

# 数据要素流通与颠覆性创新: 数据要素质量、边际生产力的视角

王海森<sup>1,2</sup>, 李纲<sup>1,2</sup>, 赵浩权<sup>3</sup>, 梁镇涛<sup>1,2</sup>

- (1. 武汉大学信息资源研究中心, 湖北 武汉 430072;
2. 武汉大学信息管理学院, 湖北 武汉 430072;
3. 武汉大学经济与管理学院, 湖北 武汉 430072)

**摘要:** 数据要素成为最具时代特征的生产要素, 其价值释放取决于能否有效流通。本文将数据要素存量区分为数据要素生成量和数据要素交易量, 纳入包含颠覆性创新的动态随机一般均衡模型中, 并采用 N 元语法模型词性模板、大语言模型语义分析结合词频—逆文本频率指数/点互信息方法抽取技术术语, 通过维基百科词条目录验证技术特征, 测算企业颠覆性创新。而后借助交错双重差分模型, 探讨数据要素市场化如何重构现有技术创新范式。结果表明, 数据要素市场化降低了数据交易摩擦, 扩大了企业数据资源转化为有效数据要素的比例, 减缓了因数据要素原始积累而陷入数据要素陷阱现象。这一过程主要通过数据要素质量和数据要素边际生产力两种路径发挥作用。为保障数据要素市场化的创新效应有效发挥, 关键在于加强高质量数据供给和夯实数字基础设施建设。本文研究结果为激励技术创新路径提供了新机制, 为完善数据要素市场提供了新启示。

**关键词:** 数据要素市场化; 数据要素质量; 数据要素边际生产力; 颠覆性创新

**中图分类号:** G350      **文献标识码:** A      **文章编号:** 1005-0566(2025)11-0019-14

## Data element circulation and disruptive innovation: From the perspective of data element quality and marginal productivity

WANG Haisen<sup>1,2</sup>, LI Gang<sup>1,2</sup>, ZHAO Haoquan<sup>3</sup>, LIANG Zhentao<sup>1,2</sup>

- (1. Centre for Studies of Information Resources, Wuhan University, Wuhan 430061, China;
2. School of Information Management, Wuhan University, Wuhan 430061, China;
3. Economics and Management School, Wuhan University, Wuhan 430061, China)

**Abstract:** Data elements have become the most characteristic production factors of the times, and their value release depends on whether they can be effectively circulated. This article divides the stock of data elements into data element generation volume and data element transaction volume, and incorporates them into a dynamic stochastic general equilibrium model that includes disruptive innovation. The N-gram model part of speech template, big language model semantic analysis, and word frequency inverse text frequency index/point mutual information method are used to extract technical terms. The technical features are verified through the Wikipedia entry directory, and the disruptive innovation of enterprises is calculated. Then, using the interleaved double difference model, we explored how the marketization of

收稿日期: 2025-08-29      修回日期: 2025-11-05

**基金项目:** 教育部人文社会科学研究青年基金项目(25YJC790099); 博士后科学基金面上项目(2025M773690); 国家资助博士后研究人员计划(GZC20252439); 中央高校基本科研业务费专项资金(1206-413000345); 国家自然科学基金专项项目(72442030)。

**作者简介:** 王海森(1997—), 男, 安徽安庆人, 武汉大学信息管理学院助理研究员, 经济学博士, 研究方向为数字经济、技术创新。

data elements can reconstruct the existing technological innovation paradigm. The results indicate that the marketization of data elements reduces data transaction friction, expands the proportion of enterprise data resources converted into effective data elements, and slows down the phenomenon of falling into data element traps due to the original accumulation of data elements. This process mainly works through two paths: data element quality and data element marginal productivity. To ensure the effective utilization of innovative effects in the marketization of data elements, the key lies in strengthening the supply of high-quality data and consolidating the construction of digital infrastructure. This article provides a new mechanism for incentivizing technological innovation paths and offers new insights for improving the data element market.

**Key words:** data element marketization; data element quality; marginal productivity of data elements; disruptive innovation

数据要素已跃居为当今数字经济时代的核心基石之一,与劳动力、资本、技术、土地等传统生产要素并立,共同构筑起推动经济社会发展的新引擎。当特斯拉通过海量车辆数据训练出纯视觉自动驾驶方案时,当 DeepMind 利用数百万篇生物学文献构建 AlphaFold3 时,这些颠覆性创新都在印证一个真理:数据的自由流动是数字时代创新的“氧气”。然而,我国作为全球数据资源最丰富的国家之一,却在数据流通领域面临“富矿难掘”的困境。工业互联网平台的大部分数据处于“沉睡”状态,政府公共数据开放程度和利用水平仍有不足,企业间数据交易成本过高。数据要素的流通悖论,正在成为我国颠覆性创新突破的关键瓶颈。以往研究已深刻洞见数据要素在加速信息流转效率、催化产业结构跃迁以及激发创新潜能方面的核心价值<sup>[1-2]</sup>,不仅重塑了生产关系的内涵,更从根本上奠定了数据作为战略性资源在生产要素体系中的核心与引领地位。但在实践中,数据要素流通却面临诸多挑战。一方面,由于资本投入主体更倾向于自用数据要素,而非促进流通<sup>[3]</sup>;另一方面,数据要素在市场化配置过程中的增值路径尚不明朗,赋能应用的深度与广度均有待提升,掣肘了数据对经济社会发展的推动作用<sup>[4]</sup>。因此,学术界主要关注数据要素市场生态功能的识别定位问题、互联互通的价值协同问题以及合规健康的多边共治问题<sup>[5-6]</sup>,而对于数据要素流通可能造成的创造性破坏以及如何变局等问题尚未进行充分讨论。当数据要素突破企业边界、跨越行业壁垒,甚至跨越国界,将如何颠覆传统的创新范式?在这一变革的理论逻辑背后,既蕴含着培育新质生产力的历史机遇,也暗藏着颠覆传统

产业生态的巨大潜力。

阿罗替代效应揭示了大企业的泊松流密度较高,在短期内通过迭代式创新可以吸纳的用户数量更多<sup>[7-8]</sup>,在此过程中生成的数据要素又能反过来增加大企业进行迭代式创新的能力,形成逻辑闭环,逐渐挤出小企业的市场份额,造成“大企业更大而小企业更小”的情形<sup>[9-10]</sup>。当数据要素成为新兴生产要素时,进一步加剧了阿罗替代效应,形成了“数据要素存量—泊松流密度—迭代式创新—新数据要素”的数据要素“陷阱”闭环<sup>[11-12]</sup>。数据要素市场化进一步降低了企业获取数据要素的门槛,按照阿罗替代效应的理论逻辑,这会进一步为大企业带来短期内的稳定收益与产值增长,但也激励小企业寻求新的创新路径与生存策略,在一定程度上会激发颠覆性创新的动机。因此,构建一个不依赖于既有假设,而是融合快速发展的数字技术管理理论新框架,对于深入理解数据要素流通如何赋能颠覆性创新尤为重要。

本文探索性地将企业数据要素分为数据要素原始积累部分、数据要素交易获取部分以及数据要素折旧,而后代入包含颠覆性创新的动态随机一般均衡模型中,刻画了数据要素市场化对数据要素陷阱的缓解作用。结果表明,数据要素市场化虽然增加了企业的数据要素存量,但却并未引发数据要素陷阱,数据要素交易量增加在促进企业数据要素积累的同时,还通过影响产品质量增加了企业最终产出,导致企业单位应用技术价值增加,提高了企业对进一步技术创新的需求,创新所带来的生产效率提升扩大了生产活动,也增加了企业从生产活动中积累(生成)的数据要素,即

便考虑一定程度的数据要素折旧,数据交易也通过激励企业颠覆性创新促进了数据要素的积累。最终形成“数据要素原始积累/数据要素交易量—产品质量—最终产出—技术创新价值与需求—技术创新—最终产出—新数据要素积累—原始数据要素积累”的螺旋式创新范式闭环。为解决数据要素流通领域的现实问题,应当以促进数据要素合规高效流通使用为主线,聚焦于加强高质量数据供给、夯实数字基础设施建设两条路径,推动完善数据要素市场化进程。

