**引用格式**：黄桂鑫，钱娟，朱智猛，等. 解锁数据红利：数据要素开放驱动企业创新——基于公共数据平台上线的证据[J]. 技术经济，2025，44(12)：77-92.

Huang Guixin，Qian Juan，Zhu Zhimeng，et al. Unlocking data dividends：how data element openness drives corporate innovation—Evidence from the launch of public data platforms[J]. Journal of Technology Economics，2025，44(12)：77-92.

# 解锁数据红利：数据要素开放驱动企业创新

## ——基于公共数据平台上线的证据

黄桂鑫[1,2,3]，钱 娟[1,2,3]，朱智猛[1]，郭雪颖[1,2,3]

（1. 新疆大学经济与管理学院，乌鲁木齐 830046；2. 西北能源碳中和教育部工程研究中心，乌鲁木齐 830046；3. 新疆宏观经济高质量发展研究中心，乌鲁木齐 830046）

**摘 要**：在数据要素上升为国家基础性战略资源的背景下，解构公共数据开放这一中国式制度创新的经济效应，成为阐释数字经济时代“有为政府与有效市场”辩证关系的关键切口。通过中国217个地级市政府数据平台上线数据，构建准自然实验场景下的多期双重差分(DID)识别框架，在因果推断层面实证检验公共数据要素共享对企业创新行为的结构性重塑机制。研究发现，公共数据开放能够促进企业创新。从制度性约束层面，公共数据开放通过优化营商环境、改善政企关系促进企业创新。从要素供给层面，公共数据开放通过企业运营能力的提升、企业间交流的增加与研发不确定性的降低促进企业创新。在数字经济政策供给强度高、社会信用高的地区，以及对于民营企业、小型企业和战略性新兴产业企业，公共数据开放对企业创新的影响更大，且该影响只存在于基础良好和高质量数据的开放中。拓展分析发现，公共数据对企业管理者的决策行为具有修正效应，而对企业的创新行为具有掩蔽效应。研究结果为阐释中国式现代化进程中数据治理的制度优势提供理论注脚，为全球数字经济发展贡献东方智慧方案。

**关键词**：公共数据开放；企业创新；修正效应；掩蔽效应；要素供给；制度约束

**中图分类号**：F423.3 **文献标志码**：A **文章编号**：1002-980X(2025)12-0077-16

**DOI**：10.12404/j.issn.1002-980X.J25063017

## 一、引言

随着新一轮科技革命和产业变革深入发展，数据作为关键生产要素的价值日益凸显。公共数据是由政府部门、经依法授权具有管理公共事务职能的组织在依法履行职责或者提供公共服务过程中收集、产生的数据[1]。充分挖掘数据要素报酬递增、可低成本复用等特质，是赋能实体经济发展、优化资源配置的重要抓手。这些独有的价值属性，正是实现经济社会高质量发展所期冀的数据红利的核心来源。具体而言，数据红利的释放，不仅体现在生产效率的全面提升，更在于优化资源配置、催生新业态新模式及促进社会福利改善等多个层面。数据要素的非竞争性、低边际成本特性使其能够在多场景、多主体间复用，大幅降低信息搜寻成本和交易成本，从而实现生产要素的优化组合，激发创新活力。此外，数据红利还包括基于海量数据分析所带来的洞察力提升，助力企业精准决策，降低市场不确定性，进而驱动技术创新和产业升级。公共数据作为数据要素中占比最大的种类，科学引领公共数据发展具有重要意义[2]。2015年，国务院发布的《促进大

**收稿日期**：2025-06-30

**基金项目**：新疆维吾尔自治区自然科学基金面上项目“碳交易制度促进新疆工业低碳转型的路径选择”(2022D01C369)；国家社会科学基金“治理现代化下偏向型技术进步促进工业绿色转型升级的机理与路径研究”(20XJL012)；中国博士后科学基金“统一碳交易市场促进我国工业绿色低碳转型的机理与路径研究”(2021M702603)；新疆维吾尔自治区社科联新时代党的治疆方略理论与实践研究课题“新疆增强重大装备产品及关键器件生产配套能力路径研究”(2024ZJFLY34)

**作者简介**：黄桂鑫(1998—)，新疆大学经济与管理学院博士研究生，研究方向：数据要素；(通信作者)钱娟(1987—)，博士，新疆大学经济与管理学院教授，研究方向：能源与环境经济学；朱智猛(1996—)，新疆大学经济与管理学院博士研究生，研究方向：能源利用；郭雪颖(2000—)，新疆大学经济与管理学院硕士研究生，研究方向：虚拟集聚。

数据发展行动纲要》明确推动政府数据开放共享,促进社会事业数据融合和资源整合。截至2024年7月,已有243个省市政府上线了数据开放平台,省级平台增加至24个,城市平台数量达到219个。各地数据开放的有效数据集总量显著增长,从2017年的8000余个增至2024年的37万余个,无条件开放的数据量从2019年的15亿条飙升至2024年的679亿条①。可见,公共数据开放已成为数据要素开放共享与使用的重点工作。

创新作为强化国家战略科技力量、塑造发展优势的必由之路[3],是赋能实体经济、优化社会资源配置的重要驱动力。熊彼特创新理论提出创新的本质是在生产中引入未曾使用过的生产要素和生产条件的组合,以实现对生产要素或生产条件的新组合。要实现这个过程需要经过三个基本步骤:第一,创意的生成;第二,将想法发展成一个可用的概念;第三,该概念的成功应用[4]。在此过程中须考虑可能面临的问题:一是与外部宏观市场环境的摩擦[5-6],其主要体现在地方保护主义倾向、政企目标错位等方面;二是与内部微观企业的摩擦[7],其主要体现在信息不对称、生产要素的错配、创新风险管理不善、管理者的短视主义等。数据开放作为数字时代的"包容性制度",能打破传统要素市场的路径依赖,为创新营造更为广阔且自由的发展空间。具体而言,从资源可获得性视角出发,公共数据开放以标准化、结构化的供给降低信息搜寻、匹配的交易成本,提升信息的可比性与可信度。在此基础上,企业得以实现流程的可视化与数据驱动的运营优化,研发环节的不确定性与试错成本随之下降,协同层面的共同语义与统一口径又进一步扩展跨区域、跨行业的知识交流边界,激活网络外部性与知识再组合效应。从制度性约束视角出发,公共数据开放将规则可见、程序可预期与责任可追溯嵌入治理过程,弱化行政裁量并降低政策执行的误差,提升规则的一致性与市场的可争性。由此,选择性执法与关系型交易的空间被压缩,政企互动的交易摩擦与关系成本下降;政府承诺的可核验性与回应性增强,企业所面临的制度性不确定性与合规噪声随之缓解,企业的跨期预期得以稳固;研发项目所需的风险补偿门槛的降低,对高不确定性、探索式投入的容忍度提高,从而为创新活动的持续推进与质量提升提供稳健的制度外部性。因此,在加快新质生产力发展的背景下,探明公共数据开放能否促进企业创新具有重要现实意义与理论意义。

从有关公共数据开放的相关研究来看,现有文献借助省市级公共数据平台上线的准自然实验,探究公共数据开放制度的社会经济效应,发现其对于促进企业投资[8-9]、拓展资本跨区域流动[10-11]、缓解区域发展不协调[12]、激发区域创业活力[13]、提升全要素生产率[11]与降低当地城投债的发行利差[14]等具有积极作用。有关公共数据开放与技术创新的研究不断推陈出新,公共数据开放对于促进企业创新投入[15]、数字技术创新[16-17]、高技术企业创新[18]等具有积极作用。李长英等[19]、谢贞发和冉芙小[20]认为改善市场竞争环境与降低经营风险是公共数据开放促进企业创新的主要渠道。Chen等[21]研究发现城市间的竞争和学习强化了公共数据开放促进企业创新的空间溢出效应。上述文献主要从资源的可获得角度探究公共数据开放对企业的专利申请、获得和研发投入等的影响和作用机制,正如不完全契约理论所提出的"可验证性决定可执行性",当交易或政策状态难以被第三方验证时,合约与承诺易受到事后再议问题的侵蚀;公共数据开放通过将原本不可观察的过程与结果转化为可被第三方核验的信息,提高规则与承诺的可执行性,抑制选择性执法与关系型交易,进而在制度层面降低执行失误与制度性不确定性。由此,公共数据开放的经济效应不仅来自信息与资源的更可得及其所承载的价值信息效应,亦可被视为一种由规则可见、程序可预期与责任可追溯所形成的制度性约束。然而,鲜有文献系统地将此制度性视角纳入分析框架。本文据此从"资源可获得性-制度性约束"的双重视角出发,探讨公共数据开放对企业创新的影响,以期在理论上补足制度机制的阐释,在实证上提供与既有信息供给路径相互印证的证据。

基于此,本文将政府公共数据平台上线作为外生冲击,基于2008—2022年沪深A股2749家上市企业和217个城市的非平衡面板数据,构建多期双重差分(difference-in-differences,DID)模型,系统检验公共数据开放与企业创新的因果关系。可能的边际贡献有以下三点:一是丰富了数据要素共享与创新领域的相关研究成果,为深入理解数据要素开放所能创造的社会价值提供经验证据。二是从资源可获得性和制度性约束的

