

“东数西算”工程如何赋能数据要素 ——基于要素化、市场化和服务化的分析框架

董 竹^{1,2}, 董思莹², 马鹏飞^{1,2}

(1. 吉林大学 数量经济研究中心, 吉林 长春 130012; 2. 吉林大学 商学与管理学院, 吉林 长春 130012)

摘 要:算力与数据要素是中国经济增长的新动能, 异地算力部署已经成为促进企业数据要素开发利用的重要举措。为了检验算力基础设施建设对企业数据要素开发利用的影响, 文章借助 BERT 和 DeepSeek 大语言模型对数据要素相关政策文件以及企业年报中管理层讨论与分析部分进行文本分析, 构建了企业层面数据要素开发利用指标体系。文章基于 2014—2023 年中国沪深 A 股上市公司数据, 以“东数西算”工程作为准自然实验, 实证检验了该政策对企业数据要素开发利用的效果。研究发现, “东数西算”工程能够显著提高企业数据要素开发利用水平, 这一促进作用是通过算力供给提升、算力运载协同、算力普惠渗透和算力应用赋能四个机制来实现。进一步分析发现, 该促进作用在数据治理能力更强的企业、大型企业和算力需求敏感的行业更为显著, 在市场化水平更高、数据开放程度更高以及营商环境更好的地区表现更明显。此外, “东数西算”工程能够促进区域经济增长、缓解区域发展不平衡并降低碳排放强度, 从而实现区域绿色协调发展。文章揭示了算力基础设施的微观赋能机制, 量化了“东数西算”工程的实施效果, 对全国统筹推进算力网络体系建设、优化算力资源布局及充分释放数据要素价值具有重要的理论和实践意义。

关键词:东数西算; 算力基础设施; 数据要素开发利用

中图分类号:F276 **文献标识码:**A **文章编号:**1001-9952(2026)07-0004-15

DOI: 10.16538/j.cnki.jfe.20260610.401

一、引 言

全球数字经济扩张浪潮下, 数据成为关键生产要素, 而算力是处理和传输数据的关键能力, 算力和算力基础设施的发展与国家竞争力密切相关。然而, 数据价值的充分释放仍面临瓶颈问题: 算力资源的空间配置与数据应用需求之间存在结构性失衡。局部区域的算力供给可能无法满足社会需求。当高昂的算力成本与有限的算力获取渠道成为普遍约束时, 企业很难对海量数据资源进行开发利用。为了缓解算力供需的结构性失衡, 应构建不同架构、主体和区域算力资源的互联互通网络, 因此“东数西算”工程应运而生。“东数西算”与“南水北调”“西电东送”工程类似, 其本质是通过构建覆盖全国且要素流动高效的算力基础设施网络体系, 对算力进行空间优化配置。根据《国家数据基础设施建设指引》, 网络设施和算力设施作为国家数据基础

收稿日期: 2025-12-08

基金项目: 国家自然科学基金项目(72402075); 吉林大学研究生创新训练计划项目(2026CX095)

作者简介: 董 竹(1968—), 女, 吉林长春人, 吉林大学数量经济中心、商学与管理学院教授, 博士生导师;

董思莹(2002—)(通讯作者), 女, 吉林长春人, 吉林大学商学与管理学院硕士研究生;

马鹏飞(1989—), 男, 吉林长春人, 吉林大学数量经济中心、商学与管理学院副教授, 硕士生导师。

设施的关键组成部分，与其他部分深度融合、紧密相关，共同为数据要素的高效流通和价值释放提供支撑保障。现有文献关于网络基础设施、数据基础设施的经济后果研究已相对完善（Goldfarb 和 Tucker, 2019; 张涛和李均超, 2023; 刘秉镰等, 2025; Lv 等, 2025），然而对算力基础设施的探究仍处在起步阶段。那么，算力基础设施能否促进企业数据要素开发利用？通过什么机制实现？关于这些问题的探讨对促进数据要素开发利用、驱动经济高质量发展具有重要意义。

2019 年，党的十九届四中全会将数据确立为新型生产要素。近年来，《企业数据资源相关会计处理暂行规定》等一系列政策文件相继出台，初步形成了数据要素政策体系。自从数据要素被确立为新型生产要素以来，学术界对数据要素的确权、交易和定价（Liang 等, 2018; 熊巧琴和汤珂, 2021; 刘涛雄等, 2023）以及非竞争性、非排他性等特性进行了深入研究，这些传统生产要素所不具备的特性是数据要素实现价值创造能力倍增的关键（Jones 和 Tonetti, 2020; 蔡跃洲和马文君, 2021; Tang, 2025）。关于数据价值形态演进的研究，李海舰和赵丽（2021）将其划分为四个层级，依次为数据资源、数据资产、数据商品和数据资本；尹西明等（2022）则构建了“动态整合价值化理论模型”，将数据价值化过程概括为汇聚、确权、治理、交易和应用五个环节。然而，从数据的汇聚处理到场景应用均离不开算力基础设施的底层支撑。那么，探究算力基础设施如何赋能数据要素开发利用全过程？这是一个重要问题。

“东数西算”工程是推进算力基础设施建设、促进数据要素的高效流通与价值释放的重要举措，其构建的算力服务体系包括基础层、平台层、应用与优化层以及政策与战略层。首先，基础层建设了数据中心、智算中心等硬件设施，直接推动了网络基础设施的优化。算力依赖于网络来调度任务和传输数据，而数据价值的实现依赖于算力进行处理并依赖网络进行流通，因此，“东数西算”工程会提升企业的算力供给水平和算力运载效率，进而提升企业数据要素化水平。其次，在平台层以及应用与优化层，阿里云、腾讯云等头部云服务商搭建了全球化云计算平台，通过算力网络的高速互联和智能调度机制为用户提供弹性、便捷的计算资源和服务，企业无需投入巨资自建数据中心，可按需租用计算资源，极大降低了企业数据开发利用的成本，提升了算力普惠水平。最后，在政策与战略层面，政府投资和政策引导有效带动算力产业链上下游的协同发展，促进了数据要素的采集、流通与应用，从而推动了企业的数据要素市场化和服务化进程。综上所述，从数据要素的形态演进过程出发，算力基础设施对企业数据要素开发利用的影响及作用机制是一个重要的实证问题。然而，关于“东数西算”工程对企业层面价值创造效应的实证分析还处于起步阶段，而且现有文献多集中探究数据要素的经济后果，缺乏对其生成机制的分析，“东数西算”工程对企业数据要素开发利用的影响缺乏系统性研究。

因此，本文以 2014—2023 年中国沪深 A 股上市公司为研究样本，实证检验了“东数西算”工程对企业数据要素开发利用水平的影响。研究发现：“东数西算”工程能够显著提高企业数据要素开发利用水平，且经一系列内生性和稳健性检验后依然成立；机制检验表明，其赋能效应主要通过提升算力供给水平、算力运载能力以及算力普惠水平和应用赋能深度来提高企业数据要素开发利用水平；异质性分析显示，“东数西算”工程在数据治理能力更强的企业、大型企业和对算力需求敏感的行业更为显著，在市场化水平更高、数据开放程度更高以及营商环境更好的地区也表现更明显；进一步分析表明，“东数西算”工程能够促进区域经济增长、缓解区域发展不平衡并降低碳排放强度，实现区域绿色协调发展。