与既有文献相比,本文的边际贡献为:第一,为拟合场内数据交易赋能企业颠覆性创新的运作模式,本文基于以往研究<sup>[13]</sup>,进一步将数据交易以及数据要素积累纳入包含颠覆性创新的动态随机一般均衡模型中,讨论了数据生成、积累和折旧的波动性以及过渡动态中的周期性影响,从理论层面回答了数据要素实现价值释放的内在机理,验证了数据要素市场化在加剧阿罗替代效应后的破局思路,拓展和深化了已有理论研究;第二,本文从宏观视角切入,定量测算了数据要素在企业端的配置效率,用以阐释数据要素市场化何以赋能企业颠覆性创新的深层原因,验证了创造性破坏也能成为颠覆性创新的“催化剂”,既对要素资源配置效率领域研究的重要补充,也为政府进一步优化数据要素市场提供了经验证据;第三,本文基于中国全量发明专利文本数据、被引数据、新闻媒体报道数据、企业招聘数据,依托维基百科页面这一外部知识源,运用跨数据源融合分析、技术生命周期模型、专利“兴起年”等方法与概念,采用自动化算法和人工审核协同的方式,最大限度地保留了更广泛和更详细的颠覆性技术信息,最终整理得出一套完整的企业颠覆性创新测算方法体系。

### 一、理论模型

本文将数据交易与数据要素积累纳入包含颠覆性创新的动态随机一般均衡模型中,以研究数据要素市场化对企业的数据交易与颠覆性创新的影响。企业需要借助数据要素提升自身的产品质

量,进而扩大生产收益。数据要素的积累一部分来自自身生产活动的生成,另一部分来自数据交易。然而,由于存在交易成本等,企业购买的数据无法完全转化为有效的数据要素,数据的交易存在摩擦。在借助数据要素提升产品质量之外,企业还可以通过颠覆性技术创新增加其新产品品类以扩大生产收益,而生产收益的提升也激励了企业进一步的颠覆性创新。因此,数据要素能通过提高产品质量扩大技术创新的收益,进而促进企业的颠覆性创新,而企业因颠覆性创新促进的生产规模也能促进数据要素的积累。本文在此部分尝试在一般均衡框架下探讨企业的数据交易、数据要素积累与颠覆性创新之间的联动,并研究当数据要素市场化降低数据交易摩擦时,会对企业数据交易与颠覆性创新产生何种影响,主要涉及的部门为代表性家庭与企业。

#### (一)代表性家庭

家庭部门提供低技能与高技能两类劳动力,其中低技能劳动力  $L_t^l$  在企业的生产部门从事生产工作,高技能劳动力  $L_t^s$  在颠覆性创新部门负责技术研发与应用。此外,家庭作为企业客户,在消费企业产品的同时,也通过数据交易市场为企业提供数据信息<sup>①</sup>。代表性家庭通过选择自身消费  $C_t$  与安全资产  $B_t$ , 提供劳动力  $L_t^l$ 、 $L_t^s$  与自身数据信息  $D_t$  以最大化终身贴现效用。其效用函数为:

$$U_t = \max_{C_t, L_t^l, L_t^s, B_t, D_t, K_t} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \ln C_t + \zeta_t B_t - \varphi^l \frac{(L_t^l)^{1+\sigma_L}}{1+\sigma_L} - \varphi^s \frac{(L_t^s)^{1+\sigma_L}}{1+\sigma_L} - \frac{\kappa_d}{2} D_t^2 \right\} \quad (1)$$

式(1)中,  $\beta$  是家庭部门贴现因子;  $\varphi^l$  和  $\varphi^s$  分别是两类劳动供给参数;  $\sigma_L$  为劳动力的 Frisch 替代弹性的倒数。

参考以往研究<sup>[14]</sup>,  $\zeta_t$  为家庭的流动性需求冲击。当  $\zeta > 0$  时,会产生流动性溢价,导致安全资产收益率与资本收益率之间存在利差。此外,参

<sup>①</sup> 本文将数据交易市场刻画为家庭与企业之间的数据交易媒介,由于数据交易并未完全市场化,市场交易过程中存在数据交易成本,其主要作用在于体现交易过程中的摩擦。

考以往研究<sup>[15-16]</sup>,家庭在提供数据信息时,会因泄露隐私而损失效用。其预算约束式为:

$$C_t + B_t + Q_t K_t = R_{t-1} B_{t-1} + R_t^k Q_{t-1} K_{t-1} + w_t L_t + w_t^s L_t^s + Q_t^d D_t \quad (2)$$

式(2)中,  $Q_t$  和  $Q_t^d$  分别为资本品价格与数据信息价格,作为企业客户,家庭提供自身的数据信息以获取免费的数字服务,企业可以借助这些客户信息更新数据库,进而提升自身产品质量<sup>②</sup>。

令  $u_c$  为家庭最优化问题的拉格朗日乘子,决定最优数据交易量  $D_t$  的一阶条件为<sup>③</sup>:

$$\kappa_d D_t = u_c Q_t^d \quad (3)$$

## (二) 企业部门

企业部门中包括生产部门与颠覆性创新部门,分别负责制定生产决策和技术研发与应用。生产部门除了需要购买资本品  $K_{m,t}$ 、雇佣劳动力  $L_{m,t}$  以生产中间品外,还需要在数据交易市场购买数据信息以形成数据要素,用于提高自身产品质量<sup>④</sup>。

### 1. 生产部门

生产部门包括中间品部门与最终品部门,分别负责中间品的生产与最终品的合成,并决定两类最终品的相对价格。每个企业对单类中间品  $j$  的生产函数为 CD 函数形式:

$$Y_{m,t}(j) = A_t [K_{m,t}(j)]^\alpha [L_{m,t}(j)]^{1-\alpha} \quad (4)$$

式(4)中,  $A_t$  代表企业的产品质量,参考以往研究的形式<sup>[16]</sup>。产品质量与企业的要素存量正相关:  $A_t = \Omega_t^\eta$ 。其中,  $\Omega_t$  为企业的数据要素存量;  $\eta$  为数据要素提升产品质量的边际效率,  $0 < \eta < 1$  代表其边际效益递减。由于数据存在非竞

争属性,数据要素可以为所有中间品提升产品质量,用于最终品的合成。数据要素的积累形式为:

$$\Omega_t = (1 - \delta^d) \Omega_{t-1} + x_t + \iota_t D_t \quad (5)$$

每期有  $\delta^d$  比率的数据要素因数据质量、时效性等原因而淘汰,而  $x_t = \omega K_t^\alpha L_t^{1-\alpha}$  代表新的数据要素积累:一部分来自经济活动生成的有效数据,  $\omega$  代表能够有效生成数据要素的比例<sup>⑤</sup>;另一部分来自企业在数据交易市场从数据所有者购买的数据信息,可能来自产品客户或者其他企业。参考以往研究<sup>[17]</sup>,由于存在数据交易壁垒、交易成本等原因,企业购买的单位数据信息  $D_t$  只有  $\iota < 1$  的比例能够转化为有效提升产品质量的数据要素<sup>⑥</sup>。

企业根据式(4)生产中间品  $Y_{m,t}(j)$  后,中间品部门对同类中间品进行合成。当同类中间品同质时,每个企业的中间品数量决定为:

$$Y_{m,t} = \left( \int_0^{T_t} [Y_{m,t}(j)]^{\frac{1}{\theta}} dj \right)^\theta = T_t^\theta Y_{m,t}(j) \quad (6)$$

式(6)中,企业应用技术存量  $T_t$  可以增加成功应用于生产的中间品品类,代表技术的研发与应用可以扩大企业的中间品产出与收益。

参考以往研究<sup>[18]</sup>,每个企业  $i$  采用  $Y_{m,t}(i)$  单位中间品以生产最终品:  $Y_t(i) = Y_{m,t}(i)$ 。假设所有企业为一个测度为 1 的连续统,当企业的同类中间品同质时,最终品的加总表示为:

$$Y_t = \left( \int_0^1 [Y_t(i)]^{\frac{1}{\theta}} di \right)^\theta = Y_{m,t} = T_t^\theta Y_{m,t}(j) \quad (7)$$

求解利润最大化问题可以得到中间品价格与最终品价格的关系为  $p_{m,t} = T_t^{\theta-1} p_t$ , 定义有效劳动力与有效资本品为:

