

大数据综合试验区设立对区域创新能力的影响研究

卢福财¹, 曾鑫², 黄婷¹

(1. 江西财经大学数字经济学院; 2. 江西财经大学经济学院)

摘要:随着新一轮科技革命和产业变革的孕育和兴起,科技创新已成为世界主要国家发展战略的重心。在这一背景下,大数据作为数字经济时代的关键生产要素,其通过赋能区域创新能力推动经济高质量发展的机制亟待深入探讨。基于 2007—2021 年中国 281 个城市面板数据,以大数据综合试验区的设立作为一项准自然实验,实证检验国家大数据综合试验区设立对区域创新能力的影响及其作用机制。结果表明:第一,试验区设立显著提升了区域创新能力,且这一结论在经过多个稳健性检验后依然成立。空间杜宾模型进一步揭示,试验区政策存在显著的正向空间溢出效应,其对邻近地区创新能力的带动作用甚至超过本地效应,凸显了数据要素跨区域流动的创新外溢价值。第二,机制检验显示,试验区设立主要通过增加创新主体、集聚创新资源和优化创新环境 3 种渠道提升区域创新能力。第三,异质性结果显示,试验区设立对数字经济发展水平较高的地区、市场化水平较高的地区和东部地区的创新能力的促进作用更加明显。调节效应结果显示,加强知识产权保护能够增强试验区对区域创新能力的促进作用。

关键词:大数据综合试验区;区域创新能力;创新主体;创新资源;创新环境

创新作为推动经济增长和社会发展的核心驱动力,已成为全球各国战略竞争的焦点。现有研究表明,尽管中国创新产出数量快速增长,但创新质量与发达国家仍存在显著差距^[1],这一现象主要体现为:一是核心专利占比偏低,知识产权贸易逆差持续扩大^[2];二是专利转化率不足,平均寿命较短;三是出口技术复杂度相对落后^[3]。这些事实表明,提升区域创新能力是实现经济高质量发展的关键所在^[4]。在这一背景下,创新驱动发展上升为国家战略:《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》明确提出,要将

“坚持创新在我国现代化建设全局中的核心地位”^[5];《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》强调,要“提升国家创新体系整体效能,全面增强自主创新能力”^[6];2025 年底举行的中央经济工作会议进一步提出,要“坚持创新驱动,加紧培育壮大新动能。”^[7]这一战略导向与内生增长理论和创新系统理论的核心观点高度契合,即创新能力的提升需要系统性的制度安排和政策支持。随着全球信息化步伐的加速推进,大数据作为数字经济时代的重要资源,在发现新知识、创造新价值和提高创新能力等方面发挥着关键作用,

引用本文: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.

基金项目:国家社会科学基金重大招标项目(23&ZD073);国家自然科学基金青年项目(72201114);江西省自然科学基金管理科学项目(20252BAA100059)

作者简介:卢福财(1963—)男,教授,博士,主要从事数字经济研究。E-mail:lu-fucan@263.net

正引领新一轮科技创新浪潮。2014年,大数据首次被写入政府工作报告;2015年,大数据发展被纳入国家战略框架;2020年,数据被正式确立为关键生产要素之一,与传统四大要素并列,并参与分配。通过采集、整理、聚合、分析和处理,数据转化为具有使用价值的数据资源。数据要素价值的实现依赖于大数据技术的收集、处理和共享,大数据发展对集聚数据资源、培育数据要素市场至关重要。《促进大数据发展行动纲要》(以下简称《纲要》)指出,国家大数据资源统筹发展工程的重要任务之一是整合分散的数据中心资源,运用云计算技术优化规模小、效率低、能耗高的数据中心,构建布局合理、规模适度、支撑力强且绿色低碳的政务数据中心体系^[8]。大数据综合试验区(以下简称试验区)的设立,有效促进了大数据资源与资金等生产要素的高效融合,成为推动数字经济发展的关键引擎,为经济绿色发展描绘了宏伟蓝图。因此,在大数据发展背景下,如何利用试验区优化要素之间的关系,提升企业、科研单位和高等教育机构等区域性组织的价值共创能力,对我国经济高质量发展具有重要意义。

基于中国特色的制度创新——试验区的设立,从创新主体、创新资源和创新环境3条路径考察试验区设立对区域创新能力的带动作用。可能的边际贡献在于:一是研究内容上,将试验区设立和区域创新能力纳入同一分析框架,深入剖析试验区设立对区域创新能力的直接影响,并使用双重机器学习模型等方法提高研究结论的可信度,进而丰富试验区设立与区域创新能力的相关研究。二是研究视角上,现有关于试验区设立与创新能力的研究主要围绕人才^[9]、技术和资本^[10]等创新资源的影响机制展开,较少关注创新主体和创新环境对其的影响。基于区域创新生态系统的视角,本研究从创新主体增加效应、创新资源集聚效应和创新环境优化效应3个方面厘清了传导机制。三是研究内容上,探究了试验区设立影响区域创新能力的空间溢出效应和知识产权保护的调节作用,为理解区域间数据要素的联动效应提供了科学依据。

一、政策背景与理论分析

1. 政策背景和文献评述

在新一轮科技革命和产业变革的全球浪潮中,如何更好地应用数据要素已成为重塑国际经济格局、提升国家核心竞争力的关键举措。分析现实情境可知,大数据在我国的发展与应用已初具规模,展现出显著的市场优势和广阔的发展前景,但其中产业链不完整、衍生基础和配套不足、数据开放程度不够、创新应用领域偏窄等问题依然是需要长期攻关的难题。为全面推进我国大数据发展和应用、加快建设数据强国,2015年国务院印发《纲要》,明确将大数据发展纳入国家战略框架,提出未来5~10年在社会治理、经济运行、产业生态等领域的发展目标,而试验区正是这一战略的核心落地载体,旨在通过区域试点探索可复制、可推广的大数据发展路径,为全国数字经济发展提供“试验田”。同年9月,贵州大数据综合试验区建设正式启动,旨在推动区域数据基础设施整合与数据资源的集聚应用;次年10月,为推进国家大数据战略的实施、贯彻落实《纲要》要求,包括京津冀、珠江三角洲、上海、河南、重庆、沈阳大数据综合试验区及内蒙古大数据基础设施统筹发展类的第二批大数据综合试验区获批设立。从试验区选择来看,8个大数据综合试验区在资源条件、经济实力、信息化程度及生态环境等方面存在异质性,呈现多元化发展趋势^[11]。从政策目标来看,尽管各试验区特色各异,但其核心目标是一致的,即突破数据资源壁垒,加强基础设施统筹,推动大数据产品创新,培育骨干企业,建设众创空间,培养专业人才,通过挖掘数据价值,提升制度和科技创新能力。

现有针对试验区设立经济效应的研究主要基于企业、产业和地区等视角,围绕其对经济增长的促进作用展开:一是生产率效应。在试验区设立背景下,数据要素通过创新赋能、运营赋能和投资赋能推动企业全要素生产率的提升^[12]。通过提高试点地区企业对大数据技术的应用能力,数据要素能够促进技术进步、提高企业创新水平、优化经营效率,从而对全要素生产率产生积极影响^[13-14]。二是创新效应。相

关研究表明,大数据技术的应用与发展能够显著降低企业的信息获取成本和决策成本,有效缓解信息不对称问题并拓展企业的经营范围,从而为企业创新提供重要驱动力^[15-16]。数据要素集聚通过优化劳动力资源配置、推动产业结构升级、提升创业活跃度、提升人力资本水平和研发投入水平等途径对科技创新产生正向影响^[17]。试验区通过充分释放数据要素赋能效应,能提高绿色技术创新水平,助力企业绿色化转型^[18]。

总的来说,目前关于试验区的研究主要关注经济增长效应,但对其创新影响机制的分析尚不系统,特别是在以下三方面存在不足:其一,多数研究仅考察创新资源投入的影响^[14],忽视了创新主体和创新环境的协同作用;其二,对空间效应的研究存在“溢出效应”与“虹吸效应”的理论分歧;其三,缺乏对政策异质性的深入分析,难以为差异化政策制定提供依据。

2. 试验区设立对区域创新能力的直接影响效应

区域创新能力是指特定地理空间内各类创新主体(如企业、高校、科研机构等)通过系统性协同,将知识、技术等要素转化为新产品、新工艺及新服务的能力。这一概念植根于创新系统理论,强调创新活动的网络化特征与空间集聚效应。从理论内涵看,区域创新能力不仅体现为创新产出(如专利、新产品收入等),更反映了区域创新生态系统的整体效能,其核心在于创新要素的配置效率与主体间的协同机制^[19]。影响区域创新能力的因素可归纳为创新主体、创新资源和创新环境3个维度,以下先从试验区的设立背景及其对区域创新能力的直接影响进行分析。

试验区作为国家推动数据要素市场化配置与产业数字化转型的重要制度安排,其核心目标在于通过制度创新与政策赋能,破解数据资源壁垒、优化数据基础设施布局、激发市场主体创新活力。试验区通常会赋予试点地区财税优惠、行政审批简化、数据开放共享等制度性特权,构建激励相容的政策环境,能有效降低企业制度性交易成本与创新风险,为区域创新系统注入新的制度动力^[11]。一是从要素配置视角

