

DOI: 10.16538/j.cnki.fem.20260312.401

殊途同归不同效: 数据要素赋能企业创新质量 及其对资源配置效率的影响研究

方修园, 吴金璇, 李元旭

(复旦大学 管理学院, 上海 200433)

摘要:新一轮科技革命和产业革命决定了数据成为赋能企业创新质量的关键生产要素。目前,关于企业如何通过内外部数据要素的有效组合实现高创新质量以及不同实现路径是否会带来差异化结果仍缺乏系统证据。本文整合资源基础理论和开放式创新理论,采用2017—2023年沪深A股上市企业平衡面板数据,运用动态QCA方法,系统分析企业内部数据要素(数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现)与外部数据要素(公共数据开放、数据要素市场化、数据要素集聚)对创新质量的组态影响,运用倾向得分匹配法(PSM)进一步检验不同路径形成的高创新质量对企业资源配置效率的异质性作用。研究发现,企业可通过三类路径实现高创新质量,分别为“内外部数据资源协同型”“内部数据资源主导型”“内部数据能力驱动—外部数据资源补偿型”,并呈现出明显的“殊途同归”特点;进一步分析发现,实现高创新质量的不同路径对企业资源配置效率的影响具有显著差异,通过内部价值闭环或按需获取外部数据资源的企业,能够有效降低资源冗余和配置摩擦成本,在实现高创新质量后显著提升资源配置效率,呈现出“殊途同归”但“不同效”的特点。本研究有助于企业识别、选择和实施正确的数据要素组合方案以实现高质量创新,进而提升其资源配置效率。

关键词:数据要素;创新质量;动态QCA;资源配置效率

中图分类号:F270 **文献标识码:**A **文章编号:**1001-4950(2026)06-0022-18

一、引言

近年来,随着我国经济从高速发展转向高质量发展,创新也由注重数量转向注重质量。创新质量是企业构建可持续竞争优势和提升资源配置效率的重要基础。然而,我国整体创新水平虽然得到快速提升,但创新质量的提升远滞后于创新数量,呈现出明显的“数量跃进、质量滑坡”特征(陈强远等,2024)。企业是国家创新体系的重要微观主体,在创新质量仍有待提升的背

收稿日期:2025-10-27

基金项目:国家自然科学基金面上项目(71872052)

作者简介:方修园(1989—),女,复旦大学管理学院博士后;

吴金璇(1995—),女,复旦大学管理学院博士(通信作者,jxwu20@fudan.edu.cn);

李元旭(1966—),男,复旦大学管理学院教授,博士生导师。

景下(张杰和郑文平,2018;刘斐然等,2020),系统识别推动企业高创新质量的路径尤为重要。2017年,习近平总书记在主持实施国家大数据战略第二次集体学习时提出要“构建以数据为关键要素的数字经济”,并强调数据要素具有“基础资源作用”和“创新引擎作用”。此后,我国数字经济快速发展,数据要素的作用日益凸显。数据要素作为信息与知识的载体,是数字经济时代最核心的战略资源(陈丽莉等,2024;沈坤荣和林剑威,2025)。根据新华社数据显示,2024年我国数据市场交易规模已超1600亿元,同比增长30%以上。数据作为一种与劳动、资本、土地、知识、技术和管理并列的新型生产要素(洪永森和史九领,2024),具有价值性、稀缺性、不可模仿性和可组织性(魏江和应瑛,2025),能够有效减少信息不对称性,促进新知识创造(郭家堂,2025),已成为企业创新的重要引擎(王艳和杨达,2024)。因此,主动利用数据要素的赋能作用,探索实现高创新质量的多元路径,有助于企业提升价值创造能力和资源配置效率,进而助力国家经济高质量发展。

现有研究已关注到数据要素在提高创新水平方面的作用(陈丽莉等,2024;戴魁早等,2024;赵丽和胡植尧,2024),但主要聚焦于对企业创新数量的影响,忽视了其对创新质量的重要作用。更重要的是,现有研究侧重于考察单一数据要素的“净效应”,即基于单一视角独立展开,探索单个因素对企业创新的影响,如研究发现数据要素市场化能够促进企业技术创新的“增量提质”(虞涛和王兰芳,2025);较少有研究将企业内、外部数据要素置于同一分析框架中。事实上,数据要素不仅存在于企业内部(如在研发、生产、销售和管理中持续生成和积累的內部数据要素),也广泛存在于企业外部(如通过公共数据开放等途径获取的重要数据资源)。资源基础理论与开放式创新理论为理解这两类数据要素的作用提供了互补视角。资源基础理论认为企业的长期竞争优势来自企业内部有价值的、稀缺的、不可模仿且可组织的资源(Barney,1991),为理解内部数据要素的创新引擎作用提供了理论基础。然而,企业仅依靠内部资源进行封闭式创新已难以适应快速变化的市场环境。开放式创新理论认为企业应突破组织边界,获取和整合外部资源来加速内部创新和提升创新绩效(Chesbrough,2003)。企业创新质量受到内外多种因素的共同影响,仅关注单一数据要素的独立作用,忽视不同类型数据要素的互动关系及其组合效应,难以系统揭示多要素共同作用实现企业高创新质量的复杂机理。此外,已有研究关于数据驱动商业应用(陈丽莉等,2024;刘亚洲等,2026)、数据要素市场化(杨艳等,2021;郑国强等,2024)、数据要素集聚(史丹和孙光林,2024;肖文雪和范德成,2025)与创新之间的关系尚未形成一致结论,这可能是由不同因素之间的匹配与互动导致探讨单个因素的净效应并不能充分揭示其影响。因此,本文试图基于资源基础理论和开放式创新理论,探索企业内部数据要素(数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现)与外部数据要素(公共数据开放、数据要素市场化、数据要素集聚)形成的不同路径是否均能形成企业的高创新质量,从而揭示企业高创新质量的因果复杂性。

熊彼特创新理论认为企业通过对生产要素进行重新组合,能够引发“创造性破坏”,催生高质量创新,从而打破低效率的生产结构和提升资源配置效率,这一过程是经济发展的核心动力(Schumpeter,1942)。资源配置效率是企业确保成本控制、提升生产力和增强竞争力的关键(徐晔等,2025)。已有研究表明,数据要素能够通过促进企业创新进而提升其资源配置效率(洪永森和史九领,2024)。高创新质量的形成路径存在差异,不同内外部数据要素组合方式可能导致创新过程中的成本结构、风险承担和组织调整方式不同,进而使得不同路径形成的高创新质量在资源配置效率上呈现出差异化结果。目前较少有研究实证检验创新质量对资源配置效率的影响。因此,本文进一步探索不同路径形成的高创新质量对资源配置效率是否存在同样影响,即“殊途同归”是否“同效”,以帮助企业识别、选择和实施正确的数据要素组合方案以实现高创

新质量,进而提升资源配置效率。

本文基于中国沪深A股上市企业2017—2023年的平衡面板数据,利用动态QCA和PSM方法,探索企业内部数据要素(数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现)和外部数据要素(公共数据开放、数据要素市场化、数据要素集聚)对创新质量的组态影响,以及不同组态形成的高创新质量对企业资源配置效率的影响。研究贡献主要包括三方面。第一,整合资源基础理论和开放式创新理论,突破了已有研究关于单一数据要素对创新质量“净效应”或“线性关系”的分析范式,揭示了内外部数据要素协同驱动高创新质量的多路径等效性与因果复杂性,丰富和拓展了数据要素赋能企业创新质量的研究框架,也响应了学者们对于整合资源基础理论和开放式创新理论的呼吁(Vanhaverbeke和Cloudt,2014)。第二,从组态视角为同一种数据要素对企业创新影响的不一致结果提供了新解释。与大多数研究认为这种不一致结果可能与研究情景或变量测量等有关不同,本研究认为这可能是由不同因素之间的匹配与互动导致探讨单个因素的净效应并不能充分揭示其影响。第三,将创新质量形成机制与资源配置效率相结合,有助于解析创新质量与资源配置效率之间的关系。研究揭示了不同路径形成的高创新质量对企业资源配置效率的“不同效”影响,区分了企业创新质量的成因,有助于企业识别、选择和实施正确的数据要素组合方案以实现高创新质量,进而提升其资源配置效率。

