

大数据发展对企业绿色技术选择的影响研究

——基于我国上市公司年报的文本分析

汪莉¹ 张聪¹, 汪亚楠²

(1. 华东师范大学 经济与管理学院, 上海 200062; 2. 华南理工大学 经济与金融学院, 广东 广州 510006)

摘要: 绿色技术进步是绿色经济发展的核心动力, 大数据发展对企业绿色技术选择的作用效果不断增强。采用我国 2011—2020 年上市公司面板数据, 通过 Python 软件对我国上市公司年报进行匹配与抓取, 利用变异系数法处理上市公司年报关键词并构建大数据发展综合指标, 采用绿色专利规模和绿色专利占比两个指标表征企业的绿色技术选择, 实证检验大数据发展对企业绿色技术选择的影响效应。研究发现, 大数据发展能够显著优化企业绿色技术选择, 绿色投资、智力资本和研发投入是其中的重要作用机制, 该影响效应在区域、产权、管理者短视主义等多个维度中存在较为明显的异质性。研究结论在替换核心解释变量、改变样本时间范围、进行工具变量回归等多种稳健性检验中依然成立。研究结论在一定程度上能够为激发大数据潜能和制定绿色生产决策提供重要的参考依据。

关键词: 大数据发展; 绿色技术选择; 绿色投资; 智力资本; 研发投入

中图分类号: F124.3 **文献标志码:** A **文章编号:** 1672-6049(2023)06-0079-11

一、引言

在我国, 由于长期以来对环境保护方面重视不足和缺乏行之有效的管理制度, 导致一些地区出现了“重经济、轻环境”“污染治理、生态恢复脱节”等一系列环境问题。随着国家对环境保护的高度重视, 可持续发展已成为当前最重要的主题之一。党的二十大报告明确指出, 中国式现代化是人与自然和谐共生的现代化, 尊重自然、顺应自然、保护自然是全面建设社会主义现代化国家的内在要求^①。促使经济向绿色、低碳方向发展是实现高品质发展的关键步骤, 也是全面构建绿色技术创新体系的重要保障。如今, 绿色发展成为我国经济可持续发展的主要内容, 而绿色技术是绿色经济发展的主要推动力^[1], 在企业降低碳排放中发挥着重要作用。绿色技术创新成果向绿色生产力^[2]的转变, 不仅能够推动生产方式向绿色低碳改变, 而且能够推动能源清洁生产, 从而实现节能减碳的绿色发展目标。

那么, 如何实现和兼顾绿色化与低碳化发展? 随着数字经济的不断进步, 我国大数据相关领域发展势头强劲, 技术进步明显, 取得了一系列突破, 有力推动了我国经济高质量发展和民生改善。大数据是国家和企业的重要新型战略资源, 不仅能够推进我国网络化、智慧化和精准化生产模式, 而且可

收稿日期: 2023-08-18; 修回日期: 2023-10-07

基金项目: 教育部人文社会科学研究规划基金项目“‘双碳’目标下数字化创新赋能低碳经济转型的内在机理与政策研究”(22YJA790055); 广东省哲学社会科学规划 2022 年度青年项目“碳中和目标约束下数字技术驱动广东贸易碳减排的效应评估及对策研究”(GD22YYJ06); 国家社会科学基金重大项目“新发展格局下构建开放型经济体系研究”(22ZDA063); 广州市哲学社会科学规划 2023 年度课题“广州实现高水平科技自立自强路径政策研究: 基于创新链产业链资金链深度融合视角”(2023GZQN09)

作者简介: 汪莉(1988—), 女, 安徽黄山人, 经济学博士, 华东师范大学经济与管理学院副教授, 研究方向为数字金融与风险管理; 张聪(2002—), 女, 河南南阳人, 华东师范大学经济与管理学院学生, 研究方向为绿色低碳转型; 汪亚楠(1992—), 女, 江西上饶人, 通讯作者, 经济学博士, 华南理工大学经济与金融学院副教授, 博士生导师, 研究方向为数字技术创新。

