

数字经济时代下数据要素对企业数字技术创新的影响研究

李昱谦

贵州大学经济学院, 贵州 贵阳

收稿日期: 2025年5月13日; 录用日期: 2025年5月27日; 发布日期: 2025年6月27日

摘要

在数字经济时代下, 数据要素已经成为赋能经济发展的重要生产要素。本文利用沪深A股上市公司数据, 使用文本分析法度量上市公司数据要素的使用水平, 运用面板双向固定效应模型进行实证, 分析了数据要素对企业数字技术创新的影响。研究发现, 数据要素显著提升了企业数字技术创新水平, 这一结论在经过更换被解释变量、倾向得分匹配法、控制行业固定效应等一系列稳健性检验后仍然成立。异质性分析发现, 数据要素对企业新质生产力的提升在数字经济行业和非国有企业中更为显著。因而应该从促进数据基础设施建设、提升员工的数据分析技能、激发国有企业的活力和市场竞争力等方面入手, 促进企业数字创新水平。

关键词

数字经济, 数据要素, 数字技术创新

Research on the Impact of Data Elements on Digital Technology Innovation of Enterprises in the Era of Digital Economy

Yuqian Li

School of Economics, Guizhou University, Guiyang Guizhou

Received: May 13th, 2025; accepted: May 27th, 2025; published: Jun. 27th, 2025

Abstract

In the era of digital economy, data elements have become an important production factor that empowers economic development. This paper uses the data of listed companies on the Shanghai and Shenzhen

文章引用: 李昱谦. 数字经济时代下数据要素对企业数字技术创新的影响研究[J]. 电子商务评论, 2025, 14(6): 2945-2953. DOI: 10.12677/ec.2025.1462072

A-share stock exchanges and employs the text analysis method to measure the utilization level of data elements by listed companies. This paper uses the panel bidirectional fixed effects model for empirical analysis to investigate the impact of data elements on digital technology innovation of enterprises. The research finds that data elements significantly enhance the level of digital technology innovation of enterprises. This conclusion still holds after undergoing a series of robustness tests such as changing the dependent variable, propensity score matching method and controlling for industry fixed effects. Heterogeneity analysis reveals that the improvement of data elements on digital technology innovation of enterprises is more significant in the digital economy industries and non-state-owned enterprises. Therefore, efforts should be made to promote the construction of data infrastructure, enhance employees' data analysis skills, stimulate the vitality and market competitiveness of state-owned enterprises, and thereby promote the level of digital technology innovation of enterprises.

Keywords

Digital Economy, Data Elements, Digital Technology Innovation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在数字经济时代，数据要素已经成为驱动技术革新的核心动能。2020年，中共中央、国务院首次明确了数据作为新型生产要素的战略地位，2022年中共中央、国务院发布“数据二十条”，在产权、流通、收益分配等方面为数据要素市场化提供制度保障。2024年5月，第七届数字中国峰会以“释放数据要素价值，发展新质生产力”为主题于福建召开，旨在激活数据要素潜能，促进经济高质量发展。数字技术催生了共享经济、平台经济等新形态的经济模式，推动了数实产业融合和数据要素价值释放，其已然成为数字经济发展的基础，而数字技术创新则是数字技术发展的核心驱动力。那么，数据要素是否可以提升数字技术创新水平？如果可以，其作用机理又是什么？这些都需要进一步探讨。

自从十九届四中全会将数据作为第五大生产要素后，越来越多的学者围绕这一主题展开深入探究。关于数据要素的特点方面，徐翔等(2021)总结了数据要素具有非竞争性、排他性、规模报酬递增、正外部性、产权模糊性以及衍生性[1]。关于数据要素的价值化方面，李海舰、赵丽(2021)认为，数据要素形态演进经历了数据资源、数据产品、数据商品、数据资本四个阶段，分别对应数据要素价值实现过程中的潜在价值、价值创造、价值实现和价值增值[2]。卢黎歌、李婷(2022)归纳了我国建立数据要素市场所面临的数据确权、定价机制、数据安全等方面可能存在的问题并针对其提出建议[3]。在数据要素的经济效应方面，王亚飞等(2025)发现数据要素可以对企业全要素能源效率产生正向影响[4]，赵森茂等(2024)发现数据要素可以通过提升数字技术创新水平进而促进产业结构的高级化，提升了劳动生产率[5]。