② 参考文献<sup>[17]</sup>,家庭从数据交易中所获得的收益可以理解为在数字消费中的便利,比如免费的 App,而代价是作为产品客户需要提供一定消费数据信息。

③ 家庭部门以及资本品生产商的最优化问题求解参见附录。

④ 企业积累数据信息形成数据要素有利于帮助企业更好地预测客户需求、选择性投资并优化生产路径,进而提高产品质量并扩大生产规模与收益。

⑤ 与以往研究一致<sup>[17]</sup>,此部分数据要素由生产活动直接生成,与产品质量或者生产效率无关。

⑥ 这里的  $\iota$  也可以视为因交易成本等原因,企业购买单位数据需要付出额外的成本。此外,  $\iota < 1$  也代表着因信息不对称等交易摩擦或者由于场外交易的原始数据资源往往只有较低比例可以转化为高质量数据要素,交易数据中有效信息比例较低,只有部分能提炼出并转化为数据要素。例如,企业购买数据库后,其中数据的开发利用还面临原始数据结构复杂、数据安全难保障等问题,只有其中较少比例能有效转化为数据要素用于提高产品质量。

$$L_t = \int_0^T L_{m,t}(j) dj = T_t L_{m,t}(j),$$

$$K_t = \int_0^T K_{m,t}(j) dj = T_t K_{m,t}(j) \quad (8)$$

结合中间品生产函数式(4),可得到企业加总最终产出为:

$$Y_t = T_t^{\theta-1} A_t K_t^\alpha L_t^{1-\alpha} \quad (9)$$

由式(9),除颠覆性创新部门的应用技术  $T_t$  以外,企业的数​​据交易也通过数据要素存量  $\Omega_t$  提升产品质量,进而扩大了产出。因此,企业的生产部门需要将交易数据  $D_t$  作为生产中除资本、劳动外的第三种投入品纳入生产决策,在式(9)约束下解决成本最小化问题:

$$\min_{K_t, L_t, D_t} r_t^k K_t + w_t L_t + Q_t^d D_t \quad (10)$$

可以得到三种投入品的边际成本决定的一阶条件为:

$$r_t^k = \alpha \frac{Y_t}{K_t} (1 + \eta \omega A_t^{-1/\eta} T_t^{\theta-1} K_t^\alpha L_t^{1-\alpha}) \quad (11)$$

$$w_t = (1 - \alpha) \frac{Y_t}{L_t} (1 + \eta \omega A_t^{-1/\eta} T_t^{\theta-1} K_t^\alpha L_t^{1-\alpha}) \quad (12)$$

$$Q_t^d = \eta \iota_t A_t^{-1/\eta} Y_t \quad (13)$$

由式(11)和式(12),因为生产活动能够促进数据要素的积累,资本租金率和生产劳动力工资在生产活动中的边际产出之上有额外的加成,其中  $\eta \omega A_t^{-1/\eta} T_t^{\theta-1} K_t^\alpha L_t^{1-\alpha}$  代表资本与劳动力通过数据要素对产出的额外边际贡献比例。式(13)左侧  $Q_t^d$  代表单位数据价值,右侧代表数据交易  $D_t$  对产出的边际贡献。当数据交易摩擦降低( $\iota$  提高)时,单位数据交易能积累更多数据存量,进而提升数据价值。

## 2. 颠覆性创新部门

颠覆性创新部门通过技术研发与应用,在生产过程中采纳全新技术组合引入了新的中间品类别。此类创新不仅优化了既有生产流程,更可能替代旧有要素、重构生产结构,具备改变行业技术路线和市场格局的潜力。因此,其创新过程打破了原有的生产路径依赖,带来生产率的提升,符合“颠覆性创新”的实质特征。该部门包括创新技术

研发部门(R&D部门)与创新技术应用部门两个子部门,分别负责颠覆性创新技术的研发与应用,并决定研发型劳动力  $L_{r,t}^s$  和应用型劳动力  $L_{\alpha,t}^s$  的雇佣问题。

(1) 创新技术研发部门(R&D部门)。创新技术研发部门雇佣高技能劳动力中的研发型劳动力  $L_{r,t}^s$  用于创造研发技术  $RD_t$ , 每单位研发型劳动力创造的研发技术为  $\varphi_t = \chi RD_t (L_{r,t}^s)^{\rho_z-1}$ 。其中,  $\chi$  为该部门特有的研发生产技术,新研发技术的创造与研发技术存量  $RD_t$  正相关,体现了技术研发过程中的“干中学”效应。因此,创新技术研发部门每一期创造出的新研发技术为:

$$\varphi_t L_{r,t}^s = \chi RD_t (L_{r,t}^s)^{\rho_z} \quad (14)$$

式(14)中,  $\rho_z$  为技术研发的劳动力弹性,  $\rho_z < 1$  代表研发型劳动力  $L_{r,t}^s$  的边际效率递减。

创新技术研发部门需要最大化研发技术利润以决定研发型劳动力的雇佣水平,令  $A_{t,t+1} = \beta uc_{t+1}/uc_t$  为主观贴现因子,其最优化问题为:

$$\max_{L_{r,t}^s} E_0 \{ A_{t,t+1} J_{t+1} \varphi_t L_{r,t}^s - w_t^s L_{r,t}^s \} \quad (15)$$

式(15)中,  $J_t$  代表单位研发技术的价值。得到  $L_{r,t}^s$  的一阶条件为:

$$A_{t,t+1} J_{t+1} \chi RD_t (L_{r,t}^s)^{\rho_z-1} \quad (16)$$

每一期研发技术与应用技术存在淘汰率  $\rho^T$ , 因此研发技术存量的积累形式为:

$$RD_{t+1} = (1 - \rho^T) RD_t + \varphi_t L_{r,t}^s \quad (17)$$

(2) 创新技术应用部门。创新技术应用部门雇佣高技能劳动力中的应用型劳动力  $L_{\alpha,t}^s$  用于将研发技术  $RD_t$  转化为应用技术  $T_t$ 。一项新技术成功应用到生产中需要时间,因此每一期只有一定比率的研究技术被转化为应用技术,单位研发技术的应用率为  $\lambda_t = \kappa_\lambda (RD_t L_{\alpha,t}^s)^{\rho_\lambda}$ , 其中  $\kappa_\lambda$  为技术应用效率,  $\rho_\lambda < 1$  代表应用技术的转化边际效率递减。

当研发技术成功转化为应用技术  $T_t$  后,能够增加中间品品类数量,即提升中间品的合成产出,因此单位应用技术的价值  $V_t$  为:

$$V_t = p_{m,t} Y_{m,t} + (1 - \rho^T) A_{t,t+1} V_{t+1} \quad (18)$$

创新技术应用部门需要考虑在研发技术应用

率  $\lambda_t$  与支付给应用型劳动力工资的应用成本的情况下,最大化研发技术价值  $J_t$ ,以决定应用型劳动力的雇佣水平。其最优化问题为:

$$J_t = \max_{L_{\alpha,t}^s} E_0 \{ (1 - \rho^T) \Lambda_{t,t+1} [\lambda_t V_{t+1} + (1 - \lambda_t) J_{t+1}] w_t^s L_{\alpha,t}^s \} \quad (19)$$

得到  $L_{\alpha,t}^s$  的一阶条件为:

$$\rho_\lambda \frac{\lambda_t}{L_{\alpha,t}^s} (1 - \rho^T) \Lambda_{t,t+1} (V_{t+1} - J_{t+1}) = w_t^s \quad (20)$$

考虑到技术的应用率与淘汰率,应用技术存量的积累形式为:

$$T_t = \lambda_t (1 - \rho^T) (RD_{t-1} - T_{t-1}) + (1 - \rho^T) T_{t-1} \quad (21)$$

需要说明的是,研发技术  $RD_t$  是企业颠覆性创新过程中研发得到的技术存量,应用技术  $T_t$  是企业颠覆性创新过程中将技术研发应用于生产中的技术存量,两者共同反映了企业的颠覆性创新水平。