二、理论基础与研究框架

数字经济时代,数据要素是企业创新的关键资源(陈丽莉等,2024;沈坤荣和林剑威,2025)。数据要素不仅广泛存在于企业内部,还广泛存在于企业外部。资源基础理论强调了企业内部资源的战略价值(Barney,1991),开放式创新理论则指出企业可以利用外部资源加速内部创新(Chesbrough,2003)。随着市场竞争的日益激烈,企业创新质量的提升已无法仅依赖内部数据要素,还需跨越企业边界主动获取和整合外部数据要素以突破资源瓶颈和重塑创新范式。因此,本研究整合资源基础理论和开放式创新理论,从内部资源禀赋和外部资源获取双重视角出发,探究企业如何有效组合内外部数据要素以实现高创新质量,具有重要的理论意义和实践意义。在此理论视角法的基础上,本文进一步利用文献归纳法(张明和杜运周,2019),对现有数据要素相关研究进行梳理,选择重要的、具有代表性的和对创新质量有显著影响的内外部数据要素作为聚焦的数据要素类型。

内部数据要素是企业内部可控制和利用的数据要素。根据现有文献,企业内部数据要素主要包括数据要素存量、数据开发能力、数据驱动商业应用和数据价值变现(唐要家等,2022;陈丽莉等,2024),分别代表企业数据要素的基础规模、处理能力、场景嵌入程度与价值实现能力,反映了企业从原始数据要素积累到价值转化的全过程能力。数据要素存量构成企业创新的信息基础;数据开发能力提升数据的可用性与分析效率;数据驱动商业应用使数据深入嵌入业务流程,支撑创新实践;数据价值变现体现数据驱动创新成果的经济回报能力。

外部数据要素是企业外部环境中通过政策制度、市场和平台可获得的数据要素,为企业提供丰富和异质的信息输入,补充内部信息盲区。外部数据要素的可得性和数据生态环境的完善程度,直接影响企业获取外部数据要素的广度与深度,进而影响其创新的边界条件。其中,公共数据开放、数据要素市场化和数据要素集聚分别体现企业所处数据生态环境的制度、市场与区域生态支持,是目前研究较为充分且对企业创新具有重要影响的外部数据要素(沈坤荣和林剑威,2025;虞涛和王兰芳,2025;张振堃等,2025)。

综上,本文选取数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现这四类企业内部数据要素,以及公共数据开放、数据要素市场化、数据要素集聚这三类企业外部数据要

素,作为前因条件,探究企业内外部数据要素如何以不同方式组合影响企业创新质量。

(一)内部数据要素与企业创新质量

资源基础理论为理解企业内部数据要素的创新价值提供了理论依据。企业内部具有价值性、稀缺性、不可模仿性和可组织性的数据要素资源,是赋能企业创新质量的战略性资源(魏江和应瑛,2025)。企业内部数据要素作用的发挥,不仅取决于数据要素存量,还取决于企业的数据开发能力、商业应用能力和价值变现能力,构成数据要素的价值实现路径(陈丽莉等,2024)。为了系统地理解企业内部数据要素的结构及其创新效应,本文梳理了数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现四种要素对企业创新质量的影响。

1.数据要素存量对企业创新质量的影响

数据要素存量是指企业在生产经营过程中,通过数据收集、整合、处理和管理所形成并沉淀的各类数据资源(唐要家等,2022;陈丽莉等,2024)。基于资源基础理论,数据要素存量作为企业进行数据分析、知识生成与价值创造的原始素材,是企业构建高创新质量的原始资源禀赋,是企业创新全过程的信息输入和知识基础。具体来说,数据要素存量能够通过创造新知识提升企业创新质量(郭家堂,2025)。数据要素存量的丰富性与多样性为企业知识重组提供资源基础,为企业生产新产品、提供新服务和开拓新市场提供灵感(Cong等,2021),是企业生成原创性知识、推动前瞻性创新的关键源泉。由于数据要素存量是企业在长期运营中持续沉淀的结果,具备稀缺性与难以模仿性,企业据此创造的新知识,往往成为难以被竞争对手模仿的独特知识资产,从而奠定高创新质量的资源基础。

2.数据开发能力对企业创新质量的影响

数据开发能力指企业依托自动化、智能化和数字技术,对数据进行计算、分析和再开发的综合技术能力(唐要家等,2022;陈丽莉等,2024),体现了企业的信息转化和知识挖掘能力,是维持创新优势的重要来源。资源本身并不直接创造价值,企业需要具备相应能力将资源转化为竞争优势(Teece等,1997)。数据开发能力正是实现这一转换的关键机制,决定原始数据能否高效转化为有价值的信息和知识,影响数据要素在创新中的可用性(魏江和应瑛,2025)。一方面,数据开发能力能够提升数据处理效率和准确性,帮助企业剔除冗余与噪声数据,提升信息可靠性,减少基于低质数据所带来的创新风险和资源浪费,避免创新路径偏离(陈丽莉等,2024),从而提升企业创新质量。另一方面,数据开发能力赋予企业从复杂数据中挖掘深层逻辑关系、识别因果机制与模式结构的能力,推动认知模型、业务逻辑与组织知识的重构,生成更高层次的原创性知识,使其能够突破传统经验路径,形成独特的、难以模仿的知识资产,为复杂性更高和原创性更强的创新活动提供理论支撑,进而提升创新质量。

3.数据驱动商业应用对企业创新质量的影响

数据驱动商业应用是指企业以数据要素为核心投入,将数据分析与数字技术嵌入业务模式、交易方式和运营体系中,推动商业模式创新、平台化运作和运营效率提升的能力(唐要家等,2022;陈丽莉等,2024)。基于资源基础理论,数据驱动商业应用的本质是一种基于数据要素资源的整合与重构能力。数据要素本身具有非竞争性和价值不确定性(魏江和应瑛,2025),通过与业务流程的深度耦合,才能转化为支撑创新活动的独特资源。具体来说,数据驱动商业应用通过将数据分析与数字技术嵌入关键业务节点,引入实时数据反馈机制(陈丽莉等,2024),使企业动态感知市场变化、用户行为和内部流程的偏差(马淑琴等,2024),实现对创新路径的即时校准,减少“偏离式创新”和“错位式创新”等低效尝试,增强创新过程的适配性与实效性,从而提升创新质量。值得注意的是,现有关于数据驱动商业应用与创新之间关系的研究尚未形

成一致结论,有学者发现数据驱动商业应用有利于企业创新(刘亚洲等,2026),也有学者发现其影响并不显著(陈丽莉等,2024)。

4.数据价值变现对企业创新质量的影响

数据价值变现是指企业通过数据要素参与金融活动、贸易活动和新型数字资产体系实现价值创造的综合能力(唐要家等,2022;陈丽莉等,2024)。基于资源基础理论,数据价值变现是将数据要素资源转化为企业经济租金和实现竞争优势的最终环节。企业在数据价值变现过程中的市场反馈,能够提供真实、动态的价值检验信号,使企业及时识别出具备更高市场潜力的数据驱动创新路径,从而优化创新方向,提升创新的经济回报率。此外,在数据价值变现过程中,企业需要将数据要素与传统资本和劳动力等融合,重塑传统生产函数。这一过程可以创造新知识、激发技术突破和商业模式创新,促进创新“新组合”的形成(杨亚平等,2025;郑志强等,2025),进而提升创新质量。这种基于内部数据要素价值实现所触发的知识重组与创新质量提升,正是资源基础理论所强调的“资源—能力—绩效”逻辑链条的最终体现。

(二)外部数据要素与企业创新质量

数字经济时代,企业仅凭内部资源进行封闭式创新已难以适应快速变化的环境。开放式创新理论强调,企业需要突破组织边界,积极利用外部资源来加速和提升创新(Chesbrough, 2003)。公共数据开放、数据要素市场化和数据要素集聚是企业获取外部数据要素资源的重要渠道,为企业突破内部资源局限和拓展创新边界提供了关键条件。因此,本文基于开放式创新理论,识别并阐释上述三种外部数据要素赋能企业创新质量的作用机制。

1.公共数据开放对企业创新质量的影响

当前,中国数据要素中占比最大的数据资源是公共数据(方锦程等,2023;郭家堂,2025),包括政府政务数据和企业等市场主体与社会公共利益相关的数据(方锦程等,2023)。基于开放式创新理论,公共数据开放是政府通过制度供给为企业创设的低成本、高质量的外部创新资源池。随着我国逐步有序推进公共数据开放,企业得以突破传统信息渠道的局限,通过挖掘公共数据满足创新需求、激发创新活力。

公共数据开放主要通过缓解信息不对称和提供关键数据资源促进企业创新(杜善重和连立帅,2025;李荣华等,2025;沈坤荣和林剑威,2025)。首先,公共数据开放能够缓解信息不对称,降低企业在获取政策和市场信息方面的成本,降低外部环境的不确定性(杜善重和连立帅,2025),提升企业创新资源投入意愿,避免因信息匮乏而拘泥于低风险的迭代式创新(李雪琴等,2024)。其次,公共数据开放为企业创新提供了关键的数据资源(李荣华等,2025)。公共数据的非消耗性与可复用性为创新活动提供持续支撑,企业可以将内部数据与公共数据相结合,催生新应用场景和市场需求,发现契合自身战略的创新机会(沈坤荣和林剑威,2025),体现了开放式创新中“内外部知识融合”的价值创造逻辑。

