

DOI:10.3969/j.issn.1674-8131.2026.03.003

公共数据开放对数字经济企业集聚的影响研究

杨仁发, 江 灿

(安徽大学 经济学院, 安徽 合肥 230601)

摘 要:公共数据开放优化了区域数据要素供给,能够吸引数据驱动的数字经济企业进入,进而以数字经济企业集聚推动数字产业集聚。基于我国 297 个城市 2009—2022 年的面板数据,以政府数据开放平台上线为准自然实验,构建多期双重差分模型,运用 BERT 模型识别数字经济企业,研究发现:城市公共数据开放显著提升了数字经济企业集聚水平,其中,社会民生类、经济类和数字产业类公共数据开放的促进作用依次增强,而公共服务类公共数据开放对数字经济企业集聚具有负向影响;公共数据开放能够通过扩大数字产品需求、优化数字创新生态、加快数字人力资本积累 3 条渠道促进数字经济企业集聚;公共数据开放显著促进了营商环境较好、居民消费水平较高城市的数字经济企业集聚,而对营商环境较差、居民消费水平较低城市的数字经济企业集聚影响不显著。因此,应提高公共数据开放质量,改善数字创新生态和营商环境,加快数字人才培养,提升居民消费水平,促进数字经济企业高质量集聚。

关键词:公共数据开放;数字经济企业;产业集聚;数字产品需求;数字创新生态;数字人力资本;BERT 模型

中图分类号:F299.27;F129.9 文献标志码:A 文章编号:1674-8131(2026)03-0021-16

引用格式:杨仁发,江灿.公共数据开放对数字经济企业集聚的影响研究[J].西部论坛,2026,36(3):21-36.

Yang Renfa, Jiang Can. The impact of public data openness on the agglomeration of digital economy enterprises [J]. West Forum, 2026, 36 (3): 21-36.

* 收稿日期:2025-11-27;修回日期:2026-03-29

基金项目:国家社会科学基金后期资助一般项目(21FJB045);安徽省高校学科(专业)拔尖人才学术资助项目(gxbjZD2022002);安徽省社科规划一般项目(AHSKYY2023D065)

作者简介:杨仁发(1977),男,江西赣州人;教授,博士,博士生导师,主要从事产业经济与区域经济研究;E-mail:rfyang2002@163.com。江灿(2000),女,重庆人;硕士研究生,主要从事产业经济与区域经济研究;E-mail:jiangcan2023@163.com。

一、引言

《中华人民共和国国民经济和社会发展第十五个五年规划纲要》提出,要“壮大数字经济核心产业,发展新一代通信技术、云计算、区块链等产业,提升高端芯片、光电子器件、基础软件和工业软件等产业水平,打造具有国际竞争力的数字产业集群。”以集聚带动发展已成为共识,产业集聚通过促进企业、研发机构和金融机构的协同联动,加速要素流动,推动产业升级与结构转型(师磊等,2025)^[1]。数字产业集聚是指从事数字产品制造、数字产品服务、数字技术应用、数字要素驱动、数字化效率提升的企业主体和管理机构、科研机构、金融机构、基础设施供应商等相关经济主体在特定区域集聚(李天健等,2025)^[2]。企业是产业发展的微观基础,企业的集聚直接决定了产业集聚的形成与发展。数字经济企业指从事数字经济相关活动、以数字化手段增加产出和提升效率的企业,其增加值或营业收入主要来源于数字经济相关活动,是数字产品与服务的供给主体、数字技术的创新主体(周海川等,2024)^[3]、数字要素的整合主体(王如玉等,2018)^[4]。因此,数字经济企业集聚不仅是数字产业集聚的微观基础,更是驱动数字产业集聚形成与发展的核心力量。

现有文献主要聚焦于数字产业集聚,而对数字经济企业集聚的关注不足。相关实证研究从多角度考察了数字产业集聚的影响因素,分析发现:城市间创新互动、城市科技资本和科技成果积累对数字产业集聚具有显著正向影响(叶堂林等,2023)^[5],数据要素市场化通过促进数字技术创新和数字人力资本积累影响数字服务业和数字制造业集聚(赵放等,2024)^[6],服务外包政策通过人力资本积累效应、技术创新效应、数字营商环境优化效应推动数字产业集聚(桑瑞聪等,2025)^[7],智慧城市和低碳城市双试点通过推动数字技术创新和绿色技术创新、增加人力资本积累促进数字产业集聚(高志刚等,2026)^[8]。然而,鲜有文献探究数字经济企业集聚的影响因素。尽管数字经济企业集聚与数字产业集聚具有同步性,但微观层面的企业集聚与中观层面的产业集聚存在机制上的差异性。因此,有必要深入分析数字经济企业集聚的形成机制和影响因素。

随着数字经济的快速发展,数据成为驱动经济发展的核心生产要素(杨俊等,2022)^[9],其配置机制与制度安排对微观企业空间布局的影响日益凸显。公共数据包含政府内部的政务数据和从事市场活动的企业等微观主体创造的与公共利益相关的数据(方锦程等,2023)^[10],具有重要的治理价值、生产价值及社会价值,是海量数据的重要组成部分。公共数据开放增加了高质量的数据要素供给(Janssen, 2011)^[11],提高了政府信息透明度,有利于资源配置优化、市场完善和统一大市场建设(彭远怀等,2024;王生年等,2025)^[12-13],进而通过促进企业数字化转型(李佳林等,2024)^[14]、开放式创新(杨壮等,2025)^[15]等推动企业高质量发展。作为数据驱动型企业,数字经济企业的发展高度依赖高质量数据要素,公共数据开放让原本难以获取的公共数据成为企业可利用的重要资源,能够吸引数字经济企业,进而形成企业集聚效应。李晓龙和牟业洪(2025)^[16]研究发现,公共数据开放能通过降低制度交易成本、提高信息搜寻效率和增强数字技术创新促进数字产业集聚,但该研究仍聚焦于产业层面,未探讨企业层面的集聚效应。

鉴于上述,本文在已有研究的基础上,探讨公共数据开放对数字经济企业集聚的影响及其作用机制,并将城市政府数据开放平台上线作为准自然实验,构建多期双重差分模型,采用2009—2022年297个城市的数据进行实证检验。本文的边际贡献主要包括:第一,从微观企业集聚角度拓展了公共数据开放的经济效应研究,并为公共数据开放对数字经济企业集聚的促进作用提供了经验证据。第二,从数字产品需求、数字创新生态、数字人力资本3个维度探究了公共数据开放影响数字经济企业集聚的作用机

制,有助于深入认识数据赋能数字经济发展的内在机理以及企业集聚的形成机制。第三,采用 BERT 模型识别数字经济企业,以创新主体、创新平台、创新环境三大子系统的耦合协调度刻画城市数字创新生态,基于数字需求关键词的百度指数测算数字产品的潜在市场需求,为相关实证研究提供了方法借鉴。第四,进一步从营商环境和消费水平两个方面考察了公共数据开放影响数字经济企业集聚的异质性,为提高公共数据开放质量和促进数字经济企业集聚发展提供了政策启示。

