 8
P	教育	156	0.005 9	0.004 2	0.002 4	0.004 0	0.017 9
Q	卫生和社会工作	142	0.007 4	0.004 6	0.001 4	0.005 9	0.022 4
R	文化体育和娱乐业	426	0.005 6	0.002 4	0.000 7	0.005 2	0.013 7
S	综合	301	0.006 2	0.003 1	0.001 6	0.005 4	0.015 0
全部样本		29 816	0.007 8	0.004 0	0.000 2	0.007 0	0.033 9

(四) 描述统计和相关系数

变量描述性统计如表4所示,企业研发投入(*RD*)均值为0.029 5,最小值为0,最大值为0.219 0,可看出企业总体研发投入水平合理,部分企业研发投入明显偏低。企业研发文本披露(*RDTEXT*)最低占比为0.000 2,最高达0.033 9,平均占比为0.007 9;同行企业研发文本披露(*PeerRDTEXT*)最低占比为0.000 7,最高占比为0.022 8,平均占比为0.007 9,说明企业研发投入差异较大,而且研发文本披露同样差异较大。从行业均值和中位数可以看出,各个行业的企业研发投入和研发文本披露同样差异较大,说明同行企业研发文本披露对目标企业研发投入的影响具有统计意义。

表4 变量描述性统计

变量	均值	标准差	最小值	25%分位数	50%分位	75%分位数	最大值
<i>RD</i>	0.029 5	0.039 7	0	0	0.018 4	0.042 1	0.219 0
<i>PeerRDTEXT</i>	0.007 9	0.002 5	0.000 7	0.005 5	0.008 0	0.010 2	0.022 8
<i>RDTEXT</i>	0.007 9	0.004 1	0.000 2	0.004 8	0.007 1	0.010 2	0.033 9
<i>SIZE</i>	22.010 0	1.289 0	19.550 0	21.060 0	21.840 0	22.760 0	25.910 0
<i>AGE</i>	0.020 0	0.009 1	0	0.013 9	0.022 0	0.027 7	0.032 2
<i>CASH</i>	0.019 2	0.014 2	0.001 4	0.009 3	0.015 2	0.024 9	0.070 7
<i>ROA</i>	0.045 9	0.058 8	-0.169 0	0.016 2	0.041 7	0.075 0	0.226 0
<i>LEV</i>	0.042 8	0.020 9	0.004 90	0.026 0	0.042 2	0.058 8	0.088 7
<i>SUBSIDY</i>	0.157 0	0.019 2	0.099 0	0.146 0	0.158 0	0.169 0	0.202 0
<i>BOARD</i>	0.021 9	0.002 8	0.013 9	0.020 8	0.022 0	0.024 0	0.028 9

表4(续)

变量	均值	标准差	最小值	25%分位数	50%分位	75%分位数	最大值
<i>MNGHOLD</i>	0.125 0	0.201 0	0	0	0.000 8	0.210 0	0.693 0
<i>KL</i>	0.007 5	0.051 9	0	0.001 3	0.002 6	0.005 2	3.381 0
<i>TOBINQ</i>	0.020 6	0.012 9	0.009 1	0.012 7	0.016 4	0.023 5	0.085 3

表5报告了变量的相关系数。结果显示,同行企业研发文本披露(*PeerRDTEXT*)与企业研发投入(*RD*)的相关系数为0.577,在1%水平上显著正相关,说明同行企业研发文本披露越多,越有助于促使目标企业增加研发投入,初步支持假设H₁。同时,同行企业研发投入(*PeerRD*)与目标企业研发投入(*RD*)呈现显著正相关关系。此外,方差膨胀因子VIF值显示,最高为2.23,平均1.52,不超过4,不存在严重共线性问题。

表5 主要变量的皮尔逊相关系数

变量	分组1						
	<i>RD</i>	<i>PeerRDTEXT</i>	<i>RDTEXT</i>	<i>SIZE</i>	<i>AGE</i>	<i>CASH</i>	<i>ROA</i>
<i>RD</i>	1						
<i>PeerRDTEXT</i>	0.577***	1					
<i>RDTEXT</i>	0.532***	0.595***	1				
<i>SIZE</i>	-0.199***	-0.180***	-0.134***	1			
<i>AGE</i>	-0.288***	-0.233***	-0.320***	0.374***	1		
<i>CASH</i>	0.227***	0.129***	0.181***	-0.247***	-0.354***	1	
<i>ROA</i>	0.077***	0.061***	0.140***	-0.042***	-0.287***	0.305***	1
<i>LEV</i>	-0.355***	-0.270***	-0.282***	0.473***	0.399***	-0.411***	-0.383***
<i>SUBSIDY</i>	0.034***	0.035***	0.073***	0.426***	0.055***	-0.059***	0.053***
<i>BOARD</i>	-0.114***	-0.131***	-0.094***	0.242***	0.162***	-0.104***	-0.069***
<i>MNGHOLD</i>	0.360***	0.333***	0.346***	-0.312***	-0.572***	0.251***	0.200***
<i>KL</i>	-0.061***	-0.106***	-0.081***	0.106***	0.053***	-0.068***	-0.022***
<i>TOBINQ</i>	0.163***	0.138***	0.057***	-0.407***	0.058***	0.107***	0.139***
变量	分组2						
	<i>LEV</i>	<i>SUBSIDY</i>	<i>BOARD</i>	<i>MNGHOLD</i>	<i>KL</i>	<i>TOBINQ</i>	
<i>LEV</i>	1						
<i>SUBSIDY</i>	0.174***	1					
<i>BOARD</i>	0.165***	0.140***	1				
<i>MNGHOLD</i>	-0.364***	-0.079***	-0.202***	1			
<i>KL</i>	0.051***	0.050***	0.046***	-0.056***	1		
<i>TOBINQ</i>	-0.243***	-0.122***	-0.069***	-0.002	-0.037***	1	

