

公共数据开放与企业关键数字技术创新： 基于数字产业关联效应的分析

宋 建¹, 邵 蔚², 王颢琳²

(1. 南京审计大学联合研究院, 江苏 南京 211815;

2. 南京审计大学经济学院, 江苏 南京 211815)

摘 要:在推进数字经济战略与完善数据要素市场化配置背景下,公共数据开放成为释放数据价值、推动数字产业升级和重塑产业创新生态的重要抓手。本文以 2007—2023 年上市公司为样本,考察公共数据开放对企业关键数字技术突破的影响,并基于全国工商注册企业数据构建数字产业关联指数,识别其作用机制。研究发现,公共数据开放显著促进企业关键数字技术创新的“量”与“质”双提升,且经过系列稳健性检验后结论依然成立。机制检验表明,公共数据开放通过上游数字产业的“前向关联效应”、同行业数字产业的“旁侧效应”和下游数字产业的“后向关联效应”,分别经由生产要素纵向传递、同业竞争激励和下游采购需求扩张推动企业关键技术创新。在国有、规模大、年轻化、距离政府远、数字化程度高的企业,数据开放质量高的城市,以及高技术、技术资本密集型产业中,激励效应更为明显。研究结果为完善公共数据开放政策、赋能数字经济创新与产业协同升级提供了经验证据。

关键词:公共数据开放;关键数字技术创新;数字产业;产业关联效应

中图分类号:F016;F062.9;F273.1

文献标识码:A

文章编号:1005-0566(2026)06-0154-15

Public data openness and firms' breakthroughs in key digital technologies: An analysis based on digital industry linkage effects

SONG Jian¹, SHAO Wei², WANG Haolin²

(1. Joint Research Institute, Nanjing Audit University, Nanjing 211815, China;

2. School of Economics, Nanjing Audit University, Nanjing 211815, China)

Abstract: Against the backdrop of advancing the digital economy strategy and improving the market-based allocation of data factors, public data openness has become an important instrument for unlocking data value, promoting digital industrial upgrading, and reshaping the industrial innovation ecosystem. Using Chinese listed firms from 2007 to 2023 as the sample, this paper examines the impact of public data openness on firms' breakthroughs in key digital technologies. Based on national business registration data, this paper constructs a digital industry linkage index to identify the underlying mechanism. The results show that public data openness significantly improves both the quantity and quality of

基金项目:国家社会科学基金一般项目“数字化转型破解中国制造业‘卡脖子’技术困境的逻辑机理与靶向路径研究”(23BJL079);国家社会科学基金青年项目“数据要素驱动民营科技企业自立自强的‘双链对接’战略与跃迁路径研究”(25CJY035);全国统计科学研究重点项目“‘双链融合’实现高水平科技自立自强的识别框架、攻关机制与跃迁路径研究”(2024LZ022)。

作者简介:宋建(1987—),男,山东淄博人,南京审计大学联合研究院副教授,博士生导师,研究方向为数字经济与创新经济学。通信作者:邵蔚。

firms' key digital technology innovation, and the conclusion remains valid after a series of robustness tests. Mechanism tests indicate that public data openness promotes firms' key technological innovation through the "forward linkage effect" of upstream digital industries, the "lateral effect" of same-industry digital industries, and the "backward linkage effect" of downstream digital industries, operating respectively through vertical transmission of production factors, competition-induced innovation incentives, and expansion of downstream procurement demand. The effect is more pronounced among state-owned, large-scale, younger, geographically distant from government, and highly digitalized firms, as well as in cities with higher-quality data openness and in high-tech, technology and capital-intensive industries. This paper provides empirical evidence for improving public data openness policies and empowering digital economic innovation and coordinated industrial upgrading.

Key words: public data openness; key digital technology innovation; digital industry; industry linkage effects

在新一轮科技革命与产业变革加速演进、全球竞争重心持续转向关键核心技术与制度性能力竞争的背景下,关键数字技术创新已成为衡量国家科技竞争力、产业控制力和发展韧性的核心维度。“十五五”规划纲要围绕高水平科技自立自强、数字中国建设和新发展格局作出部署,明确提出健全数据要素基础制度,建设开放共享、安全可信的全国一体化数据市场,深化数据资源开发利用,激活数据要素潜能,并强调突破基础理论和核心技术。由此来看,关键数字技术创新不仅是企业层面的技术进步问题,更是数据制度供给、产业协同升级与市场循环畅通共同作用下的系统性命题。现实中,我国数字技术创新虽取得进展,但基础软件、核心器件、底层通用技术等环节仍受“卡脖子”技术约束,企业也存在高质量数据获取不足、知识协同不畅、应用验证场景不足等瓶颈。因此,公共数据开放能否供给覆盖产业链多环节、融合多维信息的数据资源,成为破解企业关键数字技术创新困境的重要突破口。

现实中,企业开展人工智能训练、数字孪生建模、工业互联网平台开发等关键数字技术创新活动,高度依赖政府治理、公共服务、基础设施运行等领域的权威性、基础性和关联性数据。这类数据多由公共部门掌握,具有覆盖范围广、标准化程度较高、跨周期可追溯等优势,能够降低企业获取高质量样本、构建行业基准模型等环节的边际成本。公共数据开放通过数据开放平台,使企业依法获取并利用政府及公共事业部门的数据资源^[1],弥补企业内部数据在广度、深度和权威性上的不足。然而,当前我国公共数据开放仍处于有

平台、缺流通,有数据、缺价值的初级阶段。开放数据在颗粒度、时效性和关联性上难以满足企业关键数字技术研发需求,跨部门、跨区域、跨层级互联互通也相对滞后,导致大量公共数据仍以“孤岛”形式沉睡。由此,公共数据开放能否以及如何转化为企业关键数字技术创新的现实动力,成为数据要素市场化配置改革中亟待验证的重要问题。围绕这一改革,中央出台《关于构建数据基础制度更好发挥数据要素作用的意见》,明确提出加强数据汇聚共享与开放开发、推进互联互通、打破“数据孤岛”;此后,《“数据要素×”三年行动计划(2024—2026年)》进一步强化了数据要素对科技创新与产业升级的乘数效应。在这一政策框架下,公共数据开放已成为检验数据制度改革创新效应的重要实践场景。

公共数据作为新型生产要素,其开放共享构成了企业数字技术创新外部知识供给与产业协同的重要制度基础。已有研究围绕数据开放度、治理透明度、公众参与度^[2]以及环境—政策—价值(EPV)框架^[3]等构建多层次评价体系,并从宏观层面验证了公共数据开放对区域协调发展^[4]和资本跨区域配置效率^[5]的影响;微观层面则发现其能够提升企业全要素生产率^[6],通过缓解信息摩擦、降低交易成本促进企业有效投资和异地投资^[7],并通过强化数据要素供给、促进数字化投资推动企业数字技术创新^[8-9]。相关研究为本文提供了重要理论和实证基础,但仍存在拓展空间:一是已有研究多以数字专利总量或一般数字技术分类衡量创新,尚未精准识别关键数字技术创新,而这类技术在知识密集度、研发风险及对公共数据

的依赖路径上不同于普通数字创新;二是现有文献多关注公共数据开放对企业创新的直接效应,较少从数字产业关联视角揭示其影响企业关键数字技术创新的内在机制。

本文边际贡献主要体现在如下:第一,拓展了公共数据开放微观经济效应研究。既有文献多关注公共数据开放对企业创新、全要素生产率等方面的影响,较少聚焦关键数字技术创新。本文进一步区分关键数字技术创新的数量与质量效应,丰富了相关研究。第二,优化了关键数字技术创新测度。本文依据《关键数字技术专利分类体系(2023)》和国际专利分类号(IPC),识别企业关键数字技术专利,并结合显性比较优势理论构建创新数量和质量指标,弥补既有测度不足。第三,深化了作用机制分析。本文从产业链上下游和同行业关联出发,检验公共数据开放通过数字产业前向关联、旁侧关联和后向关联影响企业关键数字技术创新的传导路径,并进一步考察企业、数据要素和行业层面的异质性影响,为完善公共数据开放政策、提升关键数字技术攻关能力提供经验证据。

