DOI:10.13502/j.cnki.issn1000-7636.2026.11.009

数据要素集聚、集聚外部性与城市创新能力

——基于国家级大数据综合试验区的准自然实验

郭彦丽 郭成鑫 高书丽

摘要:探索数据要素集聚对城市创新能力的影响机制与约束条件,对深化数据要素市场化改革、落实创新驱动发展战略具有重要价值。本文从集聚外部性视角出发,选取2011—2023年中国282个地级及以上城市面板数据,将国家级大数据综合试验区建设视为政策驱动型数据要素集聚的准自然实验,采用双重差分模型考察其城市创新效应。研究结果显示,政策驱动型数据要素集聚提升了城市创新能力,政策效应在实施后逐步显现并趋于稳定。机制分析表明,国家级大数据综合试验区政策通过促进数字技术人才集聚、推动数字企业专业化与多样化集聚、提高创新价值链协同来提升城市创新能力。异质性分析发现,政策效应在东北地区、长三角和珠三角城市以及大城市和经济较发达城市较为明显。本文从集聚外部性视角拓展了数据要素创新效应的研究,为完善数字基础设施、强化数字治理和差异化推进数据要素市场化配置提供了经验证据。

关键词:数据要素集聚 城市创新能力 国家级大数据综合试验区 集聚外部性 创新价值链

中图分类号:F49;F299.2 **文献标识码**:A **文章编号**:1000-7636(2026)11-0127-16

一、问题提出

随着中国经济由高速增长阶段转向高质量发展阶段,创新驱动已成为国家核心战略。自2015年以来,国家级大数据综合试验区(简称“综试区”)建设逐步成为完善数字基础设施、推动数据资源汇聚和探索数据流通制度创新的重要政策抓手[1]。与传统产业政策不同,综试区政策通过超前布局数字基础设施与先行先试数据交易制度,打破要素空间配置局限,推动全域分散化、碎片化的数据资源在特定城市空间高密度规模化集聚,为数字时代集聚外部性迭代升级提供实践场景。2020年,数据被正式列为关键生产要素;2025年,国家数据局发文指出“要建设高价值数据应用场景,推动数据科技创新成果高效转化”。在国家战略部署的驱动下,中国城市创新能力实现了跨越式提升。世界知识产权组织(WIPO)发布的《2025年全球创新指数报告》显示,中国排名首次跻身全球前十;中国科学技术发展战略研究院发布的《国家创新指数报告2025》显示,中国创新指数综合排名升至世界第9位,数字经济与创新活动的融合趋势日益显现。在城市层面,部分试点地区呈现出数据要素集聚、数字产业生态和创新产出相互促进的典型特征。以深圳为例,该城市依托

收稿日期:2026-05-02;修回日期:2026-10-08

基金项目:北京联合大学2026年高水平基金孵化项目“人机融合提升IT服务供应链韧性的机理研究”(ZK20260104)

作者简介:郭彦丽 北京联合大学商务学院教授,北京,100025;

郭成鑫 北京联合大学商务学院硕士研究生,通信作者;

高书丽 北京联合大学商务学院教授。

作者感谢审稿快线专家的评审意见。

数据交易平台、数字产业园区和市场化制度等优势,推动数据流通、场景应用和产业链协同,促进数字产业集聚和创新主体活跃度提升,国际专利申请量长期居全国首位。这一事例为考察政策驱动型数据要素集聚依赖何种主体、产业和制度条件转化为城市创新能力提供了现实问题导向。国家级大数据综合试验区政策实施过程见表 1。

表 1 大数据综合试验区政策实施过程

出台时间	政策	相关政策内容
2015 年 9 月	《促进大数据发展行动纲要》(国发〔2015〕50 号)	明确提出推进大数据综合试验区建设
2016 年 2 月	《国家发展改革委、工业和信息化部、中央网信办关于同意贵州省建设国家大数据(贵州)综合试验区的函》(发改高技〔2016〕338 号)	贵州获批建设首个国家大数据综合试验区
2016 年 10 月	《国家发展改革委、工业和信息化部、中央网信办关于同意在部分区域推进国家大数据综合试验区建设的函》(发改高技〔2016〕2123 号)	京津冀、珠江三角洲、上海、河南、重庆、沈阳和内蒙古获批
2020 年 4 月	《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》(中发〔2020〕9 号)	首次将数据列为新型生产要素
2023 年 12 月	《“数据要素×”三年行动计划(2024—2026 年)》(国数政策〔2023〕11 号)	推动数据要素在多领域深度融合应用
2025 年 1 月	《国家数据基础设施建设指引》(发改数据〔2024〕1853 号)	推进算力网络与数据流通设施一体化布局

试点城市在创新上已取得一些实效,但整体而言,中国城市创新体系仍受结构性失衡与效率瓶颈的深层影响。城市创新面临的关键约束并非单纯缺少数据,而是数据分散,专业人才、数字企业和技术需求之间难以形成有效连接。同时,面对全球产业链重构与关键核心技术封锁,原始创新能力不足的短板日益凸显,呈现出部分城市市场创新活跃度与核心技术攻关能力错配、区域创新发展分化明显的典型特征。传统要素驱动的创新增长模式已无法满足新的增长需求[2]。数据要素作为数字经济时代的核心生产力[3],凭借其非竞争性、低边际成本、强渗透性和规模报酬递增等独特属性,成为关键战略资源。数据要素集聚能够打破时空限制,大幅降低创新主体的信息搜索与供需匹配成本,加速隐性知识溢出、技术重组与资源协同迭代[4],连接数据孤岛并破除创新壁垒,为解决城市创新发展的系统性困境提供新的突破口。

从经典理论来看,传统产业集聚外部性主要划分为马歇尔专业化外部性与雅各布斯多样化外部性。马歇尔外部性强调同一产业空间集聚形成的专业化分工、技术共享与劳动力蓄水池效应;雅各布斯外部性侧重不同产业、多元创新主体的空间交汇。这两种外部性均基于传统生产要素特征产生,难以适配数字经济时代数据要素的集聚逻辑与创新赋能机制[3]。数据要素具备可复制、可共享、零损耗流转、规模报酬递增的专属特征,其空间集聚能够突破产业边界、行业壁垒与区域限制,形成区别于传统集聚的数据要素专属外部性,通过全域知识溢出、全产业链赋能、全主体协同联动重构城市创新体系,成为新时代集聚外部性理论的重要研究对象。

数据要素的经济价值已得到学术界的普遍认可,现有文献多从理论层面探讨数据要素对经济增长、产业升级的赋能作用[5]。近年来围绕综试区、公共数据开放和数据要素市场化的研究快速增加,同一政策对区域创新能力的直接效应已被多篇文献识别,但已有研究多停留于多种要素赋能的机制层面,尚未充分解析数据要素的集聚外部性[6-7]。因此,本文将重点关注数据要素集聚外部性的结构及其实现条件:一是区分数字人才、数字产业和技术市场三个层次的外部性;二是识别专业化集聚与多样化集聚的差异化作用。

本文将综试区的设立视为一项准自然实验,选取 2011—2023 年中国 282 个地级及地级以上城市的面板数据,利用熵权 TOPSIS 法测度城市创新能力综合指数,并构建双向固定效应双重差分模型,立足传统集聚

