DOI:10.20069/j.cnki.DJKX.2026036

政策驱动、数据要素集聚与企业关键数字技术创新

江平川[1]，郑明波[2]

1. 西安交通大学 经济与金融学院，陕西 西安 710049
2. 长安大学 经济与管理学院，陕西 西安 710064

移动端阅读

摘要：随着全球科技竞争加剧，关键数字技术创新水平已成为国家竞争力的重要体现。在构建企业关键数字技术创新指标的基础上，以大数据综合试验区设立为准自然实验，采用中国2009—2022年上市公司数据，从数据要素集聚视角评估政策驱动对企业关键数字技术创新的影响。研究发现，大数据综合试验区设立能够显著提升试验区内企业的关键数字技术创新水平。异质性分析显示，大数据综合试验区对非国有企业、东部地区企业以及高竞争行业企业的关键数字技术创新促进作用更为突出。机制分析表明，大数据综合试验区通过财税激励促进效应、基础研究驱动效应、数据共享协同效应和创新要素配置效应推动企业关键数字技术创新。据此提出继续完善各地区数字基础设施建设，多措并举提升对企业数字技术创新的政策支持力度，深化数字技术领域产学研合作等政策建议。

关键词：数据要素集聚；关键数字技术创新；大数据综合试验区；双重机器学习

文献标识码：A **文章编号：**1002-2848-2026(02)-0058-14

一、问题提出

在全球新一轮科技革命和产业变革加速演进的背景下，关键数字技术创新成为驱动经济高质量发展、构筑国家竞争新优势的核心引擎。人工智能、高端芯片、量子信息、区块链和工业互联网等关键数字技术，是数字化转型与产业升级的战略支点，更是实现中国式现代化不可或缺的科技支撑。党的二十届四中全会明确提出要“加快人工智能等数智技术创新，突破基础理论和核心技术”，凸显了关键数字技术在推进高水平科技自立自强、培育新质生产力中的基石作用。党的十八大以来，中国加快推进数字中国建设，在人工智能、量子信息等领域取得若干标志性成果，但在一些关键数字技术领域仍面临“卡脖子”难题。突破关键数字技术的“卡脖子”难题，不仅需要算法算力的协同进步，也离不开数据要素的底层支撑。以智能驾驶为例，百度Apollo平台通过在全国多个城市部署自动驾驶测试车队，收集超过7 000万千米的真实道路行驶数据用于感知算法训练，不断提升自动驾驶系统对复杂路况的识别精度，展现了数据要素在推动人工智能等关键数字技术创新中的重要作用。对企业而言，数据的高效获取、存储与应用能力是把握数据要素机遇、实现关键数字技术突破的核心前提。为释放数据要素乘数效应，中国政府先后出台《促进大数据发展行动纲要》《“十四五”数字经济发展规划》等多项政策，并批复建设国家大数据综合试验区（以下简称“大数据试验区”），引导数据要素在特定区域集聚，推动数据要素的流通、整合与应用，为企业开展技术研发尤其是关键数字技术研发创造了重要机遇。

与本文相关的文献主要集中在两个方向，一是关于数据要素的经济效应及其对技术创新的赋能作

基金项目：国家自然科学基金青年项目（72304047）。

作者简介：江平川，男，通信作者，西安交通大学经济与金融学院博士研究生，研究方向为技术创新，电子邮箱为pingchuan.jiang@stu.xjtu.edu.cn；郑明波，男，长安大学经济与管理学院副教授，研究方向为技术创新与绿色发展。

用，二是关于关键技术创新的驱动因素。在数据要素研究方面，Jones等[1]指出数据具有非竞争性、复用成本低和规模报酬递增等属性。同一组数据可被多个主体同时使用而不产生损耗，并在反复应用中持续放大产出效应，从而带来显著的社会收益。部分研究进一步探讨数据如何转化为创新动能。谢康等[2]提出，数据是一种“桥梁型生产要素”，需与劳动、资本、技术、知识和管理等传统要素相结合，才能从潜在资源转变为现实生产力赋能企业创新。杨俊等[3]则将大数据纳入内生增长理论框架，论证其通过乘数作用驱动技术进步与经济增长。在关键技术创新领域，现有研究从政策支持视角展开分析。吴超鹏等[4]区分了关键技术创新与一般技术创新，发现政府引导基金能激励前者，但对后者影响有限，表明政策工具需精准匹配创新类型。郑世林等[5]强调基础研究平台的重要性，指出国家重点实验室在突破“卡脖子”技术中发挥着不可替代的支撑作用。

既有文献为本文奠定了研究基础，但仍存在进一步探讨空间。一是现有研究虽已对关键技术创新的驱动因素进行初步分析，但多聚焦传统制造业或广义技术范畴，对关键数字技术创新领域的关注不足。关键数字技术具有鲜明的数字原生特征，不仅依赖长期技术积累，更离不开高质量数据的持续供给以及算法与算力的协同演进。上述特性意味着关键数字技术的创新机制可能与其他关键技术存在差异，需要开展针对性研究[6]。二是尚未有文献深入探讨数据要素集聚影响关键数字技术创新的机制和效应。有鉴于此，本文基于国家大数据综合试验区设立的政策冲击，利用2009—2022年中国上市公司数据，构建企业关键数字技术创新指标，并采用双重差分模型和双重机器学习模型，从数据要素集聚视角评估政策驱动对企业关键数字技术创新的影响。

本文的边际贡献体现在三个方面。第一，现有研究或关注数据要素对一般技术创新的影响，或分析广义关键技术创新的驱动因素，缺乏数据要素集聚与关键技术创新尤其是与关键数字技术创新关系的经验证据。本文从数据要素集聚视角出发，探讨大数据试验区政策如何赋能企业关键数字技术创新，丰富了数据要素与关键数字技术创新的相关文献。第二，基于关键数字技术创新面临的多重约束，从财税激励促进效应、基础研究驱动效应、数据共享协同效应以及创新要素配置效应四个方面剖析了试验区政策作用于企业关键数字技术创新的内在机制，为深入理解数据要素集聚与关键数字技术创新的关系提供了研究框架。第三，借助机器学习算法在高维非参数预测中的优势，将双重机器学习模型应用于政策效果评估，部分缓解了传统计量模型在模型设定方面的偏误，提升了研究结论的可信度。