一、研究假设

(一)公共数据开放对企业关键数字技术创新的直接效应分析

公共数据开放是政府或公共机构依法整合公共数据资源,并通过开放平台向社会提供公共数据服务的重要举措,其核心在于释放数据要素价值,推动数据要素市场化配置^[4]。与一般数据市场交易不同,公共数据具有公共性、非竞争性、低边际复制成本和较强共享属性^[10],能够以较低成本扩大企业可获取的数据边界,为企业数字技术创新提供基础要素支撑。从信息经济学视角看,企业创新活动高度依赖外部信息供给。公共数据平台开放的经济建设、财税金融、信用服务、劳动就业等多领域数据,与企业生产经营和政策环境密切相关,有助于企业由经验决策转向数据决策,提升技术路线选择和资源配置的科学性,降低机会成本^[11]。同时,公共数据开放能够减少政企之间和市场主体之间的信息摩擦,降低企业在数据

搜寻、采集、验证和业务协同等环节的制度性交易成本^[1],从而缓解创新过程中的不确定性,提高创新资源配置效率。

与一般数字技术相比,工业互联网核心平台、高性能计算芯片设计等关键数字技术具有技术复杂度高、研发周期长、路径依赖性强和不确定性高等特征,其突破更依赖大规模、高质量数据支撑、持续研发投入和产业生态协同。公共数据开放虽然能够为企业提供训练数据、试验验证环境和决策参考依据,但无法直接破解底层算法瓶颈、算力约束和生态适配等核心难题。因此,公共数据开放对关键数字技术创新的直接影响,更多体现为基础性赋能,而非决定性驱动因素。具体而言:一方面,公共数据开放降低了关键数字技术领域的研发准入门槛,为中小企业和科研机构提供原本难以获取的数据资源,拓宽关键数字技术创新的参与主体范围;另一方面,公共数据所反映的宏观趋势、行业标准和政策导向,有助于企业更准确判断技术方向,减少研发试错成本。由此,公共数据开放能够通过拓展数据来源、降低信息摩擦和弱化研发不确定性,为企业关键数字技术创新提供基础性支撑。基于以上分析,本文提出以下假设。

假设 H1:公共数据开放能够为企业关键数字技术创新提供基础性支撑,促进企业关键数字技术创新。

(二)公共数据开放激发企业创新的数字产业关联效应分析

公共数据开放平台为数字产业间的信息共享、技术交流和要素流动提供了重要载体,也为数字型新企业进入和成长提供了关键创新资源。随着数字经济分工体系不断深化,产业部门之间的协同依赖程度显著提升,公共数据开放能够促进数据要素在产业链体系中流动,推动企业与上下游及同行业企业之间的数字化联动^[12],并借助投入产出关联形成传导机制,引发产业部门间的连锁反应。因此,公共数据开放对企业关键数字技术创新的影响,不仅体现为数据资源的直接供给,还可能通过数字产业关联产生间接溢出效应。本文从产业链空间关联维度将数字产业关联效应划

分为纵向产业链关联和横向同业关联。其中,纵向产业链关联包括上游数字产业的“前向关联效应”和下游数字产业的“后向关联效应”;横向同业关联则体现为同行业数字产业的“旁侧效应”。公共数据开放通过推动数据要素在纵向产业链和横向同业网络中高效流通,依托上述三类产业关联渠道,差异化影响企业关键数字技术创新。

(1)数字产业关联的上游企业“前向关联效应”渠道。上游数字产业的“前向关联效应”主要体现为上游数字产业基础支撑和生产要素纵向传递。一方面,公共数据开放能够为上游产业提供基础数据资源,推动原材料、零部件、基础工艺、算法模型等环节优化资源配置,降低创新门槛,强化产业协同。依托公共数据开放平台,企业在协作与对接过程中能够实现多维知识传播和共享,加快技术知识在企业间扩散,提升企业获取外部技术溢出的能力,进而提高关键数字技术创新效率。另一方面,公共数据开放通过增强数字产业关联,促进生产要素沿产业链纵向传递。公共数据开放政策通过构建公共数据资源池,推动上游产业在生产流程中系统应用数据资源和数字技术,提升生产效率与产品质量,并将创新势能沿产业链向下游渗透。在这一过程中,基于知识的串联流通效应,企业能够形成“创新匹配”效应,进而推动关键数字技术创新。对于关键数字技术,上游前向关联效应尤为重要,因为其底层突破往往依赖基础算法、核心芯片、关键工艺等上游环节的长期积累,而公共数据开放能够为上游基础研究提供大规模、高质量的训练数据和验证环境,从而加快从0到1的原创性突破。基于以上分析,本文提出以下假设。

假说 H2a:公共数据开放能够激活上游数字产业,依托上游数字产业基础和生产要素纵向传递促进企业关键数字技术创新,即存在上游数字产业的“前向关联效应”。

(2)数字产业关联的同行业企业“旁侧效应”渠道。数字产业关联的同行业企业“旁侧效应”主要体现为同行业数字产业基础以及同业竞争引发的“逃离竞争效应”和“熊彼特效应”。同行业数字产业基础反映产业系统内部的要素循环与协同配

置能力。公共数据开放推动数据要素在产业内部横向流动,能够形成较为成熟的内部要素共享网络,为数字型企业提供低成本、高效率的创新投入渠道。就成本节约机制而言,当产业内要素关联度较高时,通用性中间产品和技术服务能够通过产业内多次交易实现成本分摊,边际使用成本趋于下降,从而降低企业开展关键数字技术创新的投入门槛。

从“逃离竞争效应”看,公共数据开放可能通过强化竞争压力激励企业创新。在同行业竞争加剧的背景下,企业为摆脱同质化竞争、获取差异化优势,会倾向于增加研发投入、强化数据资产管理,并将公共数据资源嵌入智能决策、个性化服务和商业模式创新之中,形成“技术—任务互补性”的价值跃迁^[13]。同时,公共数据资源的可得性提升了企业对数据建模、算法开发等复合型人才的需求。当企业高技能人才占比提高时,其数据处理、技术吸收和创新决策能力随之增强,有助于在数字技术领域形成难以复制的竞争优势^[10],进而推动关键数字技术创新。从“熊彼特效应”看,公共数据开放也可能产生一定抑制作用。关键数字技术创新通常具有研发周期长、投入强度高和不确定性大的特征^[14],企业需要在数据平台搭建、高端人才引进和技术落地应用等方面持续投入^[15]。若公共数据开放加剧行业竞争,压缩企业利润空间 and 市场份额,企业可能因资金约束而减少研发投入,从而削弱关键数字技术创新动力。

两类效应的相对强弱,取决于产业竞争格局、技术扩散阶段和公共数据开放的政策属性。首先,“逃离竞争效应”更容易出现在竞争较强且差异化空间尚未饱和的市场中,企业可通过数据驱动创新开辟新赛道;而“熊彼特效应”则更可能在市场集中度较高、创新收益易被在位大企业攫取的环境中占据主导。其次,人工智能、区块链、工业互联网等关键数字技术仍处于技术扩散的早期至中期阶段,应用场景和差异化路径尚未完全饱和,企业借助公共数据实现错位竞争的空间较大。最后,公共数据具有非竞争性和普惠性,其开放主要降低企业特别是中小企业的信息获取成本,而

非显著提高行业固定成本或准入门槛,因此对利润率的挤压相对有限。综合来看,竞争与创新之间可能存在倒 U 型关系^[16],但在本文样本期内,我国数字产业差异化空间仍较充裕,行业竞争尚未超出促进创新的合理区间,“逃离竞争效应”的激励作用可能强于“熊彼特效应”的抑制作用。因此,本文提出以下假设。

假说 H2b:公共数据开放能够激活同行业数字产业,同业市场竞争中的“逃离竞争效应”和“熊彼特效应”共同影响企业关键数字技术创新,即存在同行业数字产业的“旁侧效应”;同时,基于数字产业当前竞争特征,“逃离竞争效应”占据主导地位。