2.数据要素市场化对企业创新质量的影响

数据要素流通受阻会导致“数据孤岛”,抑制其价值发挥(杨亚平等,2025)。基于开放式创新理论,数据要素市场化本质是构建基于市场机制的开放、共享和可交易的外部数据资源流通体系,使企业获取外部数据的方式从“关系化、偶然化”转向“契约化、常态化”,为企业创新提供稳定支撑(刘涛雄等,2024),赋能企业高质量创新(虞涛和王兰芳,2025)。

数据要素市场化通过缓解信息不对称、促进数据共享和提供多元化数据资源等方式提升企业创新质量。首先,市场化机制为数据要素交易提供制度保障,打造规范有序的流通环境,使企业能够获取高质量的市场前瞻信息,降低研发不确定性和创新风险(戴魁早等,2023;虞涛和王兰芳,2025)。其次,数据要素市场促进企业间数据共享与知识流动,提升研发合作意愿,实现

跨组织的资源精准匹配与合作创新(徐晔等,2025;杨亚平等,2025),推动企业突破传统封闭式创新模式。最后,数据要素市场化汇聚跨地区、跨行业数据资源,拓展企业“数据资源池”,帮助企业低成本获取多元化知识和技术,降低搜索与交易成本(梁孝成等,2025;王海森和李纲,2025)。需要说明的是,虽然大部分现有研究认为数据要素市场化有利于企业创新(戴魁早等,2023;王海森和李纲,2025;杨亚平等,2025;虞涛和王兰芳,2025),但是也有研究发现其影响不显著(杨艳等,2021;郑国强等,2024)。

3.数据要素集聚对企业创新质量的影响

数据要素集聚是指在特定区域内对分散的数据资源进行系统性整合,形成规模化、协同化的数据资源体系(张振堃等,2025)。基于开放式创新理论,数据要素集聚的本质是在地理空间中构建外部数据资源网络,通过空间邻近、关系嵌入与生态协同等方式,重塑企业获取外部创新资源的方式。

数据要素集聚通过提供数据资源、促进知识溢出和缓解融资约束等方式提升企业创新质量。首先,数据要素集聚打破“数据孤岛”,拓展外部创新资源的可及性边界,帮助企业低成本、多渠道获取创新知识与信息(张朝华和徐鹏杰,2024;郑威和罗润风,2024),推动数据产品和服务开发,提升创新质量。其次,数据要素集聚带来创新知识溢出效应,强化开放式创新的生态协同。一方面,集聚推动数字基础设施建设,吸引高新企业集聚(张朝华和徐鹏杰,2024),增强企业间合作与知识交流,拓展技术创新深度与广度。另一方面,高质量人力资本随产业集聚汇入,进一步推动知识共享与技术突破,产业集聚带来的竞争压力也倒逼企业加快创新步伐(张朝华和徐鹏杰,2024)。最后,数据要素集聚有助于提升信息透明度,缓解企业与金融机构之间的信息不对称,优化融资方案(徐怀宁和刘莎莎,2025),提升企业融资可得性,为创新活动提供关键资金支持。目前,有研究发现数据要素集聚显著促进创新(史丹和孙光林,2024),也有研究发现数据要素集聚对创新存在先抑制后促进的非线性影响(肖文雪和范德成,2025)。

(三)研究框架

资源基础理论揭示了内部数据资源的创新引擎价值,开放式创新理论则强调了企业需突破组织边界,获取与整合外部数据资源以驱动创新。数据要素是数字经济时代企业创新的关键资源(陈丽莉等,2024;沈坤荣和林剑威,2025),其创新价值释放既依赖于企业内部数据资源禀赋与能力,也依赖于外部数据资源的获取与整合。然而,鲜有研究探讨企业内外部多种数据要素的不同组合对企业创新质量产生的复杂影响。基于上述数据要素赋能创新质量的前因条件及其作用机制分析,本文整合资源基础理论和开放式创新理论,提出了企业内外部数据要素赋能创新质量的前因组态研究框架(如图1所示),从而构建数据要素赋能企业高创新质量的多元路径。

三、研究设计

(一)研究方法:动态QCA方法

传统计量经济学方法侧重于变量间的线性关系与“净效应”,强调对称性与因果推断,常用面板数据应对内生性与反向因果问题。相比之下,QCA方法关注因果复杂性,强调条件组态对结果的必要性与充分性,适用于处理多重并发性、因果等效性与非对称性(孙洁和梅静怡,2026)。然而,传统QCA多基于截面数据,难以揭示因果关系的动态演化。为此,Gracia-Castro和Ariño(2016)提出动态QCA方法,引入“一致性调整距离”以捕捉组态关系在时间和案例维度的变化趋势。该方法保留了QCA对复杂因果结构的识别优势,同时实现了其在时间序列与面板数据中的应用,适用于分析因素间的联动效应与组态演化路径。因此,本文采用动态QCA方法

来分析内外部数据要素对企业创新质量的影响,以弥补传统计量经济学方法难以揭示因果复杂性的问题(孙洁和梅静怡,2026),并克服传统QCA方法的时间盲区(陈元和贺小刚,2025)。

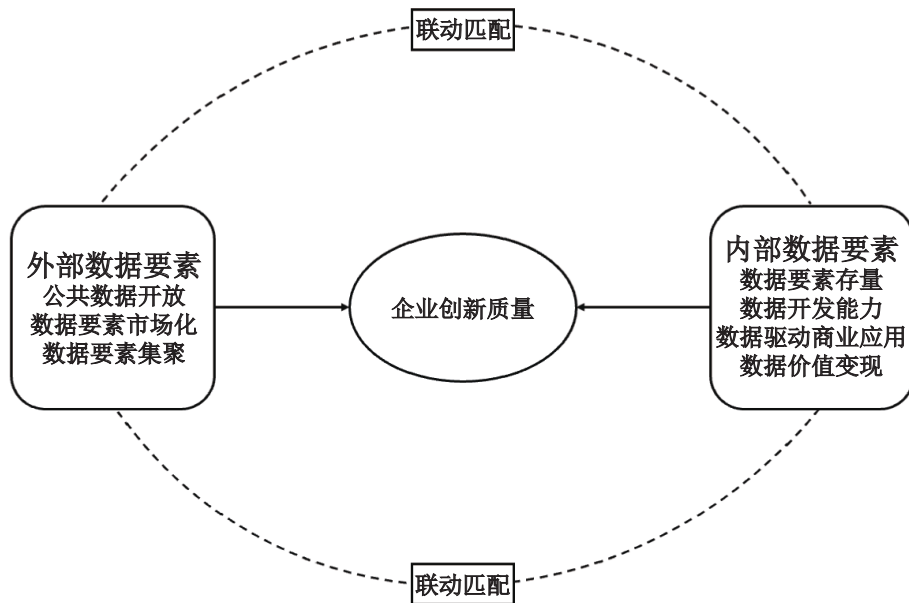


图1 理论模型:数据要素影响企业创新质量的分析框架

(二)样本选择与数据来源

本文的研究对象为2017—2023年沪深A股上市企业。选取2017年为起始年份主要基于以下考虑:2017年,习近平总书记在主持实施国家大数据战略第二次集体学习时强调“构建以数据为关键要素的数字经济”,并指出数据要素具有“基础资源作用”和“创新引擎作用”。此后,我国数字经济快速发展,数据要素的作用日益凸显。本研究前因条件的时间周期为2017—2021年,考虑到数据要素的影响存在滞后性,创新质量数据使用 $t+1$ 期,即创新质量的时间周期为2018—2022年。同样,考虑到不同路径形成的高创新质量对资源配置效率影响的滞后性,进一步分析中的资源配置效率使用 $t+2$ 期,即资源配置效率的时间周期为2019—2023年。借鉴已有研究,本文按照以下标准对样本数据进行筛选:首先,剔除属于金融行业的企业;其次,剔除被特别处理(ST)及退市风险警示(*ST)的企业;最后,剔除数据缺失样本。最终获得1158家企业共5790个观测值。本文的创新质量数据来源于中国研究数据服务平台,上市企业年报文本数据来源于巨潮资讯网,其他数据来源于国泰安(CSMAR)数据库。