注: *、**、*** 分别表示 10%、5%、1% 的显著性水平,括号内为 *t* 值。后表同。

四、回归检验分析

(一) 同行企业研发文本披露对目标企业研发投入的溢出效应

为检验基准假设 H_1 , 采用基准模型进行回归分析。被解释变量采用滞后一期 ($F.RD$) 表示, 以缓解潜在的内生性问题, 滞后一期使得模型回归观测数为 25 230 个, 检验结果如表 6 所示。从表 6 列(1)可看出, 同行企业研发文本披露 ($PeerRDTEXT$) 对目标企业研发投入 ($F.RD$) 的影响系数为 4.722 6, 在 1% 水平上正向显著。列(2)、列(3)显示, 在分别增加目标企业控制变量后, 同行企业研发文本披露对目标企业研发投入的影响系数为 1.793 1、1.774 1, 依然在 1% 水平上正向显著; 列(4)、列(5)显示, 在继续加入行业和年份虚拟变量后, 同行企业研发文本披露对目标企业研发投入的影响系数为 0.389 4、0.371 2, 分别在 5% 水平上正向显著。可以看出, 同行企业披露的研发文本信息越多, 目标企业的研发投入增加越多, 存在溢出效应。

基准模型检验表明, 同行企业研发文本披露确实存在溢出效应, 假设 H_1 得到验证。相比数字形式的财务信息来说, 丰富多样的文本信息能够提供更多与技术研发创新相关的重要描述与内容表达。同行企业披露的研发文本是目标企业进行研发决策信息组合的一个重要参考内容, 为了减少创新不确定性, 目标企业需要不断与同行业竞争企业比较, 获取行业对手研发文本信息来调整自身的研发投入。

表 6 同行企业研发文本披露对目标企业研发投入的溢出效应

变量	(1)	(2)	(3)	(4)	(5)
$PeerRDTEXT$	4.722 6*** (17.856 0)	1.793 1*** (11.091 1)	1.774 1*** (10.930 4)	0.389 4** (2.487 5)	0.371 2** (2.358 0)
$RDTEXT$		0.418 8*** (8.154 4)	0.417 9*** (8.085 5)	0.332 8*** (8.054 5)	0.332 2*** (8.042 1)
RD		0.537 4*** (98.635 2)	0.536 9*** (98.179 5)	0.841 8*** (100.878 5)	0.840 9*** (100.033 4)
$SIZE$		0.001 3*** (5.185 1)	0.001 5*** (5.465 4)	-0.000 3** (-2.537 3)	-0.000 1 (-0.939 9)
AGE		-0.288 7*** (-7.489 2)	-0.308 8*** (-7.558 4)	-0.036 9*** (-2.632 0)	-0.045 5*** (-2.826 1)
$CASH$		0.024 1** (2.064 6)	0.024 7** (2.110 9)	0.046 2*** (4.456 8)	0.046 4*** (4.461 6)
ROA		0.010 8*** (4.455 7)	0.010 1*** (4.015 9)	0.015 8*** (6.157 3)	0.014 6*** (5.475 7)
LEV		-0.032 0*** (-2.924 5)	-0.034 7*** (-3.140 8)	-0.027 4*** (-3.601 1)	-0.026 3*** (-3.451 9)
$SUBSIDY$		0.023 9*** (3.016 0)	0.024 5*** (3.085 2)	0.034 8*** (5.807 6)	0.035 0*** (5.813 9)

表6(续)

变量	(1)	(2)	(3)	(4)	(5)
<i>BOARD</i>			-0.041 3 (-0.834 3)		-0.064 9* (-1.741 5)
<i>MNGHOLD</i>			0.000 1 (0.063 5)		0.000 4 (0.537 0)
<i>KL</i>			-0.000 4 (-0.136 8)		-0.000 1 (-0.105 2)
<i>TOBINQ</i>			0.023 6* (1.831 5)		0.030 0** (2.045 2)
行业	控制	未控制	未控制	控制	控制
年份	控制	未控制	未控制	控制	控制
观测数	25 230	25 230	25 230	25 230	25 230
$\overline{R^2}$	0.434	0.425	0.425	0.847	0.847

(二) 内生性问题

针对受行业共同因素影响引发的问题,参考已有研究^[10],本文进一步引入股票特质收益率(*Idoret*)工具变量,采用系统广义矩估计(SYS-GMM)和差分 GMM 估计进行内生性检验。