二、政策背景与理论分析

（一）政策背景

2014年国务院政府工作报告首次将大数据纳入国家治理体系。2015年，国务院发布《促进大数据发展行动纲要》，为数据驱动的经济转型提供制度支撑。在此基础上，8个国家大数据综合试验区于2016年相继设立，覆盖贵州、京津冀、珠三角、上海、河南、重庆、沈阳和内蒙古等重点区域，通过跨区域联动探索数据资源整合与应用的可行路径。《“十四五”数字经济发展规划》进一步强调释放数据要素价值，引导试验区聚焦关键核心技术攻关。在政策指引下，各试验区结合本地禀赋探索差异化实施路径。贵州着力构建全产业链生态，率先布局国家级大数据工程实验室；京津冀依托中关村创新资源，打造数据—算力—算法协同平台；珠三角在深化跨境数据流动试点的同时，探索粤港澳大湾区数据流通机制；内蒙古则发挥能源优势，重点培育绿色数据中心集群。尽管发展定位存在差异，但各试验区核心目标始终围绕数据要素的集聚与应用，试验区的数据要素集聚效应逐步显现。以贵州为例，截至2023年，贵州全省在建及运营的重点数据中心达47个，总算力达40.55 EFLOPS，当地已初步形成以云服务为核心的数据产业集群①。贵州数字经济增速连续9年位居全国前列，区内大数据企业研发投入强度显著高于行业平均水平。综上，大数据试验区的设立体现了顶层设计对数据要素集聚的战略引导，也为本文实证研究提供了具有较强外生性的政策冲击。

① 参见贵州省大数据局指导、贵州省信息中心编撰的《国家大数据（贵州）综合试验区发展报告（2023）》。

(二)理论分析

相较于其他关键技术创新,关键数字技术创新展现出不同特征。首先,关键数字技术创新并非单一技术突破,而是多种前沿数字技术深度交叉融合的结果。其次,关键数字技术的成果多以模型或算法形式呈现,具备可复制、易扩散的属性,能够在快速反馈中持续优化。此外,关键数字技术创新的外部性显著,可通过数据共享与知识溢出带动上下游乃至跨行业的协同发展[6]。然而,关键数字技术创新也面临诸多挑战。一是研发投入高、周期长、风险大,需要持续资金支持;二是基础研究与产业应用脱节,技术突破动力不足;三是数据资源分散、协同不畅,存在数据孤岛现象和算力资源的碎片化分布,导致外部性难以有效释放[7];四是创新要素错配问题突出,人才、资本与技术难以形成合力,企业在创新要素不足的情况下易陷入局部最优困境,难以实现突破性创新。上述因素共同削弱了关键数字技术的创新效率。新经济地理学与要素集聚理论指出,生产要素在特定区域的集聚能够通过规模经济、知识溢出与协同效应提升创新效率。作为新型生产要素,数据具有非竞争性、低复用成本与规模报酬递增等特性,其空间集聚能够释放更多价值[8]。在此视角下,大数据试验区的设立构成推动数据要素集聚的重要政策实践,为企业开展关键数字技术研发创造有利条件。因此,本文提出研究假设H1:大数据试验区政策能够促进企业关键数字技术创新。

大数据试验区政策为企业提供财税激励,缓解了企业在关键数字技术研发中面临的资金约束。作为重要的财税激励手段,政府补贴为企业带来了现金流,能够缓解外部融资压力。以高端芯片领域为例,由于研发周期长、资金投入大、技术壁垒高,并且面临激烈的国际竞争与技术封锁,企业仅凭自身资金难以承担巨额研发成本和市场风险。政府补贴不仅能为企业提供直接资金支持,还能间接引导社会资本向该领域企业聚集,缓解企业在创新活动中的资金压力[9]。税收优惠是财政支持的另一重要形式,可直接降低企业创新活动的税收成本[10]。与政府补贴相比,税收优惠作为事后激励手段,产生的交易费用更低[11]。同时,税收优惠的普惠性为企业营造可预期的政策环境,增强了企业开展长期创新活动的信心,促使企业更愿意将资金投入关键数字技术研发领域。据此,本文提出研究假设H2:大数据试验区政策能够通过财税激励促进效应,促进企业关键数字技术创新。

关键数字技术突破离不开基础研究的支撑。现实中,基础研究与产业应用的脱节导致企业关键数字技术创新动能不足。大数据试验区政策通过构建产学研协同创新平台,推动了高校、科研院所与企业之间的数字技术创新合作。高校、科研院所在大数据试验区政策的支持下,获得了企业的真实数据与应用场景,能够提升数字技术研究的针对性。企业则借助高校的基础理论研究能力,深化数字技术攻关[12]。以人工智能领域为例,高校科研团队聚焦底层算法优化,企业则提供多模态场景数据并反馈技术需求。在大数据试验区政策推动下,校企联合研发模式既发挥了高校在基础理论研究方面的优势,也借助企业在技术转化和市场洞察上的经验,加速基础研究成果向关键数字技术创新转化。据此,本文提出研究假设H3:大数据试验区政策能够通过基础研究驱动效应,促进企业关键数字技术创新。

关键数字技术创新依赖高质量数据的积累与流通。在传统运营模式中,由于存在部门间数据壁垒,企业在研发、生产和销售等环节的信息被分散存储,数据要素价值难以被充分挖掘。同时,企业间存在数据孤岛,使得企业难以充分利用外部数据要素。大数据试验区政策通过推动数据开放共享、完善数据标准体系,促进数据在企业内部及外部的高效流动。一方面,企业内部的数据整合将联通分散在各业务模块的数据,打破从研发到市场的信息壁垒,使企业能够实时监控生产环节与资源使用状况,提升关键数字技术创新效率[13];另一方面,企业间、行业间的数据协同打破了单个企业资源与能力边界,推动生态网络中的知识溢出与数字技术创新[14]。例如,蚂蚁集团依托AQ医疗大模型与全国数千家医疗机构开展合作,整合电子病历、影像报告等多源医疗信息,通过跨机构的数据标准化处理,提升了人工智能辅助诊断的准确性,表明跨企业、跨行业的数据协同能够在复杂应用场景中激发关键数字技术的突破潜力。据此,本文提出研究假设H4:大数据试验区政策能够通过数据共享协同效应,促进企业关键数字技术创新。

关键数字技术创新活动的成效依赖创新要素的高效匹配。大数据试验区政策通过推动数字资本投资与人力资本结构优化,帮助企业改善创新要素配置效率。在数字资本方面,试验区政策通过完善数字基础设施、构建数据要素流通环境,引导企业增加在数据存储平台、智能管理系统、数字研发工具等方面

的资本投入，为企业开展算法优化、智能决策系统研发等关键技术创新活动提供物质基础，完善企业的资本要素配置。在人力资本方面，试验区政策通过搭建人才引育平台，营造良好的人才发展生态，促进企业优化数字人才配置[15]。大数据试验区的政策优势吸引高水平科研人员入驻试验区内企业，同时激励企业开展针对性的人力资源培训项目，提升员工的数字知识技能，从而为企业关键数字技术研发提供智力支撑，增强团队关键数字技术创新实力[16]。基于此，本文提出研究假设H5：大数据试验区政策能够通过创新要素配置效应，促进企业关键数字技术创新。