(三)测量与校准

1.结果变量:创新质量

本文借鉴已有研究(张杰和郑文平,2018;刘斐然等,2020),以专利知识宽度信息测量企业创新质量。具体公式如下:

$$Patentknowledge_{it} = 1 - \sum_{n=1}^N \left(\frac{Z_{int}}{Z_{it}} \right)^2 \quad (1)$$

其中, i 表示企业, t 表示年份, n 表示IPC专利分类中的大组, $n=1,2,\dots,N$ 。 N 为企业 i 在 t 年所涉及的IPC大组总数。 Z_{int} 是企业 i 于 t 年在第 n 个IPC大组下申请专利的累计数目, Z_{it} 是企业 i 于 t 年在全部IPC大组下申请专利的累计数目,两者之比即为专利分类号中各大组分类号所占比例。

$Patentknowledge_{it}$ 的值越大,表明企业的专利知识宽度越大,创新质量越高(张杰和郑文平,2018;刘斐然等,2020)。

2.前因条件

企业内部数据要素分为数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现。参考已有文献(唐要家等,2022;陈丽莉等,2024),本文利用文本分析法,首先,获取2017—2023年沪深A股上市企业年报,将其转换为文本文件,提取MD&A章节分词。其次,参考唐要家等(2022)开发的关键词词表。最后,运用Python软件对上市企业年报MD&A中相关关键词进行检索和匹配,剔除否定含义表述的关键词,得到关键词的有效词频数,并对其加1取自然对数,以反映企业的数据要素存量、数据开发能力、数据驱动商业应用和数据价值变现的水平。

公共数据开放:地方政府公共数据开放平台是公共数据最权威和主要的来源(郭立祥等,2025)。与大多数研究一致(方锦程等,2023;郑志强和何佳俐,2025),若企业注册地设有城市公共数据平台,则赋值为1,否则为0。

数据要素市场化:数据交易平台建设是数据要素市场化的重要举措,通过衔接数据要素供给方和需求方,提供互动、匹配与交易场所,促进数据要素高效流通(杨亚平等,2025),反映数据要素市场化发展程度(戴魁早等,2023)。与现有研究一致(梁孝成等,2025;杨亚平等,2025),若企业注册地位于设立有数据交易平台的城市,则赋值为1,否则为0。

数据要素集聚:为了发挥数据要素集聚作用,国家于2016年在贵州省设立了首个国家级大数据综合试验区。国家级大数据综合试验区打通了数据要素流通路径,强化数据要素的统筹与共享,是数据要素集聚的重要载体(杨亚平等,2025)。与现有研究一致(徐怀宁和刘莎莎,2025),若企业注册地位于国家大数据综合试验区内,则赋值为1,否则为0。

3.校准

本文校准结果变量和内部数据要素的四个前因条件。参考已有研究(张明等,2020),本文采用直接校准法,分别以0.95分位点、0.50分位点和0.05分位点作为完全隶属、交叉点和完全不隶属的3个定性锚点。

四、研究结果

(一)单个条件必要性分析

本文首先进行必要性分析,检验是否存在单一条件(包括其非集)构成企业创新质量的必要条件。单个条件必要性分析是通过检验结果集合是否为某一条件集合的子集来加以判断的(张明等,2019)。在传统QCA研究中,当一致性水平高于0.9时,可将该条件认定为必要条件(张明等,2019;张明和杜运周,2019)。对于动态QCA,若一致性水平小于0.9但组间一致性调整距离大于临界值0.2(Gracia-Castro和Ariño,2016),则还需进一步分析前因条件与结果的必要性因果关系,以保证汇总一致性的可靠性(孙洁和梅静怡,2026)。

表1的必要条件分析结果显示,所有条件的一致性均小于临界值0.9。然而,高数据要素存量与高创新质量、高数据要素存量与非高创新质量、非高数据要素存量与高创新质量、高公共数据开放与高创新质量、高公共数据开放与非高创新质量、非高公共数据开放与高创新质量、非高公共数据开放与非高创新质量的组间一致性调整距离大于临界值0.2,需要进一步分析。

表2为组间一致性调整距离大于0.2的高数据要素存量、非高数据要素存量、高公共数据开放、非高公共数据开放分别与企业高与非高创新质量的组间效应分析,各年组间一致性均小于0.9,因此,高数据要素存量、非高数据要素存量、高公共数据开放、非高公共数据开放均不构成企业创新质量的必要条件。

表1 必要条件分析

条件变量	高创新质量				非高创新质量			
	汇总一致性	汇总覆盖度	组间一致性调整距离	组内一致性调整距离	汇总一致性	汇总覆盖度	组间一致性调整距离	组内一致性调整距离
高数据要素存量	0.566	0.717	0.235	0.545	0.556	0.557	0.246	0.511
非高数据要素存量	0.650	0.649	0.214	0.409	0.717	0.567	0.142	0.341
高数据开发能力	0.683	0.738	0.090	0.375	0.676	0.578	0.119	0.375
非高数据开发能力	0.609	0.704	0.154	0.477	0.693	0.634	0.090	0.375
高数据驱动商业应用	0.628	0.705	0.093	0.443	0.664	0.591	0.113	0.409
非高数据驱动商业应用	0.636	0.705	0.125	0.443	0.669	0.587	0.090	0.409
高数据价值变现	0.752	0.774	0.026	0.204	0.810	0.659	0.029	0.136
非高数据价值变现	0.669	0.816	0.084	0.307	0.722	0.698	0.023	0.239
高公共数据开放	0.658	0.567	0.336	0.579	0.635	0.433	0.336	0.613
非高公共数据开放	0.342	0.542	0.614	1.090	0.365	0.458	0.620	1.022
高数据要素市场化	0.554	0.555	0.014	0.886	0.561	0.445	0.014	0.886
非高数据要素市场化	0.446	0.562	0.017	1.124	0.439	0.438	0.017	1.124
高数据要素集聚	0.434	0.554	0.014	1.124	0.442	0.446	0.017	1.124
非高数据要素集聚	0.566	0.562	0.012	0.886	0.558	0.438	0.014	0.886

表2 调整距离大于0.2的组间数据

因果组合情况			年份				
			2017	2018	2019	2020	2021
情况a	高数据要素存量与高创新质量	组间一致性	0.452	0.480	0.525	0.612	0.732
		组间覆盖度	0.680	0.713	0.729	0.724	0.723
情况b	高数据要素存量与非高创新质量	组间一致性	0.443	0.467	0.551	0.624	0.738
		组间覆盖度	0.676	0.603	0.562	0.537	0.480
情况c	非高数据要素存量与高创新质量	组间一致性	0.784	0.733	0.685	0.609	0.474
		组间覆盖度	0.581	0.613	0.675	0.690	0.733
情况d	高公共数据开放与高创新质量	组间一致性	0.367	0.562	0.686	0.792	0.826
		组间覆盖度	0.490	0.536	0.580	0.580	0.598
情况e	高公共数据开放与非高创新质量	组间一致性	0.376	0.561	0.677	0.788	0.844
		组间覆盖度	0.51	0.464	0.420	0.420	0.402
情况f	非高公共数据开放与高创新质量	组间一致性	0.633	0.438	0.314	0.208	0.174
		组间覆盖度	0.500	0.535	0.570	0.575	0.629
情况g	非高公共数据开放与非高创新质量	组间一致性	0.624	0.439	0.323	0.212	0.212
		组间覆盖度	0.500	0.465	0.430	0.425	0.371

综上,单一条件不构成企业创新质量的必要条件,应充分研究各前因条件联动匹配的组态效应对企业创新质量的影响。

(二)条件组态充分性分析

为提升组态识别的严谨性与可解释性,本研究设定一致性阈值为0.80, PRI一致性阈值为0.75,案例频数阈值为1,以排除伪组态并增强结果稳健性(Ragin, 2008; 张明和杜运周, 2019)。在反事实分析中,因缺乏统一理论预期,采取“存在或缺席皆可”策略,不设方向性假设,以全面探索因果路径。

组态解通常分为三种类型:复杂解、中间解和简约解。鉴于中间解有助于区分组态中的核心条件与边缘条件,本文以中间解作为主要分析依据(张明和杜运周, 2019)。具体而言,同时出现在中间解和简约解的条件为核心条件,而仅出现在中间解的条件为边缘条件。QCA研究中

的一致性指标衡量组态作为充分条件的可靠性,覆盖度则反映其对结果的解释力,类似于回归模型的 R^2 (张明等,2019)。本研究最终识别出五个高一一致性、高覆盖度的有效组态(见表3),其一致性均在0.866以上,整体解一致性达0.897,显示出强因果充分性,表明这些组态对高创新质量具有显著解释力。