股票特质收益率工具变量构建步骤如下:

$$r_{ijt} - rf_{it} = \alpha + \beta(\bar{r}_{ijt} - rf_{it}) + \beta MKT + \beta SMB + \beta HML + \eta_{ijt} \quad (2)$$

式中, r_{ijt} 表示 j 行业的 i 公司在 t 月的股票收益率; rf_{it} 表示 t 月的无风险收益率; $r_{ijt} - rf_{it}$ 表示公司 i 的同行公司在 t 月的平均股票收益率。*MKT*、*SMB*、*HML* 分别表示法玛模型中的市场、规模、账面市值比三个因子。在每年的年初,使用前 36 个月的数据进行回归,得到回归系数。在年度内的每个月,使用相同的回归系数,计算每只股票每月超额收益率的期望值 $\widehat{r_{ijt} - rf_{it}}$ 和股票特质收益率 $\widehat{\eta_{ijt}}$, 计算方法如下:

$$\widehat{r_{ijt} - rf_{it}} = \alpha + \beta(\widehat{r_{ijt} - rf_{it}}) + \beta MKT + \beta SMB + \beta HML \quad (3)$$

$$\widehat{\eta_{ijt}} = (r_{ijt} - rf_{it}) - (\widehat{r_{ijt} - rf_{it}}) \quad (4)$$

然后,将每个月的特质收益率复合,得到某年度的特质收益率。对样本内的每只股票在每个年度进行相同操作,得到所有股票每年度的特质收益率。

选取股票特质收益率作为工具变量的原因在于:一是股票特质收益率仅包含个股的特有信息,不含能够影响整个市场和行业的因素,具有工具变量所需的外生性;二是企业个股股票价格会对企业创新决策产生影响,与内生变量具有相关性。

同时,滞后期被解释变量(*F. RD*)和解释变量(*PeerRDTEXT*)高度相关,符合面板数据广义矩估计的条件,适合采用 GMM 估计法,包括 SYS-GMM 和差分 GMM。首先,利用差分 GMM 进行工具变量检验,如表 7 列(1)所示,同行企业研发文本披露(*PeerRDTEXT*)对目标企业创新投入(*F. RD*)的影响系数为 4.572 6,显著为正。其次,利用 SYS-GMM 估计检验。汉森(Hansen)检验的 P 值大于 0.1,说明选取的工具变量不存在过度识别问题;IV 类型工具变量检验的 P 值大于 0.1,通过 SYS-GMM 估计的工具变量外生性检验。如表 7

列(2)所示, SYS-GMM 估计的系数显著为正。上述两个 GMM 工具变量回归结果说明, 在排除同行企业受到行业外部冲击同时影响的前提下, 同行企业研发文本披露确实正向显著影响目标企业的研发投入, 存在溢出效应, 均支持假设 H_1 。

表 7 工具变量的 GMM 估计

变量	差分 GMM	SYS-GMM
<i>PeerRDTEXT</i>	4.572 6*** (5.423 1)	4.575 3*** (5.428 8)
控制变量	控制	控制
行业	控制	控制
年份	控制	控制
观测数	18 534	21 609

同时, 本文采用两种方法来缓解其他可能的内生性问题。一是采用残差法回归方法。对同行企业研发文本披露(*PeerRDTEXT*)和目标企业研发文本披露(*RDTEXT*)进行回归, 得到同行企业研发文本披露残差(*PeerRDTEXTres*), 表示去除行业基本研发文本信息后的同行企业研发文本信息增量。将该残差放入基准回归模型进行回归, 同行企业研发文本披露残差的系数依然显著为正, 如表 8 列(1)所示。二是采用公司固定效应回归方法。为解决遗漏变量所造成的可能内生性问题, 在模型(1)的回归中加入公司固定效应, 替换掉行业固定效应。如表 8 列(2)所示, 同行企业研发文本披露的影响系数为 1.774 1, 在 1%水平上显著, 这表明在控制了公司固定效应之后, 同行企业研发文本披露与目标企业研发投入之间的正向关系仍然成立, 遗漏变量问题并不影响研究结论。

表 8 内生性检验

变量	残差法回归	公司固定效应
<i>PeerRDTEXTres</i>	0.433 2*** (2.592 5)	
<i>PeerRDTEXT</i>		1.774 1*** (10.213 0)
控制变量	控制	控制
行业	控制	控制
年份	控制	控制
观测数	25 230	25 230
\bar{R}^2	0.847	0.884

(三) 稳健性检验

1. 替换解释变量

对管理层分析和展望(MD&A)部分进行研发文本分析, 单独采集以获得研发文本披露指标 *PeerRDTEXT_MDA*。将解释变量变更为 *PeerRDTEXT_MDA*, 替换掉解释变量 *PeerRDTEXT*, 代入基准回归模型, 其他控制