三、研究设计

（一）模型设定

基于上述理论分析，本文以大数据试验区设立为准自然实验，构建双重差分模型，从数据要素集聚视角考察政策驱动对企业关键数字技术创新的影响。模型设定为：

$$inn_{j,t}=a_0+a_1bd_{j,t}+a_2X_{j,t}+\mu_j+\epsilon_t+\varepsilon_{j,t} \quad (1)$$

其中，下标j和t分别代表企业与年份；被解释变量$inn_{j,t}$刻画企业关键数字技术创新水平；政策处理变量$bd_{j,t}$表示大数据试验区政策的实施状态；控制变量集$X_{j,t}$涵盖企业特征与区域特征，用以捕捉潜在混杂因素的影响；μ_j和ϵ_t分别控制企业与年份固定效应；$\varepsilon_{j,t}$是随机扰动项；a_1是本文关注的核心系数，用于衡量政策实施对企业关键数字技术创新的影响效果，若a_1显著为正，则表明政策实施促进了企业关键数字技术创新。

（二）变量和数据说明

1.被解释变量

本文基于企业关键数字技术发明专利授权数量，测度企业关键数字技术创新水平（*inn*）。选用发明专利授权数量作为衡量标准，可规避未授权专利的技术不确定性问题，以及实用新型专利创新层级不足的局限。在专利识别上，本文参照师磊等[17]的研究，依据国家知识产权局2023年发布的《关键数字技术专利分类体系》确定关键数字技术范围。该体系在综合参考全球主要经济体数字技术发展战略、产业政策及技术标准的基础上编制而成，涵盖人工智能、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个关键数字技术领域，共计585项技术分支。本文基于专利分类体系提供的关键数字技术领域国际专利分类号（IPC）信息，与企业专利的IPC编码进行匹配。如果企业专利的IPC编码属于关键数字技术领域，则认定该专利为关键数字技术领域的创新成果。

2.解释变量

根据《促进大数据发展行动纲要》的战略部署，国务院于2016年2月批复贵州建设首个国家大数据综合试验区，并于2016年10月将试验区扩展至京津冀、珠三角、上海、河南、重庆、沈阳以及内蒙古等7个区域，形成两批次共8个国家大数据综合试验区。本文将上述试验区作为处理组，其余地区作为控制组，以此构建大数据试验区政策的准自然实验框架。核心解释变量为地区处理状态与政策实施时间的交互项$bd_{jt}=T_j\times P_t$。其中，T_j为地区处理状态虚拟变量，试验区所在城市取值为1，其他地区为0；P_t为政策时点虚拟变量，第一批次和第二批次地区均从2016年起取值为1。

3.控制变量

本文构建包括企业层面和区域层面的控制变量体系。在企业层面，控制变量主要涵盖组织特征、财务绩效和治理结构三个维度，具体包括：企业规模（*size*），为企业总资产的自然对数；企业年龄（*age*），为企业当前年份减去上市年份加1后的自然对数；资产负债率（*lev*），为企业总负债与总资产的比值；总资产净利润率（*roa*），为净利润与总资产的比值；现金流比率（*cfo*），为现金流量净额与流动负债的比值；固定资产占比（*fix*），为固定资产净额与总资产的比值；托宾Q值（*tq*），为市值与总资产的比值；营业收入增长率（*grw*），为本年营业收入除以上一年营业收入减1；股权制衡度（*bal*），为第二大股东持股比例除以第一大

股东持股比例；管理层持股比例（*msh*），为管理层持股数除以总股本。区域层面变量则聚焦城市经济结构差异，包括：经济发展水平（*gdp*），为企业所在城市人均国内生产总值（GDP）的自然对数；产业结构（*ind*），为第二产业增加值占城市GDP的比例。

4.样本选择和数据来源

本文选取2009—2022年A股上市公司作为研究样本，企业相关数据主要来源于国泰安数据库（CSMAR）、中国研究数据服务平台（CNRDS）和国家知识产权局等，地区层面数据来自历年《中国统计年鉴》、国家统计局及各省份统计年鉴等。此外，大数据试验区政策信息来源于国家及各省级相关部门网站。在此基础上，研究样本经过如下处理：剔除金融类和房地产类企业；剔除ST和*ST类企业；剔除主要变量缺失的企业。经过筛选，最终保留4 036家企业共33 116个观测值。为控制极端值干扰，本文对连续变量进行双侧1%缩尾处理。主要变量的描述性统计结果见表1。

表1 变量描述性统计结果

变量	观测值数	均值	标准差	最小值	最大值
inn	33 116	0.901	3.321	0	36.000
bd	33 116	0.308	0.462	0	1.000
size	33 116	22.164	1.307	19.317	26.496
lev	33 116	0.414	0.208	0.026	0.930
roa	33 116	0.043	0.068	−0.373	0.254
cfo	33 116	0.046	0.069	−0.224	0.290
fix	33 116	0.198	0.155	0.002	0.769
bal	33 116	0.370	0.287	0.001	1.000
tq	33 116	2.053	1.368	0.802	16.647
age	33 116	2.020	0.955	0	3.401
grw	33 116	0.167	0.394	−0.657	4.330
msh	33 116	14.450	20.118	0	70.947
gdp	33 116	11.456	0.519	8.599	12.456
ind	33 116	0.401	0.112	0.117	0.897

四、实证结果分析

（一）基准回归结果

大数据试验区政策影响企业关键数字技术创新的基准回归结果见表2，回归均采用聚类在城市层面的稳健标准误。第（1）列仅加入核心解释变量（*bd*），估计系数为0.958，且在1%的水平下显著，表明大数据试验区建设对企业关键数字技术创新存在正向影响。第（2）列在第（1）列基础上加入企业层面和城市层面控制变量，用于控制潜在混杂因素，估计系数略有下降，但仍在1%的水平下显著。第（3）列则在第（2）列基础上纳入企业固定效应和年份固定效应，以消除个体差异导致的模型估计偏误，并捕捉时间相关的不可观测因素，结果显示核心解释变量系数依然显著为正。进一步分析表明，试验区建设后，区内企业的关键数字技术专利增加0.306个，相当于样本期均值（0.901）的33.96%，提升效果明显，验证了研究假设H1。

表2 基准回归估计结果

变量	被解释变量：*inn*		
	（1）	（2）	（3）
bd	0.958***	0.419***	0.306**
	（0.223）	（0.158）	（0.153）
控制变量	否	是	是
企业固定效应	否	否	是
年份固定效应	否	否	是
观测值数	33 116	33 116	33 116
R^2	0.018	0.069	0.716