表3 创新质量条件组态结果

视角	条件变量	组态1a	组态1b	组态2a	组态2b	组态3
内部数据资源	数据要素存量	●	●	●	●	
	数据开发能力	⊗		●	●	●
	数据驱动商业应用		⊗		⊗	⊗
	数据价值变现	●	●	●	●	●
外部数据资源	公共数据开放	⊗	⊗	⊗	⊗	●
	数据要素市场化	●	●	⊗		●
	数据要素集聚	●	●	●	●	⊗
	一致性	0.901	0.900	0.866	0.906	0.911
	<i>PRI</i>	0.750	0.768	0.756	0.787	0.766
	覆盖度	0.011	0.013	0.019	0.026	0.048
	唯一覆盖度	0.001	0.000	0.004	0.000	0.048
	组间一致性调整距离	0.029	0.046	0.043	0.035	0.052
	组内一致性调整距离	0.034	0.034	0.068	0.068	0.068
	总体一致性			0.897		
	总体 <i>PRI</i>			0.762		
	总体覆盖度			0.081		

注:●表示核心条件存在,⊗表示核心条件缺失,空格表示存在与否均可。

本研究识别出五种能够实现企业高创新质量的前因条件组态,展现了企业在不同的内外部数据要素组合下实现高创新质量的多样性与等效性。结合五个组态的具体维度和特征,本文将分为三类路径,组态1a和1b为“内外部数据资源协同型”,组态2a和2b为“内部数据资源主导型”,组态3为“内部数据能力驱动—外部数据资源补偿型”。

第一,内外部数据资源协同型。组态1a和1b在组态结构上高度相似,核心条件包括内部的高数据要素存量、高数据价值变现能力、外部的高数据要素市场化程度、高数据要素集聚水平,以及低数据开发能力/低数据驱动商业应用能力、低公共数据开放水平。这表明,尽管缺乏内部技术能力或制度性公共数据供给,但通过雄厚的内部数据资源基础,结合成熟的数据交易机制与良好的数据要素集聚等市场环境,企业依然能够实现数据价值的有效转化,进而推动创新质量的提升。从组态逻辑来看,该路径通过内部数据资源与外部数据资源之间的交互协同共同驱动企业实现高创新质量。其中,内部充足的数据要素存量为企业创新提供了稳定且可持续的原始投入;外部成熟的数据要素市场化与较高的数据集聚,通过促进数据流通、交易与跨主体协作,使内部数据要素存量在企业的技术开发或数据驱动商业应用方面尚不成熟的情况下,亦能够借助外部数据市场机制实现数据价值的有效转化,进而推动高质量创新产出。通过内外部数据资源协同,该路径为内部数据资源丰富但能力尚未成熟的企业提供了一种现实可行的高创新质量模式。该路径的典型案例有2018—2020年的蒙草生态(300355)。蒙草生态将“种质资源+生态大数据”作为核心战略,整合近30年的气象、土壤和草原盖度等指标创建了草原生态修复数据平台;“三北”生态种源数据集入选国家第一批行业高质量数据集,内部数据资源丰富。蒙草生态注册地位于呼和浩特市,该市设立有数据交易平台,且位于国家大数据综合试验区内,为企业提供了成熟的数据交易机制和良好的数据要素集聚环境。

第二,内部数据资源主导型。组态2a和2b在组态结构上高度相似,核心条件包括内部的高

数据要素存量、高数据开发能力、高数据价值变现能力、外部的高数据要素集聚水平,以及低公共数据开放、低数据要素市场化水平/低数据驱动商业应用能力。这表明,企业内部丰富的数据要素存量、较强的数据开发与变现能力,以及外部数据集聚环境提供的外部知识溢出与协同合作的空间,使企业能够将内部数据要素高效转化为创新要素,提高企业创新质量。从组态逻辑来看,该路径以内部数据资源为主导,通过内部数据资源的存量和能力之间的协同共同驱动企业实现高创新质量。其中,内部充足的数据要素存量为创新活动提供了数据基础,而较强的数据开发与变现能力则使企业能够对数据进行深度挖掘和重组,将其转化为高价值的创新要素。这种内部资源之间的协同降低了企业对外部数据资源的依赖,使其在外部数据资源相对不足的情况下,仍可依托内部数据资源实现高质量创新。该路径为外部数据资源供给不足但内部数据资源充足且能力扎实的企业提供了一种可行的高创新质量路径选择。该路径的典型案例有2021年的多氟多(002407)。多氟多将多年来在氟、锂、硅三个元素领域的知识与经验进行数据化,构成了其核心的数据要素存量,建立了一体化智慧协同平台、氟基新材料工业互联网平台和基于数实融合的氟化工生产安全综合管控平台;并向客户提供工业APP以及技术成果转让。多氟多位于焦作市,该市位于国家大数据综合试验区内,为企业提供了良好的数据要素集聚环境。

第三,内部数据能力驱动—外部数据资源补偿型。组态3的核心条件包括内部的高数据开发能力、高数据价值变现能力、外部的高公共数据开放、高数据要素市场化水平,以及低数据驱动商业应用能力、低数据要素集聚水平。这表明,当外部的公共数据资源丰富、数据市场交易机制完善,且企业具有较强的数据开发和变现能力时,企业能够将外部数据转化为高价值的创新成果,从而实现创新质量的提升。与其他组态相比,该路径的核心并不在于内部数据要素存量和外部数据要素集聚,而在于外部公共数据开放这一重要外部数据资源,以及数据开发能力这一关键内部能力。企业内部较强的数据开发能力使企业能够有效吸收和重构公共数据等外部数据资源,并在数据要素市场化机制的支持下实现价值转化。由此,外部数据资源对内部数据要素存量不足形成补偿,推动企业实现高创新质量。该路径适用于外部数据资源充足且内部数据能力突出的企业。该路径的典型案例有2020—2021年的永创智能(603901)。永创智能持续进行高强度的研发投入,将工业知识与数据技术融合,开发了包装机器人、智能包装生产线和智慧工厂工业软件等,为客户创造价值。永创智能位于杭州市,该市具有公共数据平台和数据交易平台,为企业提供了丰富的公共数据资源和完善的数据市场交易机制。

为检验上述组态的时间稳定性,本文绘制了2017—2021年间五个组态的年度一致性变化趋势,以衡量各条件组态在不同年份是否为高创新质量的充分条件,如图2所示。结果显示,五个组态的年度一致性均高于0.8的判断标准,且未呈现持续上升或下降的单向趋势,说明上述组态不存在明显的时间效应。这意味着,企业的高创新质量具有路径依赖特征,相关组态未发生结构性变化,具有较强的时间稳定性。

(三)稳健性检验

为增强研究结果的稳健性,本文还进行了条件组态的稳健性检验分析。首先,将原始一致性阈值由0.80调整为0.85,得到的研究结果是企业高创新质量的五个条件组态与原结果一致。

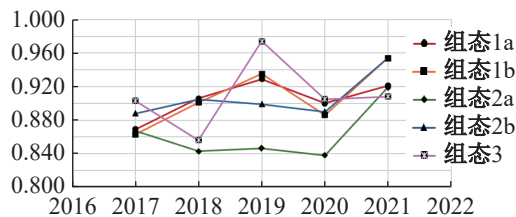


图2 组间一致性变化趋势

其次,将案例频数阈值由1调整为2,得到的企业高创新质量的条件组态同样与原结果一致。最后,将PRI一致性阈值由0.75调整为0.73(孙洁和梅静怡,2026),得到8个条件组态,其中五个组态与原结果一致,说明原结果是该结果的子集。以上结果说明本文研究结果具有稳健性^①。

五、进一步分析

通过上述组态分析可知,企业实现高创新质量有五条不同的路径,表明高创新质量具有殊途同归的特点。熊彼特创新理论认为,企业通过对生产要素进行重新组合来实现高质量创新,能够打破原有低效率生产结构,进而提升资源配置效率,是经济发展的重要推动力(Schumpeter,1942)。因此,本文进一步考察不同实现路径所形成的高创新质量是否会对企业资源配置效率产生相同的影响,即“殊途同归”是否“同效”。

考虑到规模、行业 and 所有权性质等特征对企业创新质量和资源配置效率的影响存在差异,为了更好地探究不同组态对企业资源配置效率的影响,本文将未实现高创新质量的剩余案例视为控制组来源,将实现高创新质量的案例作为实验组,采用倾向得分匹配法(PSM)进行反事实分析(Rosenbaum和Rubin,1983)。PSM可以自动匹配企业特征相似的样本进行比较,匹配出的样本在企业年龄、规模、行业类型、所有权性质和资产负债率等方面的倾向一致,唯一区别在于是否实现了五种高创新质量,从而得以控制企业个体差异对资源配置效率的干扰,以更准确地识别不同组态对企业资源配置效率的影响,给出更可靠的处理效应估计。