外部性与数据要素专属外部性的融合视角,评估数据要素赋能城市创新的影响、作用机制。本文的边际贡献主要包括三个方面。第一,基于集聚外部性视角,解释数据资源集聚如何通过共享、匹配和学习机制转化为城市创新优势。第二,在作用机制方面,分别以数字技术人才集聚、数字企业专业化与多样化集聚和创新价值链协同为切入点,从全域知识溢出、全产业链赋能和全主体协同三个层面考察数据要素赋能城市创新的实现路径。第三,在概念识别方面,明确区分数据资源集聚与数字产业空间集聚,避免将数据要素的集聚效应与数字产业的空间集聚效应混同,从而深化对数据要素赋能城市创新内在机制的认识。

二、文献综述

城市创新能力是指城市通过空间载体吸引知识、技术、资本、人才等创新要素高度集聚,并通过网络化协同机制实现要素优化配置,最终驱动区域创新产出跃升与产业结构升级的系统能力与潜力[8]。相关研究主要围绕创新能力的测度和影响因素两个方面展开。

在城市创新能力测度方面,专利申请量和授权量被广泛使用。陈婷等以城市专利申请数为主要指标,以专利授权数为替代指标,考察数据要素市场化对城市创新能力的影响[9];胡雪萍和王谊巧以每百万人专利授权数衡量城市创新水平[10]。但专利难以同时覆盖创新资源投入、创新成果形成和创新环境支撑等可以反映城市创新系统整体性特征的指标。部分研究开始采用熵权法或 TOPSIS 法,从创新投入、创新产出和创新环境等维度构建综合评价指标[11],这为识别数字政策对城市创新体系整体运行状态的影响提供了测度基础。

数据要素集聚是指在特定空间或平台上,数据资源及其相关能力的集中、汇聚与流通状态,本质是通过整合区域内分散的数据资源,形成规模化、协同化的配置体系,以提升经济活动的灵活性与创新能力[12]。关于城市创新能力影响因素的探讨,学术界经历了一个由关注传统物理要素向关注新型数据要素演进的过程[13]。传统要素的流动与集聚能够通过专业化分工、劳动力匹配、共享投入和知识溢出影响城市创新[2]。随着数字经济的发展,研究重点逐步转向数字技术、数字基础设施和数据制度建设的创新效应[14]。已有研究发现数字基础设施建设、公共数据供给及数据市场化配置能够改善信息环境、降低创新成本并促进城市或企业创新[15-17]。陈婷等从数字金融、产业结构优化和信息人才集聚角度解释数据要素市场化的城市创新效应[9];胡雪萍和王谊巧则从有效市场与有为政府视角考察公共数据开放的作用机制[10];刘传明等、卢福财等从不同政策场景视角提供了数字政策促进创新的经验证据[6-7]。上述文献较为充分地回答了数据开放、数据流通和数字政策能否促进创新的问题,但其研究重点在于信息供给改善、交易成本降低、产业结构调整等,对数据政策如何改变创新要素的空间组织形态关注不足。综试区不同于单一的数据开放平台或数据交易场所,其政策工具同时涵盖数据资源归集共享、基础设施建设、数字产业培育、应用场景开放和制度创新,能够推动数据资源、数字企业、专业人才、技术平台及服务机构在特定城市集中。因而,综试区不仅构成数字政策冲击,也可能诱发一种政策驱动型的数据要素集聚过程。

集聚外部性理论能够为城市数据要素集聚的过程提供解释。同类产业集中可以通过专业化分工、共享投入和劳动力匹配形成马歇尔外部性,异质产业共聚则能够通过知识交叉和技术重组形成雅各布斯外部性[18]。在数字化背景下,平台连接和数据复用降低了信息搜寻成本,延展了知识溢出和中间投入共享的空间边界;但数字活动仍依赖专业人才、产业生态和算力设施等互补性资产,因而不能由“数据汇聚”自动推出“创新提升”[19-20]。因此,数据要素对城市创新的影响更可能表现为数据特有外部性与传统集聚外部性共同作用的复合过程。集聚外部性主要通过知识溢出、劳动力池效应、中间投入品共享、市场需求扩大和竞争创新激励等机制[21]产生规模报酬递增效应,也体现出数据要素可复制、可共享、跨场景流通和低边际成本等专属特征。一方面,数字技术具有技能偏向特征[22],可能增加对高技能劳动者的需求;另一方面,数据要素价值释放不仅取决于数据资源本身,还依赖产业生态、算力基础和制度环境等互补性条件[23]。

综上,现有研究为本文提供了基础,但仍存在拓展空间。第一,已有数据政策研究多以专利数量衡量创

新产出,对创新投入、产出和环境共同构成的城市创新系统关注不足。第二,既有文献尚未充分揭示政策如何引导数据及其互补性要素共同集聚以重构城市创新要素的空间配置方式。第三,尚缺乏统一的集聚外部性框架以解释数字人才专业化集聚、数字产业集聚和多元创新主体协同如何分别形成全域知识溢出、全产业链赋能与全主体协同。因此,本文将综试区界定为政策驱动型数据要素集聚场景,从全域知识溢出、全产业链赋能和全主体协同三个层次开展机制分析。

三、理论分析与研究假设

(一)数据要素集聚对城市创新能力的影响

综试区通过超前布局数字基础设施,促使海量、多源、异构的数据要素在特定空间内实现高密度集聚,形成了区别于传统要素集聚的数据要素专属外部性,强化了数据资源跨主体、跨行业和跨场景流动所产生的联动效应,使城市创新活动转向多主体协同创新。第一,数据要素集聚能够降低知识搜寻与技术匹配成本,提升城市产业链协同能力。数据的非竞争性与可复制性特征[24]降低了要素流动中的知识交互阻力。例如,北京打造数智北京创新中心,推出公共数据“先创新、再授权”模式,有效降低了知识搜寻成本,激发了创新活力。第二,数据要素集聚能够强化共享投入效应。综试区建设能够降低企业尤其是中小企业使用数字基础设施和数据资源的门槛,减少投入成本。例如,上海市北高新园区打造可信数据流通共享平台,支持数字产业尤其是中小数字企业开展创新活动。第三,数据要素集聚能够促进知识溢出与创新反馈[25]。数据资源的高频流动有助于连接技术供给端和市场需求端,加快创新循环。例如,长三角打造跨省政务数据共享底座,使试点走向制度化创新。数据要素集聚通过降低知识搜寻成本、强化共享投入、促进创新反馈,重构城市创新体系。

据此,本文提出假设 1:数据要素集聚能够提升城市创新能力。

(二)数据要素集聚对城市创新能力的作用机制

1. 全域知识溢出:数字人才集聚通道

数字技术人才是数据采集、处理、分析和场景应用工作的关键主体。综试区建设通过增加数字基础设施投资、数据应用项目和数字企业需求,提高数字技术岗位的预期回报,吸引相关专业人才向试点城市集中,产生了定向的人才虹吸效应[26]。一方面,通过技能偏向型技术进步[22]扩大对研发人员、数字技术人员和平台运营人员等群体的需求。另一方面,数据要素的流动和应用会提高创新活动的信息透明度和资源匹配效率,使城市更容易吸引企业、人才和创业主体进入。从集聚外部性视角看,创新主体集聚具有典型的劳动力池效应和知识溢出效应。创新主体在空间上更加集中,不仅能够提高面对面交流和隐性知识传播的频率,也能够依托数字平台拓展跨组织、跨行业的信息交互边界。创新主体之间更密切的互动,有助于推动知识重组、技术扩散和创新灵感生成,进而提升城市创新能力。因此,数据要素集聚通过发挥人才集聚的虹吸效应与知识溢出效应,可以增加城市智力资本储备,进而提高城市创新能力。