在数字技术创新的经济效应方面，夏范社等(2025)发现数字技术创新在推动企业履行环境社会责任方面可以发挥积极作用[6]。方慧等(2024)发现，数字技术创新可以通过提升信息交互、降低交易成本等机制提升产业链韧性[7]。王东和李晓(2025)通过实证检验发现，数字技术创新可以降低交易成本，提高企业供应链效率，且这种作用有地区市场化程度、股权集中度、企业规模方面的异质性[8]。在数字技术创新的影响因素方面，潘丹丹(2025)等通过研究发现，风险投资持股可以提升企业数字技术创新水平，且这种提升作用主要通过内部监管的加强、融资约束的降低等途径实现[9]。夏后学(2025)采用双重差分法，发现

信息消费试点政策对数字技术创新有显著提升作用,且数据要素供给在其中起到了中介作用[10]。李延喜等(2025)通过文本分析方法分析上市公司年报,测度竞争压力感知指标,证明了企业竞争压力可以显著促进数字创新的提升,且这种促进作用受到企业高管专业背景的影响[11]。

总结已有文献,数据要素相关的理论内涵与实证研究较为成熟,企业数字技术创新的影响因素仍有讨论的空间。基于此,本文运用文本分析方法计算上市公司年报词频以测度企业数据要素水平,并利用企业数字技术相关专利数量衡量数字技术创新水平,探究数据要素对数字技术创新的影响,为企业应用数据要素以及政府政策实施方面提供经验证据。

2. 研究假设

首先,数字技术创新需要购置设备、招聘人才,而这些活动往往需要大量的资金投入。企业的交易、供货等经营活动产生了大量的“数字足迹”,这些“数字足迹”能够被银行和金融科技公司获取用以提升信用评估模型预测效率,促进信贷供给者间的市场竞争,提高低风险借款者获取融资的可能性以及降低贷款利率,减少融资约束和创新成本,为创新主体提供金融支持,激发数字创新动力[12]。

其次,数字技术创新需要高技术人才。一方面,随着数据要素在企业的广泛应用,劳动力市场涌现一些高技能岗位,且企业对这些岗位的需求逐渐增加,例如数据分析师、数据科学家、数据挖掘员等岗位[13],这些岗位必然会由高技术劳动力所胜任。另一方面,数据要素的利用会减少企业对低技能、高重复性岗位的需求。另外,数据要素的应用加强了劳动力的学习效应。具体而言,更大规模的数据对数据处理人员的技能提出了更高的要求,劳动者为了保证不被淘汰会持续学习以保证数据分析技能的更新。

再次,数字技术创新是一种高风险的活动,其涉及前沿技术的应用,而且数字创新产品和服务的市场需求往往难以准确预测,一旦创新失败将对企业造成打击[14]。而企业可以利用个人消费数据对目标用户进行精准的用户画像,准确研判数字技术市场的情况,抓住创新机会,降低研发失败对企业造成的潜在损失,这在主观上提升了企业进行数字创新的意愿。

最后,数据要素有效地打破了企业内部和外部的信息孤岛,加强了部门间、企业间的沟通,促进了协同创新,加快了数字技术创新效率。在内部,数据要素能连接企业研发、生产、销售等环节,增加了内部沟通的透明度,促进经营信息共享与部门间共同研发,优化了资源配置,缩短了数字技术创新周期。在外部,数据要素增加了企业与供应商、客户的信息共享,通过数据共享和知识交互,多元主体之间可以更便利地开展合作,联合攻克数字技术难题,这促进了跨地域、跨领域数字技术创新,实现了数字技术创新效率的提升[15]。