本文的研究贡献主要体现在以下几方面：（1）以“东数西算”工程为准自然实验，本文考察算力基础设施对数据要素开发利用的影响，并提出其微观作用机制，表明“东数西算”工程可显著促进区域经济增长、缓解区域发展不平衡并降低碳排放强度，从而实现区域绿色协调发展，

验证了算力基础设施在数字经济时代的重要作用,为其价值创造路径提供了新视角。(2)本文拓展了大语言模型在金融文本分析中的应用场景。本文融合了 BERT 的语义理解能力与 DeepSeek 的技术优势,利用数据要素相关的权威政策文件与企业年报文本,量化了企业数据战略与国家顶层设计的契合度,构建了企业层面数据要素开发利用的指标体系,为数据要素指标的构建提供了新方法。(3)本文从数据要素“供给—流通—应用”的完整价值链视角切入,构建了数据的“要素化、市场化、服务化”框架,为“东数西算”工程促进数据要素开发利用提供了新的理论解释。

二、理论分析和研究假说

(一)“东数西算”工程与数据要素开发利用

在数据价值实现过程中,算力、算法和数据形成了有机整体,算法决定价值取向,而算力则赋予企业释放数据价值的能力(许诺等,2025)。“东数西算”工程本质在于构建全域联动的算力资源调度体系:该体系以分布式算力中心为节点,实时监测算力资源的供需与运行状态,并根据任务需求进行智能分配与统一调度,最终形成一个全域联动的算力网络,实现算力和数据资源的汇聚和共享。“东数西算”工程通过构建全国一体化算力网络,将东部算力需求引导至西部,有助于缓解区域资源错配,使西部的资源禀赋与东部的算力需求形成互补效应,算力资源能够快速、便捷地在区域间流动,从而提升企业数据要素开发利用水平。一方面,算力基础设施建设具备典型的规模经济特征,契合梅特卡夫法则与边际成本递减规律。算力基建投入完成后,随着用户规模的扩张,单位算力建设成本不断下降,呈现边际成本递减特征。同时,算力网络的覆盖范围与接入主体不断扩大,推动资源调度效率提升,形成边际收益递增的叠加效应。这种成本递减和收益递增的双向优化有效降低了企业的算力使用成本,也为数据要素开发利用提供关键支撑。另一方面,作为新型信息基础设施的一部分,算力基础设施网络的搭建极大提高了信息传送以及知识外溢的速度,降低了企业信息获取成本,加速了数据、信息在不同群体之间交换,通过知识溢出效应产生规模报酬递增效应(张杰和付奎,2021),从而使企业更有动力进行数据要素的开发利用。基于此,本文提出以下假说:

假说 1:“东数西算”工程能够提高企业数据要素开发利用水平。

(二)“东数西算”工程与数据要素开发利用的作用机制

关于数据的价值形态演进过程,国内外学者大多从如下两方面来分析:一方面,从数据生命周期的角度,将其概括为从数据生成到利用并创造社会经济价值的一系列环节。Kriksciuniene 等(2015)提出了数据价值化的四环节模型,而 Curry(2016)提出了五环节模型。另一方面,从数据形态特征的角度,李海舰和赵丽(2021)提出“数据资源—数据资产—数据商品—数据资本”的演进过程,中国信息通信研究院发布的《数据价值化与数据要素市场发展报告(2021 年)》提出数据价值化需经历资源化、资产化和资本化三个阶段。在此基础上,金骋路和陈荣达(2022)基于数据的衍生金融属性,即数据的流动性和增值特征,提出了数据要素的金融化。马克思主义政治经济学提出的社会再生产理论认为,任何社会的生存和发展都必须通过生产要素在生产、分配、流通、消费环节的有机循环来实现价值创造。作为重要的生产要素,数据也同样遵从这一基本经济规律。本文从社会再生产理论视角切入,结合数据要素自身具备的特点,认为数据要素开发利用需经历“供给—流通—应用”三个阶段,“供给”即要素化,指原始数据经过归集、整理、加工等系列操作转化为可投入社会再生产的数据要素(陈刚等,2024)。具体来看,原始数据历经汇集与加工处理后,首先形成具备经济社会价值的数据资源;其次,依据数据资产确认条件进

行识别并予以确认；最后，在劳动、资本等传统生产要素的协同参与下，进一步将其研发转化形成数据产品。数据资源、数据资产和数据产品都是数据要素化过程中的不同形态（陈刚等，2024）。“流通”即市场化，指将数据作为关键生产要素纳入市场体系，使数据要素能够遵循市场规律进行确权、评估、流通和共享的过程。在能够为数据要素市场提供稳定供给后，还需要相对完善的数据要素市场体系让数据能够流通交易。市场化的最终目的是将数据要素配置到各个产业，实现数据赋能。若缺乏完善的流通渠道，数据要素的价值也无法有效传递，市场化是连接要素化和服务化的关键桥梁。“应用”即服务化，指将数据要素与其他生产要素进行深度融合，在具体应用场景中释放数据要素价值。该阶段的核心是将数据要素融入各类产业的生产、运营和管理全流程，从而推动产业生产方式、运营模式的变革和生产率的提升（李治国和王杰，2021）。基于此，本文提出了数据要素开发利用的“要素化—市场化—服务化”框架，并进一步提出支撑该框架的作用机制，具体如图1所示。

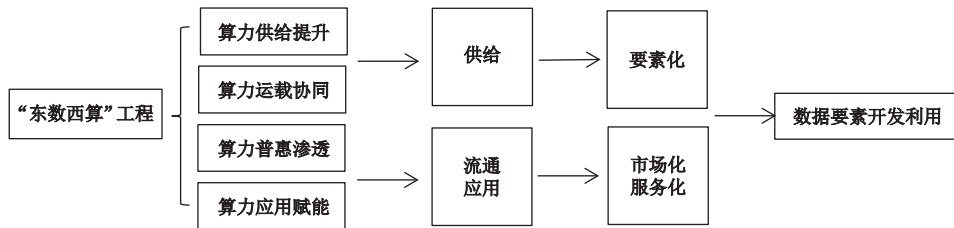


图1 逻辑框架

算力供给提升是数据要素开发利用的基础。依托全国一体化算力网络体系，“东数西算”工程破除跨区域算力壁垒，实现了多源异构算力资源的集约整合和灵活调度。该工程不仅推动大型企业、科研机构在枢纽节点建设数据中心集群，更创新性地搭建了算力交易平台，使中小企业及个人用户能够以即需即用的方式获取算力资源，大幅提升了算力的配置效率和利用水平。算力运载能力决定了算力的时效性，算力需要运力才能发挥价值。数据在光纤中传输需要时间，从东部到西部之间有几千公里的物理距离。而“东数西算”工程通过建设大规模数据中心集群、多层级算力调度平台和算力互联互通体系，便解决了物理距离带来的时间延迟问题。枢纽间和城区间网络传输效能的提升满足了企业快速、就近、灵活、高效的算力联接需求，根据《算力基础设施高质量发展行动计划》的技术标准，城区重要算力基础设施间时间延迟不高于1ms，国家枢纽节点内重要算力基础设施时间延迟不高于5ms，这为数据要素的开发利用提供了坚实的网络支撑。