(一)计算企业实现高创新质量的倾向得分

进行PSM的第一步是估计企业实现高创新质量的倾向得分。本文采用Logit回归模型进行预测,分别以企业高创新质量的五条实现路径作为被解释变量,以企业规模、企业年龄、行业类型、所有权性质和资产负债率作为解释变量纳入模型。其中,企业规模以企业员工总数取自然对数进行测量;企业年龄则以企业自成立之年起至样本观测年份之间的存续年数加以测量;行业类型依据中国证监会行业分类指引2012版(SIC三级行业代码)进行分类;所有权性质为虚拟变量,若企业的实际控制人(或实际控制人之一)属于国有企业、行政机关、事业单位以及中央或地方政府相关机构,则将该企业界定为国有企业,并赋值为1;其余情形统一认定为非国有企业,赋值为0;资产负债率为负债总额/资产总额。倾向得分的Logit回归结果如表4所示。

(二)样本匹配效果

本文以企业实现高创新质量的倾向得分作为匹配依据开展样本匹配,并以组态3为示例,展示其基于一配三原则的最近邻匹配方法所得到的匹配结果,具体见表5。结果表明,所有控制变量在匹配后的偏倚度(|Bias|)均明显缩小,且实验组与控制组之间的差异检验均不再显著。与此同时,匹配模型的LR- Chi^2 值从26.31降到0.33,对应 p 值从0.000增长到0.997。这表明匹配效果非常理想,实验组和控制组在企业规模、年龄、行业、所有权性质和资产负债率等方面不存在系统性差异,两组间的主要区别仅体现在是否实现高创新质量。图3a和图3b进一步展示了在匹配前后,实验组与控制组的核密度函数图。

(三)不同类型高创新质量对企业资源配置效率的影响分析

参考现有研究(毛捷和管星华,2022),本文使用投资效率测量企业资源配置效率,分别借鉴Richardson(2006)和Biddle等(2009)的方法估算企业资源配置效率。

本文采用基于一配三的最近邻匹配方法进行平均效应值(ATT)估计。表6展示了不同类型高创新质量对企业后续资源配置效率的分析结果。结果表明,在完成最近邻匹配后,不论采用哪种方式测量企业资源配置效率,组态2b和组态3的企业在实现高创新质量后的资源配置效率

^①由于篇幅限制,稳健性检验结果留存备案。

表4 Logit模型的估计结果

变量	模型1 组态1a	模型2 组态1b	模型3 组态2a	模型4 组态2b	模型5 组态3
企业年龄	0.018 (0.030)	-0.014 (-0.031)	0.068*** (0.016)	0.004 (-0.029)	-0.012 (-0.022)
企业规模	-0.002 (0.156)	-0.007 (-0.147)	0.287*** (0.103)	0.121 (-0.140)	-0.478*** (-0.119)
行业类型	0.003 (0.012)	0.014 (-0.011)	0.005 (0.008)	-0.016 (-0.012)	0.003 (-0.008)
所有权性质	0.752** (0.370)	0.937*** (-0.349)	0.240 (0.259)	0.275 (-0.356)	-0.319 (-0.305)
资产负债率	0.026 (1.073)	-0.654 (-1.018)	-0.959 (0.764)	-0.078 (-1.006)	2.344*** (-0.698)
常数项	-5.789*** (1.346)	-5.239*** (-1.269)	-7.903*** (0.915)	-5.454*** (-1.253)	-1.376 (-0.958)
N	5 790	5 790	5 790	5 790	5 790
Pseudo-R ²	0.013	0.018	0.035	0.009	0.030

注：*、**、***分别表示在10%、5%、1%的水平上显著相关，括号内为标准误。

表5 组态3倾向得分匹配平衡性检验

变量匹配	处理组	控制组	偏差	偏差绝对值	t值	$p> t $	V(T)/V(C)
企业年龄U	18.61	19.367	-15.2	—	-1.26	0.206	0.7
企业年龄M	18.61	18.561	1	93.6	0.06	0.951	0.66
企业规模U	7.4159	7.8875	-43.4	—	-3.46	0.001	0.57*
企业规模M	7.4159	7.3559	5.5	87.3	0.4	0.688	0.89
行业U	38.268	36.253	17	—	1.34	0.181	0.51*
行业M	38.268	37.841	3.6	78.8	0.23	0.821	0.49*
所有权性质U	0.18293	0.26524	-19.8	—	-1.68	0.093	—
所有权性质M	0.18293	0.1748	2	90.1	0.14	0.893	—
资产负债率U	0.42135	0.39492	15.3	—	1.35	0.178	0.92
资产负债率M	0.42135	0.40837	7.5	50.9	0.48	0.631	0.94

注：U是匹配前，M是匹配后。

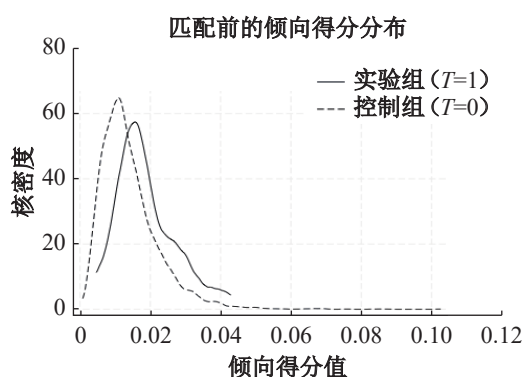


图3a 匹配前核密度函数

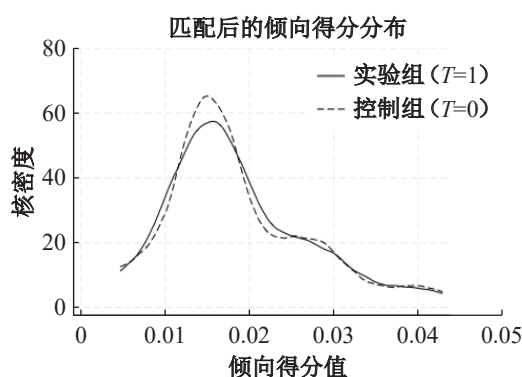


图3b 匹配后核密度函数

显著地高于控制组企业的资源配置效率(所有ATT所对应的 p 值均小于0.05),而组态1a、1b、2a的企业在实现高创新质量后的资源配置效率均与控制组企业的资源配置效率之间没有显著差异(所有ATT所对应的 p 值均大于0.10)。

具体而言,组态3以数据开发能力为核心,借助公共数据开放与数据要素市场化机制按需

表6 不同创新质量对企业资源配置效率影响的PSM分析

测量方法	组态1a		组态1b		组态2a		组态2b		组态3	
	ATT	t值	ATT	t值	ATT	t值	ATT	t值	ATT	t值
方法一	0.003	0.470	0.004	0.760	-0.002	-0.460	-0.013**	-2.080	-0.010**	-2.410
方法二	0.001	0.150	0.000	0.070	0.001	0.230	-0.013**	-1.990	-0.009**	-2.020

注:方法一采用Richardson(2006)的方法计算企业资源配置效率,方法二采用Biddle等(2009)的方法计算企业资源配置效率。*表示 $p<0.1$,**表示 $p<0.05$,***表示 $p<0.01$ 。

获取外部数据资源,减少了对内部数据要素存量的依赖,降低了内部数据资源冗余成本,因而可以提高资源配置效率;组态2b则通过在企业内部同时配置较高的数据要素存量、数据开发能力和数据价值变现能力,形成从数据要素存量、数据开发到价值实现的内部闭环,使资源投入能够在较短周期内得到反馈与再配置,且减少了跨主体配置与制度摩擦,从而可以提高资源配置效率。相比之下,组态1a、1b的高创新质量实现过程依赖内外部资源协同整合,这种跨主体的协调和整合成本相对较高,因而在资源配置效率上未表现出显著优势。虽然组态2a与组态2b相似,但两者在数据驱动商业应用能力和数据要素市场化方面有所区别。高数据驱动商业应用会使企业在创新之外,涉及更多其他尝试(如场景探索与应用扩展),使得投入与回报之间存在较强不确定性和时间滞后,而且可能带来对前沿和战略型业务的“过度投资”,从而表现出资源配置路径发散、协调成本上升,因而不能显著提高资源配置效率。反之,低数据驱动商业应用能力反而促使企业更加聚焦于已有业务的效率优化。由于组态2a未像组态2b那样通过低数据驱动商业应用能力对应用扩展形成约束,再加上组态2a以低数据要素市场化水平为核心条件,企业面临数据流通、交易与跨主体协作的限制,在借助外部数据市场机制实现数据价值的有效转化方面存在不足,更倾向于通过潜在的多场景探索与应用扩展实现数据价值转化,从而拉长资源投入回收周期,削弱资源配置效率,因而未能显著提升资源配置效率。