变量不变。回归结果(不再详细列示, 备索)均表明, 同行企业研发文本披露对目标企业研发投入的影响系数显著为正, 结果与基准假设一致。

2. 变更被解释变量指标计算方式

变更衡量指标为: 研发投入/期末总资产(RD_A); 用研发投入的行业和年度中位数填补缺失值(RD_M)。将被解释变量($F.RD$)替换为 $F.RD_A$ 、 $F.RD_M$, 分别代入基准回归模型, 其他控制变量不变。两个回归结果(不再详细列示, 备索)均表明, 同行企业研发信息对目标企业研发投入的影响系数显著为正, 结果与基准假设一致。

3. 采用被解释变量不同滞后期

采用当期(RD)、未来一期($F.RD$)、未来二期($F2.RD$)、未来三期($F3.RD$)的研发投入作为被解释变量。回归结果(不再详细列示, 备索)显示, 不像财务数据存在历史年份的勾稽关系, 研发文本信息存在跳跃性变化, 同行企业研发文本披露对目标企业后续年份研发投入的溢出效应逐年递减, 从另一个侧面支持基准假设的结论。

4. 增加宏观解释变量

借鉴黎文靖和郑曼妮(2016)^[37]的研究, 增加国内生产总值($\ln GDP$)、财政支出额($\ln FExp$)、居民消费价格指数(CPI)等宏观经济指标变量, 加入基准模型进行回归检验。结果(不再详细列示, 备索)表明, 同行企业研发文本披露对目标企业研发投入的影响系数均显著为正, 结果与前述假设一致。

五、溢出效应产生机制检验

(一) 产生机制检验: 市场竞争机制

采用中介效应模型来检验假设, 以市场竞争度($COMP$)为中介变量建立回归模型:

$$RD_{ijt+1} = \alpha_0 + \alpha_1 PeerRDTEXT_{kjt} + \sum ControlVariables_{ijt} + \sum Year + \sum Industry + \varepsilon_{ijt} \quad (5)$$

$$COMP_{ijt+1} = \alpha_0 + \alpha_1 PeerRDTEXT_{kjt} + \sum ControlVariables_{ijt} + \sum Year + \sum Industry + \varepsilon_{ijt} \quad (6)$$

$$RD_{ijt+1} = \alpha_0 + \alpha_1 PeerRDTEXT_{kjt} + \alpha_2 COMP_{ijt+1} + \sum ControlVariables_{ijt} + \sum Year + \sum Industry + \varepsilon_{ijt} \quad (7)$$

式中, 市场竞争度($COMP$)采用市场信用(NTC)、行业地位($INDP$)衡量。其中, 市场信用表示企业应收账款与营业收入之比; 行业地位用当期企业营业收入与其所在行业收入之比表示。市场信用、行业地位指标数值越小, 意味着行业竞争越激烈。市场竞争机制检验采用中介效应模型, 分别以市场信用、行业地位为中介变量进行检验。

从表9列(2)、列(3)可以看出, 同行企业研发文本披露对市场竞争度(市场信用)、市场竞争度(市场信用)对企业研发投入的影响系数均显著为正, $Sobel-P$ 值检验结果均显著, 即存在明显的中介效应, 说明行业市场竞争激烈程度在同行企业研发文本披露与目标企业研发投入之间起着部分中介作用。企业面临的市场竞争越激烈, 目标企业研发决策信息组合受同行企业研发文本披露的影响越大, 市场竞争激烈程度较高有助于企业研发活动互动与企业研发行为市场化。从表9列(5)、列(6)可以看出, 同行企业研发文本披露对市场竞争度(行业地位)、市场竞争度(行业地位)对企业研发投入的影响系数也都显著为正, $Sobel-P$ 值检验结果说明存在明显的中介效应。这两个中介效应检验均表明, 市场竞争度在同行企业

研发文本披露与目标企业研发投入之间起着中介作用,行业市场的竞争机制是形成同行企业研发文本披露的溢出效应的内在机制之一,市场竞争越激烈,同行企业研发文本披露越会显著影响目标企业研发投入,支持假设 H₂。

表 9 市场竞争机制检验

变量	<i>F</i> _{RD}	<i>NTC</i>	<i>F</i> _{RD}	<i>F</i> _{RD}	<i>INDP</i>	<i>F</i> _{RD}
	(1)	(2)	(3)	(4)	(5)	(6)
<i>PeerRDTEXT</i>	0.371 2** (2.358 0)	0.490 4*** (9.775 6)	0.513 9*** (2.934 3)	0.371 2** (2.358 0)	2.210 8*** (7.100 7)	0.512 8*** (2.928 0)
<i>NTC</i>			0.033 0*** (4.025 8)			
<i>INDP</i>						0.035 7*** (3.962 5)
控制变量	控制	控制	控制	控制	控制	控制
行业	控制	控制	控制	控制	控制	控制
年份	控制	控制	控制	控制	控制	控制
观测数	25 230	26 137	22 282	25 230	26 137	22 282
$\overline{R^2}$	0.847	0.138	0.824	0.847	0.150	0.824
<i>Sobel-P</i>		0.008			0.005	