(3)数字产业关联的下游企业“后向关联效应”渠道。数字产业关联的下游企业“后向关联效应”主要体现为下游数字产业基础支撑和下游采购规模扩大。一方面,公共数据开放能够促进下游产业数字化转型,以需求升级倒逼焦点企业开展关键数字技术创新。公共数据释放的数据要素价值为下游产业发展提供基础支撑,并通过数据流通与共享推动产业链各环节协同整合,形成对企业关键数字技术创新的需求牵引。下游应用场景的复杂性和高标准要求,是推动上游核心技术突破的重要动力。公共数据开放有助于下游企业更精准识别技术瓶颈和性能短板,并向上游提出

更具挑战性的技术标准,这种“干中学”和“用中学”的反馈循环,对于关键数字技术迭代升级尤为重要。

另一方面,公共数据开放能够通过扩大下游采购规模,强化对上游企业关键数字技术创新的拉动作用。政府开放的公共数据资源为下游产业提供关键创新要素,下游企业在应用公共数据过程中提升数字技术能力,并通过两条路径向上游传导:一是对中间投入品形成技术标准倒逼,要求中间产品提升数字化水平和质量层次;二是通过产业链后向传导路径,将技术创新压力逐级传递至上游企业。由此,上游企业为适应下游需求升级,将加快生产流程再造和数字技术研发投入,持续优化数字化产品供给体系,推动产业链协同创新。与一般数字技术相比,关键数字技术的后向关联效应具有更强的技术门槛跳跃特征。下游需求升级往往要求上游企业在核心算法、精密制造、基础软件等环节实现根本性突破,而非简单参数优化。因此,下游需求牵引和采购规模扩张更可能形成原创性创新压力,推动企业关键数字技术突破。因此,本文提出以下假设。

假说 H2c:公共数据开放能够激活下游数字产业,依托下游数字产业基础,并通过扩大下游采购规模促进企业关键数字技术创新,即存在下游数字产业的“后向关联效应”。

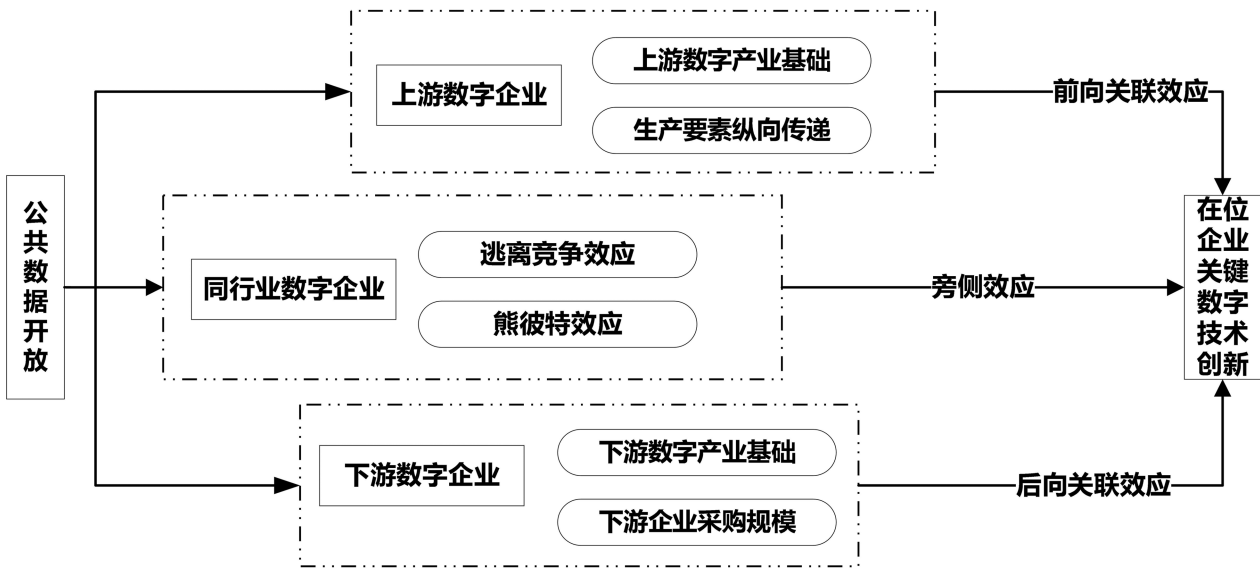


图1 公共数据开放影响企业关键数字技术创新逻辑图

二、研究设计

(一) 计量模型设定

本文以地级市公共数据开放平台上线作为准自然实验,构建多时点双重差分模型,检验公共数据开放对企业关键数字技术创新的影响。模型设定为:

$$Diginno_{ict} = \beta_0 + \beta_1 DID_{ct} + \beta_2 Controls_{ict} + \lambda_i + \eta_c + \mu_t + \varepsilon_{ict} \quad (1)$$

式(1)中,下标*i*、*c*、*t*分别表示企业、城市 and 年份。被解释变量 $Diginno_{ict}$ 表示企业关键数字技术创新。核心解释变量 DID_{ct} 表示公共数据开放政策变量,若企业*i*所在城市*c*在*t*年及以后已上线公共数据开放平台,则取值为1,否则为0。具体而言,本文依据复旦大学数字与移动治理实验室编制的《中国地方公共数据开放利用报告—城市(2023年度)》所记载的城市数据开放平台上线时序,梳理各地级市公共数据开放平台上线年份,并结合华中师范大学信息管理学院《中国政府开放数据利用研究报告(2022)》相关记录进行交叉核验。同时,平台识别遵循“平台域名包含 gov.cn”和“城市名称+数据资源/数据开放/公共资源”等关键词人工检索确认的标准,以确保平台运营主体和上线时间的准确性。 λ_i 、 η_c 、 μ_t 分别表示企业固定效应、城市固定效应和时间固定效应; ε_{ict} 表示随机扰动项。

本文选取如下企业层面控制变量 $Controls_{ict}$ 。企业年龄(*Age*,企业成立年限加1取自然对数)、企业规模(*Size*,企业总资产取自然对数)、总资产收益率(*Roa*,净利润与资产总额之比)、资产负债率(*Lev*,公司负债与资产总额之比)、流动资产占比(*Liquid*,流动资产与总资产之比)、现金流量(*Cflow*,经营活动产生的现金流净额与总资产之比)、员工规模(*Labor*,员工人数加1取自然对数)、股权结构(*Top10*,前十大股东持股比例)。此外,本文在机制检验部分加入城市层面控制变量:人口规模(*Pop*,年末总人口取自然对数)、经济发展水平(*Gdp*,人均地区生产总值取自然对数)、教育支出

水平(*Edu*,教育支出取自然对数)、外商投资水平(*Invest*,外商直接投资合同项目数量)、产业结构(*IndStr*,第三产业增加值与地区生产总值之比)、基础设施(*Basic*,每百人公共图书馆藏书册)、金融发展水平(*Fin*,年末金融机构各项贷款余额与地区生产总值之比)、科学支出水平(*Sci*,科学支出取对数)。

(二) 核心变量测算

1. 企业关键数字技术创新

既有研究多基于《数字经济及其核心产业统计分类(2021)》等文件,通过文本识别方法测度企业数字技术创新^[17],但该方法容易受到关键词选择和文本匹配规则的影响,难以精准刻画关键数字技术创新。本文以关键数字技术专利申请数量和被引次数分别衡量企业关键数字技术创新的数量和质量。首先,依据国家知识产权局发布的《关键数字技术专利分类体系(2023)》^①,将其中列示的国际专利分类号与上市公司专利申请和专利被引数据进行匹配,在IPC层面识别企业关键数字技术专利及其被引次数;其次,将专利申请数量和被引次数汇总至企业一年份和行业一年份层面;最后,基于显性比较优势思想构建企业关键数字技术创新能力优势指数(RICA)。具体公式为:

$$RICA_Qn_{imt} = \frac{Digitals_{imt}/Patents_{imt}}{\sum_m Digitals_{imt} / \sum_m Patents_{imt}} \quad (2)$$

$$RICA_Ql_{imt} = \frac{Digitals_Cite_{imt}/Patents_Cite_{imt}}{\sum_m Digitals_Cite_{imt} / \sum_m Patents_Cite_{imt}} \quad (3)$$