数据要素化即通过数据的采集、整理、加工等流程，使无序、混乱的原始数据成为有序、有使用价值的数据要素的过程（王艳和杨达，2024）。该过程与算力供给和运载能力密不可分，算力和运力的提升使得企业能够灵活调用算力，从而推进数据要素化。因此，本文提出以下假说：

假说2：“东数西算”工程通过算力供给提升和算力运载协同提高企业数据要素开发利用水平。

为了提升算力服务的普惠水平，国家和地方政府协同发力，采用了“降成本”与“增补贴”的组合政策。一方面，“东数西算”工程通过建立国家枢纽节点专项网络传输计费机制，采用弹性带宽、数据快递等创新模式，并推动算力服务模式从传统机柜租赁、包年包月向随接随用、按需付费等短租模式转型，降低了用户算力使用成本。另一方面，在“十四五”期间，国家针对算力枢纽、数据中心集群以及配套网络建设每年新增投资超4000亿元，并施行减税降费，杭州、深圳以及成都等地方政府发放“算力券”，这一系列政策导向显著优化了算力的使用环境。随着算力使用成本的下降，企业数字化转型的动力被激活，从而带动产业链需求侧的扩大，并最终为数据要素的采集、流通和应用创造更活跃的生态。对于不同规模企业而言，大型企业对算力的

核心诉求更侧重于低时延、高算力,而小型企业对成本极为敏感,其对算力需求的核心在于用得起。此外,不同行业对算力的需求也存在差异,如互联网、金融等对算力需求敏感的行业侧重于低时延、高算力,以用于大数据分析、人工智能训练和实时交易处理等核心场景,而传统农业、服务业等对算力需求非敏感的行业,算力成本的下降对其业务发展的影响相对有限。本文在异质性分析部分进一步探讨该效应在不同规模企业和不同行业的影响。算力不仅是电子信息制造业、软件和信息技术服务业、互联网行业等信息技术产业快速发展的动力来源,而且不断推进工业、农业、金融等传统产业数字化转型升级。在算力的作用下,工业企业的数字化、自动化水平显著提升,“车路协同”的智能网联汽车正在快速发展,高精度、复杂场景的气象模型训练与实时推理也得以实现。数据要素的开发利用需在场景中得以实现,依托“东数西算”工程国家枢纽节点,不同行业的企业能够根据自身需求使用不同类型的算力服务,从而激活数据要素潜能,深度赋能各行各业。

数据要素市场化是指数据作为一种新的生产要素,被纳入市场体系进行确权、定价、流通以及交易的过程。在当前数据要素市场化改革中,企业拥有的数据要素经过市场化手段进行流通交易,可以赋能各行业,推动行业数字化转型。算力向各行业的加速渗透和普惠水平的提升使得企业能够以更低成本、更有针对性地结合行业自身特点使用算力,从而推动数据要素的市场化和服务化。因此,本文提出以下假说:

假说 3:“东数西算”工程通过算力普惠渗透和算力应用赋能提高企业数据要素开发利用水平。

三、研究设计

(一)样本与数据来源

2014 年,大数据被写入政府工作报告,由此开始了数据要素的市场化发展。因此,本文选取 2014—2023 年中国沪深 A 股上市公司作为研究样本,并进行了如下数据处理过程:(1)剔除房地产和金融行业的公司;(2)剔除在样本期间的 ST 公司以及退市的公司;(3)剔除数据缺失的样本;(4)为排除极端值对估计结果的干扰,对文中主要连续变量实施 1% 和 99% 水平上的缩尾处理;(5)考虑到部分上市公司的主营业务属于算力服务商,其经营目标并非通过获取算力赋能实体产业,因此本文将研究样本中可能具有算力供给属性的上市公司剔除,以确保样本企业主要为算力使用方;(6)本文对上市公司主营业务范围、企业名称以及管理层讨论与分析部分进行梳理和识别,剔除其中属于算力服务商的企业样本。经上述数据处理后最终获得 24 146 个观测值。本文数据来源包括:“东数西算”工程设立的国家算力枢纽节点所涉及的省份和城市由人工整理得到;企业年报文本及其管理层讨论与分析(MD&A)部分利用 Python 爬取巨潮资讯网收集整理得到,与数据要素开发利用相关的政策文件来自中国政府网等权威网站;企业基本特征、财务指标相关的数据主要来自 CSMAR 数据库。

(二)模型构建

本文将“东数西算”工程设置的八个国家算力枢纽节点所涉及的省级和城市级行政单位设置为处理组,非枢纽区域设置为对照组,基于双重差分法构建模型(1)进行实证检验:

$$Data_{it} = \alpha_0 + \alpha_1 Treatpost_{it} + \alpha_2 Controls_{it} + \varphi_i + \theta_t + \varepsilon_{it} \quad (1)$$

其中,下标 i 表示企业, t 表示年份; $Data_{it}$ 表示企业数据要素开发利用水平; $Treatpost_{it}$ 表示当年企业所在地区是否实施“东数西算”工程,若企业 i 所在省份或者城市在 2021 年及以后被选为“东数西算”工程示范地,则赋值为 1,否则赋值为 0; $Controls_{it}$ 包括企业特征和区域特征两方面的控制变量;本文加入了个体固定效应 φ_i 以及年份固定效应 θ_t ; ε_{it} 为随机扰动项;本文将标准误

聚类到省份层面； α_1 为本文解释变量的回归系数，表示“东数西算”工程的实施对企业数据要素开发利用水平的影响，预期 α_1 显著为正，即“东数西算”工程会提高企业的数据要素开发利用水平。