### (三) 市场出清条件与数据要素市场化冲击

与以往研究一致<sup>[18]</sup>,高技能劳动力  $L_r^s$  由研发型劳动力  $L_{r,t}^s$  与应用型劳动力  $L_{\alpha,t}^s$  构成,其中一单位应用型劳动力对应一单位待转化的研发技术。其出清条件为:

$$L_t^s = (RD_t - T_t) L_{\alpha,t}^s + L_{r,t}^s \quad (22)$$

最终品市场出清条件为:

$$Y_t = C_t + I_t \quad (23)$$

数据要素市场化冲击降低了数据交易过程中的市场摩擦,并降低了数据交易成本,在本文模型中表现为单位数据交易形成的有效信息比例  $\iota$  提升<sup>⑦</sup>。此外,由于场外交易的往往是未经处理、价值密度低的原始数据资源,其只有较低比例能够转化为高质量的数据要素,所以  $\iota$  的提升也能体现出数据要素市场化推动企业从场外交易走向场内交易的过程。因此,令  $\varepsilon_t^i$  为数据要素市场化冲击,政策冲击形式为:

$$\iota_t = \rho_i \iota_{t-1} + (1 - \rho_i) \iota + \varepsilon_t^i \quad (24)$$

⑦ 企业购买数据后,其中数据的开发利用还面临原始数据结构复杂、数据安全难保障等问题,单位数据交易形成的有效信息比例的影响因素十分复杂。然而数据交易平台的建设(即数据要素市场化冲击)可以有效降低交易壁垒、成本与信息不对称水平,进而提升企业从单位交易数据中提炼出数据要素的比例。

⑧ 参数校准的细节参见附录。

## 二、模型参数校准与数值模拟结果

### (一) 参数校准

模型中的参数分为标准模型部分、数据要素部分与颠覆性创新部分。本文参考相关文献或者根据实际数据校准以决定相关参数取值,模型中期限调整为一期对应一个季度<sup>⑧</sup>。

### (二) 数据要素市场化冲击的影响

本部分通过理论模型的数值模拟检验数据要素市场化对数据交易、数据要素积累与企业创新的影响,政策冲击的大小设定为将  $\iota$  从 0.2 提升至 0.3,代表数据要素市场化降低了数据交易摩擦并促进企业进行场内交易,导致单位交易数据的有效信息比例提升 10%。本文首先在基准情形下模拟政策冲击的宏观经济影响,之后将数据要素市场化提高企业数据要素质量、数据要素边际生产力两种情形与基准情形进行比较,以进行渠道检验。

#### 1. 基准情形结果

图 1 展示了数据要素市场化冲击下的宏观经济反应。横轴代表冲击发生后的窗口期,纵轴代表各变量相对自身稳态的偏离百分比。

由附录图 A1 可知,由于数据要素市场化降低了数据交易摩擦,促进了企业场内数据交易,带来数据交易量的大幅提升,并同时提升了数据资源的价值。更多的数据交易量扩充了企业的数据要素存量,提升了企业的产品质量并扩大了生产活动。由式(5)可知,企业生产规模的扩大也促进了自身数据要素的生成,进一步促进数据要素的积累。

数据要素市场化在促进企业数据交易与数据要素积累的同时,也通过“螺旋状”技术创新路径促进了企业的颠覆性创新:生产规模的扩大提高了企业对颠覆性创新与生产率提升的需求,表现为单位研发与应用技术的价值提升,由式(16)和式(17)可知,这提高了颠覆性创新部门对研发型与应用型劳动力的需求,进而促进企业的颠覆性

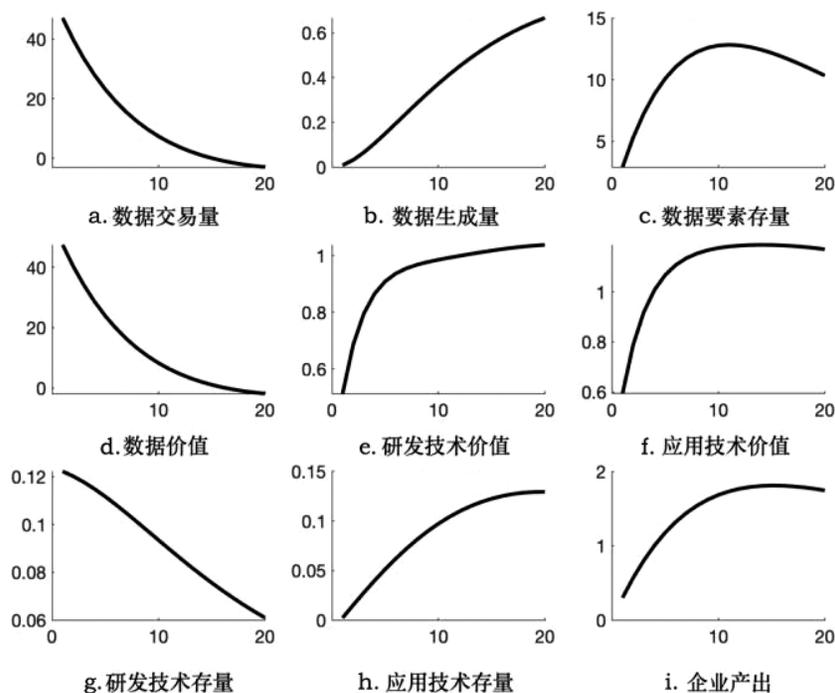


图1 数据要素市场化的影响

创新,带来了研发技术与应用技术存量的提升。根据式(9),应用到生产过程中的应用技术相当于提升了企业的生产效率,因此颠覆性创新机制进一步扩大了企业的产出。在这个过程中,进一步扩大的生产规模又通过生产活动的数据生成提高数据要素存量,进而通过产品质量使得企业最终产出得到更大的提升。因此,数据要素市场化通过促进企业的数据交易形成了该螺旋式创新范式闭环,对企业的生产与创新存在正循环的激励作用。

### 2. 提高数据要素质量的渠道检验<sup>⑨</sup>

根据式(5)中数据要素积累的决定,每一期有 $\delta^d$ 比例的数据要素因质量不足而淘汰,不再进入下一期的数据要素存量中。 $\delta^d$ 下降可以减少企业数据要素的淘汰率,因此可以反映为数据要素质量的提升。本文通过设定参数 $\delta^d$ 从基准情形的0.05下降20%至0.04作为数据要素市场化提高数据要素质量的情形,与基准情形进行比较,发现数据要素市场化可以通过提高数据要素质量进一步促进企业的颠覆性创新过程。

### 3. 提高数据要素边际生产力的渠道检验

根据产品质量的决定, $\eta$ 代表数据要素对产

品质量的边际贡献效率。根据式(13), $\eta$ 可以提升数据资源通过数据要素对产出的边际贡献,因此可以在一定程度上反映数据要素资源的配置效率。本文通过设定参数 $\eta$ 从基准情形的0.1提升20%至0.12作为数据要素市场化提高数据资源配置效率的情形,与基准情形进行比较,发现数据要素市场化可以通过提升企业数据要素边际生产力促进企业的颠覆性创新过程。

## 三、研究设计

### (一) 变量说明与测度

#### 1. 颠覆性创新

测算颠覆性创新的一个关键问题是系统性识别颠覆性技术的出现,因此首要目标是要确定哪些技术属于颠覆性技术?即构建“颠覆性创新词组库”。本文采用专利数据建立“2014—2023年间颠覆性创新技术术语库”,构建流程包括技术术语抽取、颠覆性创新技术识别、现有专利池筛选3个环节。