据此,本文提出假设 2:数据要素集聚通过促进数字人才集聚提升城市创新能力。

2. 全产业链赋能:数字产业集聚通道

数据要素集聚还会重塑数字企业的空间分布。根据马歇尔的集聚外部性理论,产业在空间上的集聚主要通过中间投入品共享、劳动力蓄水池和知识技术溢出三种机制促进区域创新[21]。一方面,综试区的政策优势能够吸引云计算、人工智能等数字企业在空间上扎堆,形成数字产业集群,实现产业链上下游的垂直整合。另一方面,数字产业集聚不断深化专业化分工,加速了制造业等实体产业新质生产力的培育与创新红利释放[27]。

此外,数字产业集聚更易产生集聚外部性,通过共享、匹配、学习机制驱动创新。首先,在共享机制上,集聚区内的企业能够共享通用的算法模型与算力底座等公共中间投入品,分摊成本,产生规模经济。其次,

在匹配机制上,数字产业集聚能够降低搜寻与匹配成本,提高创新要素的供需匹配效率,缓解信息不对称。最后,在学习机制上,集群内部频繁的交互与匹配,可以打破传统产业间的技术边界,有助于中小微企业通过发挥搭便车效应降低自身的创新门槛,进一步强化跨行业、跨领域的技术外溢[21],进而提升整个城市的创新能力。

据此,本文提出假设 3:数据要素集聚通过推动数字产业集聚提升城市创新能力。

3. 全主体协同:创新价值链协同通道

技术市场是连接企业、高校、科研机构和技术服务机构的重要交易接口。综试区建设通过改善数据供需识别、技术成果检索和交易过程留痕,提高技术供给与产业需求的可观察性,降低技术评价、议价和履约成本。一方面,数据要素集聚有助于解决创新价值链中存在的研产脱节的问题,促进各关联性创新主体的协同创新发展[28]。另一方面,数据要素集聚能够增强创新主体间要素禀赋的互补性和整体性,增大创新主体在不同地区间实现创新合作的可能性,从而推动技术转移、经验交流和新型创新模式的涌现,最终体现为城市整体创新能力的提升。

据此,本文提出假设 4:数据要素集聚通过促进创新价值链协同提升城市创新能力。

四、实证设计

(一)样本选择与数据来源

本文的样本数据是以城市为单元的平衡面板数据集,剔除了数据缺失严重的样本,排除了海南省部分市、西藏自治区部分市和港澳台地区的数据,最终得到 2011—2023 年 282 个城市共 3 666 个观测值。数据主要来源于历年《中国城市统计年鉴》《中国科技统计年鉴》与各地方统计年鉴。

(二)模型设定

综试区集中于 2016 年获批,在年度数据中不存在分期实施差异,因此采用双向固定效应双重差分模型。模型通过比较 2016 年前后处理组和控制组城市创新能力的相对变化识别政策效应。考虑试点遴选可能与既有信息化基础紧密相关,基准模型控制城市固定效应、年份固定效应及政策前信息化水平与加入线性时间趋势的交互项;稳健性分析进一步加入政策前经济发展、产业结构、交通通信与时间趋势的交互项。借鉴刘传明等[6]、何骏等[12]的做法,并参考陈诗一等[29]将前定变量与时间趋势交互项纳入模型的处理方式,设定如下双向固定效应回归模型:

$$Innov_{it} = \alpha + \beta Bdp_{it} + \boldsymbol{X}'_{it}\boldsymbol{\gamma} + \theta Pretrend_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (1)$$

其中,i 表示城市,t 表示年份,$Innov$ 为城市创新能力指数;$Bdp = Treat \times Post$ 为政策虚拟变量,$Treat$ 在处理组城市取 1,控制组取 0;$Post$ 在 2016 年及以后取 1,否则取 0。$\boldsymbol{X}'$为控制变量向量。μ 和 λ 分别为城市固定效应与年份固定效应,ε 为随机扰动项。标准误差在城市层面聚类。$Pretrend$ 为前定信息化水平与线性时间趋势的交互项:先计算各城市 2011—2015 年每百人互联网宽带用户数加 1 取自然对数的均值,在城市之间进行极差标准化,再乘以年份减 2010。该变量不缩尾,并在相关模型中覆盖整个估计期。前定信息化水平本身不随时间变化,其水平项由城市固定效应吸收。

(三)变量说明

1. 被解释变量

弗曼等(Furman et al.)[30]提出的国家创新能力经典框架指出,创新能力本质上由创新投入、创新环境与创新产出三部分构成,三者分别对应创新活动的资源供给、制度支撑与成果实现。汤等人(Tang et al.)[31]针对中国城市群的实证研究构建了包含创新资源、知识创造、创新绩效与创新环境的多维指标体系,其维度划分本质上仍可归入投入、产出、环境的分析框架。本文将上述框架中的创新资源与知识创造归并为创新

要素投入与配置维度,将创新绩效明确为创新产出与市场转化维度,并保留创新环境作为“创新环境与物质基础”维度,从而形成与弗曼等[30]的经典框架一致的三环节链条。由此,分别对应创新链条的不同阶段,形成要素投入、环境支撑、成果产出的递进关系指标体系(见表 2),并采用熵权 TOPSIS 法计算得出城市创新能力指数。

表 2 城市创新能力评价指标体系

一级指标	二级指标	指标测度
创新产出与市场转化	每万人发明专利授权量	发明专利授权数 / 年末常住人口
	新创办企业数	每百人新创企业数
创新要素投入与配置	科学研究和技术服务业从业人员占比	科学研究/技术服务业从业者
	人力资本水平	普通高等学校在校生数 / 常住人口
创新环境与物质基础	科技创新资金支持力度	地方科技财政占一般预算内支出比重
	人均地区生产总值(GDP)	ln(地区生产总值/年末常住人口)

2. 解释变量

借鉴刘传明等[6]的研究思路,以综试区政策虚拟变量作为数据要素集聚(*Bdp*)的代理变量,依据国务院及国家发展改革委相关批复文件,手工整理各综试区名单及获批时间,以 2016 年获批的两批综试区 56 个地级及地级以上城市作为处理组,其余 226 个地级及地级以上城市作为控制组。变量构造见前序模型设定说明。

3. 控制变量

本文选择以下控制变量:(1)消费零售水平(*Consume*),以社会消费品零售总额占 GDP 比重衡量;(2)政府干预程度(*Gov*),以一般公共预算支出占 GDP 比重衡量;(3)金融发展水平(*Finance*),以金融机构贷款余额占 GDP 比重衡量;(4)对外开放程度(*Trade*),以进出口贸易差额占 GDP 比重衡量;(5)投资水平(*Invest*),以固定资产投资额占 GDP 比重衡量;(6)城镇化水平(*Urban*),以城镇常住人口占年末常住人口比重衡量;(7)*Pretrend* 以互联网普及率与(年份-2010)的乘积衡量,计算办法见模型设定说明。