注：1.**和***分别表示在5%和1%的水平下显著。

2.（）内为聚类在城市层面的稳健标准误，控制变量估计结果留存备索。

（二）事前趋势检验①

双重差分模型的有效性依赖事前趋势假设，即政策干预前处理组与控制组应保持相同变动趋势。本文采用事件分析法构建模型，检验政策前的平行趋势和政策后的动态效应。为避免多重共线性问题，本文选取大数据试验区设立前2～5年以及后5年时间段，将试验区设立前1年作为基准年份，构建式（1）的动态形式，如下所示：

$$inn_{j,t}=\alpha+\sum_{\tau=2}^{5}\gamma_{-\tau}bd_{j,t}^{p(\tau)}+\gamma_0 bd_{j,t}^{c}+\sum_{\tau=1}^{5}\gamma_{+\tau}bd_{j,t}^{q(\tau)}+\beta X_{j,t}+\mu_j+\epsilon_t+\varepsilon_{j,t} \tag{2}$$

其中，$bd_{j,t}^{p(\tau)}$表示试验区设立前第τ期虚拟变量（$\tau=2,\cdots,5$），$\gamma_{-\tau}$为试验区设立前各年份政策效应；$bd_{j,t}^{c}$表示试验区设立当期虚拟变量，γ_0反映试验区设立当期政策效应；$bd_{j,t}^{q(\tau)}$表示试验区设立后第τ期虚拟变量（$\tau=1,\cdots,5$），$\gamma_{+\tau}$为试验区设立后各年政策效应。

政策效应系数的变化及其95%的置信区间结果显示，在大数据试验区政策实施前，回归系数未通过显著性检验，表明处理组与控制组在试验区设立前的企业关键数字技术创新水平变化趋势基本一致，未拒绝事前平行趋势的假设。在政策实施当年及此后几年，回归系数整体呈现上升趋势并转为正值，表明处理组与控制组在企业关键数字技术创新水平上出现显著差异，验证了大数据试验区政策的动态效果。

（三）安慰剂检验②

尽管基准回归分析已控制企业和区域层面特征变量，基于试验区政策的数据要素集聚效应仍可能受到难以观测因素的潜在干扰。因此，本文采用安慰剂检验间接验证估计结果有效性。具体而言，在保持处理组样本量不变的前提下，通过蒙特卡洛模拟随机重分配处理状态，生成500组虚拟处理样本，再对每组样本采用与基准回归相同的模型设定进行估计。理论上，随机抽取的分组不会影响企业创新水平，其系数应为0。随机模拟结果的核密度分布结果显示，从中可见随机抽样估计系数以0为中心呈正态分布，其置信区间与基准回归结果中的估计值0.306存在显著区隔。这一结果表明，未被观测到的企业特征未对本文的识别框架和估计结果产生显著影响。

（四）内生性检验

为缓解潜在的内生性问题，本文借鉴洪俊杰等[18]的研究，选取地形起伏度作为工具变量。地形起伏度是刻画区域地形特征的宏观指标，由自然地理条件决定，满足外生性要求。同时，地形条件对数字基础设施建设具有约束作用。地形起伏度较低的地区更易部署通信网络与数据中心，因而更可能被选为大数据试验区，符合工具变量的相关性条件。鉴于地形起伏度本身不随时间变化，本文将其与一个时变变量相乘，构造具有时间维度的交互型工具变量。具体而言，以上一年全国互联网宽带接入端口数作为时变权重，与地形起伏度相乘，构建适用于本文实证分析的工具变量（*iv*）。表3报告了两阶段最小二乘法（2SLS）的回归结果。第一阶段回归中，工具变量系数在1%的水平下显著为负，表明地形起伏度越大的地区设立大数据试验区的概率越低，符合预期；F统计量远大于经验法则建议的临界值10，初步排除了弱工具变量问题。第（2）列结果进一步显示，Cragg-Donald Wald F统计量高于Stock-Yogo弱工具变量检验在10%水平下的临界值，拒绝弱工具变量原假设；Kleibergen-Paap rk LM统计量在1%的水平下显著，拒绝工具变量识别不足原假设。在此设定下，核心解释变量系数仍在1%的水平下显著为正，说明在控制内生性因素后，大数据试验区政策对企业关键数字技术创新仍具有显著促进作用，验证了基准回归结果的稳健性。此外，第（3）列将工具变量直接纳入主回归方程，在控制试验区变量后，工具变量系数不再显著，表明该工具变量未通过其他渠道直接影响企业关键数字技术创新，支持外生性假设。

① 篇幅有限，结果留存备索。

② 篇幅有限，结果留存备索。

表3 内生性检验估计结果

变量	第一阶段	第二阶段	
	(1) *bd*	(2) *inn*	(3) *inn*
iv	−0.058***		−0.118
	(0.008)		(0.075)
bd		2.857**	0.798***
		(1.357)	(0.145)
F统计量	55.810***		
Cragg-Donald Wald统计量		216.039	
		[16.380]	
Kleibergen-Paap rk LM统计量		32.707***	
观测值数	20 500	20 500	20 500

注：1. **和***分别表示在5%和1%的水平下显著。

2. 企业和年份固定效应已控制，控制变量估计结果留存备索。

3. ()内为聚类在城市层面的稳健标准误，[]内为Stock-Yogo弱工具变量检验在10%水平下的临界值。

(五)双重机器学习

模型设定偏误仍可能影响回归结果的稳健性。本文进一步采用部分线性回归框架下的双重机器学习方法[19]进行检验。该方法通过正交化处理缓解高维控制变量引发的多重共线性问题，并借助机器学习算法拟合非线性关系，从而降低模型设定偏误[20]。表4第(1)列报告了以随机森林为基础学习器的估计结果，核心解释变量系数为0.874，在1%的水平下显著为正，表明大数据试验区政策提升了企业关键数字技术创新水平。本文从三个方面对结果进行敏感性分析。其一，将样本分割折数由5折调整为8折(训练集与测试集比例由4∶1变为7∶1)，通过增加每次迭代中训练数据占比，使模型更充分地学习整体数据分布。如第(2)列所示，核心解释变量系数仍在1%的水平下显著为正。其二，将重复交叉验证次数由1次增加至2次，以提高数据利用率。第(3)列结果显示，核心解释变量系数仍在1%的水平下显著为正。其三，将随机森林替换为Lasso进行算法适应性检验。第(4)～(6)列结果显示，核心解释变量系数均显著为正。综合来看，无论调整模型参数还是更换算法，政策效应的显著性均稳定在1%的水平，验证了基准回归结果的稳健性。

表4 基于部分线性回归模型的双重机器学习估计结果

变量	被解释变量：*inn*					
	(1)随机森林	(2)随机森林	(3)随机森林	(4)Lasso	(5)Lasso	(6)Lasso
bd	0.874***	0.872***	0.878***	0.646***	0.647***	0.646***
	(0.049)	(0.049)	(0.049)	(0.047)	(0.047)	(0.047)
观测值数	33 116	33 116	33 116	33 116	33 116	33 116
交叉验证折数	5	8	5	5	8	5
重复交叉验证次数	1	1	2	1	1	2