二、理论分析与研究假说

1. 公共数据开放对数字经济企业集聚的影响

企业选址需要考虑多方面的因素。作为数据驱动型企业,数字经济企业的生产经营决策、核心算法训练、场景创新落地等均高度依赖多维度、大规模、高质量的数据要素,并需要相应的数字信息平台和数据交易平台等数字基础设施的支持;同时,数字产品生产和数字服务供给通常需要多主体协同与多技术融合,产业链配套和延伸条件也是决定数字经济企业发展潜力的重要因素。因此,地区数据要素配置、数字基础设施、产业链状态是影响数字经济企业选址决策的三个关键因素。

第一,公共数据开放促进数据要素集聚、流动与共享。数据要素分散产生并存储于不同部门和主体,若数据流动和共享受到阻碍,将导致数字经济企业面临数据可得性不足、数据获取成本高、数据质量参差不齐等问题(杨俊等,2022)^[9]。公共数据具有标准化、高可信度、覆盖范围广等特点,公共数据开放会推动跨部门、跨区域的数据流动和共享,打破“数据孤岛”,并提高其数据要素质量,从而扩大数字经济企业对高质量数据要素的获取渠道。这种数据供给优势为区域内数字经济企业的创立与发展提供了更好的数据资源基础(刘阳等,2023)^[17],有利于区域数字经济企业集聚水平提升。

第二,公共数据开放促进数字基础设施建设。公共数据开放要求各级党政机关、企事业单位加强公共数据的汇聚共享和开发开放,地方政府为实现公共数据合规开放,会优先建设完善数字政务平台、政务数据共享平台等(郑志强等,2025)^[18],并优化平台的数字技术适配功能。建设和完善数字政务平台和政务数据共享平台,不仅有助于减少政府机构间的信息壁垒,扩大政务服务覆盖范围、提升政务服务效率,还会带动其他数字基础设施建设,为数字经济企业的创立和发展提供更好的硬件环境,从而吸引数字经济企业进入和集聚。

第三,公共数据开放有利于数字产业链完善。一方面,公共数据开放推动相关衍生产品研发、生产与销售,能够延伸区域内数字产业链条(赵志浩等,2025)^[19]。另一方面,公共数据具有可重复利用且价值不损耗的特性(熊广勤等,2025)^[20],降低了数据采集成本,使企业能低成本切入产业链细分环节,从而丰富数字产业链层级。数字产业链的丰富和延伸为数字经济企业的创立和发展提供了更好的产业配套(邹颖等,2026)^[21]。此外,数字经济企业的创新高度依赖隐性知识的传递,但地理距离会显著削弱隐性知识的传递效率(唐露等,2024)^[22],这将促使数字经济企业在数字产业链较完善的区域集聚(阮平等,2016)^[23]。

基于上述分析,本文提出假说 1:公共数据开放能够促进数字经济企业集聚。

2. 公共数据开放、数字产品需求与数字经济企业集聚

一方面,公共数据开放有利于扩大区域数字产品和服务的市场需求。公共数据开放使原本因数据

缺失而无法实现的潜在需求转化为实际市场需求(贾俊秀等,2021)^[24],如种植补贴数据、灾害数据、区域气象数据、物流数据等的开放会激发农业数字化生产和管理需求。公共数据涵盖企业难以自行获取的真实场景信息,企业接入后可针对性优化原数字产品和服务,而用户也愿意为体验更优质的数字产品和服务付费,因此,公共数据开放不仅能扩大居民的数字产品和服务需求,还能推动数字需求迭代升级。此外,公共数据开放直接增加了政府对数字产品和服务的需求。公共数据覆盖社会民生、经济运行、产业发展等多领域,海量数据采集会产生专业传感器、物联网终端等数据采集设备需求(赵志浩等,2025)^[19];公共数据开放对计算算力、存储容量等提出更高要求,将扩大高性能服务器、分布式存储系统、云计算服务等产品与服务的需求;公共数据的敏感性与高价值属性还会驱使政府和企业数据采集、传输、处理和运营等环节建立合规体系(王东方,2025)^[25],进而推动数据安全咨询、合规技术工具等数字产品和服务需求。

另一方面,数字产品和服务需求的扩大促进数字经济企业集聚。根据经济地理学理论,在规模效应的激励下,企业倾向于在市场潜力较大的地区布局生产(李剑培等,2021)^[26]。数字产品和服务需求较大的地区更易吸引数字生产要素进入,并形成数字经济企业集聚(Krugman,1991)^[27]。服务主导逻辑理论认为数字产品和服务通过场景化体验、即时互动反馈和具身交互等方式,强化用户的情感依赖与价值认同(赵振,2015;吴瑶等,2025)^[28-29],从而促使消费者形成对本地数字产品和服务的显著偏好。这种本地偏好推动数字经济企业本地化研发与市场深耕,使数字产品和服务与本地偏好形成强绑定,降低企业外迁意愿;同时,稳定增长的数字产品和服务需求降低了市场不确定性,能够吸引外部数字经济企业进入,进一步扩大数字经济企业集聚规模。

基于上述分析,本文提出假说2:公共数据开放通过扩大数字产品需求促进数字经济企业集聚。

3. 公共数据开放、数字创新生态与数字经济企业集聚

一方面,公共数据开放有利于数字创新生态优化。数字创新生态系统由创新主体、创新平台和创新环境三个子系统构成(吴艳等,2025)^[30],三者协同适配是创新生态系统进化的关键(闫晓勇等,2023)^[31]。公共数据开放能够推动创新主体创新能力升级、创新平台功能完善,并促进创新主体、创新平台与创新环境的耦合协调发展。数字经济企业是数字技术创新的关键主体(岳晓旭等,2026)^[32],而数字技术创新通常具有研发周期长、投入规模大和失败风险高等特点,公共数据开放有助于企业选择适宜的技术创新路径、降低技术创新风险,并以数据赋能企业数字技术创新能力提升。通过整合公共数据与行业数据服务创新主体是创新平台的重要功能(马蕾等,2025)^[33],但数据量不足、数据质量低、数据不合规等问题制约着创新平台的服务水平提升。公共数据开放能够为要素市场化交易平台和工业互联网平台等创新平台提供标准化、高质量的数据资源,降低创新平台的数据获取成本与合规风险,从而强化创新平台的要素枢纽功能,更好地帮助创新主体精准获取高质量的创新资源、降低创新试错成本、缩短技术迭代周期,并加速数字技术与产业场景的融合落地,进而形成良好的数字创新生态。

另一方面,数字创新生态的优化促进数字经济企业集聚。数字创新生态的改善,不仅能促进数字经济企业的技术创新,还能通过加快要素集聚、强化平台赋能及制度激励、推动产业链协同等路径促进数字经济企业集聚。在良好的数字创新生态下,产业链合作畅通有效,企业创新和交易成本降低,供应链中断风险减小(贾卫峰等,2025)^[34],能够吸引核心数字经济企业集聚;同时,核心数字经济企业通过开放API接口、输出数字技术标准、共享轻量化数字工具等数字化溢出,帮助边缘企业降低技术迭代的研发成本与合规风险(岳晓旭等,2026)^[32],从而吸引边缘数字经济企业进入。因此,区域数字创新生态的