看,试验区设立通过数据要素的集聚与高效流动,能显著提升劳动、资本等传统要素的配置效率;在劳动力市场,大数据技术降低了人才匹配成本,缓解了结构性失业;在资本市场,试验区设立增强了信息透明度,优化了资本配置效率。二是从产业结构视角来看,试验区设立还可通过产业数字化与数字产业化双轮驱动,推动传统产业智能化转型与新兴业态的培育,构建更具韧性与活力的区域创新生态系统。

基于此,提出假设1:试验区设立能够有效促进区域创新能力的提高。

3. 理论分析框架

试验区设立并非单一措施,而是一套旨在系统性重塑区域创新生态的政策组合,其核心在于通过制度创新,破解数据要素市场化配置中的关键瓶颈,进而激发微观主体活力、优化要素配置效率并改善宏观制度环境。基于创新主体、创新资源和创新环境3个维度,构建完整的“政策-行为-绩效”理论分析框架。

(1) 创新主体增加效应

试验区设立首先通过一系列制度性特权,显著改变了市场主体的预期收益与成本结构,从而可引致创新主体在数量与能级上的双重提升。作用路径具体表现为:第一,在准入环节,行政审批的简化与负面清单管理能有效降低大数据领域新创企业的制度性进入壁垒,可催生大量专注于数据技术开发与应用的新兴技术创新主体。第二,在运营环节,针对大数据企业的税收减免、研发费用加计扣除及专项补贴等财税工具,可直接提升创新活动的边际收益,这不仅激励了存量企业扩大研发规模,也吸引了域外相关企业的迁入与再投资。第三,在协同环节,试验区通过设立产学研合作专项基金、搭建数据共享平台,可促进高校、科研机构等知识创新主体与产业界的深度融合^[20]。与此同时,公共数据资源向科研机构的定向开放极大提升了知识生产的效率与针对性,而成果转化收益分配的激励机制则进一步强化了其面向市场需求的创新导向^[21]。因此,试验区设立通过构建“降门槛、增收益、促合作”的三重驱动,系统性拓展了区域创新主体的基数并提升其质量,为知识溢出与协同创新奠定了坚实的微观基础。

基于此,提出假设2:试验区设立通过增加创新主体提高区域创新能力。

(2) 创新资源集聚效应

试验区设立的另一核心机制在于,通过确立数据作为关键生产要素的地位,引领人力资本与金融资本等传统高阶创新资源向区内集聚,并优化其配置效率^[17]。这一过程主要体现在两个层面:一是在要素引致层面,大数据产业集群的形成与发展自然会衍生出对数据科学家、算法工程师等高技能人才的庞大需求,而试验区配套实施的安家补贴、子女教育等人才政策则进一步放大了这一吸引力,形成了强大的人力资本“磁场效应”^[22]。与此同时,政府设立的产业引导基金发挥了信号传递与风险共担作用,撬动了大量社会风险资本投向大数据创新领域。二是在要素赋能层面,海量数据资源的可获得性使企业能够构建“数据驱动型”的研发模式,通过对市场趋势、技术路径的精准洞察,大幅降低研发活动的不确定性,从而提升研发经费的配置效率。数据要素与人力、资本在试验区内的交叉融合催生了如众包研发、协同过滤等新型创新范式,实现了不同要素之间的协同倍增效应,共同构筑了区域创新能力提升的资源基石。

基于此,提出假设3:试验区设立通过集聚创新资源提高区域创新能力。

(3) 创新环境优化效应

试验区设立的影响深远之处,还在于其对区域创新“硬环境”与“软环境”的系统性优化,可为创新活动提供坚实的基础支撑与稳定的制度保障。硬环境建设方面,政府通过前瞻性布局与投资,推动数据中心、云计算平台等新型信息基础设施的升级与整合,为区域提供了强大、普惠的算力与网络支持,奠定了创新的“技术公地”。软环境优化方面,政策的影响更为多维:首先,治理环境得以改善,政府利用大数据技术提升监管与服务的精准性,可营造更加透明、高效和公平的营商环境;其次,法治环境得以强化,试验区作为制度创新的“试验田”,往往率先探索数据产权界定、数据安全、隐私保护与知识产权快速维权等法律法规,为数据要素的流通与价值化提供了坚实的法律保障,显著

降低了创新成果被侵权的风险^[11];最后,经济环境得以稳固,数字经济本身带来的新税源增强了地方财政实力,使政府有能力持续增加公共研发投入与高质量公共服务供给,形成“产业发展-财政增收-创新反哺”的良性循环^[23]。这种软硬环境的协同升级,能系统性降低区域整体的创新摩擦,增强创新生态的韧性与活力。

基于此,提出假设4:试验区设立通过优化创新环境提高区域创新能力。

二、研究设计

1. 变量说明

(1) 被解释变量——区域创新系统能力

创新投入和创新产出是衡量创新水平的关键指标,其中,采用研发支出和研发人员数量等投入指标来衡量创新能力虽具有一定的合理性,但存在重复计算和数据可获得性差两大难题。创新产出能够更为直观地展现一个地区最终的创新成果,同时也反映了该地区的综合创新能力。在现有研究中,专利数量是衡量创新产出的关键指标,包括专利申请数量和专利授权数量。通常而言,专利的存续期限越长,其私人价值往往越高。单纯依赖专利数量来评估创新能力,可能会存在不准确、不合理的问题。因此,借鉴王春杨等的方法^[24],采用《中国城市和产业创新力报告》中的创新指数衡量区域创新能力。该指数基于微观发明专利授权数据,结合专利年费结构,构建价值加权系数。通过城市层面的加权计算,最终获得2001—2021年城市创新指数,该指数能够有效克服专利质量异质性带来的测量偏差。

(2) 核心解释变量

将试验区的设立作为一项准自然实验,以地区是否设立试验区作为核心解释变量:若城市 i 在第 t 年设立为试验区,则在当年及之后年份赋值为1,否则为0。

(3) 中介变量

创新主体、创新资源和创新环境为中介变量。具体而言,在数据可获取的条件下,基于熵值法,从技术创新主体和知识创新主体两个维度计算创新主体中介变量;从人力资源和财力资源两个维度计算创新资源中介变量;从市场信息技

术环境和政府经济政策环境计算创新环境中介变量。详细计算过程见机制分析部分。

(4)控制变量

为缓解遗漏变量问题对估计结果产生的影响,参考豆建民等的做法^[25],对可能影响区域创新能力的变量进行控制。具体的控制变量包括:财政规模(财政支出占 GDP 的比重)、经济发展水平(人均 GDP 的对数)、金融发展水平(金融机构存贷款余额占 GDP 的比重)、信息发展水平(国际互联网用户数的对数值)、城市产业结构(第三产业增加值占第二产业增加值的比重)、交通运输水平(人均公路货运量的对数值)、对外开放水平(进出口总额的对数值)及教育投入水平(教育支出的对数值)。

2. 模型构建

(1)基准模型构建

考虑试验区设立是分两批扩大试点城市范围的,为科学评估试点政策对区域创新能力的影响效应,构建如下回归模型:

$$Innovation_{it} = \alpha_0 + \alpha_1 Did_{it} + \alpha_2 Control + \delta_i + \mu_t + \tau_p \mu_t + \varepsilon_{it} \quad (1)$$

式中: $Innovation_{it}$ 为城市 i 在第 t 年的创新能力; Did_{it} 为是否设立试验区,若城市 i 在第 t 年设立了试验区,则 Did_{it} 取值为 1,否则取值为 0; $Control$ 为控制变量集合; δ_i 为城市固定效应; μ_t 为年份固定效应; $\tau_p \mu_t$ 为省固定效应; ε_{it} 为随机扰动项;估计系数 α_1 度量了在试验区设立前后区域创新能力的平均差异。

(2)中介效应模型

为进一步检验试验区设立影响区域创新能力的作用机制,参考江艇的做法^[26],在式(1)的基础上进一步探究可能存在的中介效应。构建的中介效应模型如下:

$$Mediation_{it} = \beta_0 + \beta_1 Did_{it} + \beta_2 Control + \delta_i + \mu_t + \tau_p \mu_t + \varepsilon_{it} \quad (2)$$

式中: $Mediation_{it}$ 为中介变量,依次代入创新主体、创新资源和创新环境。其余符号所代表的含义与式(1)相同,若 β_1 显著,说明中介效应成立。

(3)调节效应模型

为分析知识产权保护在试验区对区域创新能力影响中发挥的外部调节效应,在基准模型

中加入知识产权保护指数与试验区的交互项,回归模型如下:

$$Innovation_{it} = \alpha_0 + \alpha_1 Did_{it} + \gamma_0 IPP_{it} + \gamma_1 Did_{it} \times IPP_{it} + \alpha_2 Control + \delta_i + \mu_t + \tau_p \mu_t + \varepsilon_{it} \quad (3)$$

式中: IPP_{it} 为衡量地区 i 第 t 年的知识产权保护指数。除主回归系数 α_1 的符号和显著性, γ_1 为知识产权保护水平如何影响试验区设立与区域创新能力之间的关系,若 γ_1 显著,且与 α_1 方向一致,表明知识产权保护强化了试验区设立对区域创新能力的影响;若与 α_1 方向相反,表明知识产权保护弱化了试验区设立对区域创新能力的影响。