注：1. ***表示在1%的水平下显著。

2. 企业和年份固定效应已控制，控制变量估计结果留存备索。

3. ()内为稳健标准误。

(六)其他稳健性检验①

为确保研究结论的可靠性，本文进一步开展了多项稳健性检验。第一，排除同期实施的“智慧城市”“宽带中国”与“公共数据开放”等区位导向性政策的潜在干扰；第二，从创新质量(以关键数字技术专利被引次数衡量)和创新效率(基于DEA-Malmquist模型测算)两个维度替换被解释变量；第三，采用倾向得

① 限于篇幅，结果留存备索。

分匹配—双重差分(PSM-DID)方法缓解样本选择偏误；第四，剔除数字基础设施薄弱地区及创新资源高度集聚的直辖市，重新平衡样本结构；第五，缩短时间窗口至2013—2019年，排除特定时期外生冲击的影响。所有检验均支持基准回归的核心结果，表明大数据试验区政策对企业关键数字技术创新的促进作用具有较强稳健性。

(七)异质性分析

1.基于企业的异质性分析

鉴于大数据试验区政策可能对不同所有制企业产生差异化影响，本文将样本分为国有企业和非国有企业进行分组回归，结果见表5。第(1)列结果显示，对于非国有企业，核心解释变量系数为正，且在5%的水平下显著，表明大数据试验区政策提升非国有企业关键数字技术创新水平；第(2)列中，国有企业对应系数虽为正，但未通过显著性检验。上述差异可能源于三方面原因。首先，非国有企业对市场信号更为敏感，能够迅速将政策提供的数据要素支持转化为创新产出；而国有企业的创新决策受到多重目标约束，市场化激励相对较弱。其次，非国有企业面临较强融资约束，政策支持有助于缓解其资源短板；而国有企业本身具有资源获得便利性，政策的增量效应相对有限。此外，非国有企业灵活的风险偏好使其更易于把握政策机遇，而国有企业在创新活动中倾向于采取审慎的风险策略。

2.基于区域的异质性分析

地理区位作为创新要素空间配置的重要载体，深刻影响区域创新能力的形成路径。鉴于东部与中西部地区在自然资源禀赋、人力资本结构及政策支持力度等方面存在较大差异，本文依据东中西部划分标准①，考察大数据试验区政策对企业关键数字技术创新的区域异质性效应。表5第(3)列结果显示，东部地区核心解释变量系数为正且在10%的水平下显著；第(4)列中，中西部地区对应系数虽为正，但未达到统计显著性。这一结果表明，大数据试验区政策对东部地区企业关键数字技术创新具有促进作用，但对中西部地区的影响尚不显著，体现出明显的区域异质性。上述差异可能源于两方面原因。其一，东部地区在数字基础设施和人才储备方面具有先发优势，能高效地将数据要素红利转化为创新产出；而中西部地区受限于数字基础设施薄弱、高层次人才集聚不足等条件，政策效果的显现可能存在时滞。其二，东部地区产业集群成熟度高，企业间知识溢出与协同创新效应显著，进一步放大了数据要素的创新促进作用；而中西部地区产业集聚水平较低，创新生态尚不完善，制约政策效能发挥。

表5　异质性分析估计结果

变量	被解释变量：*inn*					
	(1)非国有企业	(2)国有企业	(3)东部地区	(4)中西部地区	(5)高行业竞争	(6)低行业竞争
bd	0.397**	0.177	0.299*	0.186	0.285**	0.263
	(0.172)	(0.205)	(0.178)	(0.252)	(0.142)	(0.162)
观测值数	21 068	12 048	23 756	9 360	16 048	15 984
R^2	0.731	0.695	0.734	0.644	0.774	0.679

注：1.*和**分别表示在10%和5%的水平下显著。

2. 企业和年份固定效应已控制，控制变量估计结果留存备索。

3.()内为聚类在城市层面的稳健标准误。

3.基于行业的异质性分析

行业竞争程度是影响企业创新策略的重要因素。本文参考李青原等[21]的做法，采用行业集中度(*cr*)衡量竞争强度，并据此划分高竞争与低竞争行业。具体而言，行业集中度定义为行业内营业收入排名前10的企业所占份额。本文以行业集中度在行业—年份层面的中位数为阈值构建虚拟变量，若行业集中

① 本文参照国家统计局关于东中西部的分类标准，将全国31个省份(不包括香港、澳门和台湾)划分为三大区域：东部地区包括北京、天津、河北、上海、江苏、浙江、福建、山东、广东、海南、辽宁；中部地区包括山西、安徽、江西、河南、湖北、湖南、吉林、黑龙江；西部地区包括内蒙古、广西、重庆、四川、贵州、云南、西藏、陕西、甘肃、青海、宁夏、新疆。

度低于阈值，则归类为高竞争行业，否则为低竞争行业。表5第(5)列结果显示，高竞争行业中核心解释变量系数为0.285，且在5%的水平下显著；第(6)列结果显示，低竞争行业对应系数虽为正，但未通过统计显著性检验，表明政策效果在不同竞争强度的行业中存在异质性。这一结果与Aghion等[22]提出的“逃逸竞争效应”理论相契合。在竞争性较强的市场中，企业为摆脱同质化竞争困境，倾向于加大关键数字技术创新投入以构建差异化优势；而在垄断性或高度集中的行业中，缺乏外部竞争压力可能削弱企业的创新动力。

五、机制分析

(一)财税激励促进效应

为验证大数据试验区政策是否通过财税激励促进效应推动企业关键数字技术创新，本文分别以政府补贴与税收优惠作为该机制的代理变量进行检验。首先，参考邵颖红等[23]的做法，采用企业当年获得的政府补助金额的自然对数衡量政府补贴水平(*sub*)，数据来源于企业财务报表附注。表6第(1)列结果表明，试验区设立有助于企业获得政府补贴。其次，借鉴柳光强[24]的研究，以现金流量表中“收到的各项税收返还”占“收到的各项税收返还”与“支付的各项税费”之和的比例构建税收优惠指标(*tax*)。第(3)列结果显示，试验区设立有助于企业获得税收优惠。综上，大数据试验区建设通过增加政府补贴与落实税收优惠，强化了财税激励促进效应，进而推动企业关键数字技术创新，验证了研究假设H2。

表6 财税激励促进效应估计结果

变量	政府补贴		税收优惠	
	(1)*sub*	(2)*inn*	(3)*tax*	(4)*inn*
bd	0.122***		0.008***	
	(0.026)		(0.002)	
sub		0.028**		
		(0.011)		
tax				0.644***
				(0.227)
观测值数	31 585	31 585	27 395	27 395
R^2	0.733	0.721	0.756	0.721