优化有助于核心数字经济企业和边缘数字经济企业基于差异化的价值获取与风险规避动机主动进入和集聚,形成“核心引领-边缘配套”的企业集聚格局。

基于上述分析,本文提出假说 3:公共数据开放通过优化数字创新生态促进数字经济企业集聚。

4. 公共数据开放、数字人力资本积累与数字经济企业集聚

一方面,公共数据开放有利于数字人力资本积累。数字人力资本指掌握关键数字技术,具备开发或应用数字解决方案、管理复杂数字系统的能力,并能直接驱动企业数字化转型和创新的专业技术人才(余典范等,2025)^[35]。数字技术的价值实现高度依赖人力资本积累,数字人才短缺将导致数据处理效率低下、技术应用不充分、生产流程优化受阻,进而影响企业价值创造(Tambe et al., 2020)^[36]。公共数据开放可以通过数据要素的知识溢出效应助力劳动者培养数字思维与技能,进而提升劳动力要素质量(赵放等,2024)^[37];公共数据开放能加速数据要素与实体产业深度融合,拓展数据应用场景,增加数据开发需求,从而推动区域数字人力资本需求增长;公共数据开放会降低求职与招聘双方的搜寻成本,缓解劳动力市场的信息匹配障碍,提高劳动力配置效率(刘满凤等,2022)^[38];公共数据开放还有助于提升居民生活质量,促进高端数字人才流入。

另一方面,数字人力资本的积累促进数字经济企业集聚。劳动力资源具有正外部性,本地人力资源升级有助于企业通过探索式创新突破关键技术瓶颈,研发出有竞争优势的新产品,并助力企业开拓新产品市场(姜海宁等,2024)^[39]。因此,地区人力资源状况是数字经济企业选址、迁移和集聚需要考虑的重要因素(刘传明等,2025)^[40]。数字经济企业的创立和发展依赖多学科交叉的技术支持和多样化的资源配置,区域数字人力资本积累能够为本地企业发展提供多样化的人才支持,进而提升数字经济企业创业成功的可能性和进一步发展的空间。数字人力资本积累还有利于提升数字经济企业与高技能劳动力的匹配效率,降低数字经济企业的劳动雇佣成本,吸引同类及配套企业集聚(王林辉等,2024)^[41],进而提升数字经济企业的集聚规模与竞争力。

基于上述分析,本文提出假说 4:公共数据开放通过加快数字人力资本积累促进数字经济企业集聚。

三、实证研究设计

近年来,我国高度重视公共数据的开发利用,连续发布一系列文件完善政策框架。在中央顶层设计下,各级地方政府积极推动公共数据开放工作。2012 年上海市和北京市率先开展公共数据开放实践。2014 年,政府数据开放平台在武汉、佛山、无锡等地陆续上线。2015 年,国务院印发《促进大数据发展行动纲要》,提出要“稳步推动公共数据资源开放”。据统计,截至 2023 年 8 月,政府数据开放平台增长至 226 个,其中省级平台 22 个,城市平台 204 个。在政府数据开放平台建设中,各地制定了详细的数据开放目录,建立了高效的数据更新机制,并不断完善数据安全保障措施,有效提升了公共数据开放的质量和效率。基于此,本文将各城市政府数据开放平台上线视为准自然实验,利用多期双重差分模型实证检验公共数据开放对数字经济企业集聚的影响及其机制。

1. 基准模型构建

构建基准模型如式(1)所示:

$$DEA_{it} = \alpha_0 + \alpha_1 DID_{it} + \gamma M_{it} + \delta_t + \mu_c + \varepsilon_{it} \quad (1)$$

其中,下标 i 和 t 分别代表城市和年份,被解释变量 DEA_{it} 为城市 i 在 t 年的数字经济企业集聚水平,核心解释变量 DID_{it} 为公共数据开放的虚拟变量(样本期内政府数据开放平台上线的城市在上线当年及之后年份取值为 1,否则取值为 0), M 为控制变量合集, δ_t 表示年份固定效应, μ_i 表示城市固定效应, ε_{it} 为残差项。

(1)城市数字经济企业集聚水平的测度。评估数字经济企业集聚水平关键在于对数字经济企业进行识别。现有文献多将属于信息传输、软件和信息技术服务业以及计算机、通信和其他电子设备制造业的企业认定为数字经济企业(杨鹏等,2025)^[42],但该方法不能准确全面地筛选出数字经济企业。因此,本文对于可以通过二级和三级分类确定是否属于数字产业的企业,依据相应分类进行确认;对于通过三级分类仍无法确认的企业,采用微调后的 BERT 模型,基于企业的经营范围对其是否属于数字经济企业进行认定。BERT 是基于深度学习的自然语言处理(NLP)模型,通过预训练和微调能高效实现自然语言处理任务。考虑到直接应用预训练模型往往难以满足特定场景的需求,本文采用监督微调方法提升模型准确性(Huang et al., 2023)^[43]。首先,依据《数字经济及其核心产业统计分类(2021)》,人工标注 4 921 家上市公司是否属于数字产业,构建专属的监督数据集。然后,将该数据集按 9:1 分为训练集和测试集,利用这些企业的经营范围文本数据对 BERT 模型进行微调。在 BERT 模型微调过程中借助截断法截取前 509 个 tokens,采用逐层降低学习率的优化策略提升模型准确率(Sun et al., 2019)^[44],微调后模型的分类准确率收敛至 85%以上,进而依据企业经营范围识别非上市企业中的数字经济企业(参见知网附件材料)。筛选出数字经济企业后,借鉴刘晴等(2022)^[45]的方法,采用基于数字经济企业数量的区位熵衡量城市数字经济企业集聚水平,即 DEA_{it} 为 i 城市 t 年数字经济企业数量占比与全样本 t 年数字经济企业数量占比之比。

(2)控制变量的选取。为减少遗漏变量对模型估计的影响,选取以下城市层面的控制变量:一是城市人口规模(POP),用对数化处理的常住人口数量(万人)衡量;二是经济规模(GDP),用对数化处理的国内生产总值(亿元)衡量;三是数字基础设施(DI),用长途光缆密度衡量;四是交通基础设施(TI),用建成区路网密度衡量;五是技术创新水平(INN),用对数化处理的发明专利数量衡量;六是产业结构(IS),用第三产业增加值占 GDP 的比重衡量;七是人力资源(HR),用每万人高等学校在校学生数衡量;八是金融发展水平(FD),用年末金融机构存贷款余额与 GDP 之比衡量;九是外商直接投资(FDI),用对数化处理的外商投资企业工业总产值(万元)衡量。