综上所述,本文提出如下假说:数据要素的应用能够对企业新质生产力水平产生正向作用。

3. 研究设计

3.1. 模型建立

本文采用双向固定效应模型估计数据要素对数字技术创新的影响,具体模型设定如下:

$$\text{Patent}_{it} = \alpha + \beta \text{Data}_{it} + \gamma \text{Controls} + \phi_i + \theta_t + \varepsilon_{it}$$

其中,等式左边为被解释变量,表示在 t 年企业 i 的数字技术创新水平,等式右边依次为常数项、核心解释变量 t 年企业 i 的数据要素水平、企业层面控制变量、企业个体固定效应、时间固定效应和随机误差项。

3.2. 变量设定

本文被解释变量为数字技术创新(Patent),将企业专利的主分类号与《数字经济及其核心产业统计分类2021》相匹配,若可以匹配上,则此专利为数字技术相关专利。将企业数字经济专利数量加一取对数

衡量企业数字技术创新水平。在稳健性检验中，将企业数字经济发明专利数量加一取对数(Patent_invent)代替原变量进行回归。

本文核心解释变量为数据要素(Data)，首先是构建数据要素相关词典，参考已有文献的做法[16]，本文以“数据要素”、“大数据”、“算力”、“信息”、“数据化”、“算法”为关键词构建词典，理由如下：“数据要素”反映出数据作为新型生产要素的战略定位，“信息”反映出数据是被用于进行决策的信息，“大数据”和“数据化”从规模维度和过程维度出发，刻画了对数据的开发强度和数据对业务的渗透深度，“算力”和“算法”表征数据处理的硬件能力和数据分析的技术内核。接着用 python 爬取上市公司年报，统计与数据要素相关的词频，并运用 python 的第三方库 jieba 将年报分词，将与数据要素相关的词频数量加一取对数衡量数据要素水平。

本文选取企业常用的财务指标、股东指标以及可能会影响数字技术创新的变量作为控制变量，具体包括：资产负债率(Lev)，以年末总负债除以年末总资产表征；总资产净利润率(ROA)，以净利润除以总资产平均余额表征；现金流比率(Cashflow)，以经营活动产生的现金流量净额除以总资产表征；上市年限(ListAge)，以当年年份减去上市年份加一取对数表征；托宾 Q 值(TobinQ)，以流通股市值加非流通股股数乘每股净资产加负债账面值三者的总和除以总资产表征；前五大股东持股比例(Top5)，以前五大股东持股数量除以总股数表征；政府补助强度(Subsidy)，以政府补助金额加一取对数表征；研发投入强度(RDratio)，以研发投入除以营业收入表征。

3.3. 数据来源与描述性统计

本文选取 2012~2023 年中国沪深 A 股上市公司数据进行分析，企业层面的被解释变量和控制变量数据来自国泰安数据库，核心解释变量文本数据来自上海证券交易所与深圳证券交易所。本文在回归前已经对数据做如下处理：剔除金融业及房地产业公司样本；剔除 ST、PT 类公司样本。下表 1 是企业变量描述性统计。

Table 1. Descriptive statistics
表 1. 描述性统计

	mean	sd	min	max
Data	4.016	0.553	0.000	6.770
Patent	2.056	1.803	0.000	8.614
Patent_invent	1.578	1.723	0.000	8.517
Lev	0.492	0.200	0.016	1.957
ROA	0.032	0.062	-0.894	0.585
Cashflow	0.049	0.066	-0.556	0.652
ListAge	2.597	0.710	0.000	3.497
TobinQ	1.811	1.330	0.625	26.818
Top5	0.550	0.158	0.132	0.992
Subsidy	16.973	1.588	6.957	22.477
RDratio	3.435	4.506	0.000	137.450