① 来源于《"数据要素×"三年行动计划(2024-2026年)》。

双重视角出发探究了公共数据开放影响企业创新的内在机理。前者强调数据供给的扩张与信息摩擦的降低所带来的要素配置效率提升，后者强调通过规则可见、程序可预期与责任可追溯所形成的治理承诺改善企业的风险-回报权衡与研发期限结构，从而把公共数据开放对创新的作用从单纯的效率改善拓展为秩序重塑。三是探究公共数据开放所带来的修正效应与掩蔽效应，挖掘公共数据作为准公共物品和制度化信息基础设施的内在特质表现，进一步突出公共数据开放的重要价值。

## 二、理论分析与假说提出

### （一）公共数据开放驱动企业创新的直接影响

公共数据开放推动企业创新的根本逻辑，在于其有效解锁并释放了蕴含在数据中的红利。这种数据红利是数据要素非竞争性、低复制成本、可复用性等内在特质在经济活动中的体现，其使得数据能够为多主体带来价值增值，并产生超越传统要素的收益增长。数据的本质是一种信息载体，是对客观感知事物或事实的描述，但数据并不等同于数据要素，只有经过系统提取、分类、清洗与应用等处理才能成为可以参与社会生产的要素。公共数据开放为企业获取并应用数据要素创造了条件，使其得以融入生产分配与交换环节，通过降低组合过程中的搜寻与匹配成本，提升整体资源配置效率，削弱时空约束中的知识与技术扩散的限制，并降低跨区域协作的信息摩擦成本进而释放数据红利。

一方面，在数据红利视角下，公共数据开放通过降低组合成本促进企业创新。传统要素受制于替代弹性有限与边际报酬递减规律，企业在扩大生产和创新时常常面临显著的成本台阶[22]。相比之下，数据要素具有非竞争性和低边际复制成本，能够在不增加额外消耗的前提下多次使用[23]。这一特性使得数据在与其他生产要素组合时，显著降低了组合过程中的边际成本，提高了资源配置效率[13]。公共数据开放通过提供标准化和结构化的数据资源，打破数据获取的障碍，促进数据与其他生产要素的高效整合。这使得创新过程中的信息流通更加便捷，减少了资源整合、信息获取和技术应用的成本，从而显著降低创新的进入门槛和实施难度[12]。因此，公共数据开放通过降低数据获取的难度和成本，使得数据要素能够更便捷与其他要素进行组合，为企业创新提供更高效的资源配置和更低成本的创新路径。

另一方面，从数据红利对时空约束的突破来看，公共数据开放能够有效削弱传统要素组合受制于地理锁定与跨区域协作摩擦的弊端。数据要素依托标准化接口与互操作技术栈，能够实现跨地域、跨组织的实时整合。公共数据开放通过增强数据的可得性与可比性，进一步降低了企业在异地搜寻信息、匹配资源和对接市场时的摩擦成本[10]。这不仅有助于缩短“创意生成-概念形成-成果应用”的创新周期[4]，还可以拓展跨区域、跨行业的协同空间，提升知识交流的频率与广度[16]。换言之，公共数据开放通过削弱时空约束，使企业创新能够在更大范围和更高效率下展开。综上，数据要素在组合成本和时空约束两个维度的嵌入，为企业提供了超越传统要素的创新驱动力。

基于此，本文提出假设：

公共数据开放有利于促进企业创新（H1）。

### （二）公共数据开放驱动企业创新的间接影响

公共数据开放兼具要素供给与制度约束的双重属性，既能通过强化资源供给改善企业创新条件，也能通过制度性约束优化降低创新过程中的外部摩擦。具体而言，其影响机制主要体现在以下几个方面。

**1. 制度约束路径**

从制度约束视角来看，企业创新高度依赖稳定透明的制度环境。根据诺斯的制度变迁理论，制度安排的稳定性和可预期性能够有效降低交易成本并提升经济绩效[24]。制度环境的不确定性会显著抑制企业的投资与研发行为。在中国情境下，地方政府在市场运行中发挥着关键作用，若地方政府行政裁量过大、选择性执法或关系型交易的倾向，不仅会削弱市场规则一致性和可预期性，还可能因政企之间严重的信息不对称，使政府难以把握企业真实需求，企业也难以及时理解和适应政策变化，从而造成政企目标错位与资源配置效率低下。在此背景下，优化营商环境与改善政企关系成为公共数据开放发挥制度约束效应的核心路径。公共数据开放在这一过程中发挥了制度创新与沟通桥梁的双重作用。一方面，公共数据开放将原本封

闭的政府数据以标准化、可核验的形式公开，能够有效提升制度的透明性与一致性，压缩行政裁量空间，减少企业所面临的制度不确定性与合规噪声，从而有效优化营商环境[9-10]。同时，公共数据开放可以通过共享机制缓解信息不对称和抑制机会主义行为，为企业营造更公平的市场竞争环境[25]；此外通过提升政府服务效率与简化行政审批流程，公共数据开放还显著降低了企业的制度性交易成本[26]。另一方面，公共数据开放通过发布涵盖企业、行业与市场的多维数据及政策制定与执行过程的非结构化信息，帮助企业更准确地理解政策导向并及时调整战略，从而缓解政企之间的认知偏差。数据开放中嵌入的溯源与问责机制也强化了政府的回应性和责任感，推动政企关系的良性互动[27]。综上，公共数据开放通过优化营商环境与改善政企关系，降低了企业创新活动中的制度摩擦与关系成本，增强了企业对长期创新回报的信心，最终激励企业更积极地开展创新投资。

基于此，本文提出假说2：

公共数据开放通过优化营商环境与改善政企关系促进企业创新（H2）。

**2. 要素供给路径**

从要素供给视角来看，企业创新不仅取决于制度环境的稳定性，还受到资源配置效率和信息获取能力的制约。企业创新活动本质上是一个由运营夯实、研发稳定到合作拓展逐层递进的过程，这与Cumming[4]所提出的创新基本步骤相契合。然而，在此过程中企业普遍面临营运效率不足、研发不确定性较高及交流合作受限等困境，从而造成要素错配与资源浪费，制约创新活力。公共数据开放以非竞争性与低成本复用的特征，为企业提供了丰富、标准化、可验证的数据资源，能够针对性地缓解上述制约。具体来说，在企业内部运营层面，公共数据开放通过提供可比性更强的市场与行业信息，帮助企业识别差距并优化流程，缓解因信息不足导致的经营低效与要素错配[28]，降低管理者短视主义与机会主义行为所带来的机会成本[29]，并提升决策的科学性与准确性，为创新创意的孕育提供了重要条件。在技术研发环节，公共数据开放通过提供前沿信息与验证支持，使企业能够更准确地把握市场需求与产业链动态[30-31]，并借鉴先进的技术理念与方法，加速创意向可行概念的转化[32]。这一作用不仅降低研发过程中的试错成本，缩短研发周期，也增强企业在高不确定性环境下推进创新的能力。在外部协作层面，公共数据开放通过共享语义与统一口径打破企业间的信息壁垒，促进跨区域、跨行业的知识流动与资源整合[33]，不仅释放了知识溢出效应与创新网络外部性[34]，还推动了合作创新成果的应用与扩散，进一步拓展了企业的创新边界。综上，公共数据开放通过“营运水平提升-研发不确定性降低-交流合作加强”的递进链条作用，有效缓解信息不对称与要素错配问题，体现了帕累托改进意义上的效率提升逻辑，从而为企业创新提供系统性与持续性的动力。

基于此，本文提出假说3：

公共数据开放通过提升企业营运水平、降低研发不确定性与促进企业交流合作，推动企业创新（H3）。

## 三、研究设计

### （一）模型设计

公共数据平台上线为探究数据要素开放共享影响企业创新行为提供了理想的准自然实验场景，为考察公共数据平台上线对企业创新水平影响的净效应，构建多期双重差分模型进行检验，模型设定如式（1）所示。

$$lrd_{itc} = \beta_0 + \beta_1 open_city_{tc} + \gamma controls_{itc} + \eta_i + \mu_i + \varepsilon_{itc} \quad (1)$$

其中：下标 $i$ 为企业；$t$ 为时期；$c$ 为城市；$lrd_{itc}$ 为企业创新的代理变量；$open_city_{tc}$ 为政府数据平台上线的虚拟变量，表示城市 $c$ 在 $t$ 时期是否已经上线公共数据平台，上线则赋值为1，未上线则赋值为0。其中，各市级公共数据平台的上线时间主要根据相关报道手工收集整理而得。$controls_{itc}$ 为一系列企业、行业和省份层面控制变量；$\eta_i$ 和 $\mu_i$ 分别为企业固定效应和年份固定效应，以此来控制回归中不随企业与年份变化的其他因素；$\varepsilon_{itc}$ 为随机误差项；$\beta_0$ 为截距项；$\beta_1$ 为核心解释变量的回归系数。在具体回归中，对标准误进行企业层面的聚类处理。主要关注系数 $\beta_1$ 的特征，如果 $\beta_1$ 显著为正，认为公共数据开放能有效促进企业的创新水平。