第一是技术术语抽取。本文从中国国家知识产权局获取我国1985—2023年发明专利申请文本及专利引用数据,专利文本内容由标题和摘要两

<sup>⑨</sup> 渠道检验的详细结果与分析参见附录。

部分构成,为减少原始专利文本中的噪声,需要对文本进行预处理。首先,采用 BERT-wwm-ext 模型优化分词,将 Jieba 分词结果与 BERT 的动态词向量结合,通过 CRF 层修正专有名词边界。同时,构建包含若干专利领域术语的补充词典,通过 LLM 的上下文学习能力动态更新词典。依据主流中文停用词表对常见的无意义词汇进行剔除,按照 Kelly 等<sup>[19]</sup>的方法,删除了诸如“的”“得”“地”等停顿词。最后,使用 spaCy 自然语言处理库对专利文本中的语词进行词性标注,为后续名词短语抽取提供数据基础,初步得到候选名词短语并进行去重。上一步中所得到的技术术语并非都属于创新技术。为此,采用相同的处理方式对 1985—2013 年的专利文本进行技术术语抽取,定义其为现存技术列表。然后,从候选技术术语列表中剔除同时存在于现存技术列表的项目,则剩余的技术术语被视为在 2014 年及之后出现的创新技术术语,每个创新技术术语代表一项创新技术。而后进行名词短语抽取,采用 N-gram 与词性组合模板的方法对长度为 1、2、3 个中文词汇的名词短语进行抽取。通过滑动窗口生成 N-gram,提取潜在关键词,并统计其出现频次。随后,利用上一步中使用 spaCy 标记后的词性,分析潜在关键词是否表示某种技术。通常而言,新技术词组中包含名词、动词、形容词、数词中的一类或几类。此外,符合描述新技术特点的词一般是以名词结尾,且一般不以动词开头。遵循这一逻辑,本文构建不同的词性排列组合模板,并引入了 LLM 辅助的上下文验证,将候选技术术语通过 LLM 中的提示工程设计问题:“该短语是否描述具体技术方案?请举例说明”,而后剔除了不满足条件的关键词。

在此基础上,本文继续筛选在专利原文有代表性的词。具体使用 3 种方法。方法一:本文根据专利分类,构建所有短语在特定发明领域中的频率分布,剔除领域内集中度低的词;方法二:本文计算每个词在其所在发明领域的专利文本库中的 TF-IDF 值,剔除 TF-IDF 值较低的词;方法三:分别计算每个关键词与发明领域的共现概率 PMI 值,剔除 PMI 值较低的词。通过综合考虑领域相关

性、TF-IDF 值和 PMI 值,本文筛选出既具有高领域相关性又在技术上下文中有较高共现概率的词作为具有代表性的候选短语。其中需要注意的是,通过筛选可能产生结果如“神经网络”和“多维神经网络”等有包含关系的词。针对这类词语,本文首先构建短语树,记录“神经网络”是“多维神经网络”的子短语,并在后续筛选中合并分析。

第二是颠覆性创新技术识别。在上一步中得到的名词短语并非全部具备技术术语的性质,根据其内涵主要可以分为两类:一是具有明显的技术特征,如“神经网络”“雷达设备”等;二是更为一般化,如“新型装置”“使用方法”等。显然,只有具备技术特征的名词短语才应当视为技术术语。为此,本文将候选名词短语与维基百科词条进行匹配,通过对应词条的目录结构来确定该短语是否具备技术特征。具体而言,首先将候选短语输入维基百科检索栏,如果该短语出现在返回最相关词条(第一条)的标题或摘要中,或在正文中出现 10 次以上,则视为与该词条成功建立匹配。在此基础上,进入该词条主页并提取其正文的目录结构。根据 Kalyani 等<sup>[20]</sup>的研究,描述技术术语的词条,其目录往往包含“应用”“使用”“类型”“操作”“特征”“设备”“功能”“技术”章节(记为章节列表 1),而对于一般术语,其目录通常仅包含“管理”“影响”“原因”“问题”章节(记为章节列表 2)。符合要求的技术术语需要同时符合以下条件:一是目录章节至少包含在章节列表 1 中的一个章节;二是目录章节不包含章节列表 2 中的任何章节。最终,本文得到符合要求的技术术语,作为候选技术术语列表。

本文使用专利被引次数作为技术颠覆性和重要性的判断依据,专利引用需要一段时间进行积累,通常为 3~5 年(REF)。在过去研究中往往采用预留时间窗口的方式,但这一做法将导致分析时效性降低。为此,本文参考以往研究<sup>[21]</sup>,使用专利的引用生命周期模型对专利在未来可能的引用分布进行估计,并对近 5 年申请的专利之被引量进行预测性调整,从而规避了现有研究中的局限性。具体而言,首先获取包含创新技术术语的对应专

利,在此基础上,剔除被引或预测被引次数少于100次的专利。在这一过程中失去所有对应专利的创新技术术语也将被剔除,剩余的技术术语作为最终的颠覆性创新技术术语,纳入后续分析,以此作为最终的“颠覆性技术词组库”。

第三是现有专利池筛选。对于颠覆性创新技术,一个关键的属性是其“兴起”的具体时间。不同于新技术的出现,一项技术的兴起要求其在短时间内具有足够的关注度。为此,本文根据与特定颠覆性创新技术相关的专利数量逐年分布情况,若连续5年包含该技术术语的专利数量年均增长率 $\geq 10\%$ ,则将该时间段的首年作为特定颠覆性创新技术的“兴起年”,在获得颠覆性创新技术列表及其相关专利后,本文进一步对省市和公司的技术创新能力进行测度。在进行具体测度前,首先将申请时间晚于特定颠覆性创新技术“兴起年”超过3年的对应专利剔除。由于晚于“兴起年”的专利更多地反映对已有技术的改进和模仿,而非原创性的突破,剔除这部分专利有助于提高测度准确性。最终,根据前文构建的“颠覆性技术词组库”和现有专利池,本文提取了包含“颠覆性创新词组库”的专利,以此来衡量企业颠覆性技术,并加总至企业层面,将专利数量作为企业技术颠覆性创新的衡量指标。为验证方法有效性,本文采集新闻媒体报道以及企业招聘数据,通过关键词匹配方法统计颠覆性创新技术的出现频率,统计各项颠覆性创新技术的出现频率及企业持有相应技术专利数量的相关性(限于篇幅并未展示,具体可联系笔者获取)。

## 2. 数据要素市场化

本文定义了 *did* 变量。若城市在当年及其之后建立数据交易平台,则 *did* 为1;反之则为0。以此作为政策冲击的一个重要前提是数据交易平台试点城市选择的随机性,尽管在设立数据交易平台之前并没有相关官方文件作为数据交易平台的筛选参考标准,但仍然有可能存在自选择行为。本文收集了截至2023年底已经设立的49个数据交易平台所在城市特征,以及数据交易平台由直辖市、省会城市逐步向地级市发展的路径,提取了

数据交易平台设立的先决因素。具体而言,影响初期数据交易平台设立的几种因素可能有:城市经济发展水平(*gdp*)、城市企业数量(*firm*)、就业率(*work*)、人力资本(*capital*)、对外开放程度(*fdi*)、数字基础设施建设(*infra*)、环境治理(*envi*)、产业结构(*indus*)等,参考蒋灵多等<sup>[22]</sup>的研究,构建Logit模型,将城市是否设立数据交易平台作为被解释变量,由于政府设立交易平台会参考城市前期的资源禀赋,因此将所有影响因素做滞后一期处理。结果发现,城市企业数量、人力资本、数字基础设施建设是影响数据交易平台设立的重要因素,因此本文将上述变量纳入交错DID模型进行回归,以控制数据交易平台的非随机性选择问题(限于篇幅并未展示,具体可联系笔者获取)。

## 3. 控制变量

参照徐翔等<sup>[23]</sup>的研究,选择了多个企业层面的控制变量,具体如表1所示。

### (二) 交错双重差分模型

由于数据交易平台政策逐步推进建设,不同期政策效果可能因城市的经济基础、产业结构和科技发展水平等因素而存在异质性,即政策是交错干预的,且不同地区在相同年份设立数据交易平台的数量并不相同,导致政策冲击的横向不可对比。此外,在逐年推行的政策情境下,交错DID模型允许不同个体在不同时间点受到政策冲击,这意味着,在某一个体受到冲击之前,那些尚未受到冲击的个体可以作为该个体的“好控制组”,其经济和技术表现更接近未受处理的状态,从而避免使用“坏的控制组”带来的潜在偏误,故本文参考以往研究<sup>[24]</sup>,构建交错DID模型进行验证,以允许估计政策效应随时间和队列差异等多种维度变化的异质性处理效应:

$$tec\_a_{i,c,t} = \gamma_0 + \gamma_1 did_{c,t} + \sum \lambda X_{i,c,t} + \mu_i + \mu_t + \varepsilon_{i,c,t} \quad (25)$$

其中,*i*和*t*分别代表企业和年份;*c*代表企业所在城市;*tec\_a*为企业颠覆性创新。*did*即企业所在城市是否参与数据要素市场化建设。*X*代表控制变量集; $\mu_i$ 和 $\mu_t$ 代表个体固定效应和年份固定效应。理论上来说,由于数据要素市场化的冲击