4. 实证分析

4.1. 基准回归

数据要素对企业全要素生产率的回归结果见表 2。其中列(1)为未加入控制变量的结果，列(2)为加入控制变量的结果，括号内为企业层面聚类稳健标准误。可以看到，无论是否加入控制变量，核心解释变量的系数均在 1%的水平上显著为正，证明数据要素的应用可以提升企业全要素生产率水平，假说得到验证。

Table 2. Regression results
表 2. 回归结果

	(1)	(2)
	Patent	Patent
Data	0.096*** (0.021)	0.089*** (0.023)
Lev		0.238*** (0.086)
ROA		0.504*** (0.105)
Cashflow		-0.412*** (0.101)
ListAge		0.105*** (0.025)
TobinQ		-0.017*** (0.006)
Top5		0.319** (0.158)
Subsidy		0.119*** (0.011)
RDRatio		0.000** (0.000)
_cons	1.217*** (0.087)	-0.995*** (0.242)
企业固定效应	是	是
时间固定效应	是	是

注：*、**、***分别表示在 10%、5%、1%的水平上显著，括号内为聚类稳健标准误。下同。

4.2. 稳健性检验

4.2.1. 替换被解释变量

为了验证前文结果的稳健性，本文将企业数字经济发明专利数量加一取对数(Patent_invent)代替原变量进行回归。检验结果如表 3 第(1)列所示。可以看到，在替换被解释变量后，核心解释变量的系数在 1%的水平上显著为正，说明原结论具有稳健性。

Table 3. Robustness test
表 3. 稳健性检验

	(1)	(2)	(3)	(4)
	Patent_invent	Patent	Patent	Patent
Data	0.096*** (0.023)	0.083*** (0.029)	0.082*** (0.022)	
L.Data				0.092*** (0.025)
Lev	0.182** (0.082)	0.239** (0.104)	0.212** (0.083)	0.198** (0.093)
ROA	0.301*** (0.097)	0.524*** (0.138)	0.465*** (0.103)	0.586*** (0.115)
Cashflow	-0.282*** (0.096)	-0.403*** (0.122)	-0.391*** (0.099)	-0.404*** (0.108)
ListAge	0.045* (0.024)	0.084*** (0.030)	0.102*** (0.025)	0.076* (0.044)
TobinQ	-0.013** (0.006)	-0.024*** (0.007)	-0.014** (0.006)	-0.013** (0.006)
Top5	0.310** (0.156)	0.330* (0.172)	0.283* (0.152)	0.387** (0.176)
Subsidy	0.116*** (0.011)	0.127*** (0.013)	0.118*** (0.011)	0.118*** (0.012)
RDratio	0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	0.011*** (0.004)
_cons	-1.311*** (0.242)	-1.062*** (0.283)	-0.908*** (0.229)	-0.963*** (0.275)
企业固定效应	是	是	是	是
时间固定效应	是	是	是	是
行业固定效应	否	否	是	否

4.2.2. 倾向得分匹配法

本文借鉴已有文献[17]，运用倾向得分匹配法进行稳健性检验。具体地，将数据要素水平按年份进行分组，将样本按中位数分为高数据水平组和低数据水平组，将高数据水平组作为处理组，将全部控制变量作为协变量，用 Logit 模型进行倾向得分匹配，匹配比例为 1:3。将匹配到的样本重新进行回归，回归结果见表 3 第(2)列，可以看到，核心解释变量的系数仍然显著为正，再次验证了原结果的稳健性。

4.2.3. 控制行业固定效应

由于不同行业可能会存在行业政策、行业环境、行业特点等方面的不同，可能会对估计结果造成影响，本文进一步控制行业固定效应之后重新进行回归，回归结果如表 3 第(3)列所示。结果显示，核心解释变量的系数仍然显著为正，证明了结论的稳健性。