2. 样本选择和数据处理数据

本文以我国地级及以上城市为研究样本,根据数据的可得性和完备性,选取 297 个城市 2009—2022 年的数据进行实证分析。公共数据平台开放的信息来自各城市政府官方网站,企业经营范围数据来自工商注册信息,上市企业数据来自国泰安数据库和企业年度报告,其余数据来自国泰安数据库、各级政府统计年鉴和政府工作报告。剔除关键变量缺失的样本,对其他缺失数据采用前向插值法进行补齐,对连续型变量进行上下 1%的缩尾处理。主要变量的描述性统计结果如表 1 所示。

表 1 主要变量描述性统计结果

变量类型	变量名称	变量符号	样本量	均值	标准差	最小值	最大值
被解释变量	数字经济企业集聚	DEA	4 158	1. 635	0. 969	0. 504	6. 649
核心解释变量	公共数据开放	DID	4 158	0. 218	0. 413	0	1

续表 1

变量类型	变量名称	变量符号	样本量	均值	标准差	最小值	最大值
控制变量	人口规模	<i>POP</i>	4 158	5. 682	0. 995	2. 727	7. 639
	经济规模	<i>GDP</i>	4 158	7. 291	0. 991	4. 986	9. 912
	数字基础设施	<i>DI</i>	4 158	0. 177	0. 316	0. 001	2. 236
	交通基础设施	<i>TI</i>	4 158	7. 407	2. 197	2. 190	15. 01
	技术创新水平	<i>INN</i>	4 158	4. 792	1. 908	0. 693	9. 619
	产业结构	<i>IS</i>	4 158	0. 429	0. 100	0. 228	0. 734
	人力资源	<i>HR</i>	4 158	173. 6	189. 8	10. 33	933. 9
	金融发展水平	<i>FD</i>	4 158	2. 254	1. 220	0. 588	6. 760
	外商直接投资	<i>FDI</i>	4 158	13. 31	2. 194	7. 797	18. 09

四、实证结果分析

1. 事前平行趋势检验与基准回归

满足事前平行趋势假设是进行双重差分检验的重要前提,本文参考张鑫宇等(2023)^[46]的做法,将政策实施前一年作为基期,采用事件研究法进行事前平行趋势检验。从图 1 可以看出,处理效应在政策实施前并不显著,在政策实施后均显著为正,说明公共数据开放对数字经济企业集聚产生了显著的政策效应,该效应并非处理组和控制组之间固有的发展趋势所致,符合事前平行趋势假设的要求。

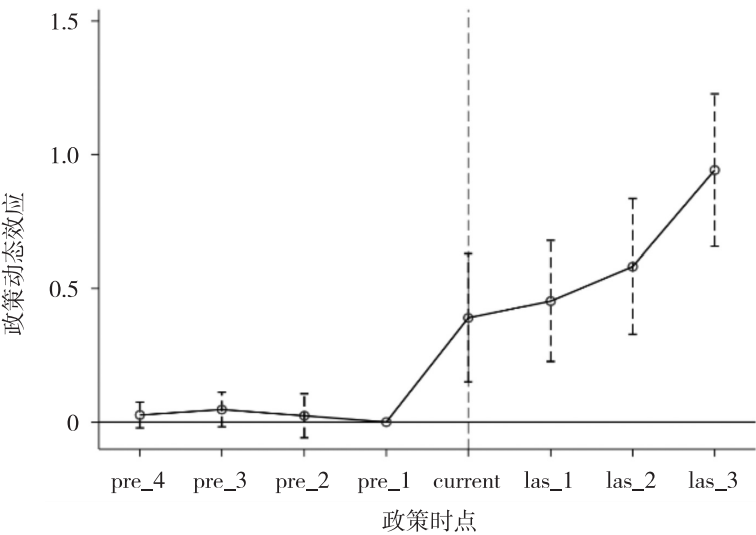


图 1 平行趋势检验结果

表 2 报告了基准回归结果,*DID* 的估计系数在 1% 的显著性水平上为正,表明城市公共数据开放显著提升了其数字经济企业集聚水平,假说 1 得到验证。考虑到公共数据有不同的类型,本文进一步分析不同类型公共数据开放对数字经济企业集聚的影响。借鉴欧阳伊玲等(2024)^[47]的研究,将公共数据分为公共服务类(*ser*)、社会民生类(*so*)、经济类(*eco*)、数字产业类(*di*) 四类。首先,采集各城市已开放的公共数据集条目,进行去重、清洗与人工核验等处理,每条数据集均包含数据名称、所属领域、发布省份、

发布城市、首次发布时间等信息;然后,根据数据集的数据名称或所属领域是否包含四类数据的关键词(见表3),确定其所属数据类型,分别加总各类型数据当年新增量;最后,将当年新增量与上年存量加总,得到城市当年已开放的各类型数据总量。以城市当年不同类型公共数据开放数量为核心解释变量,检验结果见表4。社会民生类、经济类、数字产业类公共数据的开放都显著促进了数字经济企业集聚,其中,数字产业类公共数据开放对数字经济企业集聚的促进作用最大,经济类公共数据开放次之,社会民生类公共数据开放的作用较小;而公共服务类公共数据开放对数字经济企业集聚产生了负向影响,这可能与样本期间样本城市开放的公共服务类公共数据质量不高有关。总体来看,与数字经济企业发展联系越紧密的公共数据开放,越能促进数字经济企业集聚,符合理论预期。

表2 基准回归结果

变量	DEA		
<i>DID</i>	0.477 *** (0.051)	0.184 *** (0.054)	0.133 *** (0.045)
<i>POP</i>	0.032 (0.025)	0.037 * (0.021)	-0.014 (0.020)
<i>GDP</i>	0.113 * (0.066)	0.194 *** (0.040)	0.221 ** (0.086)
<i>DI</i>	0.404 *** (0.085)	0.446 *** (0.118)	0.138 (0.087)
<i>TI</i>	0.015 (0.009)	-0.009 (0.009)	-0.003 (0.008)
<i>INN</i>	0.148 *** (0.025)	0.085 *** (0.020)	0.046 ** (0.019)
<i>IS</i>	2.402 *** (0.317)	0.514 * (0.264)	0.663 * (0.370)
<i>HR</i>	-0.000 (0.000)	0.000 ** (0.000)	0.000 (0.000)
<i>FD</i>	-0.506 *** (0.029)	0.032 * (0.017)	-0.084 *** (0.030)
<i>FDI</i>	-0.115 *** (0.036)	-0.000 (0.014)	-0.042 (0.029)
常数项	1.279 ** (0.515)	-0.806 *** (0.228)	0.304 (0.720)
年份固定效应	控制	未控制	控制
城市固定效应	未控制	控制	控制
<i>N</i>	4 158	4 158	4 158
调整 R^2	0.688	0.664	0.791