### （二）变量选择

**1. 被解释变量**

企业创新（*lrd*）。采用企业当年研发投入即研发支出与上期总资产之比衡量企业创新。上期总资产代表企业过往积累的财务实力和资源状况，是企业运营和发展的基础。用研发支出与之相比，能体现企业在已有资源基础上对创新的重视程度和投入意愿。

**2. 核心解释变量**

公共数据开放（*open_city*）。参考2024年复旦大学数字与移动治理实验室联合国家信息中心数字中国研究院发布的《中国地方公共数据开放利用报告——城市》整理出政府数字平台上线的基础数据，并在搜索引擎上通过“城市+政府数据平台”“城市+公共数据平台”“城市+公共数据开放”等关键词进行搜索验证，排除非市级行政部门以下的机构发布的信息，最终得到了截至2022年的217个城市公共数据平台开放情况。

**3. 控制变量**

（1）企业层面：企业成长性（*grow*），以营业收入增长率衡量；资金链稳定程度（*cash*），以企业现金及现金等价物余额与资产总额之比衡量；资产回报率（*roa*），以净利润与总资产之比衡量；企业规模（*size*），以总资产的自然对数衡量；企业年龄（*age*），以观测年减成立年的自然对数衡量；股权结构（*fshare*），以前十大股东持股比例衡量；独董比例（*indep*），以董事会中独立董事的占比衡量；资产负债率（*lev*），以总负债与总资产之比衡量。

（2）行业层面：行业竞争度（*hhi*），以企业所处两位码行业的 *Herfindahl-Hirschman* 指数衡量。

（3）省份层面：经济发展（*plgdp*），以各省人均GDP的自然对数衡量；人口水平（*plpop*）：以各省常住人口的自然对数衡量。

主要变量描述性统计见表1。

**表1　主要变量描述性统计结果**

| 变量 | 观测值 | 均值 | 标准差 | 最小值 | 最大值 |
|---|---|---|---|---|---|
| *lrd* | 15066 | 0.0286 | 0.0241 | 0.0001 | 0.1447 |
| *open_city* | 15066 | 0.5777 | 0.4939 | 0 | 1.0000 |
| *cash* | 15066 | 0.1649 | 0.1238 | 0.0125 | 0.6813 |
| *roa* | 15066 | 0.0454 | 0.0716 | −0.2665 | 0.2351 |
| *size* | 15066 | 22.1150 | 1.2653 | 19.9053 | 26.2313 |
| *fshare* | 15066 | 59.2936 | 14.5091 | 23.5500 | 90.3700 |
| *lev* | 15066 | 0.3944 | 0.1949 | 0.0483 | 0.8831 |
| *grow* | 15066 | 0.1657 | 0.3270 | −0.4978 | 1.8003 |
| *age* | 15066 | 2.7977 | 0.3608 | 1.6094 | 3.5223 |
| *indep* | 15066 | 37.7769 | 5.4118 | 29.1600 | 57.1400 |
| *plpop* | 15066 | 8.5638 | 0.6387 | 6.8063 | 9.4481 |
| *plgdp* | 15066 | 11.2770 | 0.4732 | 9.8889 | 12.1564 |
| *hhi* | 15066 | 0.1393 | 0.1301 | 0.0324 | 0.7978 |

### （三）数据来源与处理

本文以2008—2022年中国沪深A股上市公司作为研究对象，并对样本数据进行以下筛选和处理：①剔除主要财务指标缺失的样本；②剔除上市状态为ST（special treatment）、*ST及PT（particular transfer）的企业样本；③为避免异常值对研究结论造成干扰，本文还对相关连续型变量进行双侧1%的缩尾处理。企业及地区层面的数据主要来自国泰安数据库（CSMAR）、中国研究数据服务平台（CNRDS）和《中国城市统计年鉴》。

## 四、实证结果分析

### （一）基准回归

表2显示了公共数据开放对企业创新的回归结果。（1）列仅控制企业和年份的固定效应，结果显示公共数据开放（*open_city*）的系数为0.007且在1%的统计水平上显著，表明公共数据开放对企业创新有正向的促进效应。在（1）列的基础上加入控制变量，结果显示公共数据开放（*open_city*）的系数仍在1%的统计水平

上显著，相较（1）列的系数变动较小，说明公共数据开放与企业创新之间的因果关系受其他不可观测的潜在影响相对较小。从经济显著性上来看，当公共数据平台上线之后，企业的创新水平提高了 24.14%（即 0.007/0.029×100%）。创新是企业在不断发展和竞争的过程中的重要战略，不同行业的企业面临的市场环境、技术需求和竞争格局各不相同，企业创新行为可能也会有所差异。因此，在（2）列的基础上控制了行业固定效应，结果显示公共数据开放的系数为 0.007，与没有控制行业固定效应时的系数相差不大且在 1%的统计水平上显著，说明行业固定效应可能在解释企业创新的变化中起到较小的作用，即行业的不同不会对公共数据开放影响企业创新产生偏误。同样，在（4）列和（5）列依次加入了城市固定效应和省份固定效应，回归结果和（2）列并无较大差异。综上可知，公共数据开放可以促进企业创新，且不会受到行业、城市及省份的影响，这可能和数据要素本身特性相关，数据要素具有跨地域、跨行业的通用性与流动性，打破了传统生产要素受限于行业壁垒和地域边界的桎梏。这也进一步表明，公共数据开放作为一种新型的制度供给和资源配置方式，具有广泛的适用性和强大的赋能能力，能够在全国范围内均衡推动各类企业的创新发展，为构建创新型国家提供坚实的数据要素支撑，助力各地区、各行业企业在创新驱动发展的道路上齐头并进，缩小因行业、地域差异导致的创新能力差距，形成全国一盘棋的创新发展新格局，假设 H1 得以验证。

**表 2　基准回归结果**

| 变量 | （1） | （2） | （3） | （4） | （5） |
|---|---|---|---|---|---|
| | *lrd* | *lrd* | *lrd* | *lrd* | *lrd* |
| *open_city* | 0.007*** | 0.007*** | 0.007*** | 0.007*** | 0.007*** |
| | （12.012） | （13.227） | （13.235） | （13.459） | （13.563） |
| *hhi* | | −0.005* | −0.005 | −0.005* | −0.005* |
| | | （−1.846） | （−1.627） | （−1.794） | （−1.843） |
| *grow* | | 0.009*** | 0.009*** | 0.009*** | 0.009*** |
| | | （16.819） | （16.626） | （16.725） | （16.862） |
| *cash* | | 0.000 | 0.000 | 0.001 | 0.001 |
| | | （0.230） | （0.233） | （0.342） | （0.307） |
| *roa* | | 0.026*** | 0.027*** | 0.026*** | 0.026*** |
| | | （10.199） | （10.425） | （10.056） | （10.141） |
| *size* | | −0.001 | −0.001 | −0.001 | −0.001 |
| | | （−1.357） | （−1.156） | （−1.420） | （−1.370） |
| *age* | | −0.005 | −0.005 | −0.005 | −0.005 |
| | | （−1.385） | （−1.345） | （−1.436） | （−1.318） |
| *fshare* | | 0.000 | 0.000 | 0.000 | 0.000 |
| | | （1.106） | （1.287） | （1.036） | （1.138） |
| *indep* | | −0.000 | −0.000 | −0.000 | −0.000 |
| | | （−1.471） | （−1.341） | （−1.353） | （−1.421） |
| *lev* | | 0.001 | 0.001 | 0.001 | 0.001 |
| | | （0.292） | （0.530） | （0.405） | （0.324） |
| *plpop* | | 0.002 | 0.002 | 0.001 | 0.001 |
| | | （0.756） | （0.882） | （0.235） | （0.173） |
| *plgdp* | | −0.001 | −0.001 | −0.000 | −0.000 |
| | | （−0.543） | （−0.625） | （−0.015） | （−0.029） |
| 控制变量 | 否 | 是 | 是 | 是 | 是 |
| 企业固定 | 是 | 是 | 是 | 是 | 是 |
| 年份固定 | 是 | 是 | 是 | 是 | 是 |
| 行业固定 | 否 | 否 | 是 | 否 | 否 |
| 城市固定 | 否 | 否 | 否 | 是 | 否 |
| 省份固定 | 否 | 否 | 否 | 否 | 是 |
| *Constant* | 0.024*** | 0.049 | 0.045 | 0.046 | 0.047 |
| | （75.124） | （1.552） | （1.364） | （1.023） | （1.049） |
| *Obs* | 14687 | 14687 | 14687 | 14687 | 14687 |
| Adjust $R^2$ | 0.809 | 0.832 | 0.833 | 0.831 | 0.832 |

注：括号内数值为回归系数的 $t$ 值，标准误聚类到企业层面；***、**、* 分别表示在 1%、5%、10%的水平上显著。

## （二）稳健性检验

### 1. 平行趋势检验

理论上来说如果构建的多时点双重差分模型是有效的，那么模型需满足控制组和实验组在政策实施之前企业创新行为具有相同的趋势。因此，构建动态双重差分模型进行平行趋势检验。

$$lrd_{itc}=\beta_0+\beta_k\sum_{k=-14}^{10}open_{ct}^{k}+\gamma controls_{itc}+\eta_i+\mu_t+\varepsilon_{itc} \quad (2)$$