(4)双重机器学习模型

为保证政策效果估计的准确性,并为基准回归结果提供来自更灵活计量框架的稳健性支持,借鉴张涛等的研究^[27],采用双重机器学习模型(DML)进行辅助性估计^①。构建的部分线性 DML 模型如下:

$$Innovation_{it+1} = \theta_0 Did_{it} + g(X_{it}) + U_{it} \quad (4)$$

$$E(U_{it} | Did_{it}, X_{it}) = 0 \quad (5)$$

式中: $Innovation_{it+1}$ 为被解释变量,即区域创新能力; X_{it} 为高维控制变量集合; U_{it} 为误差项,条件均值为 0; θ_0 为处置系数。

3. 数据来源

数据主要来源于《中国城市统计年鉴》《中国科技统计年鉴》,试验区设立和数字经济政策变量来源于中国政府网及各城市政府工作报告,部分变量补充自地方统计局。在获取原始数据的基础上,主要进行如下处理:一是对于存在少数样本缺失值的控制变量和中介变量,采用线性插值法补齐;二是为构建平衡面板数据,剔除数据缺失严重的地区。最终得到 2007—2021 年度 281 个城市层面共 4 215 个面板数据(表 1)。

① 需要明确的是,DML 的核心优势在于其能够通过交叉拟合和机器学习算法有效控制高维控制变量带来的混淆偏差并缓解函数形式误设的风险,但是并不能从根本上解决多时点 DID 中因异质性处理效应或动态处理效应可能导致的估计偏误,后者仍需依赖于平行趋势假设的满足以及事件研究法等检验。因此,本研究将 DML 的定位明确为:在基准双向固定效应模型之外,提供一项重要的稳健性检验,目的在于证明,即使在考虑复杂的非线性关系和大量潜在混淆因素后,试验区政策对区域创新能力的促进效应依然存在,从而增强主要研究结论的稳健性与可信度,而非取代基准的 TWFE 估计。

表 1 描述性统计

变量	样本数量	平均值	标准差	最小值	最大值
区域创新能力	4 215	0.208 4	1.041 8	0.000 0	29.078 5
试验区设立	4 215	0.096 3	0.295 1	0.000 0	1.000 0
金融发展水平	4 215	2.358 6	1.183 4	0.560 0	21.301 5
经济发展水平	4 215	10.553 0	0.687 3	4.595 1	13.055 7
财政规模	4 215	0.188 6	0.102 7	0.042 6	1.485 2
信息发展水平	4 215	13.158 9	1.124 5	5.468 1	17.761 7
城市产业结构	4 215	0.987 9	0.564 6	0.094 3	5.348 2
交通运输水平	4 215	3.020 7	0.779 4	0.184 4	9.191 3
对外开放水平	4 215	7.296 7	2.167 1	-3.722 0	13.351 3
教育投入水平	4 215	12.900 2	0.901 5	9.282 8	16.256 0

表 1 是描述性统计分析,其中区域创新能力的平均值为 0.208 4,标准差为 1.041 8,最大值与最小值相差为 29.078 5。由此可知,样本数据中大部分地区的创新能力基本处于较低水平,且地区之间存在较大差异,表明当前区域创新能力仍然有很大提升空间。

三、实证结果分析

1. 基准回归分析

采用多时点双重差分模型估计试验区对区域创新能力的直接影响,表 2 报告了试验区设立对区域创新能力影响效应的回归结果。其中,列(1)是未考虑控制变量及固定效应的估

计结果,列(2)在此基础上进一步控制了城市、年份及省份年份联合固定效应。列(3)与列(4)是在列(1)与列(2)的基础上,添加一系列控制变量后的估计结果。由列(4)可知,试验区的回归系数为 0.684 3 且在 1%水平上显著,表明试验区设立对区域创新能力产生了显著的正向影响。从其经济显著性来看,相比于未设立试验区地区,试验区每设立一年,平均提升所在区域创新能力为所有样本均值的 3.283 6 倍,具有较强的经济显著性。此外,财政规模系数为负,可能反映了政府支出对市场研发活动的“挤出效应”;经济发展水平系数显著为负,可能意味着在控制了创新要素后,单纯的经济

表 2 基准回归结果

变量	(1)	(2)	(3)	(4)
试验区设立(是=1)	0.697 2*** (13.074 9)	0.729 3*** (14.077 1)	0.396 2* (1.864 6)	0.684 3*** (9.666 6)
金融发展水平			0.045 7 (0.874 5)	0.075 1 (0.929 4)
经济发展水平			(0.080 3)	-0.461 4*** (-2.730 9)
财政规模			-2.326 2*** (-2.762 4)	-1.394 2** (-2.012 1)
信息发展水平			-0.350 3** (-2.185 2)	-0.157 4*** (-3.005 5)
城市产业结构			0.506 1* (1.825 9)	0.096 8 (1.335 9)
交通运输水平			-0.087 9** (-2.209 8)	(0.055 7) (-1.279 8)
对外开放水平			(0.004 4)	(0.008 1)
教育投入水平			(-0.257 6)	(-0.553 7)
			0.701 2*** (3.485 3)	0.907 4*** (2.621 9)
常数项	0.141 3*** (8.535 5)	-26.830 0*** (-434.069 7)	-3.291 2*** (-5.117 5)	-30.606 6*** (-10.330 9)
固定效应	YES	YES	YES	YES
样本量	4 215	4 215	4 215	4 215
R-squared	0.039 0	0.741 1	0.157 9	0.772 7

注:***、**、* 分别表示在 1%、5%、10%的水平上显著;固定效应包括城市固定效应、年份固定效应和省份年份联合固定效应;括号内为 t 统计值;采用城市层面的聚类标准误。下同。

总量对创新效率的边际贡献下降。但总的来说,无论在何种情形下,试验区的回归系数均显著为正,这在一定程度上表明了试验区设立显著提升了区域创新能力,即假设 1 成立。

2. 稳健性检验

(1) 动态效应分析

基准回归采用多时点 DID 模型的前提在于实验组与对照组在政策发生前保持一致的变化趋势,即满足平行趋势检验假设。参考白俊红等的做法^[28],构造式(6)进行平行趋势检验:

$$\begin{aligned} Innovation_{i,t} = & \alpha + \sum_{j=1}^7 \beta_j Before(9-j)_{i,t} + \\ & \beta_{10} current_{i,t} + \beta_{11} After1_{i,t} + \\ & \beta_{12} After2_{i,t} + \beta_{13} After3_{i,t} + \\ & \beta_{14} After4_{i,t} + \beta_{15} After5_{i,t} + \\ & \gamma Control + \delta_i + \mu_t + \tau_j \mu_t + \varepsilon_{i,t} \end{aligned} \tag{6}$$

式中:时间虚拟变量为各城市批准设立试验区的前 n 年、当年和后 n 年的观测值。非试点城市的虚拟变量均为 0。由于样本期间为 2007—2021 年,而试验区的设立时间分别为 2015 年和 2016 年,即存在部分试点城市没有多于 5 期和-8 的样本值,因此将样本中多于 5 期的数据归并至第 5 期,同时将多于-8 期的数据归并至第-8 期。为避免多重共线性,剔除政策实施前一年的时间虚拟变量。从图 1 可知,在试验区设立之前,回归系数均不显著且数值较小,表明实验组与对照组在区域创新能力上无显著差异。在试验区设立后,试验区设立的影响系数显著为正并不断提升,表明试验区设立能够产生促进区域创新能力持续提升的政策效应。

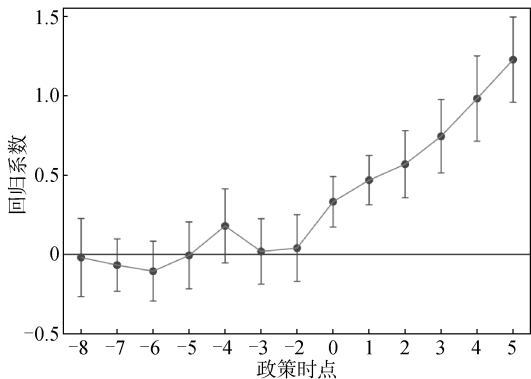


图 1 平行趋势检验

(2) 安慰剂检验

考虑到试验区设立的过程中可能会受到其他未能控制的随机因素的影响,基准回归结果可能存在一定误差,为确保区域创新能力的提升是由设立试验区这一核心因素所驱动,而非受其他未观测到的随机因素的影响,进一步采用安慰剂检验对试验区设立效果的偶然性加以识别,得到估计的 1 000 个虚拟估计系数^[29]。图 2 展示了安慰剂检验中解释变量估计系数和 P 值的核密度分布图。从图中可以看出,抽样回归的估计系数近似均值为 0 的正态分布,且所有系数均小于基准回归系数 0.684 3,大多数 P 值超过 0.10,在 10% 的水平上并不显著。上述结果表明,试验区设立提高区域创新能力的结论并非随机性因素导致。