注: *、**、***分别表示在 10%、5%、1%的水平上显著,括号内数值为城市聚类稳健标准误,下表同。

表3 各类型数据关键词

数据类型	关键词
公共服务类(<i>ser</i>)	道路交通、政务服务、公共安全、道路、交通、治安、地图、社会保障、公共管理、气象、社保、医疗、卫生
社会民生类(<i>so</i>)	住房建设、社会治理、文体体育、住房、城市建设、社会、治理、文化、娱乐、体育、教育、建设工程、养老、行政许可、处罚、权力清单、执法、监管
经济类(<i>eco</i>)	财政金融、税收、投资规划、经济发展、政府性基金、财政拨款、转移支付、公共预算、经济建设、工业、第一产业、第二产业、第三产业、市场监管、预算、融资、贷款、租赁、零售
数字产业类(<i>di</i>)	机器人、新媒体、信息传输、软件、信息技术服务、计算机、人工智能、数字、云计算、大数据、区块链、5G、基站

表 4 不同类型公共数据开放类型对数字经济企业集聚的影响

变量	DEA			
<i>ser</i>	-0.069*** (0.019)			
<i>so</i>	0.198*** (0.019)			
<i>eco</i>	0.222*** (0.022)			
<i>di</i>	0.522*** (0.107)			
<i>N</i>	4 158	4 158	4 158	4 158
调整 R^2	0.786	0.393	0.388	0.377

注:所有模型均控制了控制变量以及年份和城市固定效应,限于篇幅,控制变量和常数项估计结果略,下表同。

2. 稳健性检验

(1) 安慰剂检验。为缓解不可观测遗漏因素的影响,通过随机设定政策实施年份和处理组城市进行安慰剂检验。从 297 个城市中随机挑选 100 个城市作为“伪处理组”,假设这 100 个城市开通公共数据平台,其他城市为控制组;随机设定政策实施年份,伪处理组在该年份后的 *DID* 取值为 1,其他取值为 0。重复进行 1 000 次,检验结果如图 2 所示,伪政策变量的估计系数正态分布于 0 值附近, P 值大多大于 0.1,且基准回归结果(竖虚线标注处)属于异常值,表明基准回归得到的政策效应并非除公共数据开放以外的偶然因素引致。

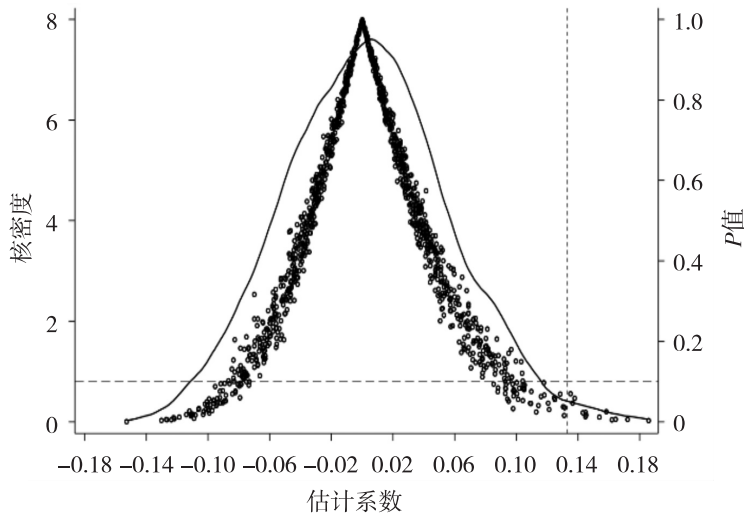


图 2 安慰剂检验结果

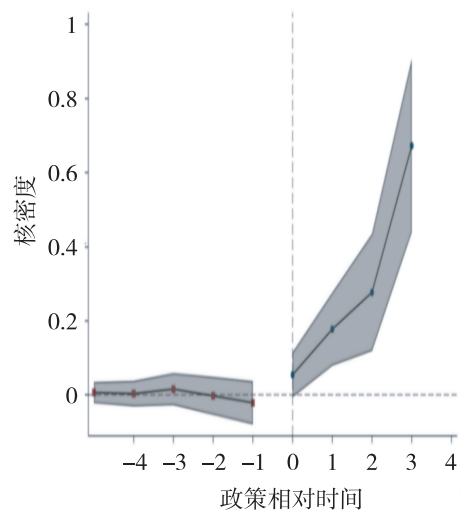


图 3 插补估计量平行趋势检验

(2) 异质性处理效应检验。多时点 *DID* 双向固定效应估计量存在异质性处理效应,即同一政策对不同个体产生的影响不仅随时间变化,还因处理时长或处理时点的不同而异 (Goodman-bacon, 2021)^[48],可能导致估计偏误。本文进行 Goodman-bacon 分解(见表 5),结果表明存在异质性处理效应。对此,采用改进的双重稳健逆概率倾斜加权最小二乘法 (DRIMP)、双重稳健逆概率加权法 (DRIPW)、逆概率加权法 (IPW) 对“组别-时期”平均处理效应进行估计 (Callaway et al., 2021)^[49],结果如表 6 所示,表明基准

回归结果稳健。进一步借鉴 Borusyak 等 (2016)^[50] 的研究,采用插补回归进行平行趋势稳健性检验,估计结果如图 3 所示,在政策实施前,处理组和控制组不存在显著差异,而政策实施后估计系数显著为正,说明公共数据开放对数字经济企业集聚产生了显著的正向影响。

表 5 Goodman-bacon 分解结果

控制组类型	权重	估计值
处理组 vs 从未受政策处理组	0.300	0.052
较早受处理组 vs 较晚受处理组	0.111	0.259
较晚受处理组 vs 较早受处理组	0.589	0.238
DID 加权估计结果	0.184	

表 6 “组别-时期”平均处理效应分析结果

变量	DRIMP	DRIPW	IPW
ATT	0.387 ^{***} (0.150)	0.114 ^{**} (0.054)	0.249 ^{***} (0.082)
N	3 626	4 130	4 130

(3) 基于双重机器学习方法的因果推断。双重机器学习方法基于机器学习算法和样本交叉拟合,能够克服引入高维度控制变量所致的多重共线性问题,实现对政策处理效应的无偏估计。借鉴张涛和李均超(2023)^[51]的方法,将样本按照 1:2、1:4、1:7 进行分割,采用一般交互式双重机器学习方法(DDML)进行回归,结果见表 7 的 Panel A, DID 的回归系数均在 1% 的水平上显著为正,表明基准回归结果稳健。

(4) PSM-DID 检验。为缓解样本选择偏差的干扰,采用倾向得分匹配方法为处理组样本匹配特征相似的控制组样本,采用匹配的样本进行回归,结果见表 7 的 Panel B,与基准回归结果一致。

表 7 稳健性检验结果 1

变量	Panel A: 双重机器学习方法			Panel B: 截面 PSM-DID 检验	
	DEA			DEA	
	1:2	1:4	1:7	权重非零样本	满足共同支撑假设样本
DID	0.336 ^{***} (0.038)	0.328 ^{***} (0.043)	0.350 ^{***} (0.034)	0.148 ^{***} (0.057)	0.122 ^{***} (0.045)
N	4 158	4 158	4 158	1 677	3 972
调整 R ²				0.790	0.790