(四)变量描述性统计

表 3 报告了描述性统计结果。城市创新能力指数的均值为 0.076 8,标准差为 0.061 3,最大值为 0.314 4,最小值为 0.017 9,表明样本期内城市创新能力存在较为明显的差异。该指数综合反映了创新投入、创新产出与创新环境三个维度,上述差异不仅体现了城市创新发展水平的不均衡,也反映出不同地区创新要素配置与创新环境建设仍存在差距。核心解释变量的均值为 0.122 2,表明政策实施后的试点城市占全部样本的 12.22%。其中,56 个城市于 2016 年进入处理组,其余 226 个城市为控制组。

表 3 描述性统计结果

变量类型	变量	观测值	均值	标准差	最小值	最大值
被解释变量	*Innov*	3 666	0.076 8	0.061 3	0.017 9	0.314 4
解释变量	*Bdp*	3 666	0.122 2	0.327 6	0	1
控制变量	*Consume*	3 666	0.381 5	0.106 2	0.140 5	0.691 8
	Gov	3 666	0.201 9	0.097 8	0.078 7	0.605 6
	Finance	3 666	2.640 3	1.216 7	1.037 3	7.094 4
	Trade×100	3 666	0.237 8	0.249 4	0.000 2	1.111 5
	Invest	3 666	0.791 5	0.328 1	0.197 4	1.870 6

表3(续)

变量类型	变量	观测值	均值	标准差	最小值	最大值
	Urban	3 666	0.400 3	0.215 2	0.102 8	1.000 0
	Pretrend	3 666	4.571 7	2.605 4	0.000 0	13.000 0

注:*Innov* 及控制变量按全样本 1%和 99%分位点缩尾,*Bdp* 和 *Pretrend* 未缩尾,后表同。

五、实证结果与分析

(一)基准回归

根据模型(1)检验试点政策对城市创新能力的影响,基准回归结果见表 4。表 4 列(1)为未控制城市和年份固定效应的混合普通最小二乘(OLS)回归结果,列(2)加入双向固定效应。

表 4 基准回归结果

变量	(1)	(2)
Bdp	0.001 0	0.006 1**
	(0.005 8)	(0.003 0)
Consume	−0.034 0**	0.011 6
	(0.014 5)	(0.008 6)
Gov	−0.227 2***	−0.062 6***
	(0.024 2)	(0.014 4)
Finance	0.027 7***	−0.001 6
	(0.002 6)	(0.001 4)
Trade	5.140 8***	0.085 7
	(0.926 6)	(0.499 1)
Invest	0.007 2	0.008 9***
	(0.005 4)	(0.002 9)
Urban	0.035 0***	0.009 8
	(0.011 6)	(0.014 2)
Pretrend	0.002 3***	0.014 2***
	(0.000 6)	(0.001 7)
常数项	0.019 8**	0.012 3
	(0.009 0)	(0.009 8)
城市固定效应	未控制	控制
年份固定效应	未控制	控制
观测值	3 666	3 666
R^2	0.590 3	0.942 4

注:***、**、* 分别表示在 1%、5%、10%水平下显著;小括号内为在城市层面聚类的稳健标准误。R^2 报告包含固定效应解释部分的决定系数,非组内 R^2,也不表示政策变量单独的解释力。

表 4 报告了数据要素集聚对城市创新能力影响的基准回归结果。列中(1)政策虚拟变量的回归系数为 0.001 0,未通过显著性检验。加入城市固定效应和年份固定效应后,列(2)中政策虚拟变量的回归系数为 0.006 1,在 5%水平下显著。这表明,在考虑城市固有特征、共同时间冲击及初始信息化水平差异所对应的变化趋势后,国家级大数据综合试验区建设对城市创新能力的促进作用存在,假设 1 得到验证。

从经济意义来看,列(2)的估计结果意味着,相较于非试点城市的同期变化,试点城市在政策实施后的创新能力指数平均提高约 0.006 1 个单位,相当于样本均值的 7.9%,约为样本标准差的 0.10 倍。尽管该变化不足以表明城市创新能力发生大幅跃升,但对于涵盖创新投入、创新产出和创新环境的综合指标而言,政策实施已表现出一定程度的创新促进效应。然而,数据资源在空间上的集中并不必然转化为创新优势,其作用发挥还取决于数据与人才、产业及技术市场等要素的有效结合。由此可见,政策驱动型数据要素集聚能够为城市创新发展带来积极影响。

(二)平行趋势假设评估

使用双重差分方法需要满足的前提条件是未拒绝事前趋势平行的假设,因此,本文

采用事件研究法进行平行趋势假设评估。图 1 展示了平行趋势假设评估的具体结果。

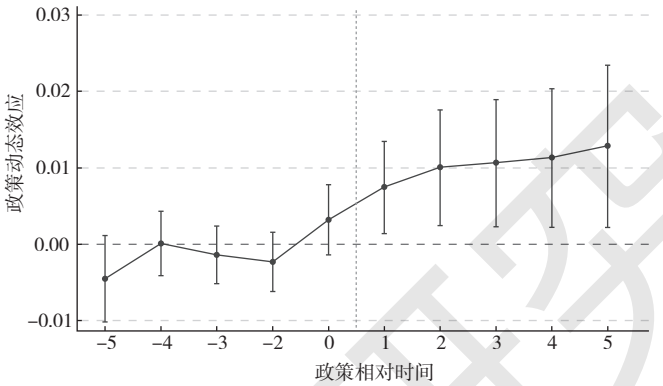


图 1　平行趋势假设评估结果

如图 1 所示,当 k <0 时(政策实施前的 5 年,-5 至-2 期),回归系数 β_k 的置信区间均穿过横轴,且整体围绕横轴上下小幅波动,表明在国家大数据综合试验区设立前,处理组和控制组的城市创新能力发展轨迹不存在系统性差异,严格满足平行趋势假设。政策实施当年效应尚不明显,实施后第 1—2 年逐步上升,此后总体保持上升趋势,表示试点短期建设期在 2 年左右,后为发展期和稳固期。该结果与政策从基础设施建设、数据汇聚到场景应用逐步落地的实施过程相一致。这一结果揭示了数据要素集聚对城市创新能力的赋能并非短期脉冲,而是具有长期性与持续内生驱动力的系统性跃迁。

(三)内生性分析

国家级大数据综合试验区试点城市并非完全随机设立,政策选择可能受到城市数字基础设施建设、经济发展水平、产业结构和行政资源等多重因素影响,这可能导致估计结果存在选择性偏误。为进一步增强识别可靠性,本文在双重差分框架下,从样本选择偏误、模型设定偏误等方面进行内生性分析。

为考察试点选择内生性对估计的影响,采用城市与光缆骨干城市的最小距离和线性时间趋势的交互项作为工具变量,进行两阶段最小二乘(2SLS)估计。工具变量定义为 Z=(距离/1 000)×(年份-2010),距离变量形成于政策实施前,较远的城市接入骨干网络和承接大数据基础设施的成本更高,因而与试点政策暴露具有相关性。在排他性方面,交通通信区位也可能直接影响知识流动和城市创新,因此本文将该工具变量作为补充识别,并控制政策前与年份趋势的交互项,检验工具变量对政策前创新趋势的预测效应。