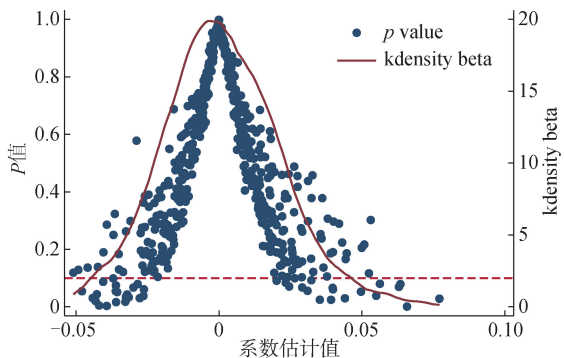


图 2 安慰剂检验

(3) PSM-DID

考虑到在试验区设立时可能会选择经济发展前景较好、创新要素集中的地区,且试验区设立并非严格意义上的自然实验,这可能导致基准估计存在样本选择问题。为缓解这一问题的干扰,运用 PSM-DID 模型进行稳健性检验,即分别采用面板数据转化法和逐期匹配法进行倾向得分匹配分析。图 3 报告了在截面 PSM 和逐年 PSM 方法下,实验组和对照组匹配前后的结果。如图所示,两种匹配方法均显著缩小了均值线的差距,两条曲线更加接近,这表明截面 PSM 和逐年 PSM 均有效降低了样本选择偏差。

表 3 的列(1)和列(2)分别展示了采用两种不同方法下的多时点 PSM-DID 回归结果。结果显示,其系数依然显著为正,且与基准回归结果相比并未出现实质性差异,这也从侧面验证了试验区设立非随机对研究结论产生的干扰有限。

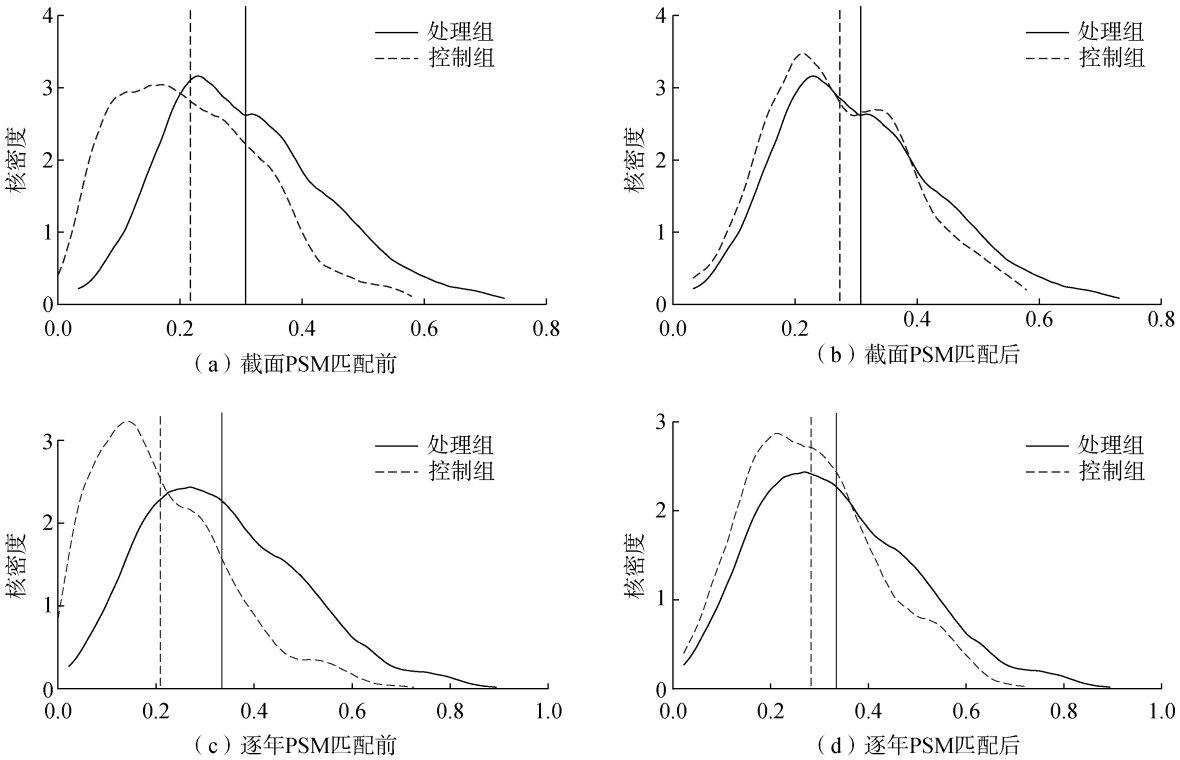


图3 PSM匹配前后核密度分布图

表3 截面和逐年PSM-DID及其他稳健性检验

变量	区域创新能力			专利申请数	专利授权数
	(1)	(2)	(3)	(2)	(3)
试验区设立(是=1)	0.645 3*** (9.708 9)	0.643 6*** (10.525 6)	0.684 3*** (9.666 6)	0.000 5*** (8.046 1)	0.000 6*** (6.788 8)
样本量	4 215	4 215	4 215	4 215	4 215
R-squared	0.598 6	0.750 8	0.772 7	0.570 1	0.419 4
控制变量	YES	YES	YES	YES	YES
固定效应	YES	YES	YES	YES	YES

(4)其他稳健性检验

①更换模型设定。试验区在全国范围内是渐进式推广的,2016年是该政策推广的峰值,只有贵州省是在2015年设立的试验区,因此参考相关研究的做法^[13],选取2016年作为政策的起始年份,使用单期DID重新估计,表3列(3)结果显示更换模型后结果与基准回归基本一致。

②替换区域创新能力的衡量指标。如前文所述,区域创新能力有多种衡量方式,每一种方式都存在优势和劣势,为此参考余泳泽等的做法^[30],选取每万人专利申请数量和每万人专利授权数量作为区域创新能力的替代指标,重新

对基准回归进行估计。表3列(4)(5)的结果显示,试验区仍然能够显著提高区域创新能力,即研究结论受被解释变量样本选择的影响有限。

③更换估计模型。在评估政策效果时,传统的双向固定效应模型面临一定的局限性,使用双重机器学习模型,提供一种在更灵活的机器学习框架下的稳健性检验,以增强主要结论的可信度^[27]。借鉴韩先锋等的研究^[31],本研究选择1:4的样本分割比例。表4列(1)为采用双重机器学习模型的回归结果,核心解释变量的系数仍显著为正,表明高维混淆变量与模型设定偏差对估计结论影响有限。

表 4 其他稳健性检验

变量	(1)	(2)	(3)	(4)	(5)
试验区设立(是=1)	0.055 8** (2.198 7)	0.682 2*** (9.587 5)	0.500 4*** (9.542 3)	0.684 2*** (9.600 6)	0.666 0*** (12.729 1)
智慧城市试点		-0.036 1 (-1.010 3)			
城市数量	281	281	281	276	281
控制变量	YES	YES	YES	YES	YES
城市固定效应	YES	YES	YES	YES	YES
样本量	4 215	4 215	3 653	4 140	4 215
R-squared		0.772 8	0.754 6	0.772 9	0.542 9

④排除其他政策影响。在试验区设立区间,2012年后分三批设立的智慧城市试点政策与本研究内容密切相关。为避免样本期间内由于相似政策干扰所产生的结果偏误,借鉴白俊红等的方法^[28],在基准回归模型中加入智慧城市试点政策的虚拟变量,重新对式(1)进行估计。表4列(2)中试验区设立的系数仍然显著且智慧城市试点的系数并不显著,表明智慧城市试点并不影响研究结论。

⑤排除重大事件干扰。为避免在样本期间所发生的重大事件影响区域创新能力,导致基准结果出现偏误,本研究删除2020年和2021年的样本数据,并重新估计。表4列(3)所示,在剔除该样本数据后,回归结果依然在1%水平上显著为正,表明估计结果依然保持稳健。

⑥剔除偏远城市。鉴于部分偏远城市在样本期间并未设立试验区,将其纳入控制组可能会增加样本异质性,进而影响结果的准确性^[32],因此剔除新疆、青海、西藏和海南等地区的样本,重新加以检验。表4列(4)的回归结果表明,基准回归的结论基本不受存在偏远城市的影响。

⑦缩尾处理。为排除数据中极端值的影响,对所有连续变量进行了上下1%的缩尾处理,表4列(5)显示试验区设立对区域创新能力的影响显著为正,只是系数变小,主要结论并未发生改变。

⑧异质性处理效应。Chaisemartin等指出,多时点DID模型可能因“异质性处理效应”导致估计偏差^[33],因此采用Stata的twowayfweights命令进行检验。结果显示,406个权重均为正值,总和

接近1(0.9999),这在一定程度上表明异质性处理效应未对基准回归结果产生显著影响,研究结论具有稳健性。

四、机制分析

基于前文理论分析,试验区设立通过增加创新主体、集聚创新资源和优化创新环境提高区域创新能力。为更好地定量分析试验区设立对区域创新能力的影响,从创新主体、创新资源和创新环境3个角度揭示其作用机制。具体而言,基于区域创新生态系统的要素构成,选取创新主体、创新资源和创新环境作为中介变量。同时,根据试验区设立对创新主体、创新资源和创新环境的影响,选取相应指标构建创新主体、创新资源和创新环境3个中介变量。具体中介变量计算方法留存备查。