式(2)中, $RICA_Qn_{imt}$ 表示企业关键数字技术专利申请数量的显性比较优势; $Digitals_{imt}$ 和 $Patents_{imt}$ 分别表示行业*m*中企业*i*在*t*年的关键数字技术专利申请数量和总专利申请数量。式(3)中, $RICA_Ql_{imt}$ 表示企业关键数字技术专利被引次数的显性

①《关键数字技术专利分类体系(2023)》以服务国家重大需求为目标,聚焦新兴数字产业和前沿技术领域,选取人工智能、高端芯片、量子信息、物联网、区块链、工业互联网和元宇宙等7类关键数字技术,运用线分类法,将关键数字技术细分为3~5个层级的技术分支。

比较优势; $Digital_Cite_{imt}$ 和 $Patents_Cite_{imt}$ 分别表示关键数字技术专利被引次数和总专利被引次数。进一步地,参考 Byun 等^[18]的做法,本文将 $RICA_Qn$ 和 $RICA_Ql$ 分别与企业研发投入(RD)相乘,构造企业关键数字技术创新数量($Diginno_{imt}^{QN}$)和质量($Diginno_{imt}^{QL}$)指标:

$$Diginno_{imt}^{QN} = RICA_Qn_{imt} \times RD_{imt} \quad (4)$$

$$Diginno_{imt}^{QL} = RICA_Ql_{imt} \times RD_{imt} \quad (5)$$

其中, RD_{imt} 为企业 i 在 t 年的研发投入,用以衡量企业当期创新努力强度。相较于研发投入存量,当期研发投入更能反映企业年度创新资源配置及其对外部环境变化的即时响应。 $RICA$ 刻画企业在关键数字技术领域的专业化程度和比较优势, RD 反映企业创新投入强度,二者相乘能够同时捕捉企业关键数字技术创新的方向优势和资源投入规模,从而综合衡量企业关键数字技术创新数量与质量。

2. 数字产业关联

在识别数字产业关联时,本文以数字型企业进入为切入点。依据《数字经济及其核心产业统计分类(2021)》,本文将数字产品制造业、数字产品服务业、数字技术应用业和数字要素驱动业认定为数字经济核心产业,并结合其与《国民经济行业分类 2017》(GB/T 4754—2017)的对应关系,识别数字经济行业。进一步地,参考黄勃等^[17]的研究,基于数字经济相关政策文本构建数字技术特征词词典,对工商注册企业“经营范围”进行文本识别,将同时属于数字经济行业且经营范围包含数字技术特征词的企业认定为数字型企业^②。本文从产业链纵向关联和同行业横向关联两个维度构建数字产业关联指标。其中,纵向关联包括上游数字产业的“前向关联效应”和下游数字产业的“后向关联效应”,横向关联体现为同行业数字产业的“旁侧效应”。

(1) 数字产业的“前向关联效应”。前向关联效应主要体现为上游数字产业基础和生产要素纵向传递。首先,本文采用数字型新进入企业占全

部新进入企业比重与上游产业直接消耗系数的乘积,构建上游数字产业基础指数。数字型新进入企业数量来源于全国工商企业注册登记数据,直接消耗系数用于衡量某产业对上游中间投入的依赖程度。构造方式为:

$$Upstream_{cmt} = \frac{\sum_c [(\sum_{m \neq n} DigNewFirm_{cmt}) \alpha_{nt}]}{\sum_c \{ \sum_c [(\sum_{m \neq n} DigNewFirm_{cmt}) \alpha_{nt}] \}} \quad (6)$$

其中, $DigNewFirm_{cmt}$ 表示处于 c 城市的产业 m 在 t 年的数字型新企业注册数量; α_{nt} 表示直接消耗系数; $Upstream_{cmt}$ 表示城市—行业—年份层面的上游数字产业基础。

其次,本文采用在位企业向其上游供应商的采购额($Purchase$)衡量生产要素纵向传递。该变量反映焦点企业从上游环节获取中间投入、技术服务和数据资源的规模,能够刻画上游生产要素沿产业链向焦点企业传导的过程。

(2) 数字产业的“旁侧效应”。旁侧效应主要体现为同行业数字产业基础以及同业竞争引致的“逃离竞争效应”和“熊彼特效应”。对于同行业数字产业基础,本文采用数字型新企业进入与行业自身消耗系数、行业自身中间需求比例进行刻画。其中,行业自身消耗系数和中间需求比例分别对应直接消耗矩阵和中间需求矩阵的对角线元素,反映行业内部自我循环和要素共享程度。构造方式为:

$$Bystander^C_{cmt} = \frac{\sum_{n=m} DigNewFirm_{cmt} \times \varphi_{nt}}{\sum_c (\sum_{n=m} DigNewFirm_{cmt} \times \varphi_{nt})} \quad (7)$$

$$Bystander^D_{cnt} = \frac{\sum_{n=m} DigNewFirm_{cnt} \times \varphi_{nt}}{\sum_c (\sum_{n=m} DigNewFirm_{cnt} \times \varphi_{nt})} \quad (8)$$

其中, φ_{nt} 表示行业自身消耗系数;系数 φ_{nt} 表示行业自身中间需求比例; $Bystander^C$ 和 $Bystander^D$

② 数字型企业识别词典限于篇幅原因未呈现,备索。

分别表示基于自身消耗和自身中间需求测算的同行业数字产业基础。

对于“逃离竞争效应”,本文从创新投入和人力资本结构两个方面刻画。一是采用企业研发总支出占营业总收入的比重($RdShare$)衡量创新投入强度;二是采用硕士及以上学历员工占比衡量人力资本结构高级化程度。对于“熊彼特效应”,本文用企业扣除研发投入支出费用化金额后的营业总成本($Operat$)衡量企业在生产经营、规模扩张和技术落地中的资源消耗,以反映竞争压力可能带来的创新抑制效应。

(3)数字产业的“后向关联效应”。后向关联效应主要体现为下游数字产业基础和下游采购规模。对于下游数字产业基础,本文采用数字型新企业进入与中间需求比例构建指标。中间需求比例用于衡量某产业产品被下游产业作为生产投入的程度,构造方式为:

$$Downstream_{cmt} = \frac{\sum_c \left[\left(\sum_{n \neq m} DigNewFirm_{cmt} \right) \chi_{nt} \right]}{\sum_c \left\{ \sum_c \left[\left(\sum_{n \neq m} DigNewFirm_{cmt} \right) \chi_{nt} \right] \right\}} \quad (9)$$

其中, χ_{nt} 表示中间需求比例,即行业 n 产品被下游产业使用的比例; $Downstream_{cmt}$ 表示城市—行业—年份层面的下游数字产业基础。

对于下游采购规模,本文采用在位企业对下游客户的销售收入($Customer$)作为测度指标。该变量反映下游客户对焦点企业产品或服务的采购需求规模,能够刻画下游数字产业扩张通过需求拉动促进焦点企业关键数字技术创新的后向关联效应。

3. 工具变量

构建思路是城市间公共数据开放和数据共享能力不仅受到制度供给影响,而且受到自然地理条件所决定的数据互联互通成本约束。自然地理条件决定了跨区域公共数据流通和协同治理的天然成本,而全国公共数据开放制度的持续推进则使这种天然差异逐步转化为不同城市公共数据开放程度的异质性。首先,利用 NASA-SRTM 数字高程模型(DEM)和 HydroSHEDS 河网数据构建全国地表摩擦系数