（三）数据要素开发利用的指标构建与分析

为了有效衡量企业数据要素开发利用程度，本文利用 BERT 以及 DeepSeek 大语言模型对数据要素相关的政策文件进行主题聚类，构建了包含四个维度、八个主题的指标体系：数据要素基础能力建设、数据要素市场化配置、政府与区域数据协同以及数据要素场景深度应用。指标构建主要分为两个核心模块，分别应用 BERTopic 和 BERT 模型从政策导向和微观企业层面识别和映射数据要素开发利用水平，生成企业层面的指标。BERTopic 是一种自然语言处理的主题建模算法，其核心原理是利用 BERT 模型强大的语义理解能力，将文档转换为包括上下文信息的语义表示，然后对这些语义使用主题建模技术进行聚类分析，并识别出文本数据中的潜在主题。其关键步骤包括：语义嵌入(BERT)、降维压缩(UMAP)、密度聚类(HDBSCAN)和主题表示(c-TF-IDF)。传统的词袋模型主要是基于 TF-IDF 方法来抽取主题，其缺点在于没有充分考虑词语的顺序、语法结构和上下文关系。因此，LDA 无法区分一词多义、同义词和近义词的情况，也无法理解词语在句子或段落中的真正含义，这也是词频法构建指标的局限性所在。在深度学习的情况下，基于预训练词嵌入的聚类方法，尤其是基于 Transformer 模型生成的嵌入向量，其核心突破在于它们是在海量文本数据上通过自监督学习训练得到的，可以充分理解句子含义，显著提高主题建模的效果。在操作过程中，BERTopic 模型选用 Sentence-BERT 库中的“paraphrase-multilingual”模型生成文档嵌入。

（四）“东数西算”工程

2021 年，国家发改委等部门联合发布了《全国一体化大数据中心协同创新体系算力枢纽实施方案》，该文件对全国算力网络枢纽节点进行布局，为“东数西算”工程的落地实施制定了清晰路径。因此，本文将 2021 年作为政策实施年份，并选择北京市、天津市、河北省、江苏省、浙江省、上海市、安徽省、广东省、成都市、重庆市、内蒙古自治区、贵州省、甘肃省、宁夏回族自治区作为政策实施地。

（五）控制变量^①

本文借鉴李雪琴等(2024)、史丹和郑玉(2024)等的做法，选取企业规模(Size)、企业年龄(Age)、财务杠杆(Lev)、盈利能力(Roa)、企业性质(Soe)、成长能力(Growth)、董事会规模(Board)、机构投资者持股比例(Ins)、两职合一(Dual)、独立董事水平(Idp)等企业层面的控制变量，并选取经济发展水平(Rgdp)、地方政府支出(Gov)和网络基础设施(Tele)作为区域层面的控制变量。

四、实证分析

（一）基准回归分析

表 1 报告了“东数西算”工程提高数据要素开发利用水平的基准回归分析结果。其中，列(1)中仅加入了个体和年份固定效应，未加入控制变量，可以发现， $Treatpost$ 估计系数显著为正。根据列(2)、列(3)的结果，解释变量 $Treatpost$ 的估计系数均在 5% 的水平下显著。根据列(3)的结果，“东数西算”工程开始实施后，示范地内企业数据要素开发利用水平提升 0.0101，

^① 限于篇幅，省略变量的描述性统计结果，留存备索。

相当于均值水平的 12.90%，表明“东数西算”工程能够显著提高企业数据要素的开发利用水平，证明了本文假说 1。

表 1 基准回归结果

变量	(1) <i>Data</i>	(2) <i>Data</i>	(3) <i>Data</i>
<i>Treatpost</i>	0.0117** (0.0050)	0.0106** (0.0048)	0.0101** (0.0040)
企业层面控制变量		控制	控制
地区层面控制变量			控制
固定效应	控制	控制	控制
<i>N</i>	24 146	24 146	24 146
<i>Adj. R</i> ²	0.5786	0.5815	0.5821

注：括号内为标准误，***、**、*分别表示在1%、5%和10%水平上显著。下表同。

(二)稳健性检验^①

1. 平行趋势检验

本文使用双重差分方法识别政策效应的基本前提为：在“东数西算”工程尚未发生的时间窗口内，处理组和控制组具有相同的时间变化趋势。为了有效区分“东数西算”的政策效应与潜在的城市固有因素，本文利用事件研究法检验实验组和控制组的平行趋势假设。具体而言，本文选取(-7, 2)的时间窗口，将-1期作为基准期，回归模型设定如下：

$$Data_{i,t} = \alpha_0 + \sum_{n=-7}^{n=2} \alpha_1 Treatpost_{i,t} + \alpha_2 Controls_{i,t} + \varphi_i + \vartheta_t + \varepsilon_{it} \quad (2)$$

根据平行趋势检验结果，在“东数西算”工程实施之前，处理组与控制组的企业数据要素开发利用水平差异并不显著，满足事前平行趋势假设。

2. 安慰剂检验

为控制其他不可观测因素对企业数据要素开发利用水平的影响，以确保本文的结论由“东数西算”工程而非随机因素驱动，本文进一步通过安慰剂检验的方式加以验证。根据图表分析结果，安慰剂检验的系数集中分布于零值附近且呈正态分布，进一步支持了基准回归结论。

3. PSM-DID

国家在选取“东数西算”工程算力枢纽节点时，可能考虑了城市自身的产业基础、资源禀赋和发展定位等因素，导致算力枢纽节点入选过程中存在样本自选择问题，对此本文采用 PSM-DID 来缓解此类内生性问题。首先，以本文微观层面的控制变量作为协变量；其次，采用 1:2 和 1:4 邻近匹配方法以及卡尺匹配方法寻找与“东数西算”工程支撑条件相似的最优对照组，从而最大限度地减少了处理组和对照组之间的系统性差异，并且通过了平衡性检验；最后，采用双重差分法估计“东数西算”工程对企业数据要素开发利用水平的影响。根据回归结果，本文的核心结论具有稳健性。

4. 熵平衡匹配

考虑到使用倾向性得分匹配法对样本进行匹配的过程中，易损失较多样本，本文参考李雪琴等(2024)的研究，利用熵平衡匹配法对实验组和控制组城市进行匹配。该方法通过对控制组

^① 限于篇幅，省略图表分析结果，留存备索。

样本重新赋权，使控制组在均值、方差及偏度等分布特征上与处理组达到一致。根据回归结果，*Treatpost* 的回归系数在 1% 的水平下正向显著，表明“东数西算”工程促进了企业数据要素开发利用。

5. 双重机器学习模型

双重机器学习的核心逻辑在于将传统的因果推断问题拆解为两个独立的预测阶段，借助机器学习算法优化参数估计过程有效提高估计结果的精准度与稳健性。在实证研究中，高维度控制变量的冗余问题容易引发“维度诅咒”，变量间的非线性关系也可能造成模型设定偏误，干扰因果识别结果。为规避上述问题，本文采用双重机器学习方法重新对基准回归进行检验，参考张涛和李均超(2023)的研究，将样本分割比例设定为 1:4 和 1:7，采用一般性交互式双重机器学习模型开展实证检验，并以随机森林作为核心预测模型。此外，本文逐步纳入控制变量的一次项、二次项及三次项以反映变量的非线性特征。根据回归结果，在借助双重机器学习方法缓解模型内生性、变量维度冗余及模型设定问题后，核心研究结论依然成立。