表 5　内生性分析回归结果

变量	第一阶段	第二阶段
Z	-0.104 0***	
	(0.026 4)	
Bdp		0.071 2***
		(0.024 3)
控制变量	控制	控制
城市固定效应	控制	控制
年份固定效应	控制	控制
Kleibergen-Paap rk *LM*	14.010 1	
	{0.000 2}	
Kleibergen-Paap rk Wald *F*	15.512 7	
	[16.380 0]	
观测值	3 666	3 666

注:大括号内为不可识别检验对应的 P 值,中括号内为 Slock-Yog 弱有效性检验在 10%水平下的临界值。

表 5 显示,在纳入控制变量、前定变量以及城市和年份固定效应,并采用城市层面聚类稳健标准误后,第一阶段工具变量的回归系数显著为负,排除工具变量的 F 统计值为 15.512 7(P=0.000 1),表明所选工具变量能够解释政策变量的变动,具有显著的相关性。Kleibergen-Paap rk LM 统计值为 14.010 1(P=0.000 2),拒绝欠识别原假设。Kleibergen-Paap rk Wald F 统计值为 15.512 7,略低于临界值(16.380 0),存在弱工具隐患。因此,进一步进行对弱工具变量稳健的 Anderson-Rubin 检验,其 Wald F 统计值为 17.706 9(P<0.000 1),拒绝政策系数为零的原假设。第二阶段政策虚拟变量 *Bdp* 的回归系数为 0.071 2,在 1%水平下显著为正,与基准回归的估

计方向一致。上述分析利用与工具变量相关的政策变动开展补充估计,为政策对城市创新能力的提升作用提供了证据。

(四)稳健性检验①

1. 安慰剂检验

为排除其他不可观测的随机因素或时间趋势对研究结果的作用,采取随机置换处理组和政策时间的方式进行安慰剂实验。通过在样本城市中随机抽取虚拟处理组,并随机设定虚拟政策时间,进行了 500 次回归模拟(结果见图 2)。

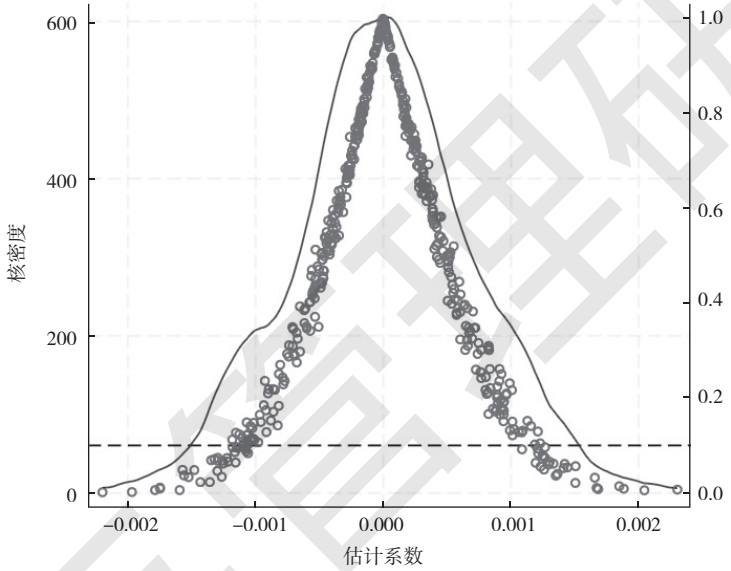


图 2　安慰剂检验结果

从图 2 可以看出,随机生成的虚假政策效应系数集中分布在 0 附近,且绝大多数回归系数的 P 值远大于 0.1,基本服从正态分布;基准回归得到的真实政策效应系数(0.006 1)作为离群值,落在分布的最右端,远离随机分布的核心区域。这一结果进一步证明数据要素集聚提升城市创新能力的结论具有较强的稳健性。

2. 控制同期政策

为考察同期信息化政策对估计结果的影响,本文在基准模型中分别加入宽带中国、智慧城市和电子商务试点变量,并同时控制三项政策,结果见表 6。其余控制变量、固定效应和标准误处理与基准模型一致。

表 6　控制同期其他信息化政策的回归结果

变量	(1)	(2)	(3)	(4)
Bdp	0.006 0**	0.006 1**	0.006 1**	0.006 1**
	(0.003 0)	(0.003 0)	(0.003 0)	(0.003 0)
Broadband	0.004 5**			0.004 3**
	(0.002 2)			(0.002 1)

① 限于篇幅,本文其他的稳健性检验结果不再详细报告,备索。

表6(续)

变量	(1)	(2)	(3)	(4)
Smartcity		-0.003 0		-0.003 6*
		(0.002 0)		(0.002 0)
Ecommerce			0.003 8	0.003 2
			(0.002 8)	(0.002 8)
控制变量	控制	控制	控制	控制
城市固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
常数项	0.012 3	0.012 9	0.012 2	0.013 0
	(0.009 8)	(0.009 8)	(0.009 8)	(0.009 7)
观测值	3 666	3 666	3 666	3 666
R^2	0.942 7	0.942 5	0.942 5	0.942 8

注:被解释变量为 *Innov*。*Broadband*、*Smartcity* 和 *Ecommerce* 分别表示宽带中国、智慧城市和电子商务试点。***、**、* 分别表示在 1%、5%、10%水平下显著;括号内为城市层面聚类的稳健标准误。

表 6 显示,分别或同时控制三项政策后,*Bdp* 的回归系数均在 5%水平下显著为正,与基准估计方向一致。该结果表明,*Bdp* 对城市创新的驱动作用具有较强的政策效应与稳健性。

3. 排除重大外部事件干扰

新冠疫情属于重大突发公共卫生事件,可能对城市经济活动和创新产出造成非正常冲击,从而干扰政策效应的评估。为了排除干扰,本文剔除了 2020 年及之后的样本并重新进行回归。表 7 结果显示,Bdp_{it} 的回归系数为 0.004 6,且在 10%水平下显著。这表明在排除了外部重大冲击的干扰后,国家级大数据综合试验区政策的创新驱动效应依然存在。

表 7 排除重大外部事件干扰的回归结果

变量	*Innov*
Bdp	0.004 6*
	(0.002 5)
控制变量	控制
城市固定效应	控制
年份固定效应	控制
常数项	0.038 3***
	(0.008 8)
观测值	2 538
R^2	0.954 6

(五)机制检验

分别从全域知识溢出(数字人才集聚)、全产业链赋能(数字产业集聚)和全主体协同(创新价值链协同)维度展开机制检验。本文设定如下影响机制模型进行检验:

$$M_{it} = \alpha + \delta Bdp_{it} + \boldsymbol{X}'_{it}\gamma + \theta Pretrend_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (2)$$

其中,M_{it} 为机制变量,分别代表数字人才集聚(*Dta*)、数字产业专业化集聚(*Dspe*)、数字产业多样化集聚(*Ddiv*)和创新价值链协同(*Ivc*)。Bdp_{it} 为核心解释变量,δ 衡量了数据要素集聚对机制变量的因果效应,若 δ 显著,则表明机制成立,其