图,将坡度、河流水系、水域和海岸线等自然地理要素转化为栅格层面的通行阻力,并采用最小累积成本路径($Minimum\ Cumulative\ CostPath, MCP$)算法计算任意城市 c 与城市 j 之间的最小天然摩擦成本:

$$MC_{cj} = \min_{p \in P_{cj}} \sum_{g \in p} F_g d_g \quad (10)$$

其中, P_{cj} 表示连接城市 c 与城市 j 的所有可行路径集合; F_g 表示栅格 g 的单位摩擦系数; d_g 为路径经过该栅格的距离。该指标反映了仅由自然地理条件决定的跨区域连通成本,摩擦成本越低意味着城市之间开展公共数据共享、数据交换和协同治理的潜在成本越低。进一步地,为刻画城市参与全国公共数据开放网络的天然优势,本文采用指数衰减函数构建城市间公共数据开放连接权重 $w_{cj} = \exp(-\lambda MC_{cj})$ 。其中, $\lambda > 0$ 为距离衰减参数。城市间天然摩擦成本越小,连接权重越高,表明两地越容易形成公共数据资源共享和数据互联互通。在此基础上,定义城市 c 的公共数据开放天然潜力指数($Public\ Data\ Opening\ Potential$)为:

$$PDO_i = \frac{1}{N-1} \sum_{j \neq c} \exp(-\lambda MC_{cj}) \quad (11)$$

其中, N 为样本城市总数。该指标完全由外生自然地理条件决定,反映城市在全国公共数据开放体系中的先天可达性和跨区域数据共享潜力。考虑到我国公共数据开放具有明显的制度推进特征,地方公共数据开放平台建设和公共数据授权运营均随着国家制度改革逐步推进,因此进一步构建全国公共数据开放制度推进项 $Policy_t$ 。本文采用 Logistic 函数描述全国公共数据开放制度的渐进扩散过程:

$$Policy_t = \frac{1}{1 + \exp[-k(t - t_0)]} \quad (12)$$

其中, t_0 表示制度扩散拐点年份, k 表示制度推进速度。该函数能够较好刻画公共数据开放由试点探索向全国推广的S型演进过程。最终,本文构建城市—年份层面的公共数据开放工具变量为:

$$IV_{ct} = PDO_c \times Policy_t \quad (13)$$

其中, PDO_c 反映城市天然公共数据开放潜力,属于固定地理禀赋; $Policy_t$ 反映全国公共数据

开放制度推进进程。二者交互形成的工具变量,刻画在全国公共数据开放制度持续深化背景下,不同城市因天然开放潜力差异而受到的异质性政策冲击。该工具变量具有较好的相关性与外生性。一方面,自然地理条件决定城市跨区域数据共享和公共数据互联互通的天然成本,天然开放潜力越高,越有利于推进公共数据开放,因而与地方公共数据开放水平具有稳定相关性;另一方面,坡度、水系和海岸线等自然地理特征长期稳定,不受企业关键技术创新活动和地方公共数据开放决策的反向影响。在控制城市固定效应和年份固定效应后,工具变量识别主要来自制度推进过程中不同城市天然开放潜力差异引致的变化。因此,该工具变量主要通过影响公共数据开放水平作用于企业关键数字技术创新,满足排他性要求。

(三)数据说明

本文以 2007—2023 年中国上市公司为研究样本,保留经营状态为“存续”和“在业”的企业。数据主要来源于三类:一是企业基本信息和财务指标,来自 CSMAR 数据库;二是专利数据,基于国家知识产权局中国专利数据库,并与《关键数字技术专利分类体系(2023)》匹配识别;三是投入产出数据,使用国家统计局发布的 2007 年、2012 年、2017 年、2018 年和 2020 年投入产出表。

三、估计结果分析

(一)基本结果

表 1 报告了公共数据开放影响企业关键数字技术创新的基准回归结果。第(1)列和第(4)列仅控制年份固定效应与个体固定效应,结果显示,公共数据开放对企业关键数字技术创新数量($Diginno^{QN}$)和质量($Diginno^{QL}$)的影响均在 1% 水平上显著为正。进一步控制企业层面变量后,第(2)列和第(5)列结果仍保持显著为正;在第(3)列和第(6)列继续加入城市固定效应后, DID 系数依然通过 1% 水平检验,表明公共数据开放能够显著促进企业关键数字技术创新。根据第(3)列和第(6)列结

果,公共数据平台上线后,企业关键数字技术专利数量提升 11.6%,专利质量提升 10.8%。由此,假说 H1 得到验证。

表 1 基准回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)
	$Diginno^{QN}$	$Diginno^{QN}$	$Diginno^{QN}$	$Diginno^{QL}$	$Diginno^{QL}$	$Diginno^{QL}$
DID	0.124 *** (0.025)	0.115 *** (0.025)	0.116 *** (0.026)	0.120 *** (0.027)	0.111 *** (0.027)	0.108 *** (0.027)
N	49 843	49 843	49 841	49 843	49 843	49 841
年份固定效应	是	是	是	是	是	是
个体固定效应	是	是	是	是	是	是
城市固定效应	否	否	是	否	否	是
企业控制变量	是	是	是	是	是	是

注:括号内数值表示聚类到企业层面的稳健标准误;***、**、* 分别表示 1%、5%、10% 的显著性水平。下同。

(二)平行趋势检验

为满足多期 DID 有效识别的前提,对公共数据开放影响企业数字创新的动态效应进行平行趋势检验^③。结果显示,政策实施前关键数字技术创新数量($Diginno^{QN}$)与质量($Diginno^{QL}$)的系数均不显著且围绕零值波动,表明处理组与对照组在事前无系统性差异,平行趋势假设成立。

(三)安慰剂检验

通过安慰剂检验^④,将处理组状态随机分配 500 次,发现随机系数集中于 0 附近且近似正态分布,而基准估计值均位于随机系数分布右侧。这表明,公共数据开放促进关键数字技术创新的效应并非随机因素或未观测变量驱动,结论稳健。

(四)其他稳健性检验

1. 剔除特殊样本

为进一步排除样本选择和特殊城市特征的干扰,本文进行两类样本剔除检验。第一,剔除 2012 年以前样本。鉴于中国首个公共数据开放平台于 2012 年正式上线,2012 年以前公共数据开放实践较少,本文剔除 2012 年以前样本后重新估计,结果见表 2 第(1)列~第(2)列。第二,剔除 4 个直辖市样本。考虑到直辖市在经济发展水平、技术基础和政策资源等方面通常优于其他城市,本文剔除直辖市样本后进行检验,结果见表 2 第(3)列~第(4)列。上述结果均表明,公共数据开放对企业关键数字技术创新的促进作用依然显著。

③ 限于篇幅,平行趋势检验图像未在正文汇报,感兴趣的读者可向笔者索取。本文还在平行趋势可能偏离的假设下进行了敏感性分析。限于篇幅,基于相对偏离程度限制及平滑限制所得的各年度处理效应检验结果未予列示,感兴趣的读者可向笔者索取。
④ 限于篇幅,安慰剂检验图像未在正文汇报,感兴趣的读者可向笔者索取。

2. PSM-DID 法

考虑到处理组和控制组企业可能存在个体特征差异,从而导致双重差分估计产生选择性偏差,本文采用 PSM-DID 方法进行稳健性检验。具体而言,以式(1)中的企业控制变量作为匹配变量,采用近邻匹配法构造可比样本,在匹配后样本基础上重新估计基准模型。结果见表 2 第(5)列~第(6)列,说明基准结论不受样本选择偏差影响,具有较好的稳健性。

表 2 稳健性检验:样本选择与 PSM-DID

变量	(1)	(2)	(3)	(4)	(5)	(6)
	剔除 2012 年以前样本	剔除直辖市样本	PSM-DID			
	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}
<i>DID</i>	0.141 *** (0.028)	0.128 *** (0.028)	0.112 *** (0.030)	0.103 *** (0.031)	0.116 *** (0.026)	0.102 *** (0.028)
<i>N</i>	41 132	41 132	39 782	39 782	45 163	45 163
固定效应	是	是	是	是	是	是
企业控制变量	是	是	是	是	是	是

注:固定效应若无特殊说明均控制个体固定效应、年份固定效应和城市固定效应。下同。

3. 高维固定效应

为进一步控制不可观测因素对估计结果的影响,本文在基准模型基础上加入“年份—省份”和

“行业—省份”交互固定效应,以控制地区年度冲击和行业区域差异。回归结果见表 3 第(1)列~第(2)列,公共数据开放的影响仍显著为正,说明基准结论较为稳健。

4. 更换被解释变量

为缓解变量测度误差可能带来的影响,本文进一步替换被解释变量进行稳健性检验。首先,根据《关键数字技术专利分类体系(2023)》中的专利小类,重新计算关键数字技术专利数量和被引次数,并参照式(4)和式(5)构造替代指标 *Diginno*^{QN}_{alter} 和 *Diginno*^{QL}_{alter},结果见表 3 第(3)列~第(4)列。其次,考虑到专利被引次数可能存在时间偏差,本文采用两类创新质量指标进行检验:一是使用知识宽度法测度关键数字技术创新质量,结果见表 3 第(5)列~第(6)列;二是基于 CNRDS 数据库专利引用信息,计算关键数字技术授权发明专利未来五年累计被引次数,以此衡量创新质量,结果见表 3 第(7)列。上述结果均表明基准结论依然成立。