注：1. **和***分别表示在5%和1%的水平下显著。
2. 企业和年份固定效应已控制，控制变量估计结果留存备索。
3. ()内为聚类在城市层面的稳健标准误。

(二)基础研究驱动效应

为验证大数据试验区政策是否通过基础研究驱动效应促进企业关键数字技术创新，本文从以下两个方面进行检验。首先，借鉴刁海璨[12]的做法，以企业与高校或科研院所合作发表的科学论文数量衡量校企联合基础研究水平。根据国家统计局定义，基础研究指不以特定应用为目的的理论性或实验性工作，旨在探索基本原理与新知识，成果多以论文、著作等形式呈现。本文以上市公司名称为作者单位，在中国知网和Web of Science中检索其与高校或科研机构合作发表的论文，构建校企合作研究指标(pap_1)。表7第(1)列结果表明，试验区建设发挥了重要桥梁纽带作用，有助于校企合作研究。其次，为捕捉校企合作的持续性，本文采用样本期内企业累计合作发表的科学论文数量构建持续合作指标(pap_2)。第(3)列结果显示，大数据试验区推动了校企长期合作研究。综上，试验区建设通过激发并维系校企基础研究合作，强化了基础研究驱动效应，进而促进了企业关键数字技术创新，验证了研究假设H3。

表7 基础研究驱动效应估计结果

变量	校企合作研究		持续合作研究	
	(1)pap_1	(2)*inn*	(3)pap_2	(4)*inn*
bd	0.258***		0.121***	
	(0.033)		(0.023)	
pap_1		0.145***		
		(0.052)		
pap_2				0.289***
				(0.079)
观测值数	8 973	8 973	8 973	8 973
R^2	0.686	0.740	0.932	0.741

注：1. ***表示在1%的水平下显著。
2. 企业和年份固定效应已控制，控制变量估计结果留存备索。
3. ()内为聚类在城市层面的稳健标准误。

（三）数据共享协同效应

为验证大数据试验区政策是否通过数据共享协同效应促进企业关键数字技术创新，本文选取企业数字化转型与产能利用率作为该机制的代理变量。首先，数字化转型水平反映企业在数据采集、存储、分析及共享方面的能力。数字化程度越高，数据在企业内部及跨组织间的流通效率通常越高。本文借鉴吴非等[25]的做法，通过提取上市公司年报文本中“数字化转型”关键词并统计词频，构建企业数字化转型指标（*dx*）。表8第（1）列结果显示，试验区设立推动当地企业数字化转型。其次，产能利用率可间接体现企业内部资源协调与数据协同的效率。较高的产能利用率意味着生产要素得到更高效的配置与利用。参考李雪松等[13]的研究，本文采用随机前沿生产函数法测度企业产能利用率（*cu*）。基于营业收入、资产总额和员工人数构建生产前沿面，以实际产出与前沿产出之比作为产能利用率指标。第（3）列结果表明，试验区设立提升企业产能利用率。综上，大数据试验区建设通过推动企业数字化转型与提升产能利用率，强化了数据共享协同效应，进而促进了企业关键数字技术创新，验证了研究假设H4。

表8　数据共享协同效应估计结果

变量	数字化转型		产能利用率	
	（1）*dx*	（2）*inn*	（3）*cu*	（4）*inn*
bd	0.058***		0.007***	
	（0.019）		（0.001）	
dx		0.144***		
		（0.017）		
cu				0.913***
				（0.307）
观测值数	33 116	33 116	31 366	31 366
R^2	0.811	0.716	0.737	0.717

注：1. ***表示在1%的水平下显著。

2. 企业和年份固定效应已控制，控制变量估计结果留存备索。

3.（）内为聚类在城市层面的稳健标准误。

（四）创新要素配置效应

为验证大数据试验区政策是否通过创新要素配置效应提升企业关键数字技术创新水平，本文选取数字资本投资与人力资本优化作为该机制的代理变量。首先，借鉴张永珅等[26]的做法，以数字化相关无形资产占无形资产总额的比例衡量企业数字资本投资（*dci*），数据来源于上市公司财务报告附注。具体而言，当无形资产明细中包含“软件”“客户端”“智能平台”“管理系统”等数字化相关内容时，将其归类为数字化无形资产。表9第（1）列结果显示，试验区设立促使企业提高数字资本投资。其次，借鉴肖土盛等[27]的做法，本文从技能结构与教育结构两个维度衡量人力资本优化。其中，将科技、市场销售及财务人员界定为高技能群体，以高技能员工占比构建技能结构指标（hum_1）；同时，以硕士及以上学历员工占比衡量教育结构（hum_2）。第（3）（5）列结果显示，试验区设立显著提高了企业高技能与高学历员工的比重。综上，大数据试验区建设通过推动企业增加数字资本投资与优化人力资本结构，强化了创新要素配置效应，进而提升企业关键数字技术创新水平，验证了研究假设H5。

表9　创新要素配置效应估计结果

变量	数字资本投资		人力技能结构		人力教育结构	
	（1）*dci*	（2）*inn*	（3）hum_1	（4）*inn*	（5）hum_2	（6）*inn*
bd	0.023***		1.195***		0.435**	
	（0.003）		（0.247）		（0.185）	
dci		0.600***				
		（0.206）				
hum_1				0.005***		
				（0.002）		
hum_2						0.116***
						（0.030）
观测值数	28 859	28 859	25 343	25 343	5 315	5 315
R^2	0.680	0.710	0.887	0.769	0.959	0.823

注：1. **和***分别表示在5%和1%的水平下显著。

2. 企业和年份固定效应已控制，控制变量估计结果留存备索。

3.（）内为聚类在城市层面的稳健标准误。

六、结论与启示

本文基于2009—2022年中国上市公司数据，构建企业关键数字技术创新指标，运用双重差分模型和双重机器学习模型，从数据要素集聚视角评估大数据试验区政策对企业关键数字技术创新的影响。研究发现，试验区政策显著提升了区内企业的关键数字技术创新水平，该结论在一系列稳健性检验后仍然成立。异质性分析表明，政策效应在非国有企业、东部地区企业以及高竞争行业企业中更为突出。机制检验表明，政策主要通过财税激励促进效应、基础研究驱动效应、数据共享协同效应和创新要素配置效应四重路径共同促进企业关键数字技术创新。本文的研究结论具有以下政策启示：