①信息来源于中国政府网, <https://www.gov.cn/zhuanti/zggcdescqgdbdh/sybgqw.htm>。

以提高企业的生产、运营和管理效率。《促进大数据发展行动纲要》《关于培育大数据产业集群的指导意见》《关于工业大数据发展的指导意见》等相关文件,进一步凸显了大数据对微观企业发展的基础性和引领性作用^[3]。面对大数据蓬勃发展的趋势和经济绿色发展的需求,在数字化技术与传统产业融合过程中,亟须思考大数据发展对微观企业绿色技术选择的影响及其作用机理。

绿色技术概念最早由 Braun and Wield^[4]提出,一方面,以大数据技术为根本的数字金融显著促进了绿色技术创新^[5],而技术创新渗透到发展、生产和消费过程中^[6],同时大数据技术应用能够系统优化生产和消费过程,实现经济全链条的绿色发展;另一方面,人工智能、云计算、大数据分析等前沿技术影响企业在绿色制造、废品制造、高效制造等方面的创新业务实践,倒逼企业进行绿色技术创新,以实现资源的最大化利用并减少对环境的污染^[7]。但问题是,现有文献多从理论分析的角度探讨大数据发展如何赋能企业低碳发展,无法说明大数据发展对企业绿色技术选择的影响程度如何,以及在区域、产权性质、管理者短视主义等层面是否存在异质性。基于此,本文利用2011—2020年2771家上市公司的面板数据,构建大数据发展指标和绿色技术选择指标,探究大数据发展是否会优化企业绿色技术选择及背后的作用路径。

总体而言,本文可能的边际贡献为:一是,采用文本分析方法抓取了2011—2020年我国A股上市公司年报中与“大数据”有关的关键词,进而构建微观企业大数据发展的指标体系。同时考虑以绿色发明专利数量和绿色发明专利占比表征绿色技术选择,探究大数据应用与微观企业绿色技术选择的互动关系与影响机理。二是,将大数据发展划分为技术和应用两个关键维度,再将技术与应用维度分为数据挖掘、算法开发、移动服务和智能制造,从不同维度探究大数据发展对企业绿色技术选择的作用差异。三是,探讨大数据发展通过何种传导路径优化微观企业绿色技术选择,并论证绿色投资、智力资本以及研发投入在其中发挥的中介作用机制。四是,考察大数据赋能微观企业绿色技术选择的异质性影响,探究不同特征条件下大数据如何优化企业绿色技术选择。

二、文献梳理与研究假说

(一) 大数据发展与绿色技术选择

随着数字化技术的迅猛发展,大数据已成为实体经济中关键的生产要素。大数据的应用改变了企业创新元素的融合方式,减少了企业在创新过程中的交易成本和治理成本,优化了企业创新表现。Braun and Wield^[4]认为绿色技术促进企业增强国际竞争力和迈向全球价值链中高端,尤其是在数字化时代,绿色技术已逐步变成企业获益并扩大市场份额的关键工具^[8]。由此可见,探讨大数据发展与绿色技术选择关系的内在逻辑演变,具有极其重要的意义。

技术创新作为企业和国民经济发展的重要因素,是促进产业转型和增强企业竞争力的根本手段。在“双碳”目标的背景下,绿色技术正成为企业追求经济、社会和环境效益的重要手段,是企业实现低碳转型的重要选择^[9]。现有研究表明,大数据发展是推动绿色技术创新的关键驱动力^[1],El-Kassar and Singh^[10]实证研究发现,企业大数据应用会影响企业的绿色创新活动。Song *et al.*^[11]认为绿色创新能够激励企业选择和升级绿色技术,大数据更是驱动企业绿色技术创新的重要途径。一方面,大数据的应用可以加快提升地区数字化水平和增强企业技术整合能力,促进企业绿色创新。通过优化生产流程和生产工艺,大数据发展可以有效提高资源采集效率,实现资源最优配置和优化绿色技术选择。另一方面,大数据技术应用不仅能够实现企业间的信息技术资源共享,而且能够降低企业在节能减排和环境治理方面的信息搜寻成本,进而有助于企业强化绿色技术创新^[12],而绿色技术创新是绿色技术进步的重要动力。基于此,本文提出假说1。