表 3 稳健性检验:模型选择与变量测度

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN} _{alter}	<i>Diginno</i> ^{QL} _{alter}	<i>Breadth</i>	<i>Breadth</i>	<i>lnCite</i>
<i>DID</i>	0.136 *** (0.045)	0.122 *** (0.045)	0.080 *** (0.027)	0.092 ** (0.042)	0.070 * (0.028)	0.064 * (0.030)	0.059 * (0.025)
<i>N</i>	49 778	49 778	49 841	49 841	45 237	45 237	49 841
个体固定效应	是	是	是	是	是	是	是
年份—省份固定效应	是	是	否	否	否	否	否
行业—省份固定效应	是	是	否	否	否	否	否
年份固定效应	否	否	是	是	是	是	是
城市固定效应	否	否	是	是	是	是	是
企业控制变量	是	是	是	是	是	是	是

5. 混淆政策冲击

研究期内可能存在与公共数据开放相近的政策冲击,包括 2016 年开始的大数据试验区(*BigData*)、2014 年开始的信息惠民国家试点(*Informed*)、2014 年开始的“宽带中国”政策试点(*Broad*)、2019 年开始的人工智能创新发展试验区(*Intell*)等。为排除相关政策干扰,本文依据官方文件和权威新闻整理各城市政策实施时间,构建相应虚拟变量并纳入模型(1)进行控制。回归结果见表 4,进一步验证了基准结论的稳健性。

表 4 稳健性检验:排除政策干扰

变量	(1)	(2)	(3)	(4)	(5)
	<i>BigData</i>	<i>Informed</i>	<i>Broad</i>	<i>Intell</i>	四个混淆政策
Panel A: <i>Diginno</i> ^{QN}					
<i>DID</i>	0.121 *** (0.026)	0.132 *** (0.026)	0.123 *** (0.026)	0.116 *** (0.026)	0.135 *** (0.026)
<i>N</i>	49 841	49 841	49 841	49 841	49 841
固定效应	是	是	是	是	是
企业控制变量	是	是	是	是	是
Panel B: <i>Diginno</i> ^{QL}					
<i>DID</i>	0.115 *** (0.027)	0.125 *** (0.027)	0.116 *** (0.027)	0.108 *** (0.027)	0.127 *** (0.027)
<i>N</i>	49 841	49 841	49 841	49 841	49 841
固定效应	是	是	是	是	是
企业控制变量	是	是	是	是	是

(五) 内生性讨论

尽管公共数据开放平台具有准自然实验特征,且前文稳健性检验已在一定程度上缓解了内生性问题,但反向因果和遗漏变量偏误仍可能存在。为此,本文借鉴 Bartik 工具变量思路,构造两类公共数据开放工具变量:*IV1* 为通行时间半衰期设定为 24 小时并采用 Logistic 型制度推进路径构造的工具变量;*IV2* 为通行时间半衰期设定为 36 小时并采用政策关键节点线性插值推进方式构造的工具变量。表 5 报告了工具变量回归结果。第一阶段结果显示,*IV1* 和 *IV2* 的系数均显著为正,表明工具变量与公共数据开放政策具有较强相关性,并通过弱工具变量检验和不可识别检验。第二阶段结果显示,*DID* 系数均在 10% 水平上显著为正,说明在控制潜在内生性后,公共数据开放仍能显著促进企业关键数字技术创新,进一步支持本文基准结论。

表 5 内生性检验:工具变量回归

变量	(1)	(2)	(3)
	第一阶段	第二阶段	
	<i>DID</i>	<i>Diginno^{QN}_{IV}</i>	<i>Diginno^{OL}_{IV}</i>
<i>DID</i>	—	0.438 * (0.229)	0.409 * (0.241)
<i>IV1</i>	0.022 * (0.012)	—	—
<i>IV2</i>	0.014 *** (0.003)	—	—
<i>N</i>	33 309	33 309	33 309
企业控制变量	是	是	是
固定效应	是	是	是
第一阶段 <i>F</i> 值	23.99	—	—
<i>Kleibergen-Paap rk LM Statistic</i>	—	127.168	
<i>Cragg-Donald Wald F Statistic</i>	—	336.206	

四、进一步分析

(一) 机制检验:城市数字产业关联渠道

1. 数字产业关联的“前向关联效应”

表 6 结果验证了假说 H2a。第(1)列显示,公共数据开放显著促进上游数字产业基础发展,表明其能够通过释放多维信息和数据资源,引导资本、人才等要素向关键领域集聚,推动上游基础能力提升。第(2)列显示,公共数据开放显著促进生产要素纵向传递,说明其有助于上游企业精准识别中下游需求,优化原材料、技术和数据等要素配

置,减少供需错配,并推动生产要素沿产业链高效流动。由此,公共数据开放能够通过上游数字产业的“前向关联效应”促进企业关键数字技术创新,假设 H2a 得到验证。

表 6 数字产业关联的“前向关联效应”分析

变量	(1)	(2)
	上游数字产业基础	生产要素纵向传递
	<i>Upstream</i>	<i>Purchase</i>
<i>DID</i>	0.008 *** (0.001)	0.517 * (0.267)
<i>N</i>	44 648	48 824
固定效应	是	是
城市控制变量	是	否
企业控制变量	否	是

2. 数字产业关联的“旁侧效应”

为检验假说 H2b,本文首先依据投入产出表,采用对角线元素衡量产业自身关联程度,并分别基于直接消耗矩阵和中间使用矩阵构造 *Bystander^C* 和 *Bystander^D*。表 7 第(1)列~第(2)列结果表明,公共数据开放显著促进同行业数字产业基础发展。其作用机制在于,公共数据开放有助于打破信息壁垒,促进数据要素在行业内部横向流动,提升资源配置效率,并形成以数据要素为纽带的创新共同体。

进一步地,本文检验“逃离竞争效应”和“熊彼特效应”。表 7 第(3)列显示,公共数据开放显著提升企业创新投入;第(4)列显示,公共数据开放促进企业人力资本结构升级,表明数据开放降低了企业获取前沿知识和技术动态的门槛,推动企业向数据驱动的科学决策模式转型,并增强其对高学历人才的需求。上述结果说明,公共数据开放能够通过提升研发投入和优化人力资本结构,激发企业“逃离竞争”的创新动力。表 7 第(5)列进一步显示,公共数据开放也可能推高企业生产经营和技术落地成本,对数字技术创新形成一定挤压效应,说明“熊彼特效应”同样存在。

为判断两类效应的相对主导关系,本文进一步借鉴 Aghion 等^[16]的竞争—创新分析框架,考察竞争强度与关键数字技术创新之间的关系。表 7 第(6)列~第(7)列中,标准化勒纳指数系数显著为正,表明样本期内竞争与关键数字技术创新总体呈正向关系;第(8)列~第(9)列加入勒纳指数

表 7 数字产业关联的“旁侧效应”分析

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	同行业数字产业基础		逃离竞争效应		熊彼特效应	主导效应			
	<i>Bystander</i> ^C	<i>Bystander</i> ^D	<i>RdShare</i>	<i>HighEdu</i>	<i>Operat</i>	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}
<i>DID</i>	0.001 * (0.001)	0.001 * (0.001)	0.001 * (0.001)	0.361 *** (0.111)	0.146 *** (0.035)	—	—	—	—
<i>LN</i>	—	—	—	—	—	0.008 *** (0.002)	0.018 *** (0.002)	0.014 (0.009)	0.010 (0.012)
<i>LN</i> ²	—	—	—	—	—	—	—	0.006 (0.007)	-0.008 (0.009)
<i>N</i>	44 648	43 139	48 270	49 841	25 093	47 764	47 764	47 764	47 764
固定效应	是	是	是	是	是	是	是	是	是
城市控制变量	是	是	否	否	否	否	否	否	否
企业控制变量	否	否	是	是	是	是	是	是	是