6. 排除其他政策干扰

样本期内，存在多个与“东数西算”相似的政策冲击，可能会对本文的回归结果产生干扰。因此，本文收集了 4 个相似的政策进行检验，包括国家大数据综合试验区政策(*Bigdata*)、“宽带中国”示范政策(*Broadband*)、数据要素市场化政策(*Datamarket*)以及公共数据开放政策(*Dataopen*)，都可能对企业数据要素开发利用水平产生一定影响。本文对上述政策设置虚拟变量，并代入回归进行检验，根据回归结果，本文基准回归结果依然稳健。

7. 战略性披露的稳健性检验

本文采用文本分析方法对企业数据要素开发利用水平进行度量，而企业年报 MD&A 文本往往包含大量的政策和战略性披露，可能出现企业描述得好但其实际开发利用水平并不高的情况，从而导致本文的回归结果出现偏误。因此，本文对被解释变量进行替代性检验。参考黄勃等(2023)和李心武等(2025)的研究，采用企业当年独立授权的数字经济专利数量加 1 取对数来表示企业的数字化程度(*Diginnov*)，数字经济专利授权基于实质审查，可以更真实客观地反映企业对数据要素的开发利用水平，从而避免企业年报对数据要素开发利用行为夸大的问题。根据回归结果，*Treatpost* 回归系数仍然正向显著，并且满足事前平行趋势假设，说明本文基准回归结果依然稳健。本文企业数据要素开发利用水平的测度有赖于企业年报文本的信息披露，可能受到企业信息披露质量和年报文本特征的干扰。因此，本文进行如下稳健性检验：(1)剔除样本期间内信息披露质量等级为不合格的企业样本；(2)剔除样本期间内信息披露违规的企业样本；(3)引入影响企业信息披露行为的相关变量。参考罗进辉和巫奕龙(2024)的研究，本文在回归模型中控制了管理层讨论与分析文本的语调特征(*Tone*)、可读性特征(*Read*)、文本相似性(*Simil*)，以缓解因信息披露差异所带来的测量误差。根据回归结果，“东数西算”工程提高了企业数据要素的开发利用水平。

8. 其他稳健性检验

本文还进行了其他稳健性检验：(1)为了排除企业因为享受算力补贴而更改企业注册地而产生的样本自选择问题，本文剔除了样本期间注册地址发生改变的企业后重新进行回归。(2)本文主回归部分通过对上市公司主营业务范围、企业名称以及管理层讨论与分析等部分进行梳理和识别，剔除其中属于算力服务商的企业样本。因此，本文进行稳健性检验，剔除可能具有算力供给属性的行业。根据回归结果，*Treatpost* 的回归系数依然正向显著，表明本文的核心结论是稳健的。(3)由于本文的基准回归将聚类层级设定在省份层面，为了缓解潜在的小样本

聚类问题,本文参考 Roodman 等(2019)和许文立和孙磊(2023)的研究,进一步采用 wild cluster bootstrap 方法进行了补充检验。根据回归结果,本文的结论不受小样本聚类问题的影响,因此,本文的结论是稳健的。

五、进一步分析

(一)机制分析

本文进一步验证该促进效果的核心作用机制。参考张国胜等(2024)的研究,本文构建如下影响机制分析模型:

$$Data_{it} = \beta_0 + \beta_1 Treatpost_{it} + \beta_2 Mech_{it} + \beta_3 Treatpost_{it} \times Mech_{it} + \beta_4 Controls_{it} + \varphi_i + \vartheta_t + \varepsilon_{it} \quad (3)$$

其中, $Mech_{it}$ 为机制变量,包括智算中心建设(IDC)、数字人才聚集($Labor$)、长途光缆线路长度密度($Distance$)、人均互联网接入端口数($Accessport$)、政府数字补贴($Subsidy$)、数字普惠金融水平($Finance$)以及数实产业融合($Techcov$),其余控制变量与基准回归模型一致,本文主要关注核心解释变量 $Treatpost_{it} \times Mech_{it}$ 的估计系数 β_3 。

如果“东数西算”工程具有提高算力供给水平和算力运载能力的作用,那么“东数西算”工程在算力供给较多和算力运载能力较强的地方将更显著,因此,算力供给水平和算力运载能力指标应当具有正向调节效应。参考许诺等(2025)的研究,本文收集并整理了工业和信息化部审批发放的互联网数据中心经营许可证数据,用企业持有的 IDC 经营许可证数量加 1 取对数表示企业的智算中心建设水平(IDC)。城市数字人才集聚水平则参照孙伟增等(2023)的研究,本文采用信息传输、计算机服务和软件业从业人员规模的对数值表示($Labor$)。数据中心建设越多、数字人才集聚水平越高,表明算力供给越多。本文采用长途光缆线路密度($Distance$)和人均互联网接入端口数($Accessport$)度量网络传输效能,表示算力运载能力。网络传输效能的提升可以降低时延和宽带成本。根据表 2 的回归结果,“东数西算”工程对数据要素的开发利用的促进效果在算力供给水平较高、算力运载能力较强的情况下更加明显,因此,假说 2 成立。

如果“东数西算”工程具有提升算力普惠水平和算力应用赋能的作用,那么“东数西算”工程在算力普惠水平较高以及算力应用赋能较强的地方更为显著,因此,算力普惠水平和算力应用赋能指标应当具有正向调节效应。本文参考蔡宏波等(2024)的研究,若上市公司财务报表附注中的“政府补助明细”中出现与数字化有关的激励项目,则认定为政府数字化激励,将金额加总并加 1 取对数后得到政府数字化激励金额($Subsidy$),同时采用北京大学编制的数字普惠金融指数($Finance$)代理算力普惠水平。政府数字化激励水平和数字普惠金融指数越高,表明区域算力资源越具有普惠性,服务可得性越强,企业的算力使用成本相应越低。本文参考黄先海和高亚兴(2023)对数实产业技术融合的定义,对企业年度数实产业技术融合累计次数加 1 后取自然对数,得到数实产业技术融合指标($Techcov$)。数实产业融合度反映了数字技术在实体企业中的渗透程度,该指标越大,表明算力应用赋能越强。根据表 2 的回归结果,算力普惠水平和算力应用水平发挥了正向调节效应,因此,假说 3 成立。

(二)异质性分析

本文接下来聚焦于数据要素开发利用的核心环节“要素化、市场化、服务化”框架以及企业和行业层面展开异质性分析。

1. 要素化的异质性

参考甄红线等(2023)和吴武清等(2024)的研究,本文将企业数字化转型程度作为企业数据治理能力的代理变量,企业数字化转型程度越高,其对数据的获取、处理与应用能力通常越强。