(二) 产生机制检验:信息学习机制

对于假设 H₃,参考陆蓉和常维(2018)^[15]的做法,按照企业规模(*Size*)、市场占有率(*Income*)、市场价值(*Market*)三类行业地位指标将行业企业划分为领先者和跟随者子样本,在基准模型的基础上分别回归进行对比检验。具体地,将企业规模、市场占有率、市场价值排名前 30%的行业内企业视为领先者,后 30%的行业内企业视为跟随者。在分析行业跟随者(领先者)对领先者(跟随者)的反应时,采用跟随者(领先者)的子样本,同时在计算解释变量同行企业研发文本披露指标(分别为 *PeerRDTEXT_Size*、*PeerRDTEXT_Income*、*PeerRDTEXT_Market*)时,仅采用行业内所有领先者(跟随者)作为同行企业,重新计算平均值,使用基准模型方程进行子样本回归。

表 10 列(1)、列(2)和列(3)的结果表明,按企业规模、市场占有率、市场价值划分的行业跟随者对领先者三个子样本检验的影响系数均显著为正,说明处于跟随地位的目标企业研发决策的确存有向行业领先企业学习的倾向,研发决策信息组合显著受到同行企业研发文本披露的影响,存在同行领先企业研发文本披露的溢出效应。由表 10 列(4)、列(5)和列(6)可知,行业领先者对跟随者的反应不显著,说明行业内领先企业研发决策更具独立性,决策信息组合更少受到同行企业研发文本披露的影响。这三个检验对比结果说明,信息学习机制是形成上市企业研发文本披露溢出效应的重要途径。信息学习需求一定程度能够凸显同行企业研发文本披露的溢出方向,即同行企业领先者创新文本信息披露增加,目标追随者企业的研发投入会相应增加,支持假设 H₃。

表 10 信息学习机制检验(规模、收入、市值)

变量	跟随者企业			领先者企业		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>PeerRDTEXT_Size</i>	0.6507** (2.3214)			0.0731 (0.3158)		
<i>PeerRDTEXT_Income</i>		0.7816*** (5.3481)			0.3042 (1.5978)	
<i>PeerRDTEXT_Market</i>			1.0039*** (3.0511)			0.5177 (1.6237)
控制变量	控制	控制	控制	控制	控制	控制
行业	控制	控制	控制	控制	控制	控制
年份	控制	控制	控制	控制	控制	控制
观测数	7533	7565	7472	7557	7566	7483
$\overline{R^2}$	0.860	0.856	0.853	0.854	0.853	0.786

六、异质性检验

(一) 异质性检验 I

在基准模型基础上,将融资约束与研发文本披露的乘积项加入,对异质性假设 H_4 进行检验。其中,融资约束指标依据 KZ 指数^①和 WW 指数^②衡量, KZ 指数和 WW 指数数值越大,融资约束越严重。若该乘积项的系数显著,则表明异质性影响成立。

回归结果显示,乘积项 $KZ \times PeerRDTEXT$ 的系数为 -0.0671 ,显著为负。这个结果表明,受限于融资能力,受到融资约束的目标企业自身缺少资金支持研发投入,缺乏条件实施研发投资,受到同行企业研发文本信息影响较弱。相对地,不受融资约束的目标企业能够筹集足够资金支持研发创新,密切关注同行业其他企业研发信息,跟进同行其他企业研发活动和创新,持续投入参与竞争。乘积项 $WW \times PeerRDTEXT$ 的系数为 -0.1259 ,显著为负,同样支持类似结论。从目标企业自身融资约束条件的两个调节作用检验结果(限于篇幅不再列示,备索)可知,目标企业融资约束程度越低,资金配套能力越强,同行企业研发文本披露对目标企业研发投入的溢出效应越强,支持假设 H_4 。

(二) 异质性检验 II

类似地,在基准模型基础上,将财政补贴与研发文本披露的乘积项纳入回归方程,对异质性假设 H_5 进行检验。其中,财政补贴指标使用企业收到的政府补助占销售收入的比例(*Subsidy*)和政府补助占总资产的

① KZ 指数的计算方法是:对全样本各个年度都按经营性净现金流/上期总资产(CF)、现金股利/上期总资产(Div)、现金持有/上期总资产($Cash$)、资产负债率($Debt$)和托宾 Q ($TobinQ$)分类构建 50%分位数,小于 50%分位数的,取值为 1,否则为 0,将五个变量的(1,0)取值加总得到 $KZindex$; $KZindex$ 对 CF 、 Div 、 $Cash$ 、 $Debt$ 和 $TobinQ$ 五个变量进行有序逻辑模型回归;在估计模型参数后,计算模型拟合值,即融资约束指数 KZ 。

② WW 指数的计算方法是: $WW = -0.091CF - 0.062DivPOS + 0.021TLTD - 0.044Size + 0.102ISG - 0.035SG$, 其中, CF 是经营性净现金流与上期总资产的比值; $DivPOS$ 是分红时取值为 1 的虚拟变量; $TLTD$ 是长期负债与总资产之比; $Size$ 是总资产的自然对数; ISG 是企业所处行业的销售增长率; SG 是企业的销售增长率。