其中：$open_{ct}^{k}$ 为市级公共数据平台上线前、后第 $k$ 年的虚拟变量；$k$ 为公共数据开放的第 $k$ 年，由于公共数据开放最早于 2012 年，样本截至 2022 年，$k$ 的最大取值为 10，最小取值为-14。为避免多重共线性问题，以公共数据开放前一年为基期，在回归中省略公共数据开放前一期，若平行趋势假设成立，则在政府数据开放前时间距离变量的回归系数均不会显著偏离如图 1 所示的 0 轴，即上文所说控制组和实验组在政策实施之前企业创新具有相同趋势。

如图 1 所示，展示了公共数据平台开放的前 6 年和后 6 年的情况，横轴是观测年度相对公共数据平台开放年份的相对距离，纵轴是回归系数，即实验组与对照组的企业创新水平的差异。在 90%的置信区间下，控制组和实验组在政策实施之前企业创新行为均与 0 轴无显著差异，满足平行趋势检验。在公共数据开放之后回归系数显著区别于 0，直到第 6 年与 0 轴有交叉，表明公共数据开放对于企业创新存在长期影响。此外，在公共数据开放之后回归系数上升到高水平，并在接下来的年份中逐渐下降。对此的解释是，公共数据平台开放之初，企业获得新的数据要素，降低创新成本，缓解企业主体与政府和市场的信息不对称性现象，但是数据要素区别于传统生产要素的一个特点是具有知识溢出效应，随着时间的推移能够很快的被模仿或者超越[35-36]，导致公共数据开放对企业创新具有递减效应。

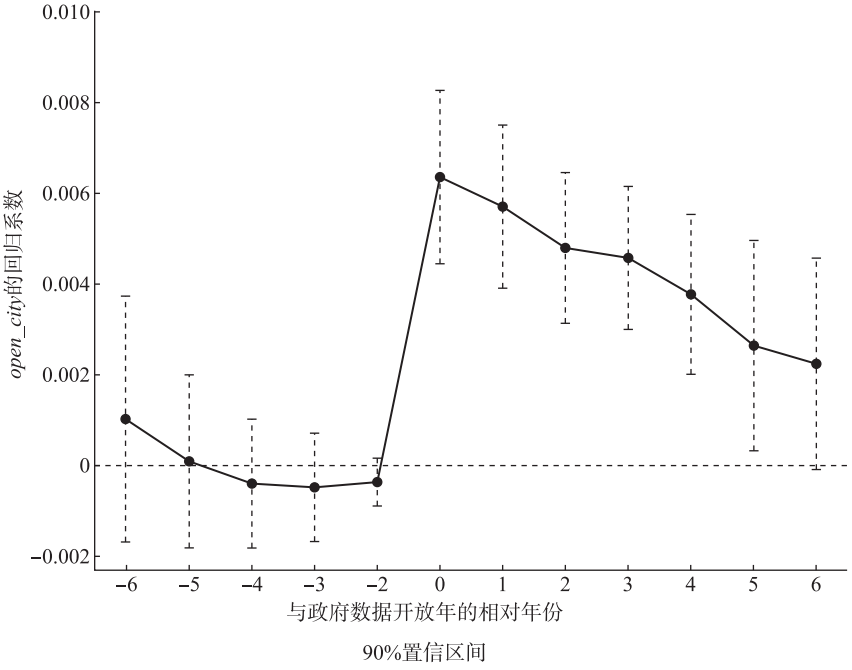



**图 1　平行趋势检验结果**

### 2. 异质性稳健估计

本文采用多期 DID 估计公共数据开放与企业创新之间的因果关系，但是多期 DID 估计结果可能会产生偏误，原因在于其是依靠多个不同处理效应的加权平均，在权重为负的情况下，估计出来的平均处理效应可能会与真实的平均处理效应方向相反[37]。因此，为了确保结果的可靠性，分别采用权重分解、模糊 DID、Callaway 和 Sant'Anna 双重差分（CSDID）和倾向性得分匹配双重差分（PSM-DID）的方法进行异质性稳健估计。

首先,借鉴 De Chaisemartin 和 D' Haultfoeuille[38]提出的分解方法进行诊断,结果见表 3 的 panel A,权重总数为 8361,其中正权重数有 5255 个,负权重数有 3106 个,负权重的比例达到 37.15%,表明 DID 估计结果不一定稳健。为此,利用模糊双重差分法(FDID)对企业所在城市是否为公共数据开放试点城市这一状态变动时的 LATE(局部平均处理效应)进行估计,结果见表 3 的 panel B,估计系数为 0.003,且在 1%的统计水平上正向显著,表明结果是稳健的。

其次,借鉴 Callaway 和 Sant'Anna[39]的研究,运用 CSDID 方法,采用逆概率加权最小二乘法(DRIPW)进行多时点稳健估计,其基本原理是对样本进行不同分组,估计分组之后各组的处理效应,在此基础上对分组之后的处理效应进行加总以降低个别组别存在偏误的权重并计算平均处理效应,结果见表 3 的 panel C。可以看出经过平均处理效应之后表现出公共数据平台上线对企业创新产生正向影响,结果稳健。

最后,采用倾向性匹配双重差分(PSM-DID)的方法进行检验。其原理是在 PSM 阶段,通过计算倾向得分为处理组中的每个个体找到对照组中相似的个体,生成匹配权重。其反映每个对照组个体在匹配过程中的重要性,随后匹配权重被用于加权回归分析中,以确保处理组和经过匹配的对照组之间的比较是有效的,可以进一步验证结果的准确性。避免了公共数据开放的城市可能存在选择偏误问题,结果见表 3 的 panel D。系数未发生显著变化,结果依旧稳健。

**表 3　异质性检验及处理**

| panel A | | | | |
|---|---|---|---|---|
| 总权重数 | 正权重数 | 负权重数 | 正权重比例 | 负权重比例 |
| 8361 | 5255 | 3106 | 62.85% | 37.15% |
| panel B | | | | |
| *ATT* | 0.003*** (3.359) | *CI* 上下限 | [0.001,0.004] | |
| panel C | | | | |
| 变量 | 简单加权平均处理效应 | | | |
| *SimpleATT* | 0.003*** (4.442) | | | |
| panel D | | | | |
| 变量 | *lrd* | | | |
| *open_city* | 0.007*** (13.284) | | | |
| 控制变量 | 是 | | | |
| 企业固定效应 | 是 | | | |
| 年份固定效应 | 是 | | | |
| *Obs* | 14687 | | | |
| Adjust $R^2$ | 0.833 | | | |

注:括号内数值为回归系数的 $t$ 值,标准误聚类到企业层面;***、**、*分别表示在 1%、5%、10% 的水平上显著。

**3. 工具变量法**

由于模型可能会受到遗漏变量或其他混淆因素影响,此处采用工具变量法进行检验。本文选取地形起伏度与上一期全国互联网宽带接入端口的自然对数交互项作为工具变量(*IV*)。原因在于:一是,通信与网络基础设施的建设与地形的崎岖程度直接相关,地形复杂的地区在通信网络基础设施建设上面临更多挑战,互联网宽带接入端口的增加在这些地区具有更为突出的边际效应,且互联网宽带的普及和接入端口的增加有助于打破地理隔阂,促进信息资源的均衡分布,推动公共数据开放和共享,满足相关性条件;二是,地区地形起伏度属于地理变量,对企业当期创新并无直接影响,上一期全国互联网宽带接入端口也不会直接影响企业创新行为,满足外生性条件。

**表 4　工具变量法**

| 变量 | (1) | (2) |
|---|---|---|
| | *open_city* | *lrd* |
| *iv* | 0.004*** (4.877) | |
| *open_city* | | 0.016* (1.810) |
| 控制变量 | 是 | 是 |
| 企业固定效应 | 是 | 是 |
| 年份固定效应 | 是 | 是 |
| *Obs* | 14687 | 14687 |
| Cragg-Donald Wald $F$ 统计量 | 75.248 | |
| Kleibergen-Paap rk LM 统计量 | 26.886 | |
| Adjust $R^2$ | 0.733 | |

注:括号内数值为回归系数的 $t$ 值,标准误聚类到企业层面;***、**、*分别表示在 1%、5%、10%的水平上显著。

表 4 的(1)列显示两阶段最小二乘法的第一阶段结果,工具变量对公共数据开放具有正的边际效应,符合理论预期;(2)列展示第二阶段的回归结果,*open_city* 的系数为 0.016 在 10%的统计水平上显著,且与基准回归系数相比相差不大,说明选取的工具变量是有效的,且 Cragg-Donald Wald $F$ 统计量的值为 75.248,说明结果依旧稳健。

**4. 其他稳健性检验②**

上文的检验在一定程度上证明了结果的稳健性,但仍旧会受到诸如模型设定、样本选择、变量衡量方式

② 由于篇幅限制,其他稳健性结果未列示,备索。

和潜在的遗漏变量问题的困扰。本文在以下几个方面进一步对结果进行稳健估计：一是，控制行业、地区时间趋势、考虑时滞问题、Oster 遗漏变量检验、排除包括“国家大数据实验区”“信息惠民”“智慧城市试点”等政策干扰，解决潜在的遗漏变量问题；二是，缩短样本时间、安慰剂检验、剔除直辖市和相邻城市，以排除样本选择的干扰；三是，从创新产出和创新质量的视角重新衡量被解释变量，以排除变量衡量方式对结果的偏误；四是，采用随机森林、套索回归和梯度提升的双重机器学习模型进行检验，以排除因模型设定而对结果的干扰。