1. 中介变量计算方法及定义

(1) 计算方法

熵值法是一种基于信息熵理论确定指标权重的研究方法,一般有6个步骤。若有 n 个样本, m 个指标,则 x_{ij} 代表第 i 个样本的第 j 个指标的数值($i=1,2,\cdots,n,j=1,2,\cdots,m$)。具体步骤如下:

第一步,对原始数据进行归一化处理,如式(7)所示。

$$x'_{ij} = \begin{cases} \frac{x_{ij}-x_{min}}{x_{max}-x_{min}}, j \text{ 为正向指标} \\ \frac{x_{max}-x_{ij}}{x_{max}-x_{min}}, j \text{ 为负向指标} \end{cases} \quad (7)$$

式中: x_{ij} 为 i 地区第 j 项指标除以常住人口后的数值; x'_{ij} 为 x_{ij} 标准化后的值; x_{max} 、 x_{min} 分别为 i 地区第 j 项指标的最大值、最小值,其中

$0 < i \leq n, 0 < j \leq m$ 。

第二步,计算第 j 项指标下第 i 个样本值占该指标的比重。

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^n x'_{ij}} \tag{8}$$

第三步,计算第 j 项指标的熵值。

$$e_j = -k \sum_{i=1}^n p_{ij} \ln p_{ij}, j = 1, 2 \cdots m \tag{9}$$

式中: $k > 0, k = \frac{1}{\ln n}$, n 为观测样本个数, $0 \leq e_j \leq 1$ 。

第四步,计算第 j 项指标信息效用值。

$$d_j = 1 - e_j \tag{10}$$

第五步,计算第 j 项指标的权重。

$$w_j = d_j / \sum_{j=1}^m d_j \tag{11}$$

第六步,计算第 i 个样本的综合得分 S_i 。

$$S_i = \sum_{j=1}^m w_j p_{ij} \tag{12}$$

(2) 创新主体

创新主体是创新生态系统中具有创新驱动功能的个体集合,主要包括以企业为代表的技术创新主体及高校、科研机构等知识创新主体^[34-36]。技术创新主体是新技术、新产品及新工艺的孕育者,其核心职责在于学习技术、推动革新、实现创造并传播技术知识,企业是承担这些职责的主要载体^[37]。知识创新主体是新知识、新思想的创造者,核心作用在于生产、传播和整合知识,以高校和研发机构为核心载体^[38]。按照创新主体的内涵,并考虑到数据的可获得性,选取创新主体的具体测量指标如表 5 所示:

表 5 创新主体测量指标

变量维度	测量指标
技术创新主体-企业	当年企业注册数量
知识创新主体-高等院校	普通高等学校数量
	普通高等学校专任教师数量

(3) 创新资源

创新资源是创新生态系统中支撑创新活动的关键要素,主要包括作为知识传播媒介的人力资本和开展研发活动所需的研发投入^[39],两者紧密相连、相辅相成,是区域创新活动中最为

关键的创新资源^[40]。从人力资源和财力资源两个维度衡量创新资源,并参考白俊红等的研究^[28],设计的创新资源的具体测量指标如表 6 所示:

表 6 创新资源测量指标

变量维度	测量指标
人力资源	科研技术服务和地质勘查业从业人员数量
	信息传输计算机服务和软件业从业人员数量
财力资源	科学技术支出
	RD 内部经费支出

(4) 创新环境

根据区域创新生态系统理论,除创新主体和创新资源,创新环境是试验区设立影响区域创新能力的重要因素。狭义的创新环境主要包括政府政策、市场环境及文化环境等,其与创新资源存在显著差异。在大数据飞速发展的背景下,创新环境包括大数据发展所需要的基础设施等条件,即硬环境。同时,大数据发展还需要具备良好且稳定的软环境,如政府财政收入、经济政策环境、市场环境、社会文化环境等。从市场信息技术环境和政府政策经济环境两个维度来构建创新环境变量:首先,信息技术环境指标选取参考赵涛等的做法^[41],采用移动电话用户数量和电信业务收入来衡量。其次,区域创新能力的提高需要政府持续增加创新投入^[42],财政收入水平决定了政府提供公共服务的规模、质量和效率^[43],因此用地方政府财政收入衡量地区政府为创新活动提供服务的能力。为衡量政府政策环境,参考孙伟增等的研究^[11],获取各城市政府工作报告中每年出现的与数字经济政策相关词频数量衡量数字经济相关政策数量。创新环境的具体测量指标如表 7 所示:

表 7 创新环境测量指标

变量维度	测量指标
市场信息技术环境	移动电话用户数量
	电信业务收入
政府经济政策环境	地方政府财政收入
	数字经济政策数量

2. 创新主体增加效应

根据前文理论分析发现,试验区设立具有示范和引领作用,能够吸引外部创新主体的加

人,同时通过降低创新成本、获得政府支持鼓励潜在创新主体升级为创新主体。进一步地,创新主体的扩大能够产生知识溢出效应,提高创新主体的知识经验学习能力,从而提高区域创新能力。因此,本研究通过实证检验试验区设立通过创新主体增加效应提高区域创新能力,主要利用式(2)进行机制检验。

表8报告了创新主体机制检验的回归结果。列(1)展示了试验区设立对创新主体的回归结果,核心解释变量 *Did* 的系数在1%的水平上显著为正,表明试验区设立能够增加创新主体规模。列(2)验证了创新主体对区域创新能力的影响,回归系数显著为正,表明创新主体规模越大,区域创新能力越高,即创新主体的扩大有利于提高区域创新能力。因此可以得出结论,试验区设立能够通过增加创新主体来提高区域创新能力,假设2成立。

3. 创新资源集聚效应

经过前文分析发现,试验区的设立会增加地区对人力资本的需求,同时大数据发展能够创造出更多的工作机会并提供更好的薪资待遇,提高人才供给。此外,大数据发展能够帮助创新主体精准决策,降低研发投入风险。综上所述,试验区设立能够从人力资本和研发投入两个方面集聚创新资源。根据前文对创新资源变量的测度,利用中介效应模型进行机制检验。由列(3)可知,核心解释变量的回归系数显著为正,表明试验区的设立能够集聚创新资源。

列(4)进一步验证了创新资源对区域创新能力的影响。*Resource* 的回归系数显著为正,表明创新资源集聚程度越高,区域创新能力越高。由此可知,创新资源集聚效应是试验区设立产生区域创新效应的重要机制之一,假设3成立。

4. 创新环境优化效应

除了通过增加创新主体和集聚创新资源两种途径来影响区域创新能力,试验区设立对创新环境的影响也极为重要。设立试验区不仅能够提升当地的大数据基础设施水平、整合相关基础设施、改善信息技术环境,还能通过引进大数据技术、提高区域数据分析能力,改善政府经济政策环境,从而提高区域创新能力。因此,根据前文对创新环境的测量,通过式(2)进行创新环境机制检验。列(5)核心解释变量的系数显著为正,说明试验区设立有利于创新环境的优化。列(6)进一步验证了创新环境优化对区域创新能力的影响,创新环境的回归系数显著为正,说明创新环境得到优化后,区域创新能力显著增强,即试验区设立确实能够通过优化创新环境提高区域创新能力,假设4成立。

最后,将3个中介变量一同纳入回归模型进行并行检验,结果如列(7)所示。创新资源集聚和创新环境优化的系数在1%的水平上依然高度显著,核心解释变量的系数虽然有所减小,但依然在1%的水平上显著,这表明政策的总效应有相当一部分确实被这两个强劲的中介

表 8 机制效应分析

变量	创新主体增加效应		创新资源集聚效应		创新环境优化效应		
	(1)创新主体	(2)区域创新能力	(3)创新资源	(4)区域创新能力	(5)创新环境	(6)区域创新能力	(7)区域创新能力
试验区设立(是=1)	0.015 5 *** (5.161 7)		0.005 6 * (1.735 3)		0.047 4 *** (13.657 1)		0.488 6 *** (5.280 4)
创新主体		6.453 7 *** (2.626 2)					3.946 4 (1.371 6)
创新资源				5.247 8 *** (4.444 3)			4.316 9 *** (3.143 8)
创新环境						4.572 4 *** (5.545 0)	2.689 4 *** (3.340 7)
控制变量	YES	YES	YES	YES	YES	YES	YES
城市固定效应	YES	YES	YES	YES	YES	YES	YES
样本量	4 215	4 215	4 215	4 215	4 215	4 215	YES
R-squared	0.476 4	0.787 9	0.469 0	0.808 7	0.474 3	0.793 2	0.805 6

变量所解释;创新主体增加的系数变得不显著。由此表明,“创新资源集聚”和“创新环境优化”是两条独立、稳健且核心的作用机制,即使在控制了其他机制后,其对创新产出的直接促进效应依然非常强劲和显著。对于“创新主体增加”效应在并行检验中不显著,可能的原因在于政策首先吸引了创新主体,而这些主体的聚集直接引发了“创新资源集聚”和“创新环境优化”这两个更直接、更终端的效应,因此,创新主体增加的效应在很大程度上被创新资源集聚和创新环境优化所中介或掩盖。