4.2.4. 内生性分析

由于数据要素和数字技术创新水平之间可能会存在反向因果问题，即不仅数据要素的使用会提升企业数字技术创新水平，而且数字技术创新水平较高的企业更倾向于使用数据要素，导致内生性问题。此外，可能存在影响数字技术创新的其他变量没有被纳入回归模型中导致遗漏变量问题。因此，本文对模型进行内生性检验。参考学界通用的做法，将核心解释变量滞后一期代入基准回归模型进行回归。滞后一期的核心解释变量在理论上与当期的核心解释变量具有相关性，而与当期的随机误差项没有相关性，满足工具变量的相关性和无关性条件。回归结果如表 3 第(4)列所示。结果显示，核心解释变量的系数仍然显著为正。

4.3. 异质性分析

4.3.1. 行业异质性分析

为验证数字经济行业与非数字经济行业的数据要素水平对企业数字技术创新水平的影响是否存在不同，本文将两组行业分组回归后进行比较。具体地，参考已有文献[17]，将 25 个大类产业定义为数字经济行业，其他行业则为非数字经济行业。回归结果如表 4 第(1)列和第(2)列所示。可以看到，非数字经济行业的回归系数为 0.052 且不显著，数字经济行业的回归系数为 0.096 且在 1%的水平上显著，说明在数字经济行业中，数据要素对数字技术创新的提升效果比非数字经济行业更强。可能的原因是，数字经济行业的数字化转型程度更高，其物联网、云计算等数字基础设施更完善，能够使企业用低成本获得海量数据，解决传统企业数据采集难、处理慢的瓶颈。此外，数字化转型程度高的企业工作流程更规范、部门间数据标准更统一，有利于部门间数据的整合分析，使企业能够更好地利用数据。以上因素使得数据要素对企业数字技术创新水平的提升作用更大。

Table 4. Heterogeneity analysis
表 4. 异质性分析

	(1)	(2)	(5)	(6)
	非数字经济行业	数字经济行业	国有企业	非国有企业
Data	0.052 (0.037)	0.096*** (0.030)	0.047 (0.042)	0.106*** (0.028)
Lev	0.291** (0.132)	0.223* (0.118)	0.451** (0.187)	0.183* (0.098)
ROA	0.581*** (0.169)	0.625*** (0.155)	0.761** (0.296)	0.457*** (0.112)
Cashflow	-0.328** (0.151)	-0.382*** (0.133)	-0.373* (0.191)	-0.460*** (0.119)
ListAge	0.032 (0.039)	0.136*** (0.033)	0.194*** (0.071)	0.158*** (0.029)

续表

TobinQ	-0.018** (0.008)	-0.013 (0.008)	-0.019 (0.014)	-0.018** (0.007)
Top5	0.393* (0.221)	0.361* (0.216)	1.034*** (0.333)	-0.132 (0.181)
Subsidy	0.088*** (0.013)	0.135*** (0.017)	0.102*** (0.017)	0.121*** (0.014)
RDratio	0.000 (0.000)	0.018*** (0.006)	0.003 (0.010)	0.000* (0.000)
_cons	-0.927*** (0.350)	-1.062*** (0.342)	-1.136** (0.475)	-0.972*** (0.291)
企业固定效应	是	是	是	是
时间固定效应	是	是	是	是

4.3.2. 产权异质性分析

为验证国有和非国有企业数据要素应用水平对数字技术创新的影响是否存在不同，本文将两组企业分组回归后进行比较。回归结果如表 4 第(5)列和第(6)列所示。可以看到，国有企业的回归系数为 0.047 且不显著，非国有企业的回归系数为 0.106 且在 1%的水平上显著，说明在非国有企业中，数据要素应用对数字技术创新的提升效果比国有企业更强。可能的原因是，非国有企业往往面临更激烈的市场竞争，数据要素会成为其优化生产、降低成本的重要工具。相比之下，国企由于市场垄断地位和政策保护，竞争压力较小，数据应用的紧迫性较低，导致数据要素在非国有企业对数字技术创新的提升更显著。

5. 结论与建议

本文选取了 2012~2023 年沪深 A 股上市公司作为样本，利用 python 语言，通过文本分析方法分析上市公司年报文本，测度数据要素，探究了数据要素对企业数字技术创新的驱动作用，并探究了作用的异质性，最终得出如下结论：