比例(*Subsidy_2*)衡量。

回归结果显示:乘积项 $Subsidy \times PeerRDTEXT$ 的系数为 5.674 5,显著为正。该结果显示,企业收到的政策资助或其他政策红利能够激励企业增加研发投入,促进了同行业研发文本披露的外溢,有助于目标企业吸收借鉴同行企业研发文本信息。乘积项 $Subsidy_2 \times PeerRDTEXT$ 的系数为 4.369 7,显著为正,同样支持类似结论。检验结果(限于篇幅不再列示,备索)表明,财政补贴更高的同行企业对研发文本披露的溢出效应越强,假设 H5 得到验证。

(三) 异质性检验 III

同样地,为验证异质性假设 H_6 ,在基准模型基础上将信息可比性与研发文本披露的乘积项纳入回归方程。其中,信息可比性选择盈余差异度(*CompAcct*)^①和盈余变化同步性(*EarnCov*)^②两个指标。

回归结果显示:乘积项 $CompAcct \times PeerRDTEXT$ 的系数为 9.136 5,显著为正。该结果显示,企业信息披露可比性增强有助于同行企业研发文本披露的外溢,有助于目标企业对比、吸收同行企业研发文本信息。乘积项 $EarnCov \times PeerRDTEXT$ 的系数为 0.370 4,显著为正,同样支持类似结论。基于信息可比性的两个调节作用检验结果(限于篇幅不再列示,备索)均表明,信息可比性更高的同行企业对研发文本披露的溢出效应越强,假设 H_6 得到验证。

七、结论、启示与展望

(一) 研究结论

基于沪深上市企业文本挖掘数据,本文检验同行企业研发文本披露是否对目标企业研发投入产生溢出效应。本文的结论包括:(1)同行企业研发文本披露显著促进了目标企业的研发投入的增加,存在溢出效应,该结论在工具变量法、残差法、固定效应回归及一系列稳健性检验中均成立。(2)产生溢出效应的作用机制主要有两个。一是市场竞争机制,即市场竞争度在同行企业研发文本披露与目标企业研发投入之间起着中介作用,企业面临的市场竞争越激烈,目标企业创新决策受同行企业研发信息披露的影响越大,行业竞争程度高的行业同行企业研发信息披露溢出效应明显;二是信息学习机制,即处于跟随地位的企业存在向行业领先者的学习需求,行业内企业规模、市场占有率、市场价值排名后 30%的企业创新决策受到排名前 30%同行领先者企业研发信息的影响,存在显著的同行领先者研发信息溢出效应。(3)异质性检验结果显示,同行信息可比性、目标企业融资约束和财政补贴对研发文本披露溢出效应具有异质性影响。具体而言,同行企业研发文本披露对融资约束低的目标企业的研发投入正向影响更为显著;获得财政补贴较多行业的同行企业研发文本披露对目标企业的研发投入正向影响更为显著;信息可比性强的同行企业研发文本披露对目标企业的研发投入正向影响更为显著。

(二) 政策启示

一是适度加强对非财务信息披露的规范监管。有效的信息披露是资本市场走向有序的必经之路,信息

① 使用连续 12 个季度市账率与股票收益率回归获得预期盈余,用两个公司预期盈余的差异绝对值平均后的相反数表示盈余差异度。盈余差异度值越大表示公司 i 与公司 j 之间的会计信息可比性越强。

② 用两公司第 t 期前连续 12 个季度的盈余为被解释变量、解释变量回归后得到的可决定系数 R^2 来表示。同一行业内两公司盈余变化同步性越高,则会计信息可比性越强。

披露规则制定和监管机构有必要对具有投资决策重大参考价值的企业研发、人力资源、商业模式、客户关系和企业未来前景等难以量化的信息,适当作出规范和披露。比如研究阶段支出在费用明细表披露,应完善财务报表附注研发信息披露,或者提供专门的研发报告。

二是制定企业创新信息公开的配套政策。存在不确定收益的科技研发信息披露具有正向外部性和行业聚焦性,即存在同行业企业之间创新行为信息的溢出效应,政府对企业创新行为信息披露应秉持鼓励原则、不限制企业选择披露方式。监管者在制定研发投入费用加计扣除优惠政策时,不仅需要考虑政策对企业研发投入的直接影响,还可考虑企业研发投入信息产生的涟漪扩散效应,制定企业创新信息公开的配套政策。

三是加强科创企业行业信息披露与可比性。研发信息溢出效应意味着资本市场上的信息悖论在一定程度上被打破。随着创新驱动发展模式及科创企业的发展,需要加强信息披露制度有效性改革,在科创板、创业板实施更广泛的研发信息披露框架理念,如针对特定科创行业出台分行业类别的研发投入信息披露指引,规范同一行业研发投入信息披露词汇,进而提高研发投入信息的可比性。

(三) 研究展望

本文的研究不足在于,研发文本信息具有叙述性和可借鉴性较差的特征,即使机器挖掘选取并人工评判,依然存有不足缺憾。未来的研究方向包括:一是采用更新的分析方法,可结合合适的加权方法从句子和段落语义分析文本信息,或采用难度更大的机器学习方法;二是研发文本研究内容和来源进一步细化,将更多的前瞻性、竞争性、研发信息等语句词汇包含于研发文本信息内涵并分类,扩展分析更多的研发文本数据来源,例如专利申请与引用、技术人员招聘信息、企业公众号、科技政策等。