余符号含义与基准模型相同。

本文选取三个指标作为传导机制的代理变量。

参考已有研究[32],人才集聚(*Dta*)采用数字人才集聚、信息技术产业从业人员区位熵衡量:

$$Dta_{it} = \frac{ITEmp_{it}/Emp_{it}}{ITEmp_{t}/Emp_{t}} \tag{3}$$

其中,$ITEmp_{it}$ 为城市 i 数字技术相关行业从业人数,Emp_{it} 为城市 i 单位从业人员总数。该指标反映城市数字专业劳动力相对于全国平均水平的集中程度。

数字产业集聚(*DIA*)分为数字产业专业化、多样化集聚[33]。

基于《数字经济及其核心产业统计分类(2021)》识别数字经济核心产业企业,采用企业数量区位熵衡量数字产业专业化集聚(*Dspe*):

$$Dspe_{it} = \frac{NDE_{it}/NE_{it}}{NDE_{t}/NE_{t}} \tag{4}$$

其中,NDE_{it} 为城市 i 在 t 年的数字经济核心产业企业数量,NE_{it} 为该城市同年的企业总数,NDE_{t} 为全国同年的数字经济核心产业企业数量,NE_{t} 为全国同年的企业总数。*Dspe* 数值越大,表明该城市数字产业相对于全国水平的集中程度越高。

采用数字经济核心产业各子行业企业数量份额的逆赫芬达尔指数衡量数字产业多样化集聚(*Ddiv*):

$$Ddiv_{it} = \frac{1}{\sum_{k=1}^{20} | s_{ikt} - s_{kt} |} \tag{5}$$

其中,s_{ikt} 为城市 i 在 t 年第 k 类数字核心产业企业数占本市数字核心产业企业总数的比重,数值越大,表示数字产业内部结构越多样;s_{kt} 为当年全国对应行业参照份额。*Ddiv* 越大,表示数字产业在城市整体产业结构中越多样。

创新价值链协同(*Ivc*)参考相关研究[28],采用技术合同成交额与 GDP 之比衡量,以表征区域技术交易强度及创新主体间的协同程度与效率。

表 8 列(1)显示,*Bdp* 提高了数字技术人才区位熵,表明试验区建设增强了数字专业劳动力相对于当地

表 8　机制检验回归结果

变量	(1)	(2)	(3)	(4)
Bdp	0.098 1**	0.070 5***	0.487 5***	0.005 1***
	(0.043 3)	(0.027 0)	(0.145 2)	(0.001 5)
控制变量	控制	控制	控制	控制
城市固定效应	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制
常数项	0.172 5	0.600 5***	1.151 2***	0.033 6***
	(0.132 5)	(0.121 2)	(0.382 1)	(0.006 4)
观测值	3 666	3 666	3 666	3 666
R^2	0.764 7	0.892 8	0.784 9	0.964 4

注:列(1)—列(4)的被解释变量依次为数字人才集聚(*Dta*)、数字产业集聚(专业化)(*Dspe*)、数字产业集聚(多样化)(*Ddiv*)和创新价值链协同(*Ivc*)。四项机制变量均在 1%和 99%分位缩尾。

总体就业的集中程度。数字劳动力市场有助于降低企业招聘和岗位匹配成本,并通过人员流动和面对面交流促进隐性知识扩散。

表 8 列(2)和列(3)显示,*Bdp* 提高了数字企业专业化集聚和多样化集聚。前者说明试验区促进了同类数字企业及专业化服务商集中,有利于共享算力、算法和中间投入并深化分工;后者说明数字核心产业内部的子行业结构更加多元,有利于异质知识组合和跨场景数据复用。产业层面的集聚效应不仅能够使企业共享基础设施和专业化供应商,降低企业的创新成本,还能通过干中学效应和技术空间与时间溢出效应,加速技术创新的迭代升级,提升城市的科技创新水平,从而强化了城市的整体创新产出,提升城市整体创新能力。

表 8 列(4)显示,大数据综合试验区对创新价值链协同的回归系数在 1%水平下显著为正。数据平台和数字化交易规则提高技术供需信息的可观察性,降低成果检索、评估和撮合成本,从而增加企业、高校和科研机构之间的技术交易与成果转化。创新价值链协同的提升意味着资源配置效率的提高和产业链协同活力的释放,为创新活动提供了更加公平、透明、高效的制度环境。

综上所述,实证结果表明数据要素集聚可以通过促进人才集聚、推动数字产业集群建设和提升市场化配置效率三条路径来提升城市创新能力,有力支持了假设 2、假设 3 和假设 4。

(六)异质性分析

考虑到不同城市在资源禀赋、战略定位及发展阶段上的差异,将样本从地理区位、城市群归属等维度进行异质性分析。其中,地理区位划分为东部、中部、西部和东北地区;城市群划分为珠江三角洲地区、长江三角洲地区、环渤海经济圈。人口规模(*Pop*)采用 2011 年全市年末常住人口的自然对数,并减去 282 个城市该对数的均值,在样本期内保持不变[34],表示人口基础在政策中对于创新能力的重要程度;经济发展水平(*High*)参考郑威和江唐洋[35]的研究,以地级市人均 GDP 中位数为标准,将样本划分为发达城市和欠发达城市。变量 *High* 在城市当年 ln(人均 GDP+1)达到或超过该年样本均值时取 1,否则取 0。通过分组回归与引入交互项的两种不同方式,进一步探究数据要素集聚对城市创新能力影响的非均衡性特征,具体结果见表 9、表 10。

表 9 不同地区的分组回归结果

变量	地理区位				城市群归属		
	东部	中部	西部	东北	长三角	珠三角	环渤海
Bdp	0.008 7	−0.003 6	−0.001 4	0.015 5***	0.031 7***	0.041 8***	0.004 4
	(0.007 1)	(0.002 8)	(0.002 1)	(0.004 5)	(0.006 2)	(0.012 1)	(0.006 9)
控制变量	控制	控制	控制	控制	控制	控制	控制
城市固定效应	控制	控制	控制	控制	控制	控制	控制
年份固定效应	控制	控制	控制	控制	控制	控制	控制
常数项	−0.004 4	0.008 9	0.043 1***	0.007 5	0.000 6	0.042 2	−0.063 3
	(0.031 2)	(0.020 3)	(0.009 2)	(0.033 3)	(0.042 8)	(0.054 4)	(0.039 5)
观测值	1 118	1 040	1 079	429	533	273	559
R^2	0.936 5	0.957 1	0.946 7	0.951 1	0.939 8	0.945 5	0.956 6

注:长三角为沪苏浙皖,环渤海为京津冀鲁辽,广东省包含全省 21 个样本城市,与处理组的珠三角 9 个城市不同。

表 10　基期全市人口规模与经济发展水平的交互项回归结果

变量	(1)	(2)
Bdp	0.002 4	0.001 9
	(0.003 1)	(0.002 7)
Bdp×Pop	0.010 7***	
	(0.004 0)	
Bdp×High		0.008 5*
		(0.004 4)
High		0.0014
		(0.002 2)
常数项	0.012 2	0.011 7
	(0.009 5)	(0.010 0)
控制变量	控制	控制
城市固定效应	控制	控制
年份固定效应	控制	控制
观测值	3 666	3 666
R^2	0.943 1	0.942 7