第一，继续完善各地区数字基础设施建设，为数据要素集聚的价值释放提供支撑。研究表明，数据要素集聚能显著促进企业关键数字技术创新，且对东部企业促进作用更强，原因在于东部地区拥有在数字基础设施、数据资源配置与技术人才供给等方面的优势。对于东部地区，要继续提高数字基础设施智能化和集约化水平，开展新型数字基础设施示范工程，提升数据要素存储、流通和交易能力。对于中西部地区，要加大数字基础设施建设力度，布局高性能计算中心、工业互联网平台和数据交易中心等设施，增强区域数据处理能力与算力供给水平。同时，应建立跨区域数据要素流通与合作机制，推动东部地区优质数据资源、科研成果与技术标准向中西部扩散，促进区域间数据要素的高效协同。

第二，多措并举提升对企业数字技术创新的政策支持力度，因企施策激发企业的主体创新活力。非国有企业与高竞争行业企业对数据要素集聚政策响应更为积极，这表明市场化机制和竞争压力在推动关键数字技术创新中发挥着重要作用。基于此，应从企业类型与行业结构两方面完善政策体系，实现对不同主体的精准支持。对于中小企业，完善研发费用加计扣除和创新投资税收抵免等支持措施，并通过设立专项信贷担保与风险补偿基金缓解企业在关键数字技术研发中的融资约束。对于国有企业，优化绩效评价体系，在传统经营指标之外增加关键数字技术研发投入、技术成果转化和数据治理效能等内容，引导资源投向半导体设备、人工智能硬件平台等战略性数字技术领域，从而发挥国有企业在关键数字技术攻关中的引领作用。针对高竞争行业，建设行业级数据共享与协同创新平台，探索以脱敏数据、安全沙箱等技术为基础的数据合作模式，促进企业间的数据互补与技术合作。在低竞争行业，可建立与关键数字技术创新绩效挂钩的税收减免机制，提高企业自主研发动力。

第三，深化数字技术领域产学研合作，加快推动数据要素开放共享，畅通数据要素集聚促进关键数字技术创新的作用渠道。具体而言，深化产学研合作，推广“企业出题、高校解题、联合攻关、成果共享”的协作模式，鼓励科研机构与企业共建创新联合体，为关键数字技术创新夯实基础。同时，加快行业级数据中枢建设，整合政务、公共与产业数据资源，建立统一标准并通过安全接口，向中小企业提供可负担、易获取的数据服务，切实提升数据要素服务普惠性。此外，完善数据资产确权、分级分类管理与入表机制，为数据要素的交易、授权和收益分配提供制度保障。

参考文献：

[1] JONES C I，TONETTI C. Nonrivalry and the economics of data[J]. American Economic Review，2020，110(9)：2819-2858.

[2] 谢康，夏正豪，肖静华. 大数据成为现实生产要素的企业实现机制：产品创新视角[J]. 中国工业经济，2020，38(5)：42-60.
XIE K，XIA Z H，XIAO J H. The enterprise realization mechanism of big data becoming a real production factor：from the product innovation perspective[J]. China Industrial Economics，2020，38(5)：42-60.

[3] 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.
YANG J，LI X M，HUANG S J. Big data，technical progress and economic growth：an endogenous growth theory introducing data as production factors[J]. Economic Research Journal，2022，57(4)：103-119.

[4] 吴超鹏，严泽浩. 政府基金引导与企业核心技术突破：机制与效应[J]. 经济研究，2023，58(6)：137-154.

WU C P，YAN Z H. Government guidance funds and breakthroughs in firm’s core technologies[J]. Economic Research Journal，2023，58(6)：137-154.

[5] 郑世林，汉馨语，郭锡栋，等. 国家战略科技力量与企业关键核心技术突破：来自国家和省级重点实验室的证据[J]. 中国工业经济，2024，42(9)：62-80.
ZHENG S L，HAN X Y，GUO X D，et al. Construction of national strategic scientific and technological power and enterprises’ key core technology breakthroughs：evidence from national and provincial key laboratories[J]. China Industrial Economics，2024，42(9)：62-80.

[6] 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.
HUANG B，LI H T，LIU J Q，et al. Digital technology innovation and the high-quality development of Chinese enterprises：evidence from enterprise’s digital patents[J]. Economic Research Journal，2023，58(3)：97-115.

[7] 冯根福，刘虹，冯照桢，等. 股票流动性会促进我国企业技术创新吗?[J]. 金融研究，2017，60(3)：192-206.
FENG G F，LIU H，FENG Z Z，et al. Does stock liquidity enhance technological innovation?[J]. Journal of Financial Research，2017，60(3)：192-206.

[8] 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，2024，42(10)：118-136.
ZHANG G S，YAN P，LI X J，et al. Big data factor agglomeration，technological capability gaps and regional disparities in productivity[J]. China Industrial Economics，2024，42(10)：118-136.

[9] 冯根福，郑明波，温军，等. 究竟哪些因素决定了中国企业的技术创新：基于九大中文经济学权威期刊和A股上市公司数据的再实证[J]. 中国工业经济，2021，39(1)：17-35.
FENG G F，ZHENG M B，WEN J，et al. What determines the Chinese firms’ technological innovation：a re-empirical investigation based on the previous empirical literature of nine Chinese economics top journals and A-share listed company data[J]. China Industrial Economics，2021，39(1)：17-35.

[10] BLOOM N，GRIFFITH R，VAN REENEN J. Do R&D tax credits work? Evidence from a panel of countries 1979-1997[J]. Journal of Public Economics，2002，85(1)：1-31.

[11] 周燕，潘遥. 财政补贴与税收减免：交易费用视角下的新能源汽车产业政策分析[J]. 管理世界，2019，35(10)：133-149.
ZHOU Y，PAN Y. Subsidization versus tax reduction：the policy choice for new energy vehicles under the constraints of transaction costs[J]. Journal of Management World，2019，35(10)：133-149.

[12] 刁海璨. 企业基础研究与新质生产力培育[J]. 数量经济技术经济研究，2025，42(3)：91-110.
DIAO H C. Analysis of the impact pathways of enterprise basic research on new quality productive forces[J]. Journal of Quantitative & Technological Economics，2025，42(3)：91-110.

[13] 李雪松，赵宸宇，聂菁. 对外投资与企业异质性产能利用率[J]. 世界经济，2017，40(5)：73-97.
LI X S，ZHAO C Y，NIE J. OFDI and corporations’ heterogeneous capacity utilization[J]. The Journal of World Economy，2017，40(5)：73-97.

[14] 孙伟增，毛宁，兰峰，等. 政策赋能、数字生态与企业数字化转型：基于国家大数据综合试验区的准自然实验[J]. 中国工业经济，2023，41(9)：117-135.
SUN W Z，MAO N，LAN F，et al. Policy empowerment，digital ecosystem and enterprise digital transformation：a quasi-natural experiment based on the national big data comprehensive experimental zone[J]. China Industrial Economics，2023，41(9)：117-135.