发生在城市层面,本应在城市层面控制固定效应,但为了进一步排除遗漏变量影响,本文选择在企业层面控制固定效应。但考虑到企业可能在不同城市间变动,此时企业固定效应不一定能够吸收城市固定效应,因此本文剔除了在不同城市之间发生选址变动的企业样本。 $\varepsilon$  代表随机误差项,模型主要使用逆概率加权结合最小二乘法的双重稳健方法来估计平均处理效应。

(三)数据来源

表2汇报了描述性统计结果,本文的数据主要来自国家知识产权局、国泰安数据库和各城市数据交易平台网站公开信息。用于验证部分的招聘数据主要从“前程无忧”“猎聘”“智联”“艾思产研数据平台”等网站爬取,新闻报道数据来自和讯网、新浪财经、东方财富网等 400 多家网络媒体文本信息。在研究数据交易平台政策对企业颠覆性创新影响时,可能会遇到企业自选择问题,即企业可能会因某城市建立了数据交易平台而选择在该城市设立子公司或入驻,从而出现自选择和互为因果问题,针对这一问题,本文剔除了各城市设立数据交易平台后新入驻的公司以及新设立子公司的母公司样本,仅保留余下样本进行估计。由于交错双重差分模型要求平衡面板数据,因此本文最终获得了 13 150 个观测值,包含 2014—2023 年 1 315 家上市公司的平衡面板数据。

表 1 描述性统计

变量名	均值	标准差	极小值	极大值	含义
<i>tec_a</i>	0.679 5	0.884 1	0.000 0	3.367 3	颠覆性创新
<i>did</i>	0.414 1	0.492 6	0.000 0	1.000 0	数据要素市场化
<i>size</i>	22.568 6	1.278 8	20.034 8	26.326 4	企业规模
<i>lev</i>	0.427 7	0.193 7	0.055 5	0.889 3	资产负债率
<i>roa</i>	0.037 6	0.059 9	-0.227 0	0.222 3	总资产净利润率
<i>cash</i>	0.052 7	0.063 7	-0.145 7	0.242 5	现金流
<i>fixed</i>	0.209 6	0.152 8	0.002 0	0.657 6	固定资产占比
<i>growth</i>	0.128 8	0.333 3	-0.537 0	2.062 1	营业收入增长率
<i>tobinq</i>	2.027 2	1.299 8	0.837 6	8.437 5	市场价值
<i>age</i>	3.020 8	0.290 1	2.079 4	3.583 5	企业年龄
<i>sa</i>	-3.899 5	0.251 7	-4.494 0	-3.203 3	融资约束
<i>green</i>	1.302 9	4.432 9	0.000 0	32.000 0	绿色技术
<i>rd</i>	0.035 9	0.045 2	0.000 0	0.280 5	研发投入

四、实证分析

(一)基准回归

本文首先使用多时期双重稳健估计量进行回

归,表2量化了数据要素市场化对颠覆性创新的影响,根据以往研究<sup>[24]</sup>,通过选择不同的权重,可以计算4种不同类型的平均处理效应,即简单加权平均处理效应、动态平均处理效应、日历平均处理效应和分组平均处理效应,即按照首次被处理的时间分组加权求和平均处理效应。按照4种不同权重设定下的平均处理效应结果,与不参与数据要素市场化的企业相比,参与数据要素市场化企业的颠覆性创新水平收到了明显的激励效果。

表 2 基准回归

变量	简单加权平均处理效应	动态平均处理效应	日历平均处理效应	分组平均处理效应
	<i>tec_a</i>	<i>tec_a</i>	<i>tec_a</i>	<i>tec_a</i>
	(1)	(2)	(3)	(4)
<i>did_att</i>	0.127 6*** (0.039 7)	—	—	—
<i>did_preavg</i>	—	-0.004 9 (0.016 9)	—	—
<i>did_postavg</i>	—	0.132 3*** (0.042 6)	—	—
<i>did_ca</i>	—	—	0.125 8*** (0.038 7)	—
<i>did_ga</i>	—	—	—	0.108 6*** (0.035 1)
控制变量、个体、年份固定效应	控制	控制	控制	控制
观测值	13 150	13 150	13 150	13 150

注:本文在个体层面上进行聚类,括号内为聚类稳健标准误,\*\*\*、\*\*、\*分别表示  $p < 0.01$ 、 $p < 0.05$ 、 $p < 0.010$  时有统计学意义,控制变量的系数估计被省略。下同。

(二)稳健性检验

为确保结论可靠性,采取以下几种方式进行稳健性检验。第一,处理组与控制组的前期差异控制。交错双重差分模型的前提是要满足平行趋势检验,即验证在没有数据要素市场化干预之前企业的颠覆性创新存在相同的变化趋势。图2汇

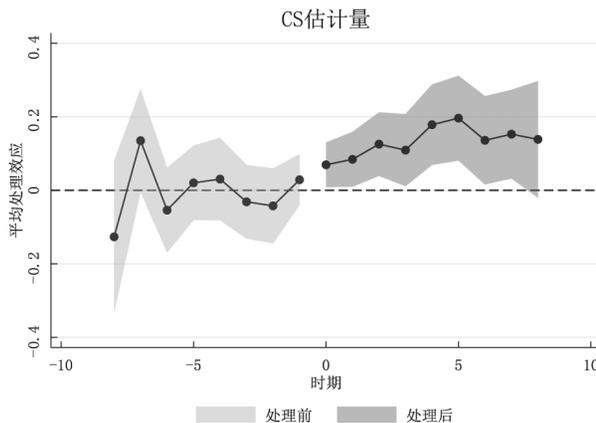


图 2 平行趋势检验

报了这一结果。从图2可以看出,政策干预时间之间,置信区间均穿过了0值,表明并不存在明显事前差异趋势。而政策干预之后,平均处理效应的置信区间连续超过0值,这也反映出数据要素市场化对于企业颠覆性创新的激励效果存在显著的长期效应。

第二,竞争溢出。与前文同理,若处理组企业与控制组企业存在行业间竞争关系或地区间竞争关系,则行业溢出效应或者地区溢出效应会干扰本文结果。针对这一考虑,Cherniwchan等<sup>[25]</sup>曾采用企业*i*在时间*t*内的潜在竞争对手集合定义为某行业在时间*t*内进行出口的工厂集合,这一做法为本文识别企业竞争对手提供了重要参考。然而,考虑到本文的被解释变量为企业颠覆性创新,依据出口行为来识别企业*i*的竞争对手显然失之偏颇。鉴于此,本文首先计算了企业所在行业的营业收入均值,当企业的营业收入大于行业营业收入均值时,表明该企业在所属行业占据一定的市场份额,因此该企业被判定为行业中其他企业的潜在竞争对手。同理,还测算了控制组中企业*i*在地区*c*的潜在竞争对手,而后剔除这两类存在竞争潜力的企业,使用余下样本进行回归。表3的回归(1)报告了这一结果,此时核心解释变量的系数估计仍然显著为正。

第三,基于LSTM模型的内生性检验。前文中,数据要素市场化被视为一项“足够外生”的政策冲击,但受实施难度、实施门槛、政策配套、市场参与者积极性等现实约束,为了提高一项政策的可行性和可复制性,政府在选择数据要素市场化试点前期也会参照地区原始的要素市场化水平。因此,本文借鉴樊纲等<sup>[26]</sup>的研究,测算了地区1997—2013年的要素市场发育程度,并借助LSTM模型预测了2014—2023年地区要素市场发育程度,并使用单期滞后数据作为数据要素市场化的工具变量。LSTM模型能够学习序列数据中的长期依赖关系,从工具变量的相关性来看,历史上的要素市场发育程度是地区原始禀赋的体现,而LSTM模型预测后的变量通常能够解决工具变量不随时变而导致的弱相关性。从外生性来

看,LSTM模型的预测误差不会直接影响过去期的预测值,这意味着LSTM模型的预测值与当前期的误差项不直接相关。表3的回归(2)报告了以LSTM模型预测的变量作为工具变量的2sls估计,此时*did*的系数估计仍然显著,结论保持稳健。