表 2 机制分析

变量	算力供给水平		算力运载能力		算力普惠水平		算力应用赋能
	(1) <i>Data</i>	(2) <i>Data</i>	(3) <i>Data</i>	(4) <i>Data</i>	(5) <i>Data</i>	(6) <i>Data</i>	(7) <i>Data</i>
<i>Treatpost</i>	0.0112*** (0.0038)	0.0056 (0.0042)	0.0070* (0.0040)	0.0072* (0.0041)	0.0099** (0.0039)	-0.0142* (0.0081)	0.0110*** (0.0034)
<i>IDC</i>	0.3910*** (0.0100)						
<i>Treatpost</i> × <i>IDC</i>	0.0817*** (0.0286)						
<i>Labor</i>		0.1779 (0.2004)					
<i>Treatpost</i> × <i>Labor</i>		0.2338*** (0.0406)					
<i>Distance</i>			0.0012 (0.0009)				
<i>Treatpost</i> × <i>Distance</i>			0.0052** (0.0019)				
<i>Accessport</i>				0.0012 (0.0008)			
<i>Treatpost</i> × <i>Accessport</i>				0.0051** (0.0020)			
<i>Subsidy</i>					-0.0004** (0.0002)		
<i>Treatpost</i> × <i>Subsidy</i>					0.0006*** (0.0002)		
<i>Finance</i>						0.0816 (0.0524)	
<i>Treatpost</i> × <i>Finance</i>						0.1079*** (0.0295)	
<i>Techcov</i>							0.0004 (0.0017)
<i>Treatpost</i> × <i>Techcov</i>							0.0097*** (0.0008)
控制变量	控制	控制	控制	控制	控制	控制	控制
固定效应	控制	控制	控制	控制	控制	控制	控制
<i>N</i>	24 148	24 146	17 653	17 653	24 138	23 281	23 888
<i>Adj. R</i> ²	0.5862	0.5832	0.5628	0.5628	0.6034	0.5836	0.5856

按照数字化转型程度的中位数，本文将企业分为数据处理能力强和数据处理能力弱两组。根据回归结果，*Treatpost* 的回归系数仅在数据处理能力较强的企业中显著，可能的原因在于：强大的数据处理能力是企业有效承接和充分利用外部算力的基础，当企业具备高效的数据处理、分析和整合能力时，“东数西算”工程所提供的低成本、高效率算力才能被充分吸收和转化。因此，企业要加强自身对数据要素的处理能力，主动融入国家算力网络，更好地挖掘数据要素的价值。

2. 市场化的异质性

本文采用王小鲁等编制的《中国分省份市场化指数报告(2024)》来衡量企业所处地区的市场化发展程度。本文以市场化指数的年度中位数为分组依据开展异质性检验。根据回归结果,在市场化水平较高的省份中, *Treatpost* 的回归系数显著为正,而另一组中并不显著,表明“东数西算”工程对数据要素开发利用的促进作用在市场化水平较高的省份中更显著。为刻画地区数据开放水平,本文选用复旦大学与国家信息中心数字中国研究院联合发布的《中国开放数林指数》作为代理指标,将样本数据按照数据开放水平分为两组,排在前 50 的城市定义为高数据开放组,其余城市定义为低数据开放组。根据回归结果,在数据开放水平较高的城市中, *Treatpost* 的回归系数显著为正,而另一组中并不显著,表明“东数西算”工程对数据要素开发利用的促进作用在数据开放水平较高的城市中更显著。

3. 服务化的异质性

产业结构是数据要素赋能产业发展的基础。为衡量产业结构高级化水平,本文参考干春晖等(2011)的研究,使用第三产业与第二产业产值的比值进行表示,产业结构升级意味着工业结构由劳动密集型向技术密集型转变。本文认为高产业结构的地区会受益于“东数西算”工程。本文按照该指标的中位数对样本进行分组回归。根据回归结果,在高产业结构的城市中, *Treatpost* 的回归系数显著为正,而另一组中不显著,表明“东数西算”工程对数据要素开发利用水平的促进作用在高产业结构地区中更显著。

4. 企业与行业的异质性

在企业层面,本文将企业总资产按照中位数划分进行回归分析,发现“东数西算”工程对企业数据要素开发利用的赋能效果在大型企业中更为显著。在行业层面,本文依据行业对算力的依赖程度,将样本企业划分为算力需求敏感行业和算力需求非敏感行业。^①由于软件和信息技术服务业以及互联网和相关服务业这两个行业可以直接销售算力,本文在全样本中剔除了具有算力供给属性的相关行业数据。根据回归结果,“东数西算”工程对企业数据要素开发利用的赋能效果,在算力需求敏感行业中更为显著。这可能是由于相较于算力需求非敏感行业,敏感行业能够更充分地利用算力供给提升和算力成本优势,提高企业数据要素开发利用水平。

(三)拓展性分析

八个算力枢纽节点可划分为两类功能主体,粤港澳大湾区、成渝地区、长三角地区、京津冀地区属于第一类主体,这四个算力枢纽节点与我国重要城市群重合,在市场、技术、资金以及人才等方面具有较好的优势;其余算力枢纽地区属于第二类主体,依托丰富的可再生能源和适宜的气候地理优势,数据中心绿色低碳发展潜力突出。^②本文分别将东部枢纽和西部枢纽进行单独的实证回归分析。在检验东部枢纽的政策效果时,剔除西部枢纽的数据中心集群涉及的企业样本,在检验西部枢纽的政策效果时,剔除东部枢纽的数据中心集群地区所涉及的企业样本。^③根据回归结果,东部地区数据中心集群的建设对当地企业数据要素开发利用水平的提升作用更强,其原因可能在于:东部地区原本的数据中心建设相对完善,能够承接企业对实时性要求较高的业务需求,同时西部数据中心的建设又能够为东部地区的数据存储、加工完善及离线分析提供服务支持,因此,东部地区对企业数据要素开发利用水平的整体促进效果更好。

① 算力需求敏感的行业主要包括:制造业、科学研究和技术服务业、电信、广播电视和卫星传输服务业和交通运输以及仓储和邮政业。

② 请参考国家发改委等部门发布的《全国一体化大数据中心协同创新体系算力枢纽实施方案》。

③ 本文剔除了八大枢纽中十大数据中心集群涉及城市的企业样本,这样处理的原因是:当检验西部枢纽的政策效果时,其算力是通过网络流向东部乃至全国的,如果剔除东部枢纽所在全部省份的企业样本,可能会导致对照组样本大量损失。

（四）经济后果分析

本文从宏观层面展开分析，深入探究“东数西算”工程对区域绿色协调发展的影响。本文选取区域经济增长($Rgdp$)、区域经济差距(Gap)和碳排放强度(CE)作为被解释变量，检验“东数西算”工程对区域绿色协调发展的影响。参考郜栋玺等(2025)的研究，本文构建如下模型：

$$Economics_{c,t} = \gamma_0 + \gamma_1 Treatpost_{c,t} + \gamma_2 Controls_{c,t} + \lambda_c + \eta_t + \varepsilon_{ct} \quad (4)$$

其中，下标 c 表示企业， t 表示年份； $Economics_{c,t}$ 表示本文的经济后果变量，包括 $Rgdp_{c,t}$ 、 $Gap_{c,t}$ 和 $CE_{c,t}$ ； $Treatpost_{c,t}$ 表示当年地级市 c 是否实施“东数西算”工程，当地级市 c 在 2021 年及以后被选为东数西算政策示范地，则赋值为 1，否则赋值为 0；本文加入的控制变量包括地方财政支出(Gov)、固定资产投资(Inv)、科技投入(Sci)、城镇化率($Urban$)、医疗卫生水平($Health$)、人口规模(Pop)、互联网普及率($Internet$)；本文还加入了城市固定效应 λ_c 和年份固定效应 η_t ； ε_{ct} 为随机扰动项，并将标准误聚类到地级市层面。