(5) 其他稳健性检验。第一,预期效应检验。考虑到公共数据开放可能存在预期效应,将政策实施时间提前一年(DID_f),重新进行检验。第二,样本调整。为避免时滞导致回归系数低估,删除 2020 年后上线政府数据开放平台的样本重新进行回归;为缓解地区经济发展水平差异带来的混淆效应,删除直辖市样本后重新进行回归。第三,控制同期相关政策影响。考虑到相关数字经济政策的实施可能干扰回归结果,在基准模型中分别加入智慧城市试点(SC)、大数据综合试验区(BD)和“宽带中国”示范城市(BB)的政策变量,重新进行回归。上述稳健性检验结果见表 8,均验证了基准回归结果的稳健性。

表 8 稳健性检验结果 2

变量	DEA					
	预期效应检验	调整样本		控制相关政策影响		
<i>DID_f</i>	0.030(0.028)					
<i>DID</i>		0.115*(0.060)	0.140*** (0.046)	0.129*** (0.045)	0.139*** (0.047)	0.130*** (0.045)
<i>SC</i>				-0.017(0.038)		
<i>BD</i>					-0.113(0.076)	
<i>BB</i>						-0.085**(0.040)
<i>N</i>	3 861	3 066	4 102	4 158	4 158	4 158
调整 <i>R</i> ²	0.787	0.786	0.783	0.791	0.792	0.792

3. 机制检验

本文构建机制检验模型如式(2)所示:

$$M_{it}=\beta_0+\beta_1DID_{it}+\gamma M_{it}+\delta_t+\mu_c+\varepsilon_{it}$$
 (2)

其中,*M*代表机制变量,其他变量同式(1)。根据前文理论分析,选取以下机制变量:第一,数字产品需求(*MR1*和*MR2*)。借鉴程刚等(2025)^[52]的方法,采用电信业务总额、电子商务交易额、信息技术服务收入3项指标,通过熵值法加权测算样本城市的数字产品市场需求(*MR1*);同时,借鉴李新等(2024)^[53]的研究,采用城市的数字需求关键词年度百度指数总和(对数化处理)衡量数字产品的潜在市场需求(*MR2*)^①。第二,数字创新生态(*DIE*)。耦合协调度模型被广泛用于评估系统间协调发展程度(梁怡彤等,2026)^[54],本文采用该模型测算创新主体、创新平台、创新环境三大子系统的耦合协调度,以此刻画样本城市的数字创新生态。构建城市数字创新生态评价指标体系如表9所示,采用熵值法测算各子系统的综合得分,再通过耦合协调度模型计算得到变量*DIE*。第三,数字人力资本水平(*DHC*)。借鉴韩奇和杨秀云(2024)^[55]的方法,采用对数化处理的信息传输、计算机服务和软件业从业人员数(万人)衡量城市数字人力资本水平。

表 9 城市数字创新生态评价指标体系

子系统	指标	衡量方法
数字创新主体	数字创新核心主体	数字独角兽企业和数字瞪羚企业总数
	数字创新互补主体	规模以上工业企业数
数字创新平台	数字基础设施	年度域名备案数
数字创新环境	数字创新投资环境	R&D经费投入
	数字创新政策环境	政府工作报告中与数字技术和数字应用相关词词频
	数字创新金融环境	数字普惠金融指数

机制检验结果见表10。*DID*对*MR1*、*MR2*、*DIE*、*DHC*的回归系数均显著为正,表明城市公共数据开

① 搜索引擎关键词选择如下:以数字经济作为种子关键词;使用百度指数需求图谱提供的相关词扩展关键词库,剔除部分搜索量过低关键词,最终的扩展关键词包括:虚拟经济、数字化转型、数字贸易、平台经济、金融科技、信息管理与信息系统、大数据管理与应用、数字乡村、新零售、区块链、新基建。

放显著推动了本地数字产品和服务的市场需求规模增长,改善了城市数字创新生态,提升了城市数字人力资本水平,进而促进了数字经济企业在本地的集聚。由此,本文提出的假说 2、3、4 得到验证。

表 10 机制检验结果

变量	扩大数字产品需求		优化数字创新生态	加快数字人力资本积累
	MR1	MR2	DIE	DHC
<i>DID</i>	0.024 *** (0.004)	0.059 * (0.034)	0.006 * (0.004)	0.877 * (0.458)
<i>N</i>	4 158	4 158	4 158	4 158
调整 <i>R</i> ²	0.819	0.932	0.927	0.857

4. 异质性分析

(1) 营商环境异质性。借鉴白钰和冯均科 (2024)^[56] 的方法计算样本城市的营商环境指数,根据其中位数将样本划分为“营商环境较好”和“营商环境较差”两组,分组回归结果见表 11 的 Panel A。公共数据开放显著提升了营商环境较好城市的数字经济企业集聚水平,但对营商环境较差城市数字经济企业集聚的影响不显著。其原因可能在于,营商环境较好的城市具备较成熟的市场机制与更高效的政务服务,数据要素流动和共享较为充分,使公共数据开放的数据赋能作用得以有效发挥;而营商环境较差的城市市场机制不完善,即便开放公共数据,也可能因数据要素流动受阻而不能有效赋能数字经济企业集聚。

(2) 居民消费异质性。居民消费水平取决于其可支配收入,本文采用城镇居民人均可支配收入衡量居民消费水平,根据其中位数将样本划分为“消费水平较高”和“消费水平较低”两组,分组回归结果见表 11 的 Panel B。公共数据开放显著提升了消费水平较高城市的数字经济企业集聚水平,但对消费水平较低城市数字经济企业集聚的影响不显著。其原因可能在于,消费水平较高的城市居民消费意愿和能力强,数字消费需求多元,公共数据开放可精准对接消费需求,催生出新的数字服务与商业模式,进而吸引大量数字经济企业入驻;而消费水平较低城市的消费体量有限,数字消费需求不足,缺乏支撑数字经济企业大规模发展的市场根基,即便开放公共数据,也难以有效带动数字经济企业集聚。

表 11 异质性分析结果

变量	Panel A: 营商环境异质性		Panel B: 居民消费异质性	
	营商环境较好	营商环境较差	消费水平较高	消费水平较低
<i>DID</i>	0.271 *** (0.069)	-0.039 (0.036)	0.148 *** (0.050)	0.008 (0.037)
<i>N</i>	2 075	2 062	2 089	2 061
调整 <i>R</i> ²	0.686	0.860	0.786	0.913

五、结论与启示

本文在理论分析的基础上,采用我国 297 个城市 2009—2022 年的面板数据,运用 BERT 模型识别数字经济企业,以政府数据开放平台上线为准自然实验,构建多期双重差分模型,实证检验城市公共数据开放对数字经济企业集聚的影响,研究结果表明:公共数据开放显著提升了城市数字经济企业集聚水平,这一结论在经过一系列稳健性检验后依然成立;社会民生类、经济类、数字产业类公共数据开放对数