### （三）异质性分析

为进一步揭示公共数据开放对企业创新的差异化效应，从政策环境、制度环境、企业特征与数据开放质量 4 个维度开展异质性分析，具体包括数字经济政策供给强度、地区信用环境、股权性质、企业规模、行业类型以及公共数据开放质量。

**1. 政策环境**

数字经济政策供给强度。在数字强国战略的引领下，各地方政府积极响应，制定并实施多项旨在促进数字经济发展的政策，着重加强数据资源的要素供给。数字经济政策供给强度直接反映政府对数字经济发展的支持力度和资源投入情况。不同强度的政策供给会对企业所处的数字化环境产生不同影响，进一步影响企业获取、利用公共数据的能力和效率。采用截至 2022 年底各城市数字经济政策文本数量作为各城市数字经济政策供给强度的代理变量，利用中位数将样本划分为数字经济政策供给强度高的地区（$Deps=1$）和数字经济政策供给强度弱的地区（$Deps=0$）。表 5 的（1）列和（2）列结果显示，相比数字经济政策供给强度高的地区，公共数据开放对数字经济政策供给强度弱的地区的企业创新行为影响更大。公共数据开放能够促进数据的流通和利用，推动数据要素市场的形成和发展。在数字经济政策供给强度弱的地区，数据资源的开发和利用往往不够充分，而公共数据的开放能够激发数据市场的活力，为企业提供更多可获取和利用数据的渠道，对企业创新促进作用更为明显。

**2. 制度环境**

信用环境。社会信用在创新活动中扮演着重要角色。信用是一种心理预期，体现了人们对他人意图和行为的积极期望[40]。在长期的商业合作和经济交易中，各方之间的行为类似于重复进行的博弈游戏[41]。在此过程中信任成为一种至关重要的因素，为参与博弈的各方提供稳定的心理预期。在低社会信用环境下，企业在遴选合作伙伴时面临搜寻与匹配困境，主体间信任机制缺失，导致合作过程充斥着高度的不确定性与信息不对称。在创新项目开展过程中，由于信任资本的匮乏，协同创新网络难以有效构建，知识共享与资源整合面临重重壁垒，导致创新项目的推进效率低下。本文采用中国城市商业信用环境指数的中位数划分高信用（$Cei=1$）和低信用（$Cei=0$）城市，结果见表 5 的（3）列和（4）列。可见，公共数据开放对于企业创新促进作用在高信用城市促进作用更强。信息不对称和代理问题是制约企业创新绩效的关键因素。社会信任的存在能够在一定程度上降低信息不对称的程度，有助于缓解创新决策过程中的代理问题。当公共数据开放后，良好的信用环境使得交易双方在合同签订、数据共享规则制定等方面的谈判更为顺利，减少不必要的时间和资源浪费，企业能够将更多精力投入到创新活动本身。

**表 5　政策与制度环境的异质性分析**

| 变量 | （1） | （2） | （3） | （4） |
|---|---|---|---|---|
| | $Deps=1$ | $Deps=0$ | $Cei=1$ | $Cei=0$ |
| | *lrd* | *lrd* | *lrd* | *lrd* |
| *open_city* | 0.006*** | 0.013*** | 0.009*** | 0.006*** |
| | （12.636） | （4.255） | （8.805） | （10.199） |
| 控制变量 | 是 | 是 | 是 | 是 |
| 企业固定 | 是 | 是 | 是 | 是 |
| 年份固定 | 是 | 是 | 是 | 是 |
| *Constant* | 0.058* | 0.121 | −0.023 | 0.051 |
| | （1.947） | （0.595） | （−0.419） | （1.306） |
| *Obs* | 12549 | 1772 | 7025 | 7262 |
| Adjust $R^2$ | 0.854 | 0.764 | 0.854 | 0.808 |

注：括号内数值为回归系数的 $t$ 值，标准误聚类到企业层面；***、**、* 分别表示在 1%、5%、10%的水平上显著。

**3. 企业特征**

（1）股权性质与企业规模。相较于民营企业和小型企业，国有企业（$Soe=1$）和大型企业（$Size=1$）在政策方针、资源倾斜方面具有天然优势，前者的创新水平可能较国有企业和大型企业低。但是也有研究表明，民营企业（$Soe=0$）和小型企业（$Size=0$）由于其获取外部融资难度的降低，创新市场预期回报的提高，会刺激

其进行创新活动[42]。因此,股权性质和企业规模的不同也会影响公共数据开放的创新效应,回归结果见表6的(1)列~(4)列。公共数据开放对民营企业和小型企业的创新活动的激励更大,与现有研究结论相同[42]。民营企业和小型企业通常面临数据资源匮乏的困境,相比大型企业和国有企业,其自身收集和积累数据的能力有限,且获取外部数据的渠道也相对较少。公共数据开放为其提供丰富的数据资源,能有效补充企业内部数据的不足,缓解信息不对称所带来的弊端,降低企业创新成本,激发企业创新动力。

(2)行业类型。作为引领未来科技创新和产业发展新方向的战略性新兴产业③,这些产业的企业具有知识技术密集、物质资源消耗少、成长潜力大、综合效益好等特点,其创新发展高度依赖数据资源。公共数据开放为战略性新兴产业的企业(*Seep*=1)提供更多市场信息及政府对产业未来发展的态度,企业从平台数据中精准识别市场需求、技术空白和竞争态势,减少创新决策中的不确定性,为企业进行创新行为打上一针"强心剂",结果见表6的(5)列和(6)列。公共数据开放共享更能显著促进战略性新兴产业企业的创新水平,符合理论预期。

**表6　企业特征的异质性分析**

| 变量 | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | *Soe*=1 | *Soe*=0 | *Size*=1 | *Size*=0 | *Seep*=1 | *Seep*=0 |
| | *lrd* | *lrd* | *lrd* | *lrd* | *lrd* | *lrd* |
| *open_city* | 0.005*** (4.954) | 0.008*** (11.920) | 0.005*** (9.034) | 0.008*** (9.724) | 0.007*** (5.129) | 0.006*** (11.837) |
| 控制变量 | 是 | 是 | 是 | 是 | 是 | 是 |
| 企业固定 | 是 | 是 | 是 | 是 | 是 | 是 |
| 年份固定 | 是 | 是 | 是 | 是 | 是 | 是 |
| *Constant* | 0.019 (0.381) | 0.071* (1.716) | 0.030 (0.786) | 0.064 (1.213) | 0.064 (0.695) | 0.038 (1.126) |
| *Obs* | 4136 | 10492 | 7239 | 7102 | 1357 | 13229 |
| Adjust $R^2$ | 0.811 | 0.830 | 0.872 | 0.833 | 0.784 | 0.836 |

注:括号内数值为回归系数的$t$值,标准误聚类到企业层面;***、**、*分别表示在1%、5%、10%的水平上显著。

**4.公共数据开放质量**

政府数据平台的上线是政府的行为表现,象征着公共数据向公众敞开大门。然而必须认识到公共数据开放共享还受到获取便捷性和数据开放的质量影响,这是衡量公共数据开放行为有效性的关键指标。如果获取开放数据质量较低或便捷度较差,只能是形式上的开放,如被列为"无条件开放"的数据却需要申请甚至申请后也无响应、对下载数据设置重重限制、数据文件无法打开或打开后无内容,平台数据出现曾经开放后又中断供给的现象,此类数据开放无法真正满足公众的需求。因此,采用复旦大学数字与移动治理实验室联合国家信息中心数字中国研究院发布的"中国开放数林指数④"(以下简称指数)进行补充分析,该指数的保障层和数据层可以满足本文的验证需求,在保障层中将大于中位数的值划分为强基础组(*sup*=1),小于中位数的值划分为弱基础组(*sup*=0);在数据层中将大于中位数的值划分为高质量组(*data*=1),小于中位数的值划分为低质量组(*data*=1)。同时,该指数也给出公共数据开放综合质量,基于此将样本划分为高综合质量组(*compre*=1)和低综合质量组(*compre*=0),结果见表7。可以看出,强基础、高质量数据和高综合质量的地区的公共数据开放共享对企业创新促进要更显著,原因在于当数据开放的基础较差或者质量较差时,企业难以获取其需要的资料和有用的信息,属于无效数据信息,无法对企业创新产生影响,符合上文的理论预期。