参考文献:

- [1] ARROW K. Economic welfare and the allocation of resources for invention[M]//National Bureau of Economic Research. The rate and direction of inventive activity: economic and social factors. Princeton: Princeton University Press, 1962: 609-626.
- [2] ANTON J J, YAO D A. The sale of ideas: strategic disclosures, property rights, and contracting[J]. The Review of Economic Studies, 2002, 69(3): 513-531.
- [3] LEARY M T, ROBERTS M R. Do peer firms affect corporate financial policy? [J]. The Journal of Finance, 2014, 69(1): 139-178.
- [4] GRENNAN J. Dividend payments as a response to peer influence[J]. Journal of Financial Economics, 2019, 131(3): 549-570.
- [5] 陆蓉, 常维. 近墨者黑: 上市公司违规行为的“同群效应”[J]. 金融研究, 2018(8): 172-189.
- [6] 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.
- [7] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [8] DURNEV A, MANGEN C. The spillover effects of MD&A disclosures for real investment: the role of industry competition[J]. Journal of Accounting and Economics, 2020, 70(1): 101299.
- [9] 底璐璐, 罗勇根, 江伟, 等. 客户年报语调具有供应链传染效应吗? ——企业现金持有的视角[J]. 管理世界, 2020, 36(8): 148-163.
- [10] 孙锦萍, 李亚飞, 侯麟科. 上市公司创新决策的区域同群效应研究[J]. 经济与管理研究, 2022, 43(7): 115-131.
- [11] VERRECCHIA R. Essays on disclosure[J]. Journal of Accounting and Economics, 2001, 32(1/2/3): 97-180.
- [12] JAMES S D, SHAVER J M. Motivations for voluntary public R&D disclosures[J]. Academy of Management Discoveries, 2016, 2(3): 290-312.
- [13] AMIHUD Y, MENDELSON H. Asset pricing and the bid-ask spread[J]. Journal of Financial Economics, 1986, 17(2): 223-249.

- [14] CHEN Z X, LI M, SONG L. Accounting disclosure, governance standards and innovation activities in emerging markets[J]. *Asian Journal of Finance & Accounting*, 2014, 6(2): 142-154.
- [15] DEDMAN E, LIN S W J, PRAKASH A J, et al. Voluntary disclosure and its impact on share prices: evidence from the UK biotechnology sector [J]. *Journal of Accounting and Public Policy*, 2008, 27(3): 195-216.
- [16] CAZIER R A, MERKLEY K J, TREU J S. When are firms sued for making optimistic disclosures? [Z]. SSRN No. 2859596, 2016.
- [17] BAYUS B L, JAIN S, RAO A G. Truth or consequences: an analysis of vaporware and new product announcements[J]. *Journal of Marketing Research*, 2001, 38(1): 3-13.
- [18] HAAN M A. Vaporware as a means of entry deterrence[J]. *The Journal of Industrial Economics*, 2003, 51(3): 345-358.
- [19] SAIDI F, ŽALDOKAS A. How does firms' innovation disclosure affect their banking relationships? [J]. *Management Science*, 2021, 67(2): 742-768.
- [20] MERKLEY K J. Narrative disclosure and earnings performance: evidence from R&D disclosures[J]. *The Accounting Review*, 2014, 89(2): 725-757.
- [21] 伊志宏, 姜付秀, 秦义虎. 产品市场竞争、公司治理与信息披露质量[J]. *管理世界*, 2010(1): 133-141.
- [22] SENARATHNE C W, WEI J G. The impact of patent citation information flow regarding economic innovation on common stock returns: volume vs. patent citations[J]. *International Journal of Innovation Studies*, 2018, 2(4): 137-152.
- [23] SENGUPTA P. Corporate disclosure quality and the cost of debt[J]. *The Accounting Review*, 1998, 73(4): 459-474.
- [24] ROYCHOWDHURY S, SHROFF N, VERDI R S. The effects of financial reporting and disclosure on corporate investment: a review[J]. *Journal of Accounting and Economics*, 2019, 68(2/3): 101246.
- [25] LAMBERT R, LEUZ C, VERRECCHIA R E. Accounting information, disclosure, and the cost of capital[J]. *Journal of Accounting Research*, 2007, 45(2): 385-420.
- [26] JONES D A. Voluntary disclosure in R&D-intensive industries[J]. *Contemporary Accounting Research*, 2007, 24(2): 489-522.
- [27] 王唯. 开放式创新下的知识治理——基于认知视角的跨案例研究[J]. *南开管理评论*, 2009, 12(3): 45-53.
- [28] 赵立雨, 张彦海. 创新网络环境下开放式创新与企业创新绩效关系研究[J]. *科技进步与对策*, 2016, 33(2): 99-102.
- [29] 连立帅, 朱松, 陈关享. 资本市场开放、非财务信息定价与企业投资——基于沪深港通交易制度的经验证据[J]. *管理世界*, 2019, 35(8): 136-154.
- [30] LIEBERMAN M B, ASABA S. Why do firms imitate each other? [J]. *Academy of Management Review*, 2006, 31(2): 366-385.
- [31] GASSMANN O, ENKEL E, CHESBROUGH H. The future of open innovation[J]. *R&D Management*, 2010, 40(3): 213-221.
- [32] BADERTSCHER B, SHROFF N, WHITE H D. Externalities of public firm presence: evidence from private firms' investment Decisions[J]. *Journal of Financial Economics*, 2013, 109(3): 682-706.
- [33] 贺小刚, 李婧, 吕斐斐, 等. 绩优企业的投机经营行为分析——来自中国上市公司的数据检验[J]. *中国工业经济*, 2015(5): 110-121.
- [34] FERRACUTI E, STUBBEN S R. The role of financial reporting in resolving uncertainty about corporate investment opportunities[J]. *Journal of Accounting and Economics*, 2019, 68(2/3): 101248.
- [35] 刘胜强, 林志军, 孙芳城, 等. 融资约束、代理成本对企业 R&D 投资的影响——基于我国上市公司的经验证据[J]. *会计研究*, 2015(11): 62-68.
- [36] 唐清泉, 巫岑. 银行业结构与企业创新活动的融资约束[J]. *金融研究*, 2015(7): 116-134.
- [37] 黎文靖, 郑曼妮. 实质性创新还是策略性创新? ——宏观产业政策对微观企业创新的影响[J]. *经济研究*, 2016, 51(4): 60-73.
- [38] BARTH M E, 李英, 叶康涛. 财务报告的全球可比性——是什么、为什么、如何做以及何时实现[J]. *会计研究*, 2013(5): 3-10.
- [39] CHEN C W, COLLINS D W, KRAVET T D, et al. Financial statement comparability and the efficiency of acquisition decisions[J]. *Contemporary Accounting Research*, 2018, 35(1): 164-202.
- [40] 袁知柱, 张小曼. 会计信息可比性与企业投资效率[J]. *管理评论*, 2020, 32(4): 206-218.
- [41] SHROFF N, VERDI R S, YOST B P. When does the peer information environment matter? [J]. *Journal of Accounting and Economics*, 2017, 64(2/3): 183-214.
- [42] 鲁威朝, 杨道广, 刘思义. 会计信息可比性、需求差异与跨公司信息传递[J]. *会计研究*, 2019(4): 18-25.
- [43] BELLSTAM G, BHAGAT S, COOKSON J A. A text-based analysis of corporate innovation[J]. *Management Science*, 2021, 67(7): 4004-4031.
- [44] 尹康, 洪丽. 系统 GMM 估计方法的比较研究[J]. *统计与决策*, 2020, 36(21): 38-42.