六、研究结论与启示

(一)研究结论

数字经济时代,数据要素如何赋能企业创新质量,成为理论研究与企业实践的重要议题。本文基于资源基础理论和开放式创新理论,采用我国沪深A股上市企业2017—2023年平衡面板数据,使用动态QCA方法探究企业内部数据要素(数据要素存量、数据开发能力、数据驱动商业应用、数据价值变现)与外部数据要素(公共数据开放、数据要素市场化、数据要素集聚)对企业创新质量的组态效应,并进一步运用PSM方法检验这些组态形成的高创新质量对企业后续资源配置效率的影响。

研究发现:首先,企业可以通过三类路径实现高质量创新,呈现出明显的“殊途同归”特点。第一,组态1a和1b的“内外部数据资源协同型”路径表明,企业可依托内部高数据要素存量、价值变现能力与外部良好的数据要素市场机制和集聚环境,实现高创新质量。第二,组态2a和2b的“内部数据资源主导型”路径表明,企业依靠内部充足的数据要素存量与数据开发能力,即便外部公共数据资源与市场机制不完善,亦能通过内生性技术整合与集聚效应,构建自主封闭的创新体系,实现高质量创新。第三,组态3的“内部数据能力驱动—外部数据资源补偿型”路径表明,企业可利用外部公共数据和市场机制这些数据资源的补偿,通过内部增强数据开发能力,弥补内部数据要素存量和外部数据要素集聚生态的不足,实现高创新质量。不同企业可以依托于不同的内外部数据资源结构,实现对创新质量的有效提升。

其次,实现高创新质量的五种不同路径对企业资源配置效率的影响具有显著差异性,呈现

出“殊途同归”但“不同效”的特点。具体而言,仅有组态2b和组态3在实现高创新质量后显著提升了资源配置效率,这主要源于二者分别通过内部价值闭环与按需获取外部数据资源的方式,有效降低了资源冗余和配置摩擦成本。相比之下,其余组态在高质量创新过程中涉及多场景应用扩展或跨主体资源整合,投入—回报不确定性较高,因而未能进一步转化为资源配置效率优势。

(二)理论贡献

本文主要有三个方面的理论贡献,具体如下:

第一,本文整合资源基础理论和开放式创新理论,系统分析企业内外部数据要素的不同组合对企业创新质量的影响,突破了以往研究聚焦单一数据要素对企业创新独立影响的局限。现有研究虽已关注到数据要素的创新效应(陈丽莉等,2024;戴魁早等,2024;赵丽和胡植尧,2024),但大多数集中于创新数量,忽视了其对创新质量的重要作用。更重要的是,现有研究侧重于单一类型数据要素的“净效应”,少有研究将企业内外部数据要素放到同一框架进行分析。本文将分析视角延伸至创新质量,系统分析了企业内部数据要素与外部数据要素对创新质量的组态影响,突破了以往研究的“线性关系”分析范式,揭示了数据要素赋能高创新质量的多路径等效性与因果复杂性,丰富了数据要素与企业创新质量之间关系的理论认识。同时,研究也响应了学者们对于整合资源基础理论和开放式创新理论的呼吁(Vanhaverbeke和Cloudt,2014)。

第二,本文从组态视角为同一种数据要素对企业创新的异质性结果提供了新解释。已有研究关于数据驱动商业应用(陈丽莉等,2024;刘亚洲等,2026)、数据要素市场化(杨艳等,2021;郑国强等,2024)、数据要素集聚(史丹和孙光林,2024;肖文雪和范德成,2025)与创新之间的关系尚未形成统一结论,通常被认为可能与研究情景或变量测量等有关。本文从组态视角为这种不一致现象提供了新解释,如组态结果显示数据驱动商业应用高或低对高创新质量并没有影响(组态1a、2a),甚至低数据驱动商业应用反而有利于企业高创新质量(组态1b、2b、3),这可能是由不同因素之间的匹配与互动导致探讨单个因素的净效应并不能充分揭示其影响。因此,本文的发现为已有研究结论不一致提供了新解释,也为后续分析数据要素对企业创新质量的影响提供了基础。

第三,本文发现了不同路径形成的高创新质量对企业资源配置效率的差异性影响,即“殊途同归不同效”,有助于解答创新质量与资源配置效率之间的关系。尽管熊彼特创新理论早就指出,通过对生产要素的重新组合引发“创造性破坏”,能够改变原有的低效率生产结构(Schumpeter,1942),目前较少有研究实证检验创新质量对资源配置效率的影响。本文进一步使用PSM方法探讨企业高创新质量不同组态形式下的资源配置效率差异,揭示了不同组态形成的高创新质量与资源配置效率之间的“不同效”特点。研究区分了企业创新质量的成因,不仅弥补了创新质量与资源配置效率之间关系的研究缺失,也有助于企业识别、选择和实施正确的数据要素组合方案以实现高创新质量,进而提升其资源配置效率。

(三)研究启示

本文研究结论具有重要实践意义,可以为企业管理者与政策制定者提供启示。第一,企业高创新质量路径呈现多元模式,企业应结合内部数据要素禀赋、外部数据要素情况,科学地选择最优路径组合,灵活配置资源与能力,避免盲目模仿或路径依赖。对政策制定者而言,应跳出“一刀切”式政策供给范式,推动形成分层、分类和精准适配的数据要素战略支持体系,结合企业规模、行业特征与区域数字生态环境,因地制宜提升政策适配性。第二,构建多元共生的数据生态系统是实现高质量创新的重要路径。数据要素集聚与数据要素市场化在多条高创新质量路径中发挥了作用。企业应积极参与区域性数据集聚平台与行业性数据交易市场建设,通过跨

组织协同实现数据共享、知识溢出与创新协作。政策制定者应布局高能级数据产业集聚区、建设多层次数据要素市场与交易平台,鼓励区域内数据资源共享、开放和流通,提升数据流动性与可得性,培育本地数据生态系统。

(四)研究局限与展望

虽然本文采用动态QCA方法以及多因素联动匹配效应框架进行较为全面和系统的研究,但仍存在一定的局限性。第一,企业创新质量不仅会受到企业内外部数据要素资源的影响,还会受到其他生产要素和企业特征等因素的影响,未来研究可以采用新的分析框架展开研究,进一步补充和丰富对数字经济时代下企业创新质量的理解。第二,本文未对企业所处行业差异进行显性区分,未能深入探讨组态逻辑在不同行业(如制造业、服务业和数字产业)之间的分化特征。未来研究可开展行业间的路径异同对比,以增强研究结论的适用边界和政策针对性。