注:两列均使用 2011—2023 年全样本。*Innov* 与控制变量按全样本 1%和 99%分位点缩尾。*Pop* 主效应由城市固定效应吸收;*High* 随年份变化,单列其主效应。

第一,政策赋能的空间非均衡性。从集聚外部性视角审视,数据要素的赋能效应在空间上呈现出鲜明的梯度分化,本质上是不同区域集聚类型与成熟度差异的反映。表 9 的分组回归结果显示,东北地区的回归系数为 0.015 5 且在 1%水平下显著,展现出专业化集聚所激发的追赶红利。数据要素以虚拟集聚形态激活主体协同、产业赋能与知识溢出,进而驱动城市创新。以沈阳为例,沈鼓集团作为“链主”企业打造智鼎工业互联网平台,贯通压缩机产业链上下游的设计、制造与运维数据,推动企业从设备制造商向全生命周期服务商的价值链升级;沈阳机床的 iSESOL 云平台接入 i5 智能机床并汇聚分散加工数据,实现制造能力的跨企业共享与供需高效匹配。综试区政策通过数据要素的虚拟集聚将外部性传导至企业创新行为,由此推动沈阳、长春等城市从传统工业基地向数据驱动的创新型城市发展。长三角与珠三角的回归系数分别为 0.031 7 和 0.041 8,说明高度成熟的数字企业多样化集聚已催生出数据要素专属集聚外部性,多元异质产业网络使数据要素能够跨行业碰撞、耦合,在完备的数字生态中实现自组织的正反馈,创新由此进入规模报酬递增轨道。环渤海等地区尽管拥有体量优势,却因偏重的产业结构与体制分割,导致多样化知识互补受阻,数据集聚停留于物理集中,难以转化为有效的互补性增益,其微弱的系数揭示了有集聚而无外部性的特征,从而在一定程度上说明了南北创新格局的分化逻辑。

第二,互补性资产约束与马太效应。为考察优质资源禀赋在数据赋能过程中产生的“政策溢价”,参考已有研究[36],本文在全样本中引入了核心解释变量与城市特征的交互项进行分析。表 10 列(1)中,*Pop* 在同一城市内不随年份变化,其主效应由城市固定效应吸收。*Bdp×Pop* 的回归系数为 0.010 7,在 1%水平下显著,表明在该设定下,基期全市人口规模越大,政策虚拟变量对应的城市创新能力增量越大。列(2)中,*Bdp×High* 的回归系数为 0.008 5,在 10%水平下显著,表明当年人均 GDP 指标达到或超过样本均值的城市,其政策虚拟变量的回归系数比人均 GDP 指标低于该均值的城市高。从企业选择视角看,数据开发利用具有固定投入和吸收能力门槛,高生产率、强研发能力和强数字化基础的企业更可能进入数据基础设施完善、人才和服务商密集的大城市;企业进入进一步扩大专业劳动力池和技术服务的市场,形成累积因果。因而,大城市和发达城市的较高政策效应既反映互补资产优势,也可能反映高能力企业的选择性进入。

六、结论与建议

本文从集聚外部性视角出发,使用 2011—2023 年中国 282 个地级及以上城市面板数据,将国家级大数据综合试验区建设视为政策驱动型数据要素集聚的准自然实验,采用双重差分模型考察其城市创新效应。本文的研究结论为:国家级大数据综合试验区建设驱动的数据要素集聚提升了城市创新能力,政策效应在实施后逐步显现并趋于稳定;数据要素集聚通过提高数字技术人才集聚、数字企业专业化与多样化集聚,提升创新价值链协同,形成知识溢出、产业赋能和协同转化三类集聚外部性;数据要素价值释放依赖产业基础、人口规模、创新网络等互补性资产;数据要素集聚的政策效应在东北地区、长三角和珠三角城市群以及大城市和经济较发达城市更为明显。

基于以上结论,本文提出如下政策建议:

第一,坚持长期系统推进,分阶段实施数据要素集聚建设。数据要素集聚对城市创新能力的赋能并非短期效应,其推进需要结合当地条件,统筹基础投入、应用扩展和效果评估。已获批综试区城市不应将数据要素建设简单等同于年度工程项目,而是需要围绕数字基础设施、公共数据开放、数据交易规则和应用场景培育建立中长期滚动推进机制。基础投入阶段重点完善数据中心、云平台、5G 网络和政务数据目录,补齐数据采集、存储和流通底座;扩容建设阶段推动公共数据、产业数据和科研数据分级分类开放,培育数据服务商、算法服务商和应用场景供给主体;稳定释放阶段重点考察创新主体活跃度、数字产业承接能力和数据要素市场化配置效率,避免以短期专利产出或单一平台建设数量评价成效。

第二,围绕全域知识溢出、全产业链赋能和全主体协同三个方向,推动数据资源与人才、企业和技术市场相互配合。在全域知识溢出方面,重视数字领域人才引育,完善跨组织知识交流和技术服务机制。在全产业链赋能方面,支持数字企业与实体企业开展合作,为中小企业参与数据相关应用创造条件,关注共享设施与专业化服务的使用效果。在全主体协同方面,完善各机构间技术供需对接与成果转化服务,增加不同主体参与技术交易的机会。政策评估应进一步考察知识扩散、产业链联系和合作关系的实际变化,避免仅以相关行业就业、企业数量或技术合同规模替代协同成效。

第三,结合本地条件实施差异化数字战略。东北地区可结合本地产业需求推进工业数据应用、生产协作和技术改造;长三角和广东省可着重完善数据交易规则、行业数据标准和跨区域数据流通机制,拓展跨行业应用场景。其他地区应依据基础设施、产业承接能力和数字公共服务的实际条件确定投入重点。中部地区可围绕制造业数字化与供应链协同配置数据资源。西部及基础薄弱地区可通过公共算力服务和跨区域协作补齐应用条件,推动数据要素市场化配置适配本地发展需求。

参考文献:

[1]陈旭东,沈利芸,秦文晋. 数字经济、财政纵向失衡与区域经济包容性增长:基于双重机器学习的因果推断[J]. 经济与管理研究,2024,45(9):61-77.

[2]于立,王建林. 生产要素理论新论:兼论数据要素的共性和特性[J]. 经济与管理研究,2020,41(4):62-73.

[3]戚聿东,肖旭. 数字经济时代的企业管理变革[J]. 管理世界,2020,36(6):135-152+250.

[4]Carlino G,Kerr W R. Agglomeration and innovation[J]. Handbook of Regional and Urban Economics,2015,5:349-404.

[5]蔡跃洲,马文君. 数据要素对高质量发展影响与数据流动制约[J]. 数量经济技术经济研究,2021,38(3):64-83.