表3 稳健性检验

变量	<i>tec_a</i>	<i>tec_a</i>
	(1)	(2)
<i>did_att</i>	0.813 0*** (0.067 4)	—
<i>did</i>	—	1.345 5** (0.653 9)
控制变量、个体、 年份固定效应	控制	控制
观测值	6 690	13 150
Kleibergen-Paap rk Wald F statistic	—	16.441***

## 五、机制检验与异质性分析

### (一)数据要素质量和数据要素边际生产力的机制识别

#### 1. 数据要素质量

数据要素质量用于衡量企业自身数据要素的质量,本文首先基于国家知识产权局的全量专利信息,并借助《数字经济核心产业分类与国际专利分类参照关系表(2023)》中涉及数字经济核心产业分类中的IPC分类号,来识别企业所有的数字经济专利。而后,基于创新和专利质量的测算思路<sup>[27]</sup>,本文使用企业所有数字经济专利的被引次数来衡量企业数据要素质量。该值越大,往往表明企业数字经济专利质量越高,则对应的企业数据要素质量越高,以*diqua*表示。

#### 2. 数据要素边际生产力

以往研究充分讨论了其他4种生产要素在经济社会发展中的配置效率问题<sup>[28]</sup>,数据要素作为中国率先提出的第五大生产要素,其合理配置关系到经济体能否打破传统生产要素局限,促进资源在更广泛范围内的优化配置。从而助力政府更有效利用数据资源,激发市场活力。本文参考以往研究<sup>[29]</sup>,以企业的数据要素资源配置效率来衡量企业的数据要素边际生产力。

#### 3. 结构方程和ACME估计

为了理清数据要素市场化与企业颠覆性创新之间的作用机制,本文纳入数据要素质量与数据要素边

际生产力,构建因果中介模型进行回归,设定如下。

$$m_{i,c,t} = \varphi_0 + \varphi_1 did_{c,t} + \sum \lambda X_{i,c,t} + \mu_i + \mu_t + \varepsilon_{2i,c,t} \quad (26)$$

$$tec_{a_{i,c,t}} = \chi_0 + \chi_1 did_{c,t} + \chi_2 m_{i,c,t} + \sum \lambda X_{i,c,t} + \mu_i + \mu_t + \varepsilon_{3i,c,t} \quad (27)$$

其中, $m$ 为中介变量,对于每一个企业,只能观察到  $tec_{a_i}(did_i, M_i(did_i))$ , 而无法观察到  $tec_{a_i}(did_i, M_i(1 - did_i))$ , 因此借鉴以往研究<sup>[30-31]</sup>, 本文采用基于准贝叶斯蒙特卡罗近似的估计方法计算得出因果中介效应, ACME 的有效估计为  $\varphi_1 \chi_2$ , ADE 为  $\chi_1$ 。至此, 本文首先对线性结构方程模型进行估计, 设定参数不确定性的准贝叶斯近似运行的模拟次数为 1000 次。进一步地, 本文基于线性结构方程系数乘积方法, 估计了 ACME、ADE 和 ATE, 如表 4 所示。可以看出, 数据要素市场化通过数据要素质量影响颠覆性创新的 ACME 为 0.03, 其对应 95% 的置信区间在 0 以上, 证明中介效应存在, ADE、ATE 分别为 0.10 和 0.12, 且置信区间也均大于 0。同理, 数据要素边际生产力的中介效应也成立。

表 4 ACME 和 ADE

	(1)	(2)
	中介变量为 <i>diqua</i>	中介变量为 <i>dra</i>
ACME	0.027 8 [0.023 5, 0.031 7]	0.002 0 [0.000 6, 0.003 4]
ADE	0.096 5 [0.070 5, 0.124 9]	0.121 9 [0.095 8, 0.150 4]
ATE	0.124 2 [0.096 2, 0.154 0]	0.124 9 [0.098 0, 0.152 8]
Proportion	0.223 9 [0.180 3, 0.288 6]	0.016 3 [0.013 2, 0.020 6]

### (二) 数据要素陷阱问题

根据徐翔等<sup>[23]</sup>的研究, 数据要素会进一步放大阿罗替代效应, 导致大企业缺乏创新动力, 而满足于通过迭代式创新增加现有用户数量, 最终将致力于突破性创新的小企业挤出市场。而数据要素市场化的实施降低了小企业获取数据要素的门槛, 拓展了小企业获取用户信息的来源, 因此为了进一步回应数据要素市场化对于数据要素陷阱的缓解作用, 本文将样本划分为大规模企业和小规模企业。结果如表 5 的回归(1)和回归(2),  $de$  代表企业原始的数据要素积累, 本文基于企业招聘数据, 以企业对数据分析人员支付的工资总和来表示。由此可以看出,  $de$  的系数估计均在 5% 的显

著性水平上显著为负, 表明阿罗替代效应确实存在, 但  $did_{de}$  的系数估计显著为正, 说明数据要素市场化显著降低了数据要素对于创新的抑制作用, 即数据要素陷阱问题得到了有效缓解。

表 5 分企业规模的异质性检验

变量	<i>tec_a</i>	<i>tec_a</i>
	大规模企业	小规模企业
	(1)	(2)
<i>did</i>	0.106 9*** (0.044 5)	-0.012 8 (0.060 3)
<i>de</i>	-0.021 1** (0.021 0)	-0.024 5** (0.028 0)
<i>did_de</i>	0.012 3** (0.012 0)	0.053 2** (0.123 0)
控制变量、个体、年份固定效应	控制	控制
观测值	6 509	6 533

## 六、何以强化数据要素市场化的创新效应

### (一) 加强高质量数据供给

要充分激活以数据要素市场化为核心的创新体系, 最大的制约在于数据流通与交易的活跃程度。数据要素市场化取决于数据处理和分析的基础能力, 从而影响数据加工的深度和质量。算力基础设施不仅可以节约数据处理和分析成本、提高数据处理和分析效率, 还可以创造更多新的创新应用场景、扩大数据应用范围, 从而形成数据供给和需求的良性循环。本文拟从算力这一核心要素出发, 分析其如何激活数据要素市场化的活力。具体而言, 本文以企业所在省份的显卡进口金额来衡量算力, 显卡属于计算密集型设备, 直接影响着高计算能力领域的算力水平。在数据产生、数据收集、数据存储、数据处理、数据分析等全生命周期过程中都需要算力的支撑, 强算力能够支撑复杂的数据加工算法, 提升数据清洗整合、数据分析等工作的效率及准确度, 保障数据质量。表 6 的回归(1)汇报了这一结果。  $meas$  代表企业所在省份算力水平,  $did$  的系数估计显著为正, 且  $meas$  与  $did$  交互项的系数估计同样显著为正, 表明算力水平强化了数据要素市场化的创新效应。

### (二) 夯实数字基础设施建设

数据流通技术和基础设施不完善、不可信是阻碍数据要素市场化实现的另一障碍, 由于缺乏完善的可信流通技术体系, 数据的高效流通受阻, 交易技术缺失导致市场活力受限。本文进一步论

证数字技术设施建设对于数据要素市场化创新效应的优化作用,以破解当前数据要素市场化面临的困境。数字基础设施作为支撑数据流动与利用的物理基础和技术框架,其完善与否直接关系到数据市场的运行效率与安全性。通过强化数据基础设施建设,提升数据传输速度、增强数据存储能力、优化数据处理算法以及建立健全数据安全与隐私保护机制,可以有效促进数据的可信流通,降低交易成本,激发市场活力,进而推动数据要素市场化的健康发展。本文参考赵星<sup>[32]</sup>的研究,从数字基础设施投入和产出两端测算了企业所在城市的数字基础设施发展水平。表6的回归(2)报告了这一结果。*infra*代表数字基础设施发展水平,*did*的系数估计显著为正,且*infra*与*did*的交互项系数同样显著为正,表明数字基础设施建设能够显著驱动数据要素市场化的创新效应。因此,加强数字基础设施建设,不仅是破解数据要素市场化阻碍的关键举措,也是推动数字经济高质量发展的内在要求。