[15] 冯根福，王珏帅，郑明波. “动态国家综合要素竞争优势理论”与中国长期经济增长[J]. 当代经济科学，2022，44(6)：1-12.
FENG G F，WANG J S，ZHENG M B. Dynamic national synthetical factors competitive advantage theory and China’s long-term economic growth[J]. Modern Economic Science，2022，44(6)：1-12.

[16] 戴魁早，李晓莉，骆苕函. 人力资本结构高级化、要素市场发展与服务业结构升级[J]. 财贸经济，2020，41(10)：

129-146.
DAI K Z, LI X L, LUO J H. Human capital structure upgrading, factor market development and service industry structure upgrading[J]. Finance & Trade Economics,2020,41(10):129-146.
[17] 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,2025,43(1):100-117.
SHI L, YANG Z, QIAN G M. Digital industrial cluster policy and radical innovation in key core technologies[J]. China Industrial Economics,2025,43(1):100-117.
[18] 洪俊杰,李研,杨曦.数字经济与收入差距:数字经济核心产业的视角[J].经济研究,2024,59(5):116-131.
HONG J J, LI Y, YANG X. Digital economy and income gap: from the perspective of core industries in the digital economy[J]. Economic Research Journal,2024,59(5):116-131.
[19] CHERNOZHUKOV V, CHETVERIKOV D, DEMIRER M, et al. Double/debiased machine learning for treatment and structural parameters[J]. The Econometrics Journal,2018,21(1):C1-C68.
[20] 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.
ZHANG T, LI J C. Network infrastructure, inclusive green growth, and regional inequality: from causal inference based on double machine learning[J]. Journal of Quantitative & Technical Economics,2023,40(4):113-135.
[21] 李青原,王露萌.会计信息可比性与公司避税[J].会计研究,2019,40(9):35-42.
LI Q Y, WANG L M. Financial statement comparability and corporate tax avoidance[J]. Accounting Research,2019,40(9):35-42.
[22] 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.
[23] 邵颖红,周恺伦,程与豪.政府补助在激励企业"卡脖子"技术创新中能否提供助力:以企业参与内循环程度为调节变量[J].科技进步与对策,2024,41(3):84-92.
SHAO Y H, ZHOU K L, CHENG Y H. Can government subsidies assist in incentivizing enterprises to make innovation in technological "bottleneck" breakthroughs? The degree of enterprise participation in domestic circulation as a moderating variable[J]. Science & Technology Progress and Policy,2024,41(3):84-92.
[24] 柳光强.税收优惠、财政补贴政策的激励效应分析:基于信息不对称理论视角的实证研究[J].管理世界,2016,32(10):62-71.
LIU G Q. Analysis of incentive effects of tax preference and financial subsidy policies: an empirical study based on the perspective of information asymmetry theory[J]. Journal of Management World,2016,32(10):62-71.
[25] 吴非,胡慧芷,林慧妍,等.企业数字化转型与资本市场表现:来自股票流动性的经验证据[J].管理世界,2021,37(7):130-144.
WU F, HU H Z, LIN H Y, et al. Enterprise digital transformation and capital market performance: empirical evidence from stock liquidity[J]. Journal of Management World,2021,37(7):130-144.
[26] 张永珅,李小波,邢铭强.企业数字化转型与审计定价[J].审计研究,2021(3):62-71.
ZHANG Y S, LI X B, XING M Q. Enterprise digital transformation and audit pricing[J]. Auditing Research,2021(3):62-71.
[27] 肖土盛,孙瑞琦,袁淳,等.企业数字化转型、人力资本结构调整与劳动收入份额[J].管理世界,2022,38(12):220-235.
XIAO T S, SUN R Q, YUAN C, et al. Digital transformation, human capital structure adjustment and labor income share[J]. Journal of Management World,2022,38(12):220-235.

编辑:郑雅妮,高原

Policy Incentives, Data Factor Agglomeration and Key Digital Technology Innovation in Enterprises

JIANG Pingchuan[1], ZHENG Mingbo[2]

1. School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710049, China
2. School of Economics and Management, Chang'an University, Xi'an 710064, China

Summary Against the backdrop of a new global technological revolution, breakthroughs in key digital technologies have become the driving force for high-quality economic development. As an important component of national competitiveness, innovation capabilities in key digital fields not only concern the security and stability of the industrial chain, but also serve as a strategic foundation for China to achieve technological self-reliance, strength, and the ability to respond to intense international competition. Unlike general technological innovation, key digital technology innovation depends on the extensive use of data factors, the combined development of computing power and algorithms, and the comprehensive support of digital infrastructure. While certain areas of key digital technologies have seen breakthroughs, significant challenges remain overall. Therefore, enhancing enterprises' innovation capabilities in key digital technologies has become a major practical issue of concern to the Chinese government.

Data factors are the core resource driving innovation in key digital technologies. The accumulation of massive data drives the continuous optimization of algorithmic models, propelling artificial intelligence technology from laboratory research to market applications. For enterprises, the ability to efficiently obtain, store, and apply data is the core prerequisite for seizing data factor opportunities and achieving breakthroughs in key digital technologies. To leverage the multiplier effect of data, the Chinese government has introduced national big data comprehensive pilot zones to guide the agglomeration of data resources within specific regions, thereby promoting the circulation and application of data factors.

Using data from Chinese listed companies between 2009 and 2022, the paper constructs an indicator for enterprise key digital technology innovation and employs a difference-in-differences (DID) model and a double machine learning (DML) approach to assess the impact of national big data comprehensive pilot zones on such innovation from the perspective of data factor agglomeration. The findings show that the pilot zones significantly enhance key digital technology innovation among enterprises within them. Heterogeneity analysis indicates that the policy has a significant positive effect on non-state-owned enterprises, those in eastern regions and those in highly competitive industries. Mechanism analysis reveals that the policy primarily facilitates key digital technology innovation through fiscal and tax support, basic research, data sharing and collaboration, and innovation factor allocation.

This research provides a micro-level empirical foundation for the precise formulation of digital economy policies, offering significant insights into how data factor agglomeration can empower corporate innovation in key digital technologies. It recommends targeted policies to narrow the innovation gap between eastern and western regions, such as improving digital infrastructure in underdeveloped areas, promoting inter-regional cooperation and supporting the development of distinctive industries. To stimulate innovation across all enterprises, the study suggests providing research and development (R&D) support to small and medium-sized enterprises (SMEs), improving innovation evaluation mechanisms for state-owned enterprises and promoting data sharing among industries. Additionally, the study proposes establishing a four-dimensional collaborative mechanism integrating fiscal and tax support, basic research, data sharing and resource allocation to enhance the digital innovation ecosystem.

Keywords data factor agglomeration; key digital technology innovation; big data comprehensive pilot zones; double machine learning