[6]刘传明,陈梁,魏晓敏. 数据要素集聚对科技创新的影响研究:基于大数据综合试验区的准自然实验[J]. 上海财经大学学报,2023,25(5):

107-121.
[7]卢福财,曾鑫,黄婷. 大数据综合试验区设立对区域创新能力的影响研究[J]. 河海大学学报(哲学社会科学版),2026,28(1):109-126.
[8]胡夏青,马海涛. 中国城市内外创新合作对创新产出的影响[J]. 地理学报,2025,80(2):377-395.
[9]陈婷,段尧清,吴瑾. 数据要素市场化能否提升城市创新能力:一个准自然实验[J]. 科技进步与对策,2024,41(3):74-83.
[10]胡雪萍,王谊巧. 公共数据开放能否提升城市创新水平:基于有效市场和有为政府的机制分析[J]. 科技进步与对策,2026,43(5):1-12.
[11]刘冬梅,陈钰,胡月,等. 长江经济带科技创新发展的现状、需求与未来思考[J]. 科技进步与对策,2025,42(2):152-160.
[12]何骏,连欣燕,田锦萱. 数据要素集聚如何赋能新质生产力?:基于国家大数据综合试验区的准自然实验[J]. 上海经济研究,2025(12):58-72.
[13]吴建军,陈乃康. 高铁开通、要素流动与城市创新能力提升:来自中国 268 个城市的经验证据[J]. 湘潭大学学报(哲学社会科学版),2023,47(3):52-62.
[14]汤凯,蔡晓培,王海杰. 数字技术对城市创新能力的影响研究[J]. 经济与管理研究,2026,47(1):83-99.
[15]彭奕潇,刘淑一,李建琴. 数字政府、制度环境与区域创新能力[J]. 浙江社会科学,2025(8):51-63+157-158.
[16]张贵,孙建华. 数智赋能:智慧城市建设如何影响城市创新韧性? [J]. 经济与管理研究,2026,47(1):100-115.
[17]李婧,郭佳玮. 数据要素集聚对城市创业活力的影响研究[J]. 管理学报,2025,22(12):2262-2270.
[18]Beaudry C,Schiffauerova A. Who's right, Marshall or Jacobs?:The localization versus urbanization debate[J]. Research Policy,2009,38(2):318-337.
[19]江小涓,靳景. 数字技术提升经济效率:服务分工、产业协同和数实孪生[J]. 管理世界,2022,38(12):9-26.
[20]张国胜,严鹏,李欣珏,等. 大数据要素集聚、技术能力缺口与生产率区域差距[J]. 中国工业经济,2024(10):118-136.
[21]吴康,刘骁啸,郭涛. 网络外部性、集聚外部性与城市经济增长[J]. 经济管理,2025,47(4):101-120.
[22]Acemoglu D,Restrepo P. The race between man and machine:implications of technology for growth, factor shares, and employment[J]. American Economic Review,2018,108(6):1488-1542.
[23]谢康,张祎,吴瑶. 数据要素如何产生即时价值:企业与用户互动视角[J]. 中国工业经济,2023(11):137-154.
[24]Jones C I,Tonetti C. Nonrivalry and the economics of data[J]. American Economic Review,2020,110(9):2819-2858.
[25]Ellison G,Glaeser E L,Kerr W R. What causes industry agglomeration?:Evidence from coagglomeration patterns[J]. American Economic Review,2010,100(3):1195-1213.
[26]童玉芬,杨艳飞. 人口空间集聚红利的概念、产生机制与实证检验:基于中国 285 个城市面板数据的分析[J]. 人口与经济,2025(2):1-15.
[27]汪晓文,朱国瑞,张子豪. 数字产业集聚与制造业新质生产力[J]. 经济与管理研究,2025,46(11):59-77.
[28]孔令丞,王悦,谢家平. 长三角区域一体化扩容、协调集聚与区域创新[J]. 财经研究,2022,48(12):34-47.
[29]陈诗一,张建鹏,刘朝良. 环境规制、融资约束与企业污染减排:来自排污费标准调整的证据[J]. 金融研究,2021(9):51-71.
[30]Furman J L,Porter M E,Stern S. The determinants of national innovative capacity[J]. Research Policy,2002,31(6):899-933.
[31]Tang K,Wang Y Y,Wang H J. The impact of innovation capability on green development in China's urban agglomerations[J]. Technological Forecasting and Social Change,2024,200:123128.
[32]Zhao L L,Ye J J. Synergizing pollution reduction and carbon mitigation:evidence from China's National AI Innovation Pilot Zones[J]. International Review of Economics & Finance,2026,110:105501.
[33]陈勇,程钰,张悦,等. 黄河流域数字经济产业集聚演化及其对绿色技术创新的影响[J]. 地理科学进展,2025,44(12):2584-2599.
[34]Ager P,Feigenbaum J J,Hansen C W,et al. How the other half died:immigration and mortality in U. S. cities[J]. The Review of Economic Studies,2024,91(1):1-44.
[35]郑威,江唐洋. 政府创新偏好、制度环境与城市绿色创新[J]. 科技进步与对策,2023,40(19):55-64.

Data Factor Agglomeration, Agglomeration Externalities, and Urban Innovation Capability

—A Quasi-Natural Experiment of China's National Big Data Comprehensive Pilot Zones

GUO Yanli, GUO Chengxin, GAO Shuli
(Beijing Union University, Beijing 100025)

Abstract: Data factor agglomeration may not raise urban innovation capacity (UIC) without skilled workers, firms, and technology markets that can turn shared information into new uses. Existing studies link data policies to innovation but give less weight to how the spatial mix of skills, firms, and trading activity shapes the gains. China's National Big Data Comprehensive Pilot Zones aim to overcome limitations in allocating spatial factors and promote the high-density and large-scale agglomeration of dispersed and fragmented data resources within specific regions, providing a practical setting for iteratively upgrading agglomeration externalities in the digital era. Therefore, exploring the mechanisms and constraints through which data factor agglomeration influences UIC holds significance for deepening market-oriented reforms of the data factor market and implementing the innovation-driven development strategy. Drawing on Marshallian and Jacobs-type externalities, this paper examines whether policy-driven data factor agglomeration raises UIC, through which links it may work, and where its effects are most apparent.

Taking the pilot policy as a quasi-natural experiment and using balanced panel data from 282 prefecture-level and above cities in China from 2011 to 2023, this paper employs the entropy-weighted TOPSIS method to measure urban innovation capacity across three dimensions: innovation inputs, outputs, and the enabling environment. The findings confirm that policy-driven data factor agglomeration enhances UIC, with the policy effects gradually emerging and stabilizing after implementation. Mechanism analyses show that promoting the agglomeration of digital technology talent, fostering the specialized and diversified agglomeration of digital enterprises, and enhancing collaboration along the innovation value chain can generate knowledge spillovers, industrial linkages, and synergistic transformation effects. The first channel accords with labor pooling and knowledge spillovers. The second captures both specialization gains from shared inputs and diversity gains from links across fields. The third is consistent with better links between research and the use of new technology. Heterogeneity analyses reveal that policy effects are more pronounced in cities in the Northeast, the Yangtze River Delta, and the Pearl River Delta, as well as in large cities and more economically developed cities. Tests based on initial city size and current income also point to possible roles for local assets.

By linking gains from shared inputs with gains from cross-sector ties, this paper gives a more precise account of digital agglomeration. Empirical evidence thus links data policy to urban innovation through a spatial view of the digital economy, while keeping policy impact distinct from its proposed channels. The findings highlight the need for sustained data infrastructure development, digital talent cultivation, technology market connectivity, and place-based policy design.

Keywords: data factor agglomeration; urban innovation capacity; China's National Big Data Comprehensive Pilot Zones; agglomeration externality; innovation value chain

编校:魏小奋;姚望春