表6 进一步分析

变量	<i>tec_a</i>	<i>tec_a</i>
	(1)	(2)
<i>did</i>	0.0613** (0.0272)	0.0661*** (0.0246)
<i>meas</i>	-0.0273 (0.0209)	—
<i>did_meas</i>	0.0620** (0.0260)	—
<i>infra</i>	—	-0.0448 (0.0351)
<i>did_infra</i>	—	0.1053*** (0.0393)
控制变量、个体、 年份固定效应	控制	控制
观测值	13150	13150
<i>R</i> <sup>2</sup>	0.6778	0.6779
调整后 <i>R</i> <sup>2</sup>	0.6414	0.6415

### 七、结论与启示

本文构建了包含颠覆性创新的动态随机一般均衡模型,在考虑数据要素折旧的基础上,将其进一步分为数据要素原始积累部分和数据要素交易部分,从理论上建构了数据要素市场化与企业颠覆性创新的内生理论框架。同时,借助2014—2023年上市公司数据,实证检验了数据要素市场化驱动技术创新范式改变的新机理。研究发现:数据要素市场化显著激励了企业颠覆性创新,缓解了数据要素引发突破性创新困境的阿罗替代效

应。此外,数据要素市场化通过提高数据要素质量,增加企业单位应用技术价值,从而提高了企业进行应用技术创新的积极性,并通过优化数据要素边际生产力,强化了“企业产出—新数据要素—产品质量”这一循环的正反馈机制。即便在考虑数据要素折旧的情形下,数据要素市场化仍缓解了大企业数据要素陷阱问题。进一步分析表明,加强高质量数据供给、夯实数字基础设施建设是有效激发数据要素市场化创新效应的重要路径。

基于上述结论,本文提出如下政策建议。

一是保障高质量数据要素供给,激活数据要素流通源头。建立健全数据质量评估与监管体系,制定严格的数据质量标准,加强对数据生产、处理、流通等全过程的监督把控,保障高质量、可信任的数据。引导培育数据创新企业的发展,运用财政政策补贴、税收优惠等政策支持企业加强对数据研发力度,完善数据分析处理能力,提升数据产品供给。营造流通的良好环境氛围。完善数据流通法律法规,保证数据流通的合法权益,激发数据流通的活力。

二是强化数字基础设施支撑,夯实数字经济发展基石。应加大对数字基础设施的资金、技术、人才等方面的投资力度,保障数字基础设施建设顺利开展。加大数字基础设施安全保障建设,完善数据保护、网络安全、隐私保护等方面的法规。推进数字基础设施开放共享,加大对偏远地区数字基础设施的建设力度,缩小数字鸿沟,促进数字经济均衡发展。加强数字基础设施与产业的融合,促进产业的数字化、智能化,提升总体经济质量和效益。

三是提高数据要素边际生产力,释放数据要素乘数效应。在现有数据要素交易平台建设的基础上,加大技术创新,确保数据要素交易平台高效运行、稳定运营。推进数据交易规则标准化建设,明晰数据产权、交易流程、数据安全等环节,提高数据交易效率、安全性。加大“数据+产业”创新应用,推进关键行业数据应用融合,创新数据引领业务。建立数据要素供给动态监测机制,对各行业、各地区的数据需求供给进行动态监测,根据需求供给情况,实施数据要素精准供给工程,重点行业、重点领域,加强数据供给保障,提高数据要素供给有效性。

## 参考文献:

- [1] 赵森茂,徐欢欢,高雯,等. 企业数据要素利用水平、数字创新和产业结构现代化[J]. 中国软科学,2024(增刊2):398-407.
- [2] 江小涓. 数字时代的技术与文化[J]. 中国社会科学,2021(8):4-34,204.
- [3] 黄尹旭,杨东. “利益—权利”二元共生:“数据要素×”的价值创造[J]. 中国社会科学,2024(2):47-64,205.
- [4] 赵放,徐熠,朱梦婷,等. 数据要素市场化配置如何影响国有企业 ESG 责任履行?:基于准自然实验的考察[J]. 中国软科学,2024(10):212-224.
- [5] 黄丽华,郭梦珂,邵志清,等. 关于构建全国统一的数据资产登记体系的思考[J]. 中国科学院院刊,2022,37(10):1426-1434.
- [6] 吴江,袁一鸣,贺超城,等. 数据要素交易多边平台研究:现状、进路与框架[J]. 信息资源管理学报,2024,14(3):4-20.
- [7] ARROW K. Economic welfare and the allocation of resources for invention[C]//The rate and direction of inventive activity: economic and social factors. Princeton University Press, 1962.
- [8] ROSEN R J. Research and development with asymmetric sizes[J]. Rand journal of economics, 1991, 22: 411-420.
- [9] 李东红,陈昱蓉,周平录. 破解颠覆性技术创新的跨界网络治理路径:基于百度 Apollo 自动驾驶开放平台的案例研究[J]. 管理世界,2021,37(4):130-159,199.
- [10] 刘海兵,刘洋,黄天蔚. 数字技术驱动高端颠覆性创新的过程机理:探索性案例研究[J]. 管理世界,2023,39(7):63-81,99,82.
- [11] EECKHOUT J, VELDKAMP L. Data and market power [R]. Columbia university working paper, 2021.
- [12] AGHION P, BERGEAUD A, BOPPART T, et al. A theory of falling growth and rising rents[J]. The review of economic studies, 2023, 90(6): 2675-2702.
- [13] JONES C I, TONETTI C. Nonrivalry and the economics of data [J]. American economic review, 2020, 110(9): 2819-2858.
- [14] FISHER J D. On the structural interpretation of the smets-wouters “risk premium” shock[J]. Journal of money, credit and banking, 2015, 47(2/3): 511-516.
- [15] CONG L W, XIE D, ZHANG L. Knowledge accumulation, privacy, and growth in a data economy[J]. Management science, 2021, 67(10): 6480-6492.
- [16] CONG L W, WEI W, XIE D, et al. Endogenous growth under multiple uses of data[J]. Journal of economic dynamics and control, 2022, 141: 104395.
- [17] FARBOODI M, VELDKAMP L. A model of the data economy[R]. NBER working paper, No. 28427, 2021.
- [18] ANZOATEGUI D, COMIN D, GERTLER M, et al. Endogenous technology adoption and R&D as sources of business cycle persistence[J]. American economic journal: macroeconomics, 2019, 11(3): 67-110.
- [19] KELLY B, PAPANIKOLAOU D, SERU A, et al. Measuring technological innovation over the long run [J]. American economic review: Insights, 2021, 3(3): 303-320.
- [20] KALYANI A, BLOOM N, CARVALHO M, et al. The diffusion of new technologies[R]. NBER working paper, No. 28999, 2024.
- [21] LERNER J, SERU A. The use and misuse of patent data: issues for finance and beyond [J]. The review of financial studies, 2022, 35(6): 2667-2704.
- [22] 蒋灵多,陆毅,张国峰. 自由贸易试验区建设与中国出口行为[J]. 中国工业经济,2021(8):75-93.
- [23] 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.
- [24] CALLAWAY B, SANT’ ANNA P H C. Difference-in-differences with multiple time periods [J]. Journal of econometrics, 2021, 225(2): 200-230.
- [25] CHERNIWCHAN J, NAJJAR N. Do environmental regulations affect the decision to export [J]. American economic journal: economic policy, 2022, 14(2): 125-160.
- [26] 樊纲,王小鲁,张立文,等. 中国各地区市场化相对进程报告[J]. 经济研究,2003(3):9-18,89.
- [27] 亢延锴,郭家宝,胡志安,等. 创新驱动、激励机制与高校科技成果转化:以省部共建国家重点实验室为例[J]. 管理世界,2025,41(3):50-76.
- [28] 刘潘,张子尧. 地方公共债务与资源配置效率:企业间全要素生产率分布差异的视角[J]. 经济研究,2023,58(10):114-133.
- [29] 王海森,李纲. 数据要素市场化对颠覆性技术的影响研究[J]. 情报学报,2025,44(5):535-548.
- [30] 杨刚强,王海森,范恒山,等. 数字经济的碳减排效应:理论分析与经验证据[J]. 中国工业经济,2023(5):80-98.
- [31] 杨刚强,王海森,岳子洋,等. 客户数字化转型、供应商碳减排与碳信息披露迎合[J]. 中国工业经济,2024(8):99-117.
- [32] 赵星. 新型数字基础设施的技术创新效应研究[J]. 统计研究,2022,39(4):80-92.

(本文责编:默 黎)