Peer Spillover Effects of Corporate R&D Text Disclosure

FU Haoliang¹, XIA Ting¹, SUN Feng'e²

(1. Guangdong University of Finance & Economics, Guangzhou 510320;

2. Beijing Forestry University, Beijing 100083)

Abstract: Information disclosure is an effective way to improve information paradox. The current treatment of expensing and capitalization of R&D expenditures makes financial information unable to fully reflect the funds of R&D investment, thereby hindering the visibility of specific directions and projects of R&D. The disclosure of diverse non-financial textual information can fill this gap and become the main carrier for external decision-makers to understand a large quantity of R&D and innovation behaviors of enterprises.

Based on the text mining of listed companies in China, this paper constructs R&D text disclosure indicators to investigate whether the R&D text of peer companies has a spillover effect on the R&D investment of target companies. It adopts the word vector technology of “seed vocabulary set+Word2Vec similar word expansion” in the field of machine learning to calculate the frequency of words that match the R&D dictionary in annual reports of listed companies on the WinGo Textual Analytics Database and conduct econometric regression tests. The results show a significant spillover effect of R&D text disclosure of listed companies in China. The target companies' R&D investment increases proportionally with the disclosure of more R&D texts by peer companies. This spillover effect can be achieved through the market competition mechanism and the information learning mechanism. The former in the industry will enhance the significant impact of R&D text disclosure by peer companies on the R&D decision information of target companies; the demand for the latter drives target companies into a follower position to learn from industry leaders. The extended research reveals that the spillover effect is significantly enhanced when the target companies' financing constraints are low, peer companies receive more financial subsidies, and the information of peer companies is highly comparable. Therefore, it is recommended to moderately increase the standardized regulation of non-financial information disclosure, formulate supporting policies for corporate innovation information disclosure, and strengthen the disclosure and comparability of information in the industry of science and innovation companies.

The contributions of this paper lie in the following aspects. First, it analyzes the annual reports of Chinese listed companies through the WinGo Textual Analytics Database to construct more direct and comprehensive R&D information indicators. The relevant methods provide a reference for subsequent measurement of China's R&D text disclosure. Second, it conducts an in-depth analysis of peer R&D information indicators, which is closer to the information needs of corporate innovation investment decisions. Third, it analyzes the mechanism and characteristics of the R&D information spillover effect among horizontal enterprises in the same industry, providing a qualitative information disclosure explanation for the horizontal information impact of enterprises.

Keywords: R&D; information disclosure; spillover effect; text analysis; information comparability

责任编辑:姚望春