③ 战略性新兴产业涵盖新一代信息技术、生物技术、新能源、新材料、高端装备、新能源汽车、绿色环保及航空航天、海洋装备等产业。

④ "中国开放数林指数"是衡量中国地方政府数据开放水平的综合性指数。从数据可得性、数据易用性、数据规模、数据更新频率等多个维度对地方政府的开放数据进行评估。

**表 7　公共数据开放质量的异质性**

| 变量 | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | *sup* = 1 | *sup* = 0 | *data* = 1 | *data* = 0 | *compre* = 1 | *compre* = 0 |
| | *lrd* | *lrd* | *lrd* | *lrd* | *lrd* | *lrd* |
| *open_city* | 0.004*** | 0.000 | 0.007*** | 0.002* | 0.006*** | 0.000 |
| | (2.951) | (0.186) | (3.174) | (1.701) | (3.206) | (0.133) |
| 控制变量 | 是 | 是 | 是 | 是 | 是 | 是 |
| 企业固定 | 是 | 是 | 是 | 是 | 是 | 是 |
| 年份固定 | 是 | 是 | 是 | 是 | 是 | 是 |
| *Constant* | −0.081 | 0.098 | −0.145 | −0.119 | −0.144 | 0.092 |
| | (−0.538) | (0.381) | (−0.725) | (−0.920) | (−0.911) | (0.729) |
| *Obs* | 1477 | 792 | 2211 | 574 | 2337 | 802 |
| Adjust $R^2$ | 0.921 | 0.903 | 0.910 | 0.890 | 0.909 | 0.900 |

注：括号内数值为回归系数的 $t$ 值，标准误聚类到企业层面；***、**、* 分别表示在 1%、5%、10% 的水平上显著。中国开放数林指数从 2017 年开始有数据可以使用，故样本量减少。

## 五、机制检验

### （一）制度约束路径

创新能力的提高需要来自市场“看不见的手”和政府“看得见的手”来合理配置公共资源[43]。良好的制度性环境是企业创新的必要条件。本文从营商环境（*BE*）和政企关系（*Intimacy*）两个方面衡量制度性约束。在优化营商环境方面，公共数据开放能保障所有企业都有机会获取相同的基础数据，避免大型企业凭借自身资源优势形成资源垄断局面，使得中小企业也能在公平的环境下竞争。在改善政企关系方面，政府开放公共数据是一种积极信号，可以增强企业对政府的信任，激励企业进行长期的创新投入。营商环境数据来自北京大学开放研究数据平台发布的《中国城市营商环境评估数据库》。政企关系来自中国人民大学国家发展与战略研究院政企关系研究中心发布的地级市-政商关系健康指数，该指数主要包括亲近指数和清白指数两个方面，此处采用亲近指数表示政府和企业之间的关系，亲近指数主要包括政府对企业的关心、服务等内容。具体而言，大于样本均值赋值为 1，否则为 0。借鉴郑志强等[16]的做法构建交互项对模型进行估计，结果见表 8。（1）列表明两者的交互项系数正向显著，这表明公共数据开放对企业创新的促进效应会随着营商环境质量的改善而增强，即良好的营商环境能够强化公共数据开放的制度约束作用。换言之，公共数据开放通过优化营商环境，降低制度性摩擦和不确定性，为企业创新创造了更稳定和可预期的制度环境。（2）列表明交互项系数正向显著，但政企关系系数负向显著，这意味着单纯紧密的政企关系可能增加企业对政府依赖而抑制创新，但在公共数据开放情境下，政企关系反而能够转化为正向效应，显著增强公共数据开放对企业创新的促进作用。这说明公共数据开放通过提升政府回应性和透明度，改善了政企互动质量，使得政企关系从潜在的“寻租依赖”转变为“合作共赢”，从而进一步激励企业创新，假设 H2 得以验证。

**表 8　优化营商环境与改善政企关系**

| 变量 | (1) | (2) |
|---|---|---|
| | *lrd* | *lrd* |
| *open_city* | 0.0060*** | 0.0050*** |
| | (7.282) | (8.280) |
| *BE* | −0.0010 | |
| | (−1.170) | |
| *BE*×*open_city* | 0.0010* | |
| | (1.826) | |
| *Intimacy* | | −0.0010*** |
| | | (−2.586) |
| *Intimacy*×*open_city* | | 0.0030*** |
| | | (4.421) |
| 控制变量 | 是 | 是 |
| 企业固定 | 是 | 是 |
| 年份固定 | 是 | 是 |
| *Constant* | 0.0452 | 0.0444 |
| | (1.438) | (1.407) |
| *Obs* | 14687 | 14687 |
| Adjust $R^2$ | 0.8320 | 0.8330 |

注：括号内数值为回归系数的 $t$ 值，标准误聚类到企业层面；***、**、* 分别表示在 1%、5%、10% 的水平上显著。

### （二）要素供给路径

从三个角度衡量要素供给路径，分别是营运水平、研发不确定性和企业交流。参考沈坤荣等[44]的做法，营运水平选取营业资金周转率（*octo*）、固定资产周转率（*fato*）、营业周期（*buscyc*）来衡量。参考孙薇和叶初

升[42]的做法,研发不确定性(*uc*)选取以企业当年申请的发明专利中最终未被授权的比例来衡量。参考蔡庆丰等[45]的做法,企业交流(*commu*)采取企业各年他引次数加 1 的自然对数来衡量。表 9 的(1)列~(3)列结果显示,公共数据开放可通过提高营业资金周转率和固定资产周转率、降低企业的营业周期来提升企业的营运水平。(4)列结果显示,研发不确定性的系数为负且在 1%的显著性水平上显著,说明公共数据开放通过企业和政府之间的信息交互降低企业的研发不确定性。(5)列结果显示,企业交流的回归系数显著为正,说明公共数据开放可通过促进企业之间的交流,拓宽企业的知识宽度,假设 H3 得以验证。

**表 9　提升微观企业发展能力**

| 变量 | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | 营运水平 | | | 研发不确定性 | 企业交流 |
| | 营业资金周转率 | 固定资产周转率 | 营业周期 | *uc* | *commu* |
| | *octo* | *fato* | *buscyc* | | |
| *open_city* | 0.036***<br>(3.665) | 13.928*<br>(1.809) | −13.854**<br>(−2.492) | −0.031***<br>(−2.920) | 0.080***<br>(3.157) |
| 控制变量 | 是 | 是 | 是 | 是 | 是 |
| 企业固定 | 是 | 是 | 是 | 是 | 是 |
| 年份固定 | 是 | 是 | 是 | 是 | 是 |
| *Constant* | 1.983**<br>(2.297) | 128.300<br>(0.488) | 1.310<br>(0.002) | −1.123<br>(−1.322) | −3.842**<br>(−2.023) |
| *Obs* | 14687 | 14687 | 14687 | 14687 | 14687 |
| Adjust $R^2$ | 0.779 | 0.091 | 0.536 | 0.586 | 0.728 |

注:括号内数值为回归系数的 *t* 值,标准误聚类到企业层面;***、**、*分别表示在 1%、5%、10%的水平上显著。

## 六、进一步拓展分析

### (一)公共数据开放的“修正效应”

企业管理者在创新过程中发挥着重要的领导和推动作用[46-47]。创新具有风险性与不确定性,其大小取决于企业管理者对公司和市场的综合判断。因此,管理者的前瞻性与其对外部环境的感知影响着企业的创新决策。公共数据涵盖了市场长期趋势、消费者长期行为变化、宏观经济数据和政策动态数据等诸多方面。对内,企业管理者通过公共数据获取长期的消费数据,看到消费者对产品需求的变化,而不仅仅是关注短期的销售波动,使管理者能够从更长远的角度来规划企业的产品布局与创新决策,避免只追求短期利润而忽视长期市场需求,或只追求创新所带来的高额回报而忽视其巨大的机会成本的短视行为。对外,公共数据可作为管理者预测政策趋势的依据,通过分析公共数据中的经济增长数据、政府的财政支出方向等,企业管理者预测货币政策和财政政策可能的调整方向,降低因对经济政策感知迟钝而带来的成本。

为此,本节探究公共数据开放对于管理者自身能力是否具有修正效应。借鉴胡楠等[48]和聂辉华等[49]的做法,分别采取管理者短视主义(*mssight*)和经济政策不确定性感知(*uncert*)作为企业管理者对企业内部环境和外部环境所作出反应的代理变量,管理者短视主义和经济政策不确定性感知是利用文本分析法从 MD&A(management discussion and analysis)获取,结果见表 10。可以看出,公共数据开放与管理者短视主义、经济政策不确定性感知的交互项为负且均在 1%的显

**表 10　公共数据的修正效应**

| 变量 | (1) | (2) |
|---|---|---|
| | *lrd* | *lrd* |
| *open_city* | 0.008***<br>(12.740) | 0.008***<br>(12.568) |
| *open_city×mssight* | −0.025***<br>(−3.887) | |
| *mssight* | 0.013***<br>(2.963) | |
| *uncert* | | 0.001<br>(0.381) |
| *open_city×uncert* | | −0.006**<br>(−2.339) |
| 控制变量 | 是 | 是 |
| 企业固定 | 是 | 是 |
| 年份固定 | 是 | 是 |
| *Constant* | 0.052<br>(1.639) | 0.050<br>(1.586) |
| *Obs* | 14687 | 14687 |
| Adjust $R^2$ | 0.833 | 0.833 |

注:括号内数值为回归系数的 *t* 值,标准误聚类到企业层面;***、**、*分别表示在 1%、5%、10%的水平上显著。

著性水平显著，说明公共数据开放可以明显降低由管理者自身对企业外部环境和内部环境刺激所作出错误决定的概率，即验证了公共数据开放在促进企业创新的过程中对管理者能力具有“修正效应”，符合上文理论预期。