字经济企业集聚的促进作用依次增强,而公共服务类公共数据开放对数字经济企业集聚具有负向影响;公共数据开放能够通过扩大数字产品需求、优化数字创新生态、加快数字人力资本积累三条渠道促进数字经济企业集聚;公共数据开放显著促进了营商环境较好、居民消费水平较高城市的数字经济企业集聚,而对营商环境较差、居民消费水平较低城市的数字经济企业集聚影响不显著。基于上述研究结论,本文得出以下启示:

第一,推进数据要素市场化改革,完善数据开放体系。政府应在加快建设5G、物联网、云计算、大数据中心等新型基础设施的同时,加大公共数据开放力度。建立动态化公共数据开放清单,并引入数字经济企业及相关专家组成数据开放监督委员会,不断提高公共数据发布的完整性和可用性;切实提高公共数据开放质量,尤其要提高公共服务类公共数据开放质量,从而更好地发挥公共数据开放对实体经济高质量发展的赋能作用。第二,优化数字创新生态,促进数字经济企业高质量发展。各级政府和企业应加大数字技术研发投入,尤其要充分利用公共数据开放的数据赋能作用优化数字创新生态,持续加快数字技术创新和应用。针对数字产业链中的薄弱环节和缺失环节,通过政策引导、资金支持和产学研结合等方式强链补链延链,为数字经济企业的集聚和发展提供更好的产业链支持。第三,加快数字人才培养,提升数字人力资本水平。应健全数字人才引育留用机制,统筹推进数字人才梯队建设,完善数字技能培训体系,以高水平数字人力资本支撑数字经济企业高质量发展。第四,着力优化城市营商环境,不断提升数字产品和服务消费水平。一方面,应完善政务服务体系,降低企业经营成本,破除要素流通壁垒,激发市场主体活力,增强区域发展竞争力与吸引力;另一方面,应加大补贴力度促进居民数字产品消费,丰富数字消费应用场景,健全数字消费配套体系,持续释放内需增长动能,助力数字经济企业集聚发展。

参考文献:

- [1] 师磊,阳镇,钱贵明. 数字产业集群政策与关键核心技术突破式创新[J]. 中国工业经济, 2025 (1): 100-117.
- [2] 李天健,李伟. 数字产业集群发展: 竞争优势重塑与政策范式重构[J]. 经济学家, 2025 (2): 35-44.
- [3] 周海川,刘帅,孟山月. 打造具有国际竞争力的数字产业集群[J]. 宏观经济管理, 2023 (7): 27-32, 48.
- [4] 王如玉,梁琦,李广乾. 虚拟集聚: 新一代信息技术与实体经济深度融合的空间组织新形态[J]. 管理世界, 2018, 34 (2): 13-21.
- [5] 叶堂林,刘哲伟,张京亮. 数字产业空间集聚影响因素探析[J]. 科技进步与对策, 2023, 40 (15): 75-82.
- [6] 赵放,李文婷,马婉莹. 数据要素市场化能否促进数字产业集聚——来自准自然实验的证据[J]. 浙江学刊, 2024 (3): 143-152.
- [7] 桑瑞聪,范剑勇. 服务外包政策促进了数字产业集聚吗——来自示范城市的证据[J]. 国际贸易问题, 2025 (12): 119-138.
- [8] 高志刚,韩宁,韩延玲. 智能化与绿色化协同对数字产业集聚的影响——基于智慧城市和低碳城市双试点的准自然实验西部论坛, 2026, 36 (2): 17-29.
- [9] 杨俊,李小明,黄守军. 大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论[J]. 经济研究, 2022, 57 (4): 103-119.
- [10] 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.
- [11] Janssen K. The influence of the PSI directive on open government data: an overview of recent developments [J]. Government Information Quarterly, 2011, 28 (4): 446-456.
- [12] 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.

- [13] 王生年,侯丹阳.政府公共数据开放能否助力商品要素自由流动?——基于企业跨地区采购的经验证据[J].现代财经(天津财经大学学报),2025,45(7):56-76.
- [14] 李佳林,王禹,赵西卜.政府公共数据开放对企业数字化转型的影响研究[J].管理学报,2024,21(10):1467-1475.
- [15] 杨壮,吴福象,龚恩泽.公共数据开放能促进企业开放式创新吗?——基于政府公共数据平台上线的准自然实验[J].经济经纬,2025,42(4):3-15.
- [16] 李晓龙,牟业洪.公共数据开放能否促进数字产业集聚?——基于双重机器学习的因果推断[J].商业研究,2026(1):1-11.
- [17] 刘阳,宫舒文,王泽宇.产业集群内裂变新创企业的外部网络关系与创业绩效研究[J].东岳论丛,2023,44(9):138-149,192.
- [18] 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.
- [19] 赵志浩,揭阳扬,雷平,等.数字经济发展路径对流动人口就业集聚的影响[J].经济地理,2025,45(8):11-19.
- [20] 熊广勤,方扶星.数据要素共享能否促进城市共同富裕?——来自公共数据开放的证据[J].财贸研究,2025,36(11):17-33.
- [21] 邹颖,石福安,祁亚.以数促联:公共数据开放与企业供应链地理布局[J].世界经济,2026,49(1):85-110.
- [22] 唐露,刘伟.企业全球研发网络的治理新机制——总部技术外派促进隐性知识跨界整合的案例研究[J].管理世界,2024,40(1):135-155.
- [23] 阮平,曹细玉.基于约束理论的集群内产业链升级——以古镇灯饰产业集群为例[J].科技管理研究,2016,36(11):162-167.
- [24] 贾俊秀,孟庆钦,吴涛.考虑电商平台潜在需求信息共享的促销及定价策略研究[J].管理学报,2021,18(10):1562-1572.
- [25] 王东方.公共数据授权运营中的数据安全风险及其体系化治理[J].图书馆,2025(9):1-7.
- [26] 李剑培,顾乃华,潘捷.中国制造业省际贸易的本地市场效应研究[J].国际经贸探索,2021,37(7):50-68.
- [27] Krugman P. Increasing Returns and Economic Geography[J]. Journal of Political Economy, 1991: 99(3) 483-499.
- [28] 赵振.“互联网,”跨界经营:创造性破坏视角[J].中国工业经济,2015(10):146-160.
- [29] 吴瑶,邹波,陈昕彤,等.寻找情感链接:满足用户情感需求的产品互动创新机制[J].外国经济与管理,2025,47(7):120-140.
- [30] 吴艳,贺正楚,周海欣.省级区域新质生产力的多元培育路径——基于数字创新生态系统的视角[J].管理学报,2025,38(6):31-46.
- [31] 闫晓勇,李烨,王刘伟,等.基于三方演化博弈的创新生态系统自组织集聚机制研究[J].科学学与科学技术管理,2023,44(3):63-79.
- [32] 岳晓旭,张煜,谢其军,等.核心企业知识基础对数字创新生态形成的影响[J/OL].科学学研究,1-20(2026-01-08). <https://doi.org/10.16192/j.cnki.1003-2053.20260107.001>.
- [33] 马蕾,万劲波,余江,等.数字创新生态系统发展的内在逻辑:技术、数据与场景的共演[J].中国软科学,2025(9):200-211.
- [34] 贾卫峰,蒋泽宇.人工智能应用对产业链韧性的影响研究——数字创新生态系统驱动视角[J].科研管理,2025,46(12):23-32.
- [35] 余典范,张宇,宋晴.数字人力投入与企业价格加成——来自招聘大数据的证据[J].中国工业经济,2025(7):62-80.
- [36] Tambe P, Hitt L, Rock D, et al. Digital capital and superstar firms[R]. NBER Working Papers, 2020, No. 28285.
- [37] 赵放,李文婷,马婉莹.数字经济视域下地理集聚与虚拟集聚的演化特征及耦合关系[J].吉林大学社会科学学报