五、进一步讨论

1. 异质性分析

(1)基于数字经济发展水平的异质性分析
数字经济正逐渐成为激发区域创新活力、驱动经济社会向高质量迈进的关键力量和新引擎,现有研究表明,数字经济的快速发展能够显著推动我国区域创新能力的提升^[44]。参考张柳钦等的研究^[45],根据《中国数字经济发展与就业白皮书(2019)》中各省市数字经济的总量排名,将排名前半的省市归为高水平的数字经济发展水平组,而其他省市则归为低水平的数字经济发展水平组。表 9 列(1)(2)报告了试验区设立对不同数字经济发展水平地区创新能力的估计结果,可以发现,相较于低数字经济发展水平地区而言,试验区对高数字经济发展水平地区创新能力提升的促进作用更加显著。这可能是因为数字经济发达的地区具有数据资源集聚的先天优势,并通过试验区设立得到了强化,进一步促进了区域创新能力的提升。

(2)基于市场化水平的异质性分析

试验区设立在提高区域创新能力的影

响方面可能会因地区市场化水平的不同而具有显著差异,因此从市场化水平方面来检验试验区设立对区域创新能力的差异性影响效果。具体来说,以樊纲等编制的市场化指数为基础,根据 2016 年市场化指数的中位数将地区划分为两类,高于或等于中位数的地区视为高市场化水平地区,而低于中位数的地区则视为低市场化水平地区。表 9 列(3)(4)报告了在不同市场化水平下,设立试验区对区域创新能力的影响。由估计系数可知,试验区设立对市场化水平较高地区创新能力的正向促进作用相较于低市场化水平地区更大,且对低市场化水平地区的促进作用不显著。这可能是因为非市场因素的过度干预会阻碍数据要素的集聚过程,进而影响数据资源的配置效率。在这种情境下,试验区内数据要素的活跃度会降低,数据要素的流动与交易也会受到阻碍,从而难以有效推动区域创新能力的提升,因此试验区设立对市场化水平较低地区的创新效应不明显。

(3)基于区位特征的异质性分析

本研究基准回归中样本数据的划分主要依据是行政区域,为进一步探究试验区设立是否会因城市地理位置特性的差异而产生不同的创新效应,进行区位特征异质性分析。根据《关于明确东中西部地区划分的意见》的划分标准,将样本分为东部、中部和西部 3 个区域子样本,并分别进行分组回归分析。区位特征异质性结果如表 9 列(5)(6)(7)所示,可以看出,东部地区的核心解释变量在 1%的统计水平上显著为正,而中部和西部地区城市的核心解释变量并不显著,结果表明相较于中西部地区,试验区对东部地区城市创新能力具有显著的促进作用。从地理角度来看,东中西三大区域的经济

发展水平、要素禀赋、经济结构和生产环境都

表 9 异质性检验

变量	(1)高数字经济	(2)低数字经济	(3)高市场化	(4)低市场化	(5)东部	(6)中部	(7)西部
试验区设立(是=1)	0.598 4*** (3.549 8)	-0.023 2 (-0.146 7)	0.592 3*** (3.441 2)	0.018 7 (0.271 3)	0.613 1*** (3.117 9)	-0.025 7 (-0.300 8)	-0.038 3 (-0.469 7)
控制变量	YES	YES	YES	YES	YES	YES	YES
城市固定效应	YES	YES	YES	YES	YES	YES	YES
样本量	2 580	1 635	2 595	1 620	1 485	1 485	1 245
R-squared	0.790 7	0.493 8	0.791 6	0.312 1	0.823 7	0.314 2	0.432 5

存在明显的异质性,一般认为东部经济发展水平和资源禀赋最高、中部次之、西部最低,因此由于东部地区的地理位置优势,试验区设立对其创新能力的促进作用显著高于中西部地区。

2. 调节效应分析

大数据技术显著提升了信息交互和知识扩散的效率。然而,与大多数创新成果类似,大数据领域的创新同样面临知识产权侵权的风险。在数字化环境下,创新成果被非法侵占将给创新主体带来更为严重的损失^[18]。在大数据发展背景下,当知识产权保护较弱时,创新成果更容易被模仿甚至侵占,数据资源和数据产品可能更容易受到非法获取和面临泄露的风险,创新者的权益无法得到充分保障,从而会降低创新者的创新动力。借鉴周泽将等的做法^[46],将原始的知识产权保护指数除以 100,再从中减去相应年度的中位数,所得结果作为衡量地区知识产权保护水平的指标。在基准回归上加入知识产权保护指数和试验区设立的交互项,使用式(3)回归得到相应结果。表 10 列(2)显示知识产权保护与区域创新能力的交互项系数估计值为 5.783 8,在 1%的统计水平上显著。具体含义是指当知识产权保护水平提高时,试验区设立对区域创新能力的促进作用更加显著,这意味着在知识产权保护增强的时候,“搭便车”和知识产权侵权现象减少,创新成果得到保护,创新主体的利益得到保障,进而提高了创新主体进行研发创新的积极性。换言之,知识产权保护通过营造良好的创新环境,增强了试验区对区域创新能力的提升作用。

表 10 知识产权保护的调节效应

变量	(1)	(2)
试验区设立(是=1)	0.684 3*** (9.666 6)	0.765 5*** (10.269 4)
IPP		-43.348 7*** (-73.722 1)
Did×IPP		5.783 8*** (10.713 1)
控制变量	YES	YES
城市固定效应	YES	YES
样本量	4 215	4 215
R-squared	0.772 7	0.772 7

3. 空间效应分析

前文的分析证实了试验区设立对区域创新能力的影响,然而对于城市之间空间相关性的忽视会导致估计结果的偏误^[47],因此进一步探究试验区设立对本地区及邻近地区的影响,即试验区设立在实施过程中的空间效应。

(1)空间效应理论分析

经济全球化的深化推动了跨区域创新合作的发展,区域知识结构理论为此背景下新经济地理学视角的区域创新增长研究提供了新分析框架。该理论提出,区域创新能力的提升是在既有知识基础上动态生成新知识的过程,既包含本地知识的开发利用,也涉及区域内的知识溢出效应。从理论层面看,依据区域创新生态系统理论,大数据发展可通过增加创新主体、集聚创新资源与优化创新环境影响区域创新。而新经济地理学指出,在试验区设立情境中,大数据发展因规模经济的向心力可能产生虹吸作用,又因运输成本、土地租金等离心力可能产生空间溢出效应。

就空间溢出效应而言,空间知识溢出视角下,数据资源集聚中的外部知识溢出、学习效应与吸收能力是创新能力提升的重要渠道^[48]。传统集聚理论认为,当集聚收益低于成本时,经济体会出现发散效应。一方面,在创新主体合作机制上,大数据能打破信息壁垒,强化主体交流,邻近地区通过“干中学”提升创新能力;数字技术突破地理限制,降低沟通成本,为知识技术传播提供高效路径,产生空间溢出。另一方面,在创新资源流动机制上,数据的低扩散成本与高扩散速度使其流动性受地理约束小,地理接近的创新主体通过共享数据可增强空间溢出效应。

从虹吸作用来看,市场化改革与要素流动机制完善加速了新兴技术与实体经济的融合,城市集聚效应增强,生产要素与经济活动向核心区域集中,边际报酬递增,推动区域创新能力提升。在创新主体合作方面,大数据平台虽促进了跨区域协同创新,但高水平地区的主体更具优势,既能精准把握市场动态,又能借助试验区提高合作效率。在创新资源流动方面,新古典经济学认为要素自由流动会使区域经济短期集聚^[49];Krugman 的模型表明,低运输成本与

规模经济促使创新主体集聚^[50],其“中心-外围”理论指出此时集聚力量占优。此外,试验区集聚人力资本存在“马太效应”,会形成自我强化机制,导致区域创新能力“强者更强”。

因此,除前文所述试验区设立对区域创新能力的直接影响,值得深入研究的是,试验区设立在提升本地区域创新能力的同时是否会对邻近地区创新能力产生空间影响,这种影响究竟是空间溢出效应还是虹吸效应?为深入分析这一问题,从空间视角出发,利用空间杜宾模型检验试验区设立对区域创新能力的空间效应。

(2)空间计量模型构建

为检验试验区设立对区域创新能力的空间效应,参考曾艺等的研究^[51],依据经济地理距离矩阵设定的基本空间计量模型如下:

$$Innovation_{it} = \alpha_0 + \rho W \times Innovation_{it} + \beta_0 Did_{it} + \theta_0 W \times Did_{it} + \beta_1 Control + \theta_1 W \times Control + \delta_i + \mu_t + \varepsilon_{it} \quad (13)$$

式中: ρ 为空间自相关系数,表示本地创新能力对邻近地区创新能力的影 响; W 为 281×281 阶的空间权重矩阵; β_0 、 β_1 分别为核心解释变量和控制变量的系数; θ_0 、 θ_1 分别为核心解释变量及控制变量空间交互项的弹性系数; δ_i 、 μ_t 分别为个体效应与时间效应; ε_{it} 为误差项。