平方项后,其系数不显著,说明竞争与关键数字技术创新未呈现倒 U 型关系。由此可见,在公共数据开放背景下,竞争加剧仍能促进企业关键数字技术创新的“量”与“质”提升,“逃离竞争效应”占据主导地位,假说 H2b 得到验证。

3. 数字产业关联的“后向关联效应”

表 8 结果验证了假设 H2c。第(1)列结果表明,公共数据开放显著提升下游需求强度,说明公共数据开放能够帮助企业精准识别用户需求,开发个性化产品和服务,进而激发数据驱动的消费场景创新。第(2)列中 *DID* 系数显著为正,表明公共数据开放能够扩大下游采购规模。其原因在于,数据共享有助于提升供应链协同水平,促进上下游产能、库存与需求的精准匹配,提高采购计划的科学性,并推动定制化产品开发和差异化采购需求增长。由此,公共数据开放能够通过下游数字产业的“后向关联效应”促进企业关键数字技术创新,假说 H2c 得到验证。

表 8 数字产业关联的“后向关联效应”分析

变量	(1)	(2)
	下游数字产业基础	下游企业采购规模
	<i>Downstream</i>	<i>Customer</i>
<i>DID</i>	0.006 *** (0.001)	0.310 *** (0.089)
<i>N</i>	44 648	49 841
固定效应	是	是
城市控制变量	是	否
企业控制变量	否	是

(二)异质性讨论

1. 企业异质性

(1)企业所有制性质。考虑到不同所有制企

业受公共数据开放影响可能存在差异,本文根据所有制类型构造虚拟变量,将样本划分为国有企业(*State* = 1)和非国有企业(*State* = 0),并构造交互项 *DID* × *State* 纳入基准模型。表 9 第(1)列 ~ 第(2)列结果表明公共数据开放对国有企业关键数字技术创新的促进作用更强。这可能是因为国有企业与政府部门联系更紧密,在数据资源获取、政策信息共享和技术资源整合方面具有优势,能够更有效地将公共数据转化为关键数字技术创新。

(2)企业规模。本文以企业资产总额衡量企业规模,并按照中位数将样本划分为大规模企业(*Scale* = 1)和小规模企业(*Scale* = 0)。表 9 第(3)列 ~ 第(4)列结果表明,公共数据开放对规模较大企业的促进作用更明显。原因在于,大规模企业在研发投入、技术积累和人才储备方面更具优势,能够更充分地吸收、处理和应用公共数据,并将其转化为关键数字技术创新成果。

(3)企业年龄。本文依据企业年龄中位数将样本划分为年轻企业(*FirmAge* = 1)和成熟企业(*FirmAge* = 0)。表 9 第(5)列 ~ 第(6)列结果显示,公共数据开放对年轻企业的促进作用更显著。年轻企业通常内部数据积累不足、资源约束较强,公共数据开放能够降低其数据获取门槛、弥补资源缺口,并增强其探索新技术和应用新数据资源的能力。

(4)企业与政府之间的地理距离。本文基于企业所在地与当地政府所在地的经纬度,采用 Haversine 公式计算地理距离,并按照中位数划分距离政府较远的企业(*Distance* = 1)和距离政府较近的企业(*Distance* = 0)。表 9 第(7)列 ~ 第(8)列结果

表明,公共数据开放对距离政府较远企业的关键数字技术创新促进作用更强,说明数据开放平台能够缓解企业与政府之间的信息不对称,降低地理距离带来的信息摩擦,从而赋能企业关键数字技术创新。

表 9 异质性分析:企业异质性

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}
<i>DID</i> × <i>State</i>	0.202 *** (0.036)	0.178 *** (0.036)	—	—	—	—	—	—
<i>DID</i> × <i>Scale</i>	—	—	0.155 *** (0.026)	0.151 *** (0.026)	—	—	—	—
<i>DID</i> × <i>FirmAge</i>	—	—	—	—	0.063 ** (0.031)	0.174 *** (0.028)	—	—
<i>DID</i> × <i>Distance</i>	—	—	—	—	—	—	0.113 *** (0.038)	0.138 *** (0.036)
<i>N</i>	49 841	49 841	49 841	49 841	49 841	49 841	49 841	49 841
固定效应	是	是	是	是	是	是	是	是
企业控制变量	是	是	是	是	是	是	是	是

2. 数字经济差异性

(1)数字型企业。本文进一步考察公共数据开放对不同技术属性企业的影响差异。依据前文数字技术关键词词典,将样本企业划分为数字型企业虚拟变量(*Defirm*)。表 10 第(1)列~第(2)列结果表明,公共数据开放对关键数字技术创新的促进作用主要集中于数字型企业,说明数字型企业更能将公共数据资源转化为关键数字技术创新,进一步验证了数字产业关联机制。

(2)数字化转型差异。本文采用上市公司年报数字化转型文本词频衡量企业数字化转型水平,并构造交互项 *DID* × *Digitaldegree*。表 10 第(3)列~第(4)列结果显示,公共数据开放对数字化转型水平较高企业的促进作用更显著。这表明,数字化基础较好的企业能够更快整合开放数据资源,提高数据利用效率,从而进一步放大公共数据开放的创新效应。

(3)公共数据开放质量。考虑到不同地区公共数据开放质量存在差异,本文依据历年《中国开放数林指数》^⑤构建公共数据开放质量虚拟变量(*Dum_Open*)。表 10 第(5)列~第(6)列结果表明,公共数据开放质量越高,其对企业关键数字技术创新的促进作用越明显。这说明,高质量开放平台能够提供更加丰富、准确的数据资源,并通过更开放透明的数据共享机制降低数据获取和利用

成本,从而提升企业关键数字技术创新效率。

表 10 异质性分析:数字经济差异性

变量	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}
<i>DID</i> × <i>Defirm</i>	0.019 ** (0.009)	0.030 *** (0.009)	—	—	—	—
<i>DID</i> × <i>Digitaldegree</i>	—	—	0.067 *** (0.023)	0.089 *** (0.023)	—	—
<i>DID</i> × <i>Dum_Open</i>	—	—	—	—	0.085 *** (0.022)	0.077 *** (0.023)
<i>N</i>	49 841	49 841	49 841	49 841	49 841	49 841
固定效应	是	是	是	是	是	是
企业控制变量	是	是	是	是	是	是

3. 行业异质性

(1)高新技术产业。考虑到高新技术产业企业对公共数据开放的响应可能更强,本文依据《战略性新兴产业分类目录》等文件识别高新技术上市公司,并构造交互项 *DID* × *HighTech*。表 11 第(1)列~第(2)列结果表明,公共数据开放对高新技术产业企业关键数字技术创新的促进作用更明显。这可能是因为高新技术企业对数据资源依赖程度较高,且具备较强的技术研发和数据处理能力,能够更有效地将开放数据转化为算法优化、产品迭代和技术突破。

(2)行业要素密集度。不同要素密集型行业对公共数据开放的利用能力存在差异。本文参考 2012 年中国证监会行业划分,将企业划分为劳动密集型、资本密集型和技术密集型行业。表 11 第(3)列~第(4)列结果表明,公共数据开放对资本密集型和技术密集型行业企业的关键数字技术创新促进作用更强。原因在于,资本密集型和技术密集型

⑤ “中国开放数林指数”是我国首个专注于评估公共数据开放水平的专业指数,由复旦大学数字与移动治理实验室制作出品。

行业对数据资源、研发投入和技术积累依赖程度更高,能够利用开放数据开展技术研发、生产优化和商业模式创新;而劳动密集型行业更多依赖人力成本优势,对复杂数据处理和技术应用需求相对较低,因此公共数据开放的直接创新驱动作用相对有限。