- 报, 2024, 64 (1): 117-132, 238.
- [38] 刘满凤, 杨杰, 陈梁. 数据要素市场建设与城市数字经济发展[J]. 当代财经, 2022 (1): 102-112.
- [39] 姜海宁, 张俊, 郑舒婷, 等. 关系——演化视角下海外归国人才对产业集群发展的影响机理[J]. 经济地理, 2024, 44 (11): 113-121.
- [40] 刘传明, 马青山. 网络基础设施建设是否重塑了人工智能企业空间布局——基于跨越地理边界的集聚视角[J]. 经济学动态, 2025 (3): 54-74.
- [41] 王林辉, 周慧琳, 徐超. 工业智能化、地理空间产业集聚和数字空间虚拟集聚[J]. 东北师大学报 (哲学社会科学版), 2024 (6): 1-12.
- [42] 杨鹏, 孙伟增, 张柳钦, 等. 数字经济政策与企业数字创新质量[J]. 统计研究, 2025, 42 (7): 70-81.
- [43] Huang Q, Tao M, An Z. Lawyer LLaMA Technical Report[R]. ArXiv (Cornell University), 2023.
- [44] Sun C, Qiu X, Xu Y, et al. How to fine-tune BERT for text classification? [J]. Lecture Notes in Computer Science, 2019, 11856: 194-206.
- [45] 刘晴, 张盼盼. 经济政策不确定性对中国出口企业升级的影响——基于贸易中介集聚视角的经验分析[J]. 北京理工大学学报 (社会科学版), 2022, 24 (1): 90-104.
- [46] 张鑫宇, 汤向俊, 刘瑞明. 政策红利、地理禀赋与增长差异——来自“长江经济带发展战略”的证据[J]. 学术月刊, 2023, 55 (7): 46-60.
- [47] 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, 47 (2): 174-203.
- [48] Goodman-Bacon A. Difference-in-differences with variation in treatment timing[J]. Journal of Econometrics, 2021, 225 (2): 254-277.
- [49] Callaway B, Sant' Anna P H C. Difference-in-differences with multiple time periods[J]. Journal of Econometrics, 2021, 225 (2): 200-230.
- [50] Borusyak K, Jaravel X, Spiess J. Revisiting event study designs: robust and efficient estimation[J]. Review of Economic Studies, 2024, 91 (6): 3253-3285.
- [51] 张涛, 李均超. 网络基础设施、包容性绿色增长与地区差距——基于双重机器学习的因果推断[J]. 数量经济技术经济研究, 2023, 40 (4): 113-135.
- [52] 程刚, 常会芳, 李旭辉. 中国数字消费的时空演变特征及影响因素[J]. 地理研究, 2025, 44 (4): 1102-1118.
- [53] 李新, 王颖, 闫相斌, 等. 面向多源数据细粒度情感挖掘的旅游需求预测[J]. 系统工程理论与实践, 2024, 44 (7): 2293-2308.
- [54] 梁怡彤, 王承云, 吴佳颖, 等. 长三角地区数字经济与绿色创新耦合协调演进及政策建议[J]. 自然资源学报, 2026, 41 (3): 928-942.
- [55] 韩奇, 杨秀云. 公共数据开放能否促进企业数字化转型? [J]. 现代经济探讨, 2024 (4): 44-59.
- [56] 白钰, 冯均科. 公共政策审计促进公平竞争的研究——基于企业竞争程度的视角[J]. 财经论丛, 2024 (5): 82-92.

The Impact of Public Data Openness on the Agglomeration of Digital Economy Enterprises

YANG Renfa, JIANG Can

(School of Economics, Anhui University, Hefei 230601, Anhui, China)

Summary: The agglomeration of digital enterprises serves as a crucial foundation for building a modern

industrial system and acts as the core engine driving high-quality economic development in China. Public data openness has emerged as a critical institutional arrangement for unlocking the value of data and advancing the development of the digital industry. However, existing research rarely integrates public data openness and digital enterprise agglomeration into a unified analytical framework to systematically examine how public data openness drives the agglomeration of digital enterprises at the micro level.

Using panel data from 297 prefecture-level cities in China from 2009 to 2022, this study takes the launch of government data platforms as a quasi-natural experiment, constructs a multi-period difference-in-differences (DID) model, and empirically tests the impact of public data openness on the agglomeration of digital enterprises and its mechanisms after identifying digital enterprises using the BERT model. The results show that: (1) Public data openness significantly enhances the level of digital enterprise agglomeration, and this conclusion remains robust after a series of tests. (2) Heterogeneity analysis indicates that the promoting effect of public data openness on digital enterprise agglomeration is more pronounced in cities with a better business environment and higher household consumption levels. Regarding public data categories, the opening of social livelihood data, economic data, and digital industry data significantly improves digital enterprise agglomeration, whereas the opening of public service data has a negative effect. (3) Mechanism analysis reveals that public data openness enhances digital enterprise agglomeration through three channels: expanding market demand for digital products, fostering the development of the digital innovation ecosystem, and accelerating the accumulation of digital human capital.

Compared with existing studies, this paper makes three main contributions. First, this paper examines the impact of public data openness on digital enterprise agglomeration from both government and market perspectives, complementing the evaluation of public data openness policies at the micro-enterprise level. Second, it systematically clarifies three transmission mechanisms through which public data openness affects digital enterprise agglomeration and identifies the boundary conditions of this effect, thereby enriching industrial agglomeration theory in the context of the digital economy. Third, it improves the measurement framework by optimizing the identification of digital enterprises using the BERT model and by refining the measurement of digital product market demand using Baidu Index data, thereby correcting the biases inherent in traditional measurement approaches.

The findings of this study reveal the underlying logic of how public data openness drives the agglomeration of digital enterprises at the micro level. These findings provide empirical evidence for deepening market-oriented reforms of data factors and improving the public data openness system in China. They also offer decision-making references for local governments to formulate differentiated support policies for the digital industry and cultivate characteristic digital industry clusters under the “Data Element ×” Action Plan.

Keywords: public data openness; digital economy enterprise; industrial agglomeration; digital product demand; digital innovation ecosystem; digital human capital; BERT model

CLC number: F299.27; F129.9

Document code: A

Article ID: 1674-8131(2026)03-0021-16

(编辑:刘仁芳;朱 艳)