(3)空间计量估计结果分析

空间相关性即“地理学第一定律”,指出所有事物都具有空间关联性,且这种关联性与距离成反比。空间相关性主要来源于溢出效应、模仿行为等。空间统计学一般利用空间自相关系数来反映空间依赖性,莫兰指数是统计学中常用来反映空间自相关的指数。由于存在学习效应,不同地区的区域创新能力可能存在空间关联性,采用 2007—2021 年城市面板数据进行分析。在进行空

间计量分析之前,运用 Moran’s I 指数方法,基于经济距离矩阵对各年度的空间效应进行测算,对研究对象是否存在空间效应进行检验。相应年份的全局 Moran’s I 指数具体如表 11 所示。

由表 11 可知,2007—2021 年区域创新能力的莫兰指数均在 1%水平上显著为正,因而可以说明各城市的创新能力具有显著的正自相关性。为此,进一步选择空间计量模型分析试验区设立影响区域创新能力的空间特征。

为选择准确合适的空间面板模型,借鉴豆建民等检验思路进行系列检验^[25],检验结果如下:首先,LM 检验结果显示 LM-error、LM-lag 和 RLM-lag 均在 1%水平上显著,而 RLM-error 不显著,表明数据存在空间相关性,且更适合采用 SAR 模型。其次,LR 和 WALD 检验均显著,表明应使用空间杜宾模型(SDM)而非空间滞后模型(SAR)或空间误差模型(SEM)。接着,Hausman 检验在 1%水平上显著,表明固定效应模型优于随机效应模型。最后,LR 似然比检验显示时空双固定效应模型比单独的个体或时间固定效应更为合适。基于以上检验结果,最终选择时空双固定的 SDM 模型进行分析。

(4)空间计量估计结果分析

为了比较和检验变量参数的稳定性,在基准回归中列入时空双固定的三类模型的结果。如表 12 所示,SAR 和 SDM 模型的 ρ 显著为正,SEM 模型的 λ 也显著为正,说明区域创新能力具有正向的空间溢出效应。在 SDM 模型中,核心解释变量的系数及其空间交互项系数均显著为正,说明试验区设立具有显著的正向空间溢出效应。综上所述,试验区设立不仅能够提高本地区的创新能力,也有利于促进邻近地区创新能力的提升,具有显著的正向空间溢出效应。

表 11 Moran’s I 指数检验结果

年份	被解释变量	P 值	年份	被解释变量	P 值
2007	0.159	0.000	2015	0.150	0.000
2008	0.157	0.000	2016	0.155	0.000
2009	0.150	0.000	2017	0.158	0.000
2010	0.146	0.000	2018	0.160	0.000
2011	0.143	0.000	2019	0.159	0.000
2012	0.144	0.000	2020	0.158	0.000
2013	0.145	0.000	2021	0.159	0.000
2014	0.146	0.000			

传统点回归方法在分析区域间空间溢出效应时可能存在估计偏差,因此,采用变量变化的偏微分方法进行进一步分析。表 13 的效应分解结果显示,试验区设立对区域创新能力的直接效应、间接效应和总效应均在 1%水平上显著为正,从空间分解的量化系数还可看出,试验区设立对区域创新能力的空间溢出效应占总效应的 75%以上,甚至超过了直接效应,说明试验区设立的正外部性较强,能够通过显著的空间溢出效应扩散至邻近城市。究其原因,主要在于城市资源禀赋的异质性分布,不同城市在科技创新资源、产业基础、人才储备和政策环境等方面存在显著差异,这种差异为区域之间的要素流动和协同创新提供了内在动力。试验区设立会加速地区之间人才的流动与交流,对周边城市吸引并汇聚人才具有积极作用,进而带动邻近城市创新水平的提升,具有较好的空间辐射效应。

六、结 语

在大数据迅猛发展的背景下,充分利用试验区设立的优势是提高区域创新能力的有效途径,本研究深入探讨试验区设立对区域创新能力的影响,旨在揭示二者之间的内在联系和作用机制。主要研究结论如下:其一,试验区设立能够显著提高区域创新能力,且这一结论在经过一系列稳健性检验后依然成立。更为重要的

是,该政策的影响并非局限于本地,空间计量分析结果表明,试验区设立具有显著的正向空间溢出效应,其不仅促进了本地的创新活动,也带动了邻近地区创新能力的提升。其二,创新主体增加、创新资源集聚和创新环境优化是试验区设立影响区域创新能力的作用机制。其三,试验区设立的创新促进效应存在明显的异质性。试验区设立对数字经济发展水平较高地区、市场化程度较高地区和东部地区的促进作用更为显著,这表明政策的有效性依赖区域特定的经济基础与制度环境。其四,调节效应结果显示,知识产权保护水平在其中扮演着关键角色。当知识产权保护力度加强时,试验区设立对区域创新能力的促进作用更为强劲,这印证了完善的制度环境是确保政策红利充分释放的重要保障。

基于中国大数据飞速发展的现实背景,结合上述研究结论,提出如下对策建议:第一,坚持试点推广与经验扩散,强化“有效市场”与“有为政府”的协同。应稳步扩大试验区覆盖范围,并将成功经验进行系统梳理与标准化推广。政府工作的重点在于构建公平竞争的市场环境与稳定的政策预期,通过深化“放管服”改革降低制度性交易成本,让市场在创新资源配置中发挥决定性作用。第二,聚焦核心传导机制,实施精准化的政策干预。政策设计应直接对应于已验证的三大作用机制:在创新主体层

表 12 基准空间计量回归结果

变量	(1)空间自回归模型	(2)空间误差模型	(3)空间杜宾模型
试验区设立(是=1)	0.304 7*** (0.000 0)	0.278 2*** (0.000 0)	0.304 7*** (0.000 0)
空间权重×试验区设立(是=1)			0.542 8*** (0.000 0)
<i>rho</i>	0.438 1*** (0.000 0)		0.379 2*** (0.000 0)
<i>sigma2_e</i>	0.363 5*** (0.000 0)	0.365 3*** (0.000 0)	0.348 6*** (0.000 0)
<i>lambda</i>		0.446 4*** (0.000 0)	
样本量	4 215	4 215	4 215
R-squared	0.259	0.275	0.005
控制变量	Yes	Yes	Yes

表 13 空间溢出效应分解结果

变量	(1)直接效应	(2)间接效应	(3)总效应
试验区设立(是=1)	0.343 4*** (0.000)	1.031 9*** (0.000)	1.375 3*** (0.000)
控制变量	Yes	Yes	Yes
样本量	4 215	4 215	4 215
R-squared	0.005	0.005	0.005

面,需对中小企业、高校、科研机构实施差异化、精细化的激励方案;在创新资源层面,核心是深化数据要素市场化改革,以数据流高效牵引人才流与资金流;在创新环境层面,需坚持“硬基建”与“软制度”双轮驱动,特别是在知识产权保护、数据安全与隐私保障等领域进行前瞻性立法与监管。第三,实施区域差异化策略,推动全域创新网络形成。必须承认并顺应发展的异质性规律。对于基础较好的东部地区,应鼓励其瞄准全球前沿,承担国家重大科技任务;对于中西部等地区,则应支持其发展特色化、场景化的数据应用模式,并通过建立跨区域创新联盟、飞地经济等模式,主动承接先进地区的知识与技术溢出,将空间溢出效应从理论可能转化为现实增长。第四,构建开放协同的创新治理体系,保障创新成果的价值实现。政策的成功离不开开放包容的生态。政府应积极搭建产学研用协同平台,促进创新要素的跨区域、跨领域流动。同时,必须将知识产权保护置于核心战略位置,通过建立快速维权通道与提高侵权成本,切实保障创新者的合法权益,从而形成“创新—保护—再创新”的良性循环。

参考文献:

[1] 黎文靖,郑曼妮. 实质性创新还是策略性创新?——宏观产业政策对微观企业创新的影响[J]. 经济研究,2016,51(4):60-73.

[2] 杜梅,任声策,操友根. 中国创新能力发展趋势、困境、差距与政策建议[J]. 科研管理,2024,45(5):1-10.

[3] 臧培华,金磊. 跨国并购与自主创新:摆脱创新双低困境的理论机制与战略选择[J]. 上海经济研究,2023(11):55-70.

[4] 李健,李程骅. 新质生产力背景下区域创新生态系统提升产业链韧性的机理与路径[J]. 河海大学学报(哲学社会科学版),2025,27(6):156-168.

[5] 中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要[N]. 人民日报,2021-03-13(01).

[6] 中共中央关于制定国民经济和社会发展第十五个五年规划的建议[N]. 人民日报,2025-10-29(01).

[7] 谷业凯. 坚持创新驱动,加紧培育壮大新动能——着力推进全年经济工作八大重点任务②

[N]. 人民日报,2026-01-06(05).

[8] 中华人民共和国国务院. 促进大数据发展行动纲要[EB/OL]. (2015-08-31). http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[9] 刘明. 大数据综合试验区对企业创新的影响研究——来自中国上市公司的经验证据[J]. 云南大学学报(社会科学版),2024,23(1):79-91.

[10] 陈蕊,王宏伟. 大数据发展与企业创新能力提升[J]. 当代经济管理,2024,46(5):30-42.

[11] 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(9):117-135.