表 11 异质性分析:技术行业与行业要素密集度

变量	(1)	(2)	(3)	(4)
	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}	<i>Diginno</i> ^{QN}	<i>Diginno</i> ^{QL}
<i>DID</i> × <i>HighTech</i>	0.309 *** (0.030)	0.142 *** (0.028)	—	—
<i>DID</i> × <i>Tech</i>	—	—	0.169 *** (0.037)	0.125 *** (0.034)
<i>DID</i> × <i>Capital</i>	—	—	0.286 *** (0.053)	0.332 *** (0.054)
<i>N</i>	49 841	49 841	49 841	49 841
固定效应	是	是	是	是
企业控制变量	是	是	是	是

(三)拓展性分析

考虑到公共数据开放对不同类别关键数字技

术创新的影响可能存在差异,本文依据《关键数字技术专利分类体系(2023)》,进一步从人工智能、高端芯片、量子信息、物联网、区块链、工业互联网和元宇宙七类技术进行细分检验。表 12 结果显示,公共数据开放对各类关键数字技术创新数量和质量总体均具有促进作用,其中对人工智能类技术创新的促进效应更为显著。可能原因在于,人工智能作为典型的数据驱动型技术,其模型训练、算法优化和场景迭代高度依赖大规模数据资源。公共数据开放能够为自然语言处理、机器学习等领域提供丰富的数据支撑,从而更直接地推动人工智能技术创新。相比之下,高端芯片、量子信息等技术更依赖硬件研发、基础理论突破和长期技术积累,对公共数据开放的直接敏感度相对较弱。因此,公共数据开放对人工智能类关键数字技术创新的赋能作用更加突出。

表 12 企业关键数字技术细分分析

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A						
	<i>Intelligence</i> ^{QN}	<i>HighChip</i> ^{QN}	<i>Quantum</i> ^{QN}	<i>Internet</i> ^{QN}	<i>BlockChain</i> ^{QN}	<i>Industrial</i> ^{QN}	<i>Metaverse</i> ^{QN}
<i>DID</i>	0.066 *** (0.022)	0.043 * (0.025)	0.051 ** (0.022)	0.058 ** (0.026)	0.033 ** (0.016)	0.047 * (0.026)	0.050 ** (0.023)
<i>N</i>	49 841	49 841	49 841	49 841	49 841	49 841	49 841
固定效应	是	是	是	是	是	是	是
企业控制变量	是	是	是	是	是	是	是
	Panel B						
	<i>Intelligence</i> ^{QL}	<i>HighChip</i> ^{QL}	<i>Quantum</i> ^{QL}	<i>Internet</i> ^{QL}	<i>BlockChain</i> ^{QL}	<i>Industrial</i> ^{QL}	<i>Metaverse</i> ^{QL}
<i>DID</i>	0.105 *** (0.026)	0.048 * (0.028)	0.064 ** (0.025)	0.051 * (0.029)	0.041 ** (0.017)	0.047 * (0.028)	0.059 ** (0.025)
<i>N</i>	49 841	49 838	49 841	49 841	49 841	49 841	49 841
固定效应	是	是	是	是	是	是	是
企业控制变量	是	是	是	是	是	是	是

五、结论与启示

在数据要素市场化配置加快推进和关键数字技术竞争日益加剧的背景下,公共数据开放不仅是政府数据治理方式的转变,也是企业获取外部知识资源、拓展创新边界的重要制度供给。本文基于 2007—2023 年上市公司数据,考察公共数据开放对企业关键数字技术创新的影响。研究表明,公共数据开放能够显著提升企业关键数字技术创新水平,其作用并非局限于数据资源的直接供给,而是通过数字产业链上下游及同行业关联形成创新溢出。即上游环节强化要素传导,同行业环节激发竞争性创新,下游环节扩大需求牵引,

共同推动企业技术突破。这一效应在资源整合能力较强、数据利用能力较高以及技术基础较好的企业和行业中更为突出。

本文提出以下政策建议。第一,提升公共数据开放平台建设质效,增强数据要素对关键数字技术创新的基础支撑。中央政府应进一步完善公共数据开放制度体系,明确数据权属、开放标准、质量规范和安全边界,降低企业依法获取公共数据的制度成本。地方政府应依托既有公共数据开放平台,对公共数据进行分类治理、精准定级和动态更新,提高开放数据的颗粒度、时效性、可机读性和可关联性。同时,应围绕人工智能、工业互联

网、数字孪生等关键数字技术研发需求,建立更加完整的公共数据资源清单,推动高价值、高质量数据有序开放,使公共数据由“可获得”进一步转向“可利用”“可创新”。第二,发挥公共数据开放的数字产业关联效应,推动产业链协同创新。应以数据要素流通为纽带,推动公共数据资源向数字产业链上下游精准配置。具体而言,应支持上游数字产业提升基础技术供给能力,促进资本、技术、人才和数据等要素纵向传导;推动同行业企业基于开放数据开展差异化创新,形成“逃离竞争效应”,避免低水平同质化竞争;同时,依托下游应用场景释放需求牵引作用,扩大数字技术产品和服务采购规模,形成“数据开放—场景应用—技术迭代”的良性循环。第三,构建差异化、精准化的公共数据开放支持机制,提升政策适配性。政府应根据企业数字化基础、行业技术属性和区域数据开放水平实施分类施策。对于数字化基础较好的企业,应提供更高质量、更专业化的数据接口和应用场景,促进其开展关键数字技术攻关;对于年轻企业和距离政府较远的企业,应通过平台化开放降低信息摩擦和数据获取门槛;对于高新技术、技术密集型和资本密集型行业,应加强公共数据、研发资金和创新场景协同供给,推动公共数据开放更有效服务关键数字技术突破。

参考文献:

- [1] 欧阳伊玲,王愉靖,李平,等. 数据要素与城投债定价:基于公共数据开放的准自然实验[J]. 世界经济,2024(2): 174-203.
- [2] LEE G, KWAK Y H. An open government maturity model for social media-based public engagement[J]. Government information quarterly, 2012, 29(4): 492-503.
- [3] ZUIDERWIJK A, JANSSEN M. Open data policies, their implementation and impact: a framework for comparison[J]. Government information quarterly, 2014, 31(1): 17-29.
- [4] 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):124-142.
- [5] 蓝发钦,胡晓敏,徐卓琳. 公共数据开放能否拓展资本跨区域流动距离:基于异地并购视角[J]. 中国工业经济, 2024(9):156-174.
- [6] 彭远怀. 政府数据开放的价值创造作用:企业全要素生产率视角[J]. 数量经济技术经济研究,2023,40(9): 50-70.
- [7] 彭远怀,胡军. 政府数据开放与资本区际流动:企业异地投资视角[J]. 数量经济技术经济研究,2024,41(10): 89-110.
- [8] 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.
- [9] 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):83-104.
- [10] GOLDFARB A, TUCKER C. Digital economics[J]. Journal of economic literature, 2019, 57(1): 3-43.
- [11] JONES C I, TONETTI C. Nonrivalry and the economics of data[J]. American economic review, 2020, 110(9): 2819-2858.
- [12] 范合君,吴婷,何思锦. 企业数字化的产业链联动效应研究[J]. 中国工业经济,2023(3):115-132.
- [13] ACEMOGLU D, RESTREPO P. Tasks, automation, and the rise in US wage inequality[J]. Econometrica, 2022, 90(5): 1973-2016.
- [14] FIRK S, GEHRKE Y, HANELT A, et al. Top management team characteristics and digital innovation: exploring digital knowledge and TMT interfaces[J]. Long range planning, 2022, 55(3): 102166.
- [15] 蔡宏波,汤城建,韩金镭. 减税激励、供应链溢出与数字化转型[J]. 经济研究,2023,58(7):156-173.
- [16] AGHION P, BLOOM N, BLUNDELL R, et al. Competition and innovation: an inverted-U relationship[J]. The quarterly journal of economics, 2005, 120(2): 701-728.
- [17] 黄勃,李海彤,刘俊岐,等. 数字技术创新与中国企业高质量发展:来自企业数字专利的证据[J]. 经济研究, 2023,58(3):97-115.
- [18] BYUN S K, OH J M, XIA H. Incremental vs. breakthrough innovation: the role of technology spillovers[J]. Management science, 2021, 67(3): 1779-1802.