### （二）公共数据开放的“掩蔽效应”

本文的回归结果表明，总体上公共数据开放对于企业创新具有促进作用，但企业创新也取决于企业创新能力，不同创新能力的企业面对相同的外部冲击会有不同的表现。一般来说，企业创新能力越高，其创新的意愿也越高。即使创新具有不确定性与风险性，创新能力强的企业具有更先进的技术或更强大的知识转换能力，能够更加容易创新成功，降低创新的不确定性与风险性，从而带来更大的收益。相反，低创新能力企业通常面临知识吸收能力不足和技术整合能力缺乏的问题，即便公共数据开放提供了丰富的外部数据资源，这些企业仍可能难以将其有效转化为实际的创新成果。此外，低创新能力企业通常更依赖于传统的生产和经营模式，对创新活动的投入意愿相对较低。这些企业面临的资源限制、风险规避倾向及对短期回报的偏好，使得它们更关注如何利用外部数据提升现有运营效率，而非通过创新实现技术突破。因此，即使公共数据开放为其提供了支持，这些企业也未必能充分利用这些资源进行创新。图 2 展示了本文中上市公司创新能力样本数据的分布情况，可以看出在样本数据中大部分企业创新能力较弱，则这些企业进行创新活动的意愿不强。

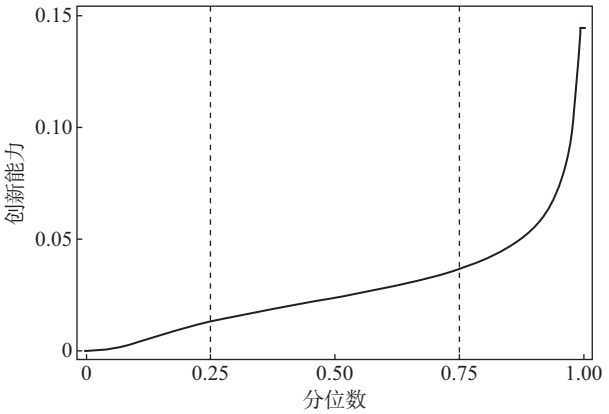



**图 2　企业创新分位数图**

“掩蔽效应”一词取自声学研究[50]。本文认为创新能力高的企业的创新效应掩盖了公共数据开放对于促进大部分企业创新的无效性，即公共数据具有掩蔽效应，其在促进高创新能力企业进行创新行为的同时掩盖了其对低创新能力企业创新的无效性。

为验证公共数据开放的掩蔽效应，采取因变量的 25 分位数样本和 75 分位数样本分别作为低创新能力企业（*lrd-low*）和高创新能力企业（*lrd-high*）的代理变量，如果公共数据开放促进高创新能力企业创新且没有或降低低创新能力企业的创新则认为公共数据具有“掩蔽效应”。与传统异质性分析不同，“掩蔽效应”不仅强调创新能力差异带来政策效果的异同，更进一步揭示企业个体行为与市场总体表现不匹配现象[50]。

**表 11　公共数据的掩蔽效应**

| 变量 | （1） | （2） |
|---|---|---|
| | *lrd-low* | *lrd-high* |
| *open_city* | 0.000<br>（1.587） | 0.012***<br>（6.104） |
| 控制变量 | 是 | 是 |
| 企业固定 | 是 | 是 |
| 年份固定 | 是 | 是 |
| *Constant* | −0.008<br>（−0.497） | −0.009<br>（−0.060） |
| *Obs* | 3522 | 3326 |
| Adjust $R^2$ | 0.711 | 0.766 |

注：括号内数值为回归系数的 $t$ 值，标准误聚类到企业层面；***、**、*分别表示在 1%、5%、10%的水平上显著。

从表 11 的（1）列可以看出，公共数据开放对于低创新能力企业的创新行为不显著。但从（2）列对高创新能力企业或高创新投入企业的创新行为在 1%的水平上显著，说明公共数据开放具有“掩蔽效应”，其对高创新能力企业创新的影响掩盖了其对促进大部分低创新能力企业的无效性。

## 七、结论与启示

### （一）研究结论

从基准回归结果来看，本文证实公共数据开放能够有效解锁数据红利，显著促进企业创新水平提升，该结论在经过一系列稳健性检验之后依然成立；从异质性分析结果来看，在数字经济政策供给强度高、社会信用高的地区，以及民营企业、小型企业和战略性新兴产业企业中公共数据开放对于企业创新的影响更大，且其影响只存在于基础良好和高质量数据的开放中；从机制分析结果来看，制度性约束层面，公共数据开放通

过优化营商环境、改善政企关系促进企业创新;要素供给层面,公共数据开放通过企业营运能力的提升、研发不确定性的降低与企业交流的增加促进企业创新;从进一步拓展分析来看,公共数据开放降低由管理者自身对企业外部环境和内部环境刺激所作出错误决定的几率,即公共数据开放具有“修正效应”;公共数据开放在促进高创新能力企业进行创新行为的同时掩盖了其对低创新能力企业创新的无效性,即公共数据开放具有“掩蔽效应”。

**(二)研究启示**

(1)加快推进数据要素开放共享和利用,以充分释放和有效捕获数据红利。在推进数字中国建设的过程中,应充分重视数据要素开放与利用的重要作用。将公共数据开放作为推动数字经济发展和企业创新的关键举措,纳入国家和地方的发展战略规划中。加大对公共数据开放基础设施建设投入,整合各部门、各层级的数据资源,建立统一的数据开放平台,打破数据孤岛,确保数据的互联互通和共享。制定符合中国国情的数据开放法律法规,明确数据开放的范围、标准、程序和保障措施,为公共数据的有效利用和数据安全提供坚实的法律基础。鼓励政府部门与企业、科研机构等建立紧密的合作关系,共同推动公共数据的创新应用。通过举办数据应用大赛、设立创新孵化基地等方式,激发企业和社会各界对公共数据的创新活力,挖掘数据潜在价值。

(2)重视数据开放的基础与质量。由于对数据开放重视程度低和开放数据质量差的地区,公共数据开放对企业创新行为较难产生激励效应,陈旧的数据和烦琐的程序在一定程度上加大了企业进行创新的难度。为此,一是在推进公共数据开放过程中建立严格的数据质量管理体系。制定数据采集、存储、处理、开放等各环节的质量标准和规范,加强数据审核和校验工作,确保数据的准确性、完整性、一致性和时效性。二是加强对数据开放平台的运营管理,建立数据更新机制,确保数据的及时更新和动态调整。同时,提高数据开放的透明度和可解释性,向公众提供数据的来源、用途、处理方法等详细信息,增强公众对数据的信任度。

(3)制定差异化的数据开放共享政策措施。由于公共数据开放因企业所处的外部环境和企业特征的差异会产生不同的刺激效果,应制定差异化的数据开放共享政策措施,充分挖掘数据要素应用场景和潜在功能,优化资源配置效率。对于民营企业和小型企业,由于其在数据获取和创新资源方面相对薄弱,应给予更多的政策倾斜和扶持。对于国有企业和大型企业,应鼓励其发挥规模优势和资源整合能力,加强与中小企业的合作,共同推动产业创新发展。支持国有企业建立行业数据共享平台,与上下游中小企业共享数据资源,开展协同创新项目,带动产业链整体升级。同时,引导国有企业加大对新兴技术领域的研发投入,发挥引领示范作用,推动行业创新发展。

(4)优化宏观市场环境,提升微观企业发展能力,提高公共数据利用效率。优化宏观市场环境和提升微观企业发展能力是公共数据开放助力企业创新的两个重要渠道。为此,一是着力优化宏观市场环境,加强市场监管,打破地方保护主义,建立公平竞争的市场秩序。简化行政审批流程,提高政务服务效率,降低企业制度性交易成本。二是加大对企业创新的支持力度,通过财政补贴、税收优惠、金融扶持等政策手段,鼓励企业增加研发投入,提升自主创新能力。

(5)建立动态监测与评估调整机制确保数据开放有效性。一是建立公共数据开放共享动态监测体系,实时跟踪数据开放共享的实施效果。通过收集企业创新指标、市场反馈、社会经济效益等多方面的数据,全面评估其有效性和存在的问题。二是根据监测评估结果,及时调整和优化公共数据开放政策及相关配套措施。对于政策实施过程中发现的问题,如数据开放范围不合理、数据质量不达标、企业利用困难等,应迅速采取针对性措施加以改进。对于某些行业对特定类型数据需求较大但开放不足,应及时调整数据开放策略,增加相关数据的供给确保能适应企业创新发展的实际需求,持续发挥推动企业创新的乘数效应。

## 参考文献

[1] 彭远怀. 政府数据开放的价值创造作用:企业全要素生产率视角[J]. 数量经济技术经济研究, 2023, 40(9): 50-70.

[2] KANG J. An analysis on the factors affecting of big data utilization in local government: Focused on the public servants perception in Seoul[J]. Journal of Governance Studies, 2017, 12(1): 161-197.