1. “东数西算”工程能否缩小数字鸿沟？

受数字技术发展水平差异的影响，东南沿海等经济发达的地区受到数字红利的影响较大(张勋等, 2021)。本文参考徐斌等(2019)和方锦程等(2023)的研究，采用全国地级市人均 GDP 的自然对数来刻画区域经济增长($Rgdp$)，以地级市内部各区县 GDP 的标准差与其内部各区县 GDP 均值的比值来衡量城市区域经济发展不平衡程度(Gap)。根据回归结果， $Treatpost$ 的回归系数均在 10% 的水平下显著，事前趋势在统计上也不存在显著差异，表明“东数西算”工程不仅能够促进区域经济增长，还能够降低区域经济发展不平衡程度，从而促进区域协调发展。

2. “东数西算”工程能否助力绿色低碳发展？

“东数西算”工程可充分发挥西部地区的气候、能源和环境方面的优势，各数据中心积极引入绿色能源，梳理算力产业链各环节的“碳足迹”，发布创新低碳产品和解决方案，促进算力全产业链节能减排。参考景国文和汪达(2024)的研究，本文采用地级市二氧化碳排放量与 GDP 的比值衡量碳排放强度(CE)。根据回归结果，“东数西算”工程会降低地区的碳排放强度，助力绿色低碳发展。

六、结论与启示

“东数西算”工程是一项国家级战略工程，该工程统筹能源禀赋、区位战略和产业基础等关键因素，通过设立八大国家枢纽节点和十个数据中心集群，构建全国一体化算力网络体系，旨在解决算力资源供需空间失衡的难题，并促进区域协调发展和绿色低碳发展。为了研究“东数西算”工程对企业数据要素开发利用的影响，本文基于 2014—2023 年中国沪深 A 股上市公司数据，借助 BERT 和 DeepSeek 大语言模型构建了企业层面数据要素开发利用的指标体系，测度了企业数据要素开发利用水平，并以“东数西算”工程作为准自然实验，实证检验了国家算力资源跨域调配政策对企业数据要素开发利用的影响及其作用机制。研究发现：“东数西算”工程能够显著提高企业数据要素开发利用水平；本文验证了算力供给提升、算力运载协同、算力普惠渗透以及算力应用赋能的作用机制。结合本文的研究结论，提出如下政策建议：

第一，加快西部算力枢纽开发进程，缩小东西部枢纽赋能差距。本文拓展性分析发现，“东数西算”工程对东部算力枢纽的赋能效果显著优于西部枢纽，这与西部算力枢纽开发起步较晚、配套设施不完善、资源利用效率不高的现实情况密切相关。因此，应加大对西部算力枢纽的政策倾斜与资源投入。一方面，完善西部枢纽配套设施建设，优化网络传输通道，提升算力传输效率；另一方面，出台针对性扶持政策，使数据处理企业向西部枢纽集聚，推动西部算力资源

充分利用。应结合西部清洁能源丰富的优势,推动算力枢纽与清洁能源深度融合,降低算力能源消耗,实现绿色低碳发展,进一步提升西部枢纽的竞争力,促进东西部算力枢纽协同发展。

第二,全面兼顾企业算力需求,推动算力普惠赋能提质增效。本文对不同规模企业的异质性分析表明,“东数西算”工程对企业数据要素开发利用的赋能效果在大型企业中更为显著,而小型企业因资金、技术和人才等短板导致算力的赋能效果不明显。因此,政府可以进一步优化政策设计,在持续发挥对大型企业赋能作用的同时,加大对小微企业的扶持力度。具体而言,可扩大地方算力券的覆盖范围和发放额度,重点向小微企业倾斜,降低其算力使用成本,以适配小微企业“用得起、用得上”的核心需求。此外,也应加强对小微企业的技术培训与人才支持,提升其数据处理、算力应用能力,帮助小微企业突破技术瓶颈,充分释放算力赋能的潜在价值,实现大型企业与小微企业的协同发展。

第三,优化算力服务生态,驱动数据协同发展。本文的机制分析发现,为实现数据要素的供给、流通与应用全链条贯通,应关注算力供给提升、运载协同、普惠渗透及应用赋能核心机制,通过优化算力服务生态,加速推进数据要素化、市场化与服务化进程。一方面,应强化算力供给基础和运载效率。本文认为政府应加大对东西部网络基础设施的投入,重点提升连接东西部的骨干网络带宽,降低网络传输时延,有效提升网络传输效能,并设立合理的网络传输计费机制,降低企业跨域算力调度成本;另一方面,应促进普惠渗透及应用赋能。当前,算力产业发展与地区资源禀赋结合不够紧密的现象仍然存在,算力对实体经济的渗透率较低。本文认为政府应当引导企业探索算力在不同行业和领域创新应用的情景,推动算力与人工智能、大数据、云计算等数字技术的深度融合,为传统产业数字化转型提供技术支持,促进算力产业链上下游企业的交流合作,形成良好的产业生态环境。

主要参考文献:

- [1]蔡宏波,张馨文,汤城建.政府数字化激励促进企业纵向并购的机制研究[J].兰州大学学报(社会科学版),2024,(4):69-82.
- [2]蔡跃洲,马文君.数据要素对高质量发展影响与数据流动制约[J].数量经济技术经济研究,2021,(3):64-83.
- [3]陈刚,颜斌斌,汤珂.数据的要素化与资产化:理论辨析与实践探索[J].国际经济评论,2024,(5):153-176.
- [4]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,(9):124-141.
- [5]干春晖,郑若谷,余典范.中国产业结构变迁对经济增长和波动的影响[J].经济研究,2011,(5):4-16.
- [6]郜栋玺,陈肖雄,张龙耀,等.大数据产业政策提升了货币政策传导效率吗?——基于国家大数据综合试验区的准自然实验[J].财经研究,2025,(5):4-18.
- [7]黄勃,李海彤,刘俊岐,等.数字技术创新与中国企业高质量发展——来自企业数字专利的证据[J].经济研究,2023,(3):97-115.
- [8]黄先海,高亚兴.数实产业技术融合与企业全要素生产率——基于中国企业专利信息的研究[J].中国工业经济,2023,(11):118-136.
- [9]金骋路,陈荣达.数据要素价值化及其衍生的金融属性:形成逻辑与未来挑战[J].数量经济技术经济研究,2022,(7):69-89.
- [10]景国文,汪达.能源转型政策的碳减排效应——基于新能源示范城市的准自然实验[J].工业技术经济,2024,(2):106-115.
- [11]李雪琴,郑酌基,韩先锋.乘“数”而上:政府数据治理赋能企业数字创新[J].数量经济技术经济研究,2024,(12):68-88.