假说1: 大数据发展有助于优化企业绿色技术选择。

(二) 绿色投资与绿色技术选择

绿色投资,作为生态发展的关键组成部分,依托科学合理使用资源和保护环境的根本准则,并采取社会责任投资作为途径,旨在实现生态均衡和居民健康,从而推动经济的持续发展。Ren *et al.*^[13]认为绿色投资和技术创新是减少环境污染、促进经济增长和减少碳排放的最有力手段。大

数据技术为推动绿色发展提供了新渠道、新选择,大数据技术深度赋能传统行业的数字化和低碳化改造。大数据技术也将激励绿色投资的增加,为绿色产业发展提供强大的资金支持。绿色投资规模的扩大稳步提升了清洁能源和绿色基础设施的覆盖率,促进节能环保产业的培育和发展,为经济可持续发展注入新动能^[14]。一方面,绿色投资作为新型的资源配置方式,促使企业将稀缺资源配置到绿色技术及清洁能源开发等领域,推动绿色科技的研发与落地应用,对企业绿色技术选择起到显著的激励作用。另一方面,绿色投资可以有效解决资金约束的问题,推动绿色技术在社会最优效率水平下稳步创新,为企业提升绿色技术提供发展根基。基于此,本文提出假说2。

假说2:大数据发展通过增加绿色投资来优化绿色技术选择。

(三) 智力资本与绿色技术选择

相较于传统企业,新经济企业通常以创新为主导,属于以智力资本为主的轻资产型公司。无形资产作为智力资本的主要形式^[15],能够创造可观的经济价值。大数据技术被广泛应用到数字经济等诸多领域,有效缓解信息不对称的问题,降低信息搜寻成本,促进企业资源利用效率的改善^[16],不断提高企业创新能力与创新能力,激励无形资产的增加。无形资产是企业自主创新成果的重要组成部分,其核心技术价值对无形资产及企业市场价值有着决定性影响^[17]。一方面,包括绿色技术研究成果、员工技能等在内的智力资本能够提升企业的绿色技术创新水平,为企业绿色技术进步提供动力;另一方面,大数据时代下良好的企业声誉有利于降低融资成本和引进技术型人才,增强企业绿色技术选择的意愿和能力。基于此,本文提出假说3。

假说3:大数据发展通过扩大智力资本规模来优化绿色技术选择。

(四) 研发投入与绿色技术选择

企业研发投入往往高度依赖外部融资渠道^[18],具有投入资金大、调整成本高等特点,导致研发投入的融资成本较高。申明浩和谭伟杰^[19]指出大数据应用能够减少信息不对称带来的逆向选择和道德风险,进而拓宽企业融资资源的获取渠道,在提高企业研发投入的同时,依托大数据的应用优势,企业能够降低研发过程中的不确定性,为研发活动累积大量的资源信息,提升企业研发积极性和产出效率^[20],进而助推产品研发模式创新。杨俊等^[21]指出更高的大数据共享程度会吸引更大的研发投入,具有十分明显的乘数效应。然而,企业研发投入是绿色创新的重要驱动力,也是影响绿色创新绩效的重要因素^[22],大量投入并且合理利用企业获得的知识存量和专利可以提高绿色技术创新能力,进而优化绿色技术选择。此外,企业会加大在更新生产设备和绿色技术工艺研发上的研发投入,以实现节约资源和能源、减少废物产生和降低有害物质排放等目的,促进绿色技术的转化和实现。基于此,本文提出假说4。

假说4:大数据发展通过促进研发投入来优化绿色技术选择。

三、数据来源、指标构建及计量模型选择

(一) 数据来源

本文以2011—2020年中国A股所有上市公司作为初始样本,依次剔除如下样本:(1)房地产、金融、保险类公司;(2)上市未超过3年的公司;(3)ST、*ST公司;(4)公司未上市时期的观测值;(5)关键变量缺失的公司。最终得到3111家上市公司和31037个样本,由于部分控制变量存在缺失值现象,使得表5和表6中的样本总量分别为31033和31028。公司层面大数据发展指标均来源于公司年报中的文本分析,其余公司层面指标均来源于Wind数据库。中国经济不确定指数来源于Baker *et al.*^[23]编制的中国经济政策不确定性指数。