- 1) 数据要素的应用能够对企业数字技术创新产生正向影响，这一结论在更换被解释变量、倾向得分匹配法、控制行业固定效应等一系列检验后仍然稳健。
- 2) 数据要素对企业新质生产力的促进作用在数字经济行业和非国有企业中更明显。

综合以上结论，本文提出如下几点建议：

- 1) 推动企业与金融机构共建数据共享平台，提升中小企业的融资可获得性。企业可为低技能员工提供数据技能培训，同时通过激励机制鼓励员工持续学习，适应岗位需求变化。引入 AI 技术分析消费数据与市场趋势，预测潜在风险并及时调整研究方向。鼓励企业与供应商、客户聚焦共性研发难题进行联合研究，缩短数字技术创新周期。
- 2) 优先支持数字经济园区建设，完善数据基础设施，加大 5G、云计算等新型基础设施投入，降低企业数据存储和算力成本，推动数据采集、标注、交易一体化平台发展。针对特定领域制定统一的数据分类标准，实现数据跨系统对接，强化数据基础设施在数据要素提升数字技术创新中的调节作用。
- 3) 对国有企业数字技术创新研发投入给予加计扣除优惠，对引入数据科学家、算法工程师的国企给

予人才补贴,在招投标中优先支持数据要素水平高的国企,倒逼国企提升数据要素水平,激发国有企业的活力和市场竞争能力,强化国企数据要素对数字技术创新的提升作用。

参考文献

- [1] 徐翔,厉克奥博,田晓轩.数据生产要素研究进展[J].经济学动态,2021(4):142-158.
- [2] 李海舰,赵丽.数据成为生产要素:特征、机制与价值形态演进[J].上海经济研究,2021(8):48-59.
- [3] 卢黎歌,李婷.我国数据要素统一大市场构建目标、存在问题与对策分析[J].理论探讨,2022(5):163-167.
- [4] 王亚飞,张璐瑶,钟敏.数据要素利用对企业全要素能源效率的影响机制研究[J/OL].软科学:1-9.
<http://kns.cnki.net/kcms/detail/51.1268.G3.20250425.1256.004.html>, 2025-04-29.
- [5] 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(S2):398-407.
- [6] 夏范社,何德旭,杨明月,等.数字技术创新与企业环境社会责任[J].统计与信息论坛,2025,40(5):72-88.
- [7] 方慧,姜春宇,李小明.数字技术创新与产业链韧性提升:基于节点企业视角的考察[J].经济学动态,2025(2):127-144.
- [8] 王东,李晓.数字技术创新驱动流通企业供应链效率提升效应与机制研究[J].商业经济研究,2025(6):147-150.
- [9] 潘丹丹,范从来,王子敏.风险投资持股与关键数字技术创新[J].现代金融研究,2025,30(3):32-43.
- [10] 夏后学.信息消费试点政策如何影响企业数字技术创新——来自中国A股上市公司的经验证据[J].经济与管理,2025,39(2):21-30.
- [11] 李延喜,杨立佳,董维刚,等.企业竞争压力感知、高管激励与数字技术创新[J].中国软科学,2025(2):140-149.
- [12] 谢丹夏,魏文石,李尧,等.数据要素配置、信贷市场竞争与福利分析[J].中国工业经济,2022(8):25-43.
- [13] Brynjolfsson, E. and McElheran, K. (2016) Data in Action: Data-Driven Decision Making in US Manufacturing. University of Toronto-Rotman School of Management.
- [14] 马震.企业ESG表现对数字技术创新影响的实证研究[J].统计与决策,2025,41(6):172-177.
- [15] 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.
- [16] 赵丽,胡植尧.数据要素、动态能力与企业全要素生产率——破解“数据生产率悖论”之谜[J].经济管理,2024,46(7):55-72.
- [17] 曾国安,苏诗琴,彭爽.企业杠杆行为与技术创新[J].中国工业经济,2023(8):155-173.