主要参考文献

- [1]陈丽莉,张若琪,戎珂.数据要素赋能企业创新:基于内外部资源视角[J].管理评论,2024,36(12):15-25.
- [2]陈强远,赵浩云,林思彤,等.中国高质量技术创新:情境叙事与测度体系[J].管理世界,2024,40(5):23-39.
- [3]陈元,贺小刚.何种公司治理促进了家族企业新质生产力:研发联盟中介下的组态分析[J].外国经济与管理,2025,47(12):59-78.
- [4]戴魁早,黄姿,梁银笛.数据要素与服务型制造发展[J].经济研究,2024,59(12):95-112.
- [5]戴魁早,王思曼,黄姿.数据交易平台建设如何影响企业全要素生产率[J].经济学动态,2023,(12):58-75.
- [6]杜善重,连立帅.公共数据开放对企业技术创新的影响机理研究[J].科研管理,2025,46(10):40-49.
- [7]方锦程,刘颖,高昊宇,等.公共数据开放能否促进区域协调发展?——来自政府数据平台上线的准自然实验[J].管理世界,2023,39(9):124-141,8.
- [8]郭家堂.公共数据开放与中国绿色全要素生产率:数据要素的视角[J].经济研究,2025,60(2):56-72.
- [9]郭立祥,岳书敬,高鹏.数实融合背景下公共数据开放对企业虚拟集聚的影响研究[J].经济学动态,2025,(2):55-73.
- [10]洪永淼,史九领.数据要素与数据经济学[J].经济理论与经济管理,2024,44(8):1-16.
- [11]李荣华,王娇娇,张磊.公共数据开放的新质生产力效应——基于企业数字创新的视角[J].财经研究,2025,51(10):50-63.
- [12]李雪琴,郑酌基,韩先锋.乘“数”而上:政府数据治理赋能企业数字创新[J].数量经济技术经济研究,2024,41(12):68-88.
- [13]梁孝成,吕康银,陈思.数据要素市场化对企业新质生产力水平的影响研究[J].科研管理,2025,46(2):12-21.
- [14]刘斐然,胡立君,范小群.产学研合作对企业创新质量的影响研究[J].经济管理,2020,42(10):120-136.
- [15]刘涛雄,张亚迪,戎珂,等.数据要素成为中国经济增长新动能的机制探析[J].经济研究,2024,59(10):19-36.
- [16]刘亚洲,扈文秀,吴邦海,等.数据要素应用对企业双元创新的影响研究[J].科技进步与对策,2026,43(6):22-31.
- [17]马淑琴,徐苗,张陈宇.数据要素应用与企业创新效应——来自中国A股上市公司的经验证据[J].商业经济与管理,2024,(2):50-67.
- [18]毛捷,管星华.地方政府纾困政策的效应研究:来自上市公司的证据[J].经济研究,2022,57(9):82-98.
- [19]沈坤荣,林剑威.链“岛”成“陆”:公共数据开放的技术创新效应研究[J].管理世界,2025,41(2):83-100,6.
- [20]史丹,孙光林.数据要素与新质生产力:基于企业全要素生产率视角[J].经济理论与经济管理,2024,44(4):12-30.
- [21]孙洁,梅静怡.新三板公司在多层次资本市场层级变动因何不同——基于机器学习重要因素挖掘与动态QCA因果关系分析[J].南开管理评论,2026,29(4):101-112,137.
- [22]唐要家,王钰,唐春晖.数字经济、市场结构与创新绩效[J].中国工业经济,2022,(10):62-80.
- [23]王海森,李纲.数据要素市场化对颠覆性技术的影响研究[J].情报学报,2025,44(5):535-548.
- [24]王艳,杨达.中国式管理会计体系变革:从数据要素到数据资产[J].管理世界,2024,40(10):171-188.
- [25]魏江,应瑛.企业数据基础观的构建[J].中国社会科学,2025,(12):53-68,200.
- [26]肖文雪,范德成.数字经济“走深入实”的力量:虚拟集聚对制造企业创新发展的影响研究[J].软科学,2025,39(2):85-

92,130.

- [27]徐怀宁, 刘莎莎. 数据要素集聚与企业债务融资——基于国家大数据综合试验区的准自然实验[J]. 经济管理, 2025, 47(4): 121-135.
- [28]徐晔, 王志超, 陶长琪. 数据要素市场化建设对企业资源配置效率的影响研究[J]. 科研管理, 2025, 46(9): 25-34.
- [29]杨亚平, 杨鑫, 陈凯. 数据要素市场化对企业数字化创新的影响[J]. 经济管理, 2025, 47(3): 5-25.
- [30]杨艳, 王理, 廖祖君. 数据要素市场化配置与区域经济发展——基于数据交易平台的视角[J]. 社会科学研究, 2021, (6): 38-52.
- [31]虞涛, 王兰芳. 数据交易平台建设何以促进企业技术创新“增量提质”?——基于市场主体之间的信息不对称[J]. 产业经济研究, 2025, (1): 114-128.
- [32]张朝华, 徐鹏杰. 数据要素集聚能提升城市经济韧性吗——来自大数据综合试验区建设的经验证据[J]. 宏观经济研究, 2024, (6): 59-76.
- [33]张杰, 郑文平. 创新追赶战略抑制了中国专利质量么? [J]. 经济研究, 2018, 53(5): 28-41.
- [34]张明, 陈伟宏, 蓝海林. 中国企业“凭什么”完全并购境外高新技术企业——基于94个案例的模糊集定性比较分析(fsQCA)[J]. 中国工业经济, 2019, (4): 117-135.
- [35]张明, 杜运周. 组织与管理研究中QCA方法的应用: 定位、策略和方向[J]. 管理学报, 2019, 16(9): 1312-1323.
- [36]张明, 蓝海林, 陈伟宏, 等. 殊途同归不同效: 战略变革前因组态及其绩效研究[J]. 管理世界, 2020, 36(9): 168-185.
- [37]张振堃, 文杰, 刘礼稳. 数据要素集聚的绿色治理效应研究——基于环保补贴有效性的视角[J]. 科研管理, 2025: 1-16.
- [38]赵丽, 胡植尧. 数据要素、动态能力与企业全要素生产率——破解“数据生产率悖论”之谜[J]. 经济管理, 2024, 46(7): 55-72.
- [39]郑国强, 张馨元, 赵新宇. 数据要素市场化能否促进企业绿色创新?——基于城市数据交易平台设立的准自然实验[J]. 上海财经大学学报, 2024, 26(3): 33-48.
- [40]郑威, 罗润凤. 数据要素集聚能否促进产业链现代化?——基于数字金融发展与数字人才集聚的双重视角[J]. 产业经济研究, 2024, (6): 43-55,69.
- [41]郑志强, 何佳俐. 唤醒“沉睡”的数据: 公共数据开放与新企业进入[J]. 外国经济与管理, 2025, 47(4): 82-97.
- [42]郑志强, 何佳俐, 陈尧. 公共数据开放与企业数字技术创新[J]. 财经研究, 2025, 51(2): 19-33.
- [43]Barney J. Firm Resources and sustained competitive advantage[J]. Journal of Management, 1991, 17(1): 99-120.
- [44]Biddle G C, Hilary G, Verdi R S. How does financial reporting quality relate to investment efficiency?[J]. Journal of Accounting and Economics, 2009, 48(2-3): 112-131.
- [45]Gracia-Castro R, Ariño M A. A general approach to panel data set-theoretic research[J]. Journal of Advances in Management Sciences & Information Systems, 2016, 2: 63-76.
- [46]Chesbrough H. Open innovation: The new imperative for creating and profiting from technology[M]. Nachdr. Boston, Mass: Harvard Business School Press, 2003.
- [47]Cong L W, Xie D X, Zhang L T. Knowledge accumulation, privacy, and growth in a data economy[J]. Management Science, 2021, 67(10): 6480-6492.
- [48]Ragin C. Redesigning social inquiry: Fuzzy sets and beyond[M]. Chicago, IL: University of Chicago Press, 2008.
- [49]Richardson S. Over-investment of free cash flow[J]. Review of Accounting Studies, 2006, 11(2): 159-189.
- [50]Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55.
- [51]Schumpeter J A. Capitalism, socialism, and democracy[M]. New York: Harper & Brothers, 1942.
- [52]Teece D J, Pisano G, Shuen A. Dynamic capabilities and strategic management[J]. Strategic Management Journal, 1997, 18(7): 509-533.
- [53]Vanhaverbeke W, Cloudt M. Theories of the firm and open innovation[A]. Chesbrough H, Vanhaverbeke W, West J. New frontiers in open innovation[M]. Oxford: Oxford University Press, 2014.

Same Destination, Different Effects: A Study on How Data Factors Empower Corporate Innovation Quality and Its Impact on Resource Allocation Efficiency

Fang Xiuyuan, Wu Jinxuan, Li Yuanxu

(School of Management, Fudan University, Shanghai 200433, China)

Abstract: The new wave of scientific and industrial revolutions has positioned data as a critical production factor that empowers corporate innovation quality. At present, systematic evidence remains limited regarding how firms achieve high innovation quality through alternative configurations of internal and external data factors, and whether heterogeneous configuration paths lead to differentiated outcomes. Drawing on the resource-based theory and the open innovation theory, this paper employs balanced panel data of Shanghai and Shenzhen A-share listed firms from 2017 to 2023 and adopts dynamic qualitative comparative analysis (QCA) to systematically analyze the configurational effects of internal data factors (data stock, data development capability, data-driven business applications, and data monetization) and external data factors (public data openness, data marketization, and data agglomeration) on innovation quality. Furthermore, it uses the PSM method to examine the heterogeneous effects of high innovation quality on resource allocation efficiency. The results reveal that firms can achieve high innovation quality through three paths: an “internal–external data resource synergy” path, an “internal data resource dominant” path, and an “internal data capability driven with external data resource compensation” path, exhibiting a clear characteristic of “different paths to the same destination”. Further analysis reveals that the impact of these paths on resource allocation efficiency vary significantly, characterized by “different paths to the same destination but with different effects”. Firms that form an internal value closed loop or acquire external data resources on demand can effectively reduce resource redundancy and allocation frictions, and thus significantly improve resource allocation efficiency after achieving high innovation quality. The study helps firms identify, select, and implement appropriate data factor configuration schemes to achieve high-quality innovation, thereby enhancing their resource allocation efficiency.

Key words: data factors; innovation quality; dynamic QCA; resource allocation efficiency

(责任编辑:王雅丽)