[3] 宋建，郑江淮. 中国企业创新与劳动技能升级：基于生产率频谱效应视角[J]. 世界经济，2022，45(10)：28-57.
[4] CUMMING B S. Innovation overview and future challenges[J]. European Journal of Innovation Management，1998，1(1)：21-29.
[5] DAHLANDER L，GANN D M. How open is innovation?[J]. Research Policy，2010，39(6)：699-709.
[6] 周铭山，张倩倩，杨丹. 创业板上市公司创新投入与市场表现：基于公司内外部的视角[J]. 经济研究，2017，52(11)：135-149.
[7] 吴超鹏，唐苟. 知识产权保护执法力度、技术创新与企业绩效——来自中国上市公司的证据[J]. 经济研究，2016，51(11)：125-139.
[8] 叶永卫，余田田，陶云清，等. 政府数据要素共享的稳投资效应：来自公共数据开放平台的证据[J]. 数量经济技术经济研究，2025，42(1)：136-156.
[9] 王海，叶帅，尹俊雅. 公共数据开放如何提振企业有效投资——基于产能利用视角[J]. 中国工业经济，2024(8)：137-153.
[10] 蓝发钦，胡晓敏，徐卓琳. 公共数据开放能否拓展资本跨区域流动距离——基于异地并购视角[J]. 中国工业经济，2024(9)：156-174.
[11] 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.
[12] 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.
[13] 蔡运坤，周京奎，袁旺平. 数据要素共享与城市创业活力——来自公共数据开放的经验证据[J]. 数量经济技术经济研究，2024，41(8)：5-25.
[14] 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.
[15] 李泽宇，王雪方，陈新芳. 政府数据开放对企业创新的影响效应及机制研究[J]. 科研管理，2024，45(7)：11-20.
[16] 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.
[17] 沈坤荣，林剑威. 链“岛”成“陆”：公共数据开放的技术创新效应研究[J]. 管理世界，2025，41(2)：83-104.
[18] 孙国良. 政府数据开放对中国高技术企业自主创新的影响[J]. 商业研究，2025(1)：11-19.
[19] 李长英，王曼，张帅. 公共数据开放对民营企业创新的影响[J]. 广东财经大学学报，2025，40(2)：4-21.
[20] 谢贞发，冉芙小. 政府数据开放对企业创新投入的影响——兼论政府数据开放体系建设[J]. 财政科学，2025(1)：29-50.
[21] CHEN Q，DING X，WANG T. Can open government data improve urban innovation? Empirical research from a dual perspective in China[J]. Government Information Quarterly，2025，42(3)：102059.
[22] ARROW K J. Economic welfare and the allocation of resources for invention[M]. Princeton：Princeton University Press，1962.
[23] JONES C I，TONETTI C. Nonrivalry and the economics of data[J]. American Economic Review，2020，110(9)：2819-2858.
[24] NORTH D C. Institutions，institutional change and economic performance[M]. Cambridge：Cambridge University Press，1990.
[25] PARK S，GIL-GARCIA J R. Open data innovation：Visualizations and process redesign as a way to bridge the transparency-accountability gap[J]. Government Information Quarterly，2022，39(1)：101456.
[26] 肖文，林高榜. 政府支持、研发管理与技术创新效率——基于中国工业行业的实证分析[J]. 管理世界，2014，30(4)：71-80.
[27] LIU D. Credit corruption，government relation and corporate innovation[J]. Finance Research Letters，2023，58：104408.
[28] MA R，GUO F，LI D. Can public data availability affect stock price crash risk? Evidence from China[J]. International Review of Financial Analysis，2024，94：103270.
[29] FARBOODI M，VELDKAMP L. A model of the data economy[R]. Cambridge，MA：National Bureau of Economic Research，2021.
[30] CHEN K，ZHANG S. How does open public data impact enterprise digital transformation?[J]. Economic Analysis and Policy，2024，83：178-190.
[31] GOLDFARB A，TUCKERR C. Digital economics[J]. Journal of Economic Literature，2019，57(1)：3-43.
[32] LIN R，XIE Z，HAO Y，et al. Improving high-tech enterprise innovation in big data environment：A combinative view of internal and external governance[J]. International Journal of Information Management，2020，50：575-585.
[33] GIBSON J，OLIVIA S，BOE-GIBSON G，et al. Which night lights data should we use in economics，and where?[J]. Journal of Development Economics，2021，149：102602.
[34] BACON E，WILLIAMS M D，DAVIES G. Coopetition in innovation ecosystems：A comparative analysis of knowledge transfer configurations[J]. Journal of Business Research，2020，115：307-316.
[35] 徐翔，赵墨非，李涛，等. 数据要素与企业创新：基于研发竞争的视角[J]. 经济研究，2023，58(2)：39-56.
[36] 赖烽辉，李善民. 共同股东网络与国有企业创新知识溢出——基于国有企业考核制度变迁的实证研究[J]. 经济研究，2023，58(6)：119-136.
[37] ATHEY S，IMBENS G W. Design-based analysis in difference-in-differences settings with staggered adoption[J]. Journal of Econometrics，2022，226(1)：62-79.
[38] DE CHAISEMARTIN C，D'HAULTFOEUILLE X. Two-way fixed effects estimators with heterogeneous treatment effects[J]. American Economic Review，2020，110(9)：2964-2996.
[39] CALLAWAY B，SANT'ANNA P H C. Difference-in-differences with multiple time periods[J]. Journal of Econometrics，2021，225(2)：200-230.

[40] LAPORTA R, LOPEZ-DE-SILANEs F, SHELEIFER A, et al. Trust in large organizations[J]. American Economic Review, 1996, 87(2): 333-338.
[41] XIE Q, WU X, DAI Q, et al. An integrated data envelopment analysis and non-cooperative game approach for public transportation incentive subsidy allocation[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 21515-21530.
[42] 孙薇, 叶初升. 政府采购何以牵动企业创新——兼论需求侧政策"拉力"与供给侧政策"推力"的协同[J]. 中国工业经济, 2023(1): 95-113.
[43] 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.
[44] 沈坤荣, 乔刚, 林剑威. 智能制造政策与中国企业高质量发展[J]. 数量经济技术经济研究, 2024, 41(2): 5-25.
[45] 蔡庆丰, 王仕捷, 刘昊, 等. 城市群人口集聚促进域内企业创新吗[J]. 中国工业经济, 2023(3): 152-170.
[46] 汪丽, 茅宁, 龙静. 管理者决策偏好、环境不确定性与创新强度——基于中国企业的实证研究[J]. 科学学研究, 2012, 30(7): 1101-1109, 1118.
[47] 唐慧洁, 罗瑾琏, 易明, 等. 科创企业创新悖论的主动应对机制——基于管理者角色身份建构的探索性案例研究[J]. 管理世界, 2023, 39(10): 189-205.
[48] 胡楠, 薛付婧, 王昊楠. 管理者短视主义影响企业长期投资吗? ——基于文本分析和机器学习[J]. 管理世界, 2021, 37(5): 139-156, 11, 19-21.
[49] 聂辉华, 阮睿, 沈吉. 企业不确定性感知、投资决策和金融资产配置[J]. 世界经济, 2020, 43(6): 77-98.
[50] HE X, DAI W. The masking effect of green innovation: A study based on carbon market shocks[J]. Energy Economics, 2025, 141: 108035.

# Unlocking Data Dividends: How Data Element Openness Drives Corporate Innovation—Evidence from the Launch of Public Data Platforms

Huang Guixin[1, 2, 3], Qian Juan[1, 2, 3], Zhu Zhimeng[1], Guo Xueying[1, 2, 3]

(1. School of Economics and Management, Xinjiang University, Urumgi 830046, China; 2. Engineering Research Center of Northwest Energy Carbon Neutrality (ERCNECN), Ministry of Education, Urumgi 830046, China; 3. Institute for Macroeconomy High-Quality Development of Xinjiang, Urumgi 830046, China)

**Abstract**: Against the backdrop of data elements being elevated to a national foundational strategic resource, the economic effects of public data openness, a unique institutional innovation in China, are deconstructed. This serves as a crucial lens for explaining the dialectical relationship between an "active government and an effective market" in the digital economy era. Utilizing data on the launch of government data platforms in 217 prefecture-level cities in China, a multi-period difference-in-differences (DID) identification framework was constructed under a quasi-natural experiment setting. The structural reshaping mechanism of public data element sharing on corporate innovation behavior was empirically examined at the causal inference level. It is found that public data openness can promote corporate innovation. From the perspective of institutional constraints, public data openness facilitates corporate innovation by optimizing the business environment and improving government-enterprise relations. From the perspective of factor supply, public data openness promotes corporate innovation through the enhancement of corporate operational capabilities, increased inter-firm exchanges, and reduced R&D uncertainty. The impact of public data openness on corporate innovation is found to be greater in regions with high digital economy policy supply intensity and high social credit, as well as for private enterprises, small enterprises, and strategic emerging industry enterprises. This effect is also observed only in the openness of well-established and high-quality data. Further analysis reveals that public data has a corrective effect on the decision-making behavior of corporate managers, while it exhibits a masking effect on corporate innovation behavior. These findings provide theoretical underpinnings for explaining the institutional advantages of data governance in the process of Chinese modernization and contribute an Eastern wisdom solution to global digital economic development.

**Keywords**: open public data; enterprise innovation; correction effect; masking effect; factor supply; institutional constraints