[12] 史丹,郑玉. 数据要素的赋能机制与企业全要素生产率提升——来自国家级大数据综合试验区的证据[J]. 改革,2024(11):1-16.

[13] 史丹,孙光林. 大数据发展对制造业企业全要素生产率的影响机理研究[J]. 财贸经济,2022,43(9):85-100.

[14] 邱子迅,周亚虹. 数字经济发展与地区全要素生产率——基于国家级大数据综合试验区的分析[J]. 财经研究,2021,47(7):4-17.

[15] 张志新,徐世超,高惠楠. 大数据发展能否推动企业绿色技术创新“质效并举”——基于“国家大数据综合试验区”的准自然实验[J]. 当代经济研究,2024(4):103-115.

[16] LYYTINEN K, YOO Y J, BOLAND J R Jr. Digital product innovation within four classes of innovation networks[J]. Information Systems Journal, 2016, 26:47-75.

[17] 张慧,易金彪,徐建新. 数字化变革如何影响城市创新——基于国家大数据综合试验区建设的经验证据[J]. 科学学研究,2023,41(8):1484-1494.

[18] 任英华,刘宇钊,胡宗义,等. 大数据发展、知识产权保护对企业绿色技术创新的影响[J]. 中国人口·资源与环境,2023,33(7):157-167.

[19] 谢会强,雷一鸣,郑淑芳. 全球价值链嵌入对区域创新能力的作用机制及溢出效应研究——基于市场整合的视角[J]. 科研管理,2024,45(6):1-11.

[20] AGRAWAL A, GOLDFARB A. Restructuring research: communication costs and the democratization of university innovation[J]. American Economic Review, 2008,98(4):1578-1590.

[21] GLAESER E L, KALLAL H D, SCHEINKMAN J A,

- et al. Growth in cities [J]. *Journal of Political Economy*, 1992, 100(6): 1126-1152.
- [22] XU Y B, WEI Y, ZENG X, et al. Big data development and labor income share: evidence from China's national big data comprehensive pilot zones [J]. *Economic Analysis and Policy*, 2024, 84: 1415-1437.
- [23] 杨彩虹, 梁宏志. 数字经济与地方财政可持续性——基于国家级大数据综合试验区的准自然实验[J]. *财经论丛*, 2024(5): 39-48.
- [24] 王春杨, 兰宗敏, 张超, 等. 高铁建设、人力资本迁移与区域创新[J]. *中国工业经济*, 2020(12): 102-120.
- [25] 豆建民, 王光丽, 马融. 数字经济发展对城市合作创新的影响——基于空间溢出效应的视角[J]. *经济管理*, 2023, 45(7): 56-75.
- [26] 江艇. 因果推断经验研究中的中介效应与调节效应[J]. *中国工业经济*, 2022(5): 100-120.
- [27] 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.
- [28] 白俊红, 张艺璇, 卞元超. 创新驱动政策是否提升城市创业活跃度——来自国家创新型城市试点政策的经验证据[J]. *中国工业经济*, 2022(6): 61-78.
- [29] 曾鑫, 黄炜, 何凡. 跨境运输的稳就业效应——来自中国上市企业的证据[J]. *世界经济文汇*, 2024(3): 21-38.
- [30] 余泳泽, 刘大勇, 龚宇. 过犹不及事缓则圆: 地方经济增长目标约束与全要素生产率[J]. *管理世界*, 2019, 35(7): 26-42.
- [31] 韩先锋, 宋文飞, 李勃昕. 互联网能成为中国区域创新效率提升的新动能吗[J]. *中国工业经济*, 2019(7): 119-136.
- [32] 黎绍凯, 朱卫平, 刘东. 高铁能否促进产业结构升级: 基于资源再配置的视角[J]. *南方经济*, 2020(2): 56-72.
- [33] CHAISEMARTIN C D, D'HAULTFOEUILLE X. Two-way fixed effects estimators with heterogeneous treatment effects [J]. *American Economic Review*, 2020, 110(9): 2964-2996.
- [34] LEE E Y, CIN B C. The effect of risk-sharing government subsidy on corporate R&D investment: empirical evidence from Korea [J]. *Technological Forecasting & Social Change*, 2010, 77: 881-890.
- [35] 柳卸林, 孙海鹰, 马雪梅. 基于创新生态观的科技管理模式[J]. *科学学与科学技术管理*, 2015, 36(1): 18-27.
- [36] JIAO H, ZHOU J H, GAO T S, et al. The more interactions the better? The moderating effect of the interaction between local producers and users of knowledge on the relationship between R&D investment and regional innovation systems [J]. *Technological Forecasting & Social Change*, 2016, 110: 13-20.
- [37] DOLOREUX D. What we should know about regional systems of innovation [J]. *Technology in Society*, 2002, 24: 243-263.
- [38] ZHAO S L, CACCIOLATTI L, LEE S H, et al. Regional collaborations and indigenous innovation capabilities in China: a multivariate method for the analysis of regional innovation systems [J]. *Technological Forecasting & Social Change*, 2015, 94: 202-220.
- [39] 杨武, 杨大飞, 雷家骥. R&D投入对技术创新绩效的影响研究[J]. *科学学研究*, 2019, 37(9): 1712-1720.
- [40] ZHANG M L, LI B Z, YIN S. Configurational paths to regional innovation performance: the interplay of innovation elements based on a fuzzy-set qualitative comparative analysis approach [J]. *Technology Analysis & Strategic Management*, 2020, 32(12): 1422-1435.
- [41] 赵涛, 张智, 梁上坤. 数字经济、创业活跃度与高质量发展——来自中国城市的经验证据[J]. *管理世界*, 2020, 36(10): 65-76.
- [42] 吴海军, 杨其静, 阳镇. 生产性政府债务与城市创新力——基于中国城市面板数据的经验研究[J]. *中国工业经济*, 2023(10): 42-60.
- [43] 李政, 杨思莹. 财政分权、政府创新偏好与区域创新效率[J]. *管理世界*, 2018, 34(12): 29-42.
- [44] 李雪, 吴福象, 竺李乐. 数字经济与区域创新绩效[J]. *山西财经大学学报*, 2021, 43(5): 17-30.
- [45] 张柳钦, 李建生, 孙伟增. 制度创新、营商环境与城市创业活力——来自中国自由贸易试验区的证据[J]. *数量经济技术经济研究*, 2023, 40(10): 93-114.
- [46] 周泽将, 汪顺, 张悦. 知识产权保护与企业创新信息困境[J]. *中国工业经济*, 2022(6): 136-154.
- [47] 徐换歌, 蒋硕亮. 国家创新型城市试点政策的效果以及空间溢出[J]. *科学学研究*, 2020, 38(12):

2161-2170.

[48]

ELLISON G, GLAESER E L, KERR W R. What causes industry agglomeration? Evidence from coagglomeration patterns [J]. American Economic Review, 2010, 100(June) : 1195-1213.

[49]

SOLOW R M. A contribution to the theory of economic growth [J]. The Quarterly Journal of Economics, 1956, 70(1) : 65-94.

[50]

KRUGMAN P. Increasing returns and economic geography[J]. NBER Working Paper # 3275, 1990.

[51]

曾艺, 韩峰, 刘俊峰. 生产性服务业集聚提升城市经济增长质量了吗? [J]. 数量经济技术经济研究, 2019, 36(5) : 83-100.

The Impact of the Establishment of Big Data Comprehensive Experimental Zones on Regional Innovation Capabilities / Lu Fucai¹, Zeng Xin², Huang Ting¹(1. School of Digital Economics, Jiangxi University of Finance and Economics; 2. School of Economics, Jiangxi University of Finance and Economics)

Abstract: With the rise of a new round of technological revolution and industrial transformation, technological innovation has become a central focus in the development strategies of major countries. In this context, big data, as a key factor of production in the digital economy era, urgently requires in-depth exploration regarding its mechanism for promoting high-quality economic development through enhancing regional innovation capabilities. Based on panel data from 281 Chinese cities from 2007 to 2021 and using the establishment of national big data comprehensive pilot zones as a quasi-natural experiment, this paper empirically examines the impact of these pilot zones on regional innovation capabilities and their underlying mechanisms. The findings are as follows. First, the establishment of big data comprehensive pilot zones significantly enhances regional innovation capabilities, a conclusion remaining valid after multiple robustness tests. The spatial Durbin model further reveals that the policy exhibits significant positive spatial spillover effects, with its stimulating effect on the innovation capabilities of neighboring regions even surpassing the local effect, highlighting the innovative spillover value of cross-regional data flow. Second, mechanism tests indicate that the establishment of big data comprehensive pilot zones primarily improves regional innovation capabilities through increasing the number of innovation entities, agglomerating innovation resources, and optimizing the innovation environment. Third, heterogeneity analysis shows that the innovation-enhancing effect of big data comprehensive pilot zones is more pronounced in regions with higher levels of digital economic development, higher marketization levels, and in eastern China. Moderating effect analysis further reveals that strengthening intellectual property protection amplifies the positive impact of big data comprehensive pilot zones on regional innovation capabilities.

Key words: big data comprehensive pilot zone; regional innovation capacity; innovation entities; innovation resources; innovation environment