- [12]李治国, 王杰. 数字经济发展、数据要素配置与制造业生产率提升[J]. 经济学家, 2021, (10): 41-50.
- [13]刘秉镰, 袁博, 刘玉海. 数字基础设施如何畅通区域间资本要素流动——基于企业注册大数据的证据[J]. 数量经济技术经济研究, 2025, (1): 72-92.
- [14]刘涛雄, 戎珂, 张亚迪. 数据资本估算及对中国经济增长的贡献——基于数据价值链的视角[J]. 中国社会科学, 2023, (10): 44-64.
- [15]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.
- [16]李心武, 卢闯, 牛煜皓, 等. 企业数字化转型中耐心资本的价值发现功能研究[J]. 数量经济技术经济研究, 2025, (11): 178-198.
- [17]罗进辉, 巫奕龙. 数字化转型与审计费用: 负能还是赋能?[J]. 会计研究, 2024, (6): 148-161.
- [18]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.
- [19]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.
- [20]王艳, 杨达. 中国式管理会计体系变革: 从数据要素到数据资产[J]. 管理世界, 2024, (10): 171-188.
- [21]吴武清, 李祁恒, 章柳漪, 等. 公共数据资源与企业全要素生产率——基于地方政府数据开放的准自然实验[J]. 系统工程理论与实践, 2024, (6): 1815-1833.
- [22]熊巧琴, 汤珂. 数据要素的界权、交易和定价研究进展[J]. 经济学动态, 2021, (2): 143-158.
- [23]徐斌, 陈宇芳, 沈小波. 清洁能源发展、二氧化碳减排与区域经济增长[J]. 经济研究, 2019, (7): 188-202.
- [24]许诺, 毛聚, 毛新述, 等. 算力部署、数据跨域流动与企业全要素生产率——来自智算中心的证据[J]. 中国工业经济, 2025, (4): 61-79.
- [25]许文立, 孙磊. 市场激励型环境规制与能源消费结构转型——来自中国碳排放权交易试点的经验证据[J]. 数量经济技术经济研究, 2023, (7): 133-155.
- [26]尹西明, 林镇阳, 陈劲, 等. 数据要素价值化动态过程机制研究[J]. 科学学研究, 2022, (2): 220-229.
- [27]甄红线, 王玺, 方红星. 知识产权行政保护与企业数字化转型[J]. 经济研究, 2023, (11): 62-79.
- [28]张国胜, 严鹏, 李欣珏, 等. 大数据要素集聚、技术能力缺口与生产率区域差距[J]. 中国工业经济, 2024, (10): 118-136.
- [29]张杰, 付奎. 信息网络基础设施建设能驱动城市创新水平提升吗?——基于“宽带中国”战略试点的准自然试验[J]. 产业经济研究, 2021, (5): 1-14.
- [30]张勋, 万广华, 吴海涛. 缩小数字鸿沟: 中国特色数字金融发展[J]. 中国社会科学, 2021, (8): 35-51.
- [31]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, (4): 113-135.
- [32]Curry E. The big data value chain: definitions, concepts, and theoretical approaches[A]. Cavanillas J, Curry E, Wahlster W. New horizons for a data-driven economy[M]. Cham: Springer, 2016: 29-37.
- [33]Goldfarb A, Tucker C. Digital economics[J]. Journal of Economic Literature, 2019, 57(1): 3-43.
- [34]Jones C I, Tonetti C. Nonrivalry and the economics of data[J]. American Economic Review, 2020, 110(9): 2819-2858.
- [35]Kriksciuniene D, Sakalauskas V, Kriksciunas B. Process optimization and monitoring along big data value chain[A]. Abramowicz W. Business information systems workshops. Cham: Springer, 2015: 75-86.
- [36]Liang F, Yu W, An D, et al. A survey on big data market: Pricing, trading and protection[J]. IEEE Access, 2018, 6: 15132-15154.

- [37]Lv K J, Zheng H Y, Ge L M, Digital empowerment to reduce air pollution: The impact of urban network infrastructure construction on PM2.5[J]. *Journal of Asian Economics*, 2025, 98: 101930.
- [38]Roodman D, Nielsen M Ø, Mackinnon J G, et al. Fast and wild: Bootstrap inference in Stata using boottest[J]. *The Stata Journal: Promoting communications on statistics and Stata*, 2019, 19(1): 4–60.
- [39]Tang M H. Data element utilization and firm value[J]. *International Review of Financial Analysis*, 2025, 102: 104004.

How does the “East-to-West Computing Resource Transfer” Project Empower Data Elements? An Analytical Framework Based on Elementization, Marketization, and Serviceization

Dong Zhu^{1,2}, Dong Siying², Ma Pengfei^{1,2}

(1. *Center for Quantitative Economics, Jilin University, Changchun 130012, China;*

2. *School of Business and Management, Jilin University, Changchun 130012, China)*

Summary: Computing power and data elements have become new drivers of China’s economic growth, and cross-regional deployment of computing power has emerged as a key measure to facilitate the development and utilization (D&U) of enterprise data elements.

In order to examine the impact of computing infrastructure construction on the D&U of enterprise data elements, this paper uses the BERT and DeepSeek large language models to conduct text analysis on policy documents related to data elements and the MD&A sections of corporate annual reports, and constructs an indicator system for the D&U of enterprise data elements. Based on data from China’s A-share listed companies from 2014 to 2023, this paper treats the “East-to-West Computing Resource Transfer” project as a quasi-natural experiment and empirically examines its impact on the D&U of enterprise data elements. The results show that the project enhances the D&U of enterprise data elements, driven by four mechanisms: improved computing power supply, coordinated computing power transportation, inclusive penetration of computing power, and application-enabled computing power. Further analysis indicates that this effect is more pronounced among firms with stronger data governance capabilities, larger firms, and industries with higher sensitivity to computing power demand; it is stronger in regions with higher levels of marketization, higher levels of data openness, and better business environments. Additionally, the project fosters regional economic growth, alleviates development imbalances, and reduces carbon emission intensity, thereby facilitating coordinated green development.

The marginal contributions of this paper are reflected in the following aspects: First, by leveraging the policy shock of the “East-to-West Computing Resource Transfer”, it examines the impact of computing power infrastructure on the D&U of enterprise data elements and proposes the micro-level enabling mechanisms. Second, by integrating BERT’s semantic understanding capabilities with DeepSeek’s technical advantages, it constructs an indicator system for the firm-level D&U of data elements and offers a new methodology for constructing data element indicators. Third, it constructs a “elementization–marketization–servitization” framework for data and aligns the micro-level enabling mechanisms of the project with the process of the D&U of enterprise data elements, providing a new theoretical explanation for how the project promotes the D&U of enterprise data elements.

Key words: “East-to-West Computing Resource Transfer”; computing power infrastructure; development and utilization of enterprise data elements

(责任编辑 顾 坚)