(二) 变量说明

1. 公司层面大数据发展指数

本文将企业大数据发展分成大数据技术普适化和大数据应用创新化两个二级指标^①。首先,对于

①限于篇幅,正文未展示关于“大数据”的159个关键词,感兴趣的读者可向作者索要。

大数据技术普适化,数据是核心,因此如何挖掘获取海量数据成为关键。算法和算力是实现大数据过程中必不可少的基础保证。因此,本文进一步在大数据技术普适化指标下设置数据挖掘和算法开发两个三级指标。其次,对于大数据应用创新化,人们对于移动办公、无接触服务、自动化生产等新型生活工作模式的需求大大提升,据此构建移动服务和智能制造两个三级指标。

对于大数据发展的测度,本文借鉴张叶青等^[24]的研究方法,对上市公司年报中披露的信息进行文本分析,提取出159个与大数据相关联的关键词。通过Python对我国上市公司年报进行匹配与抓取,构造大数据发展指标,度量我国上市公司的大数据发展情况。年报中的文本信息充分反映出公司的战略特征和未来发展方向,可以从整体上刻画公司的大数据发展状况,具有一定的可行性和科学性。为了剔除个别公司年报造假情况,本文在获取关键词词频后利用变异系数法对各维度指标进行客观赋权,最终得到总指标。

2. 绿色技术选择

参考王林辉等^[25]的研究思路,本文用绿色专利规模和绿色专利占比两个指标表征企业的绿色技术选择。绿色专利规模分别用公司申请和获得的绿色发明专利数、绿色新型实用专利数及绿色外观设计专利之和衡量,实证分析中对该指标进行对数化处理。绿色专利占比则通过公司申请和获得的绿色专利占全部专利的比值来表征。

3. 中介变量

本文涉及三个中介变量,分别为绿色投资、智力资本、研发投入。首先,大数据效应促进了绿色技术选择偏向的优化,也激励着绿色投资规模扩大,而且绿色投资可以为绿色产业发展提供强大的资金支持。本文采用企业的环境资本支出衡量绿色投资总额,进行自然对数处理。其次,本文使用上市公司年报财务报表附注中的新增无形资产衡量无形资产总额,用其自然对数衡量企业的智力资本。最后,研发投入是统计年度内企业投入到基础研究、应用研究和试验发展上面的经费,并进行自然对数处理。

(三) 计量模型构建

为了考察大数据发展对绿色技术选择的影响,本文构建如下基准回归模型:

$$GTS_{it} = \beta_0 + \beta_1 bigdata_{it} + \beta_2 X_{it} + Year + Ind + \varepsilon_{it} \quad (1)$$

在上式中, i 和 t 分别表示公司维度和时间维度。被解释变量绿色技术选择 GTS 包括公司绿色专利申请量 AGL 、绿色专利获得量 OGL 、绿色专利申请量占比 AGR 和绿色专利获得量占比 OGR 。核心解释变量表示企业大数据发展指数,二级指标为大数据技术普适化指数 $tech$ 和大数据应用创新化指数 $apply$,以及三级指标数据挖掘词频 A 、算法开发词频 B 、移动服务词频 C 、智能制造词频 D 。 X 表示控制变量, $Year$ 与 Ind 分别表示时间固定效应和行业固定效应, ε_{it} 为随机误差项。

四、实证检验结果

(一) 基准回归结果

表1汇报了大数据发展影响绿色技术选择的基准回归结果,列(1)至列(2)、列(3)至列(4)分别以绿色专利规模、绿色专利比例作为被解释变量,回归中大数据发展系数均显著为正,在1%的水平下显著。这意味着大数据发展对绿色专利规模和结构都产生了显著的促进作用,假说1得以验证。大数据迅速发展和快速渗透可以提高生产效率和优化资源配置,减少信息不对称及供需不匹配造成的资源错配现象,降低外部搜寻成本,继而增强企业绿色技术创新的积极性和驱动力。

表1 大数据发展影响绿色技术选择的基准效应

变量	绿色专利水平		绿色专利比例	
	(1) 绿色专利 申请量	(2) 绿色专利 获得量	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比
大数据发展	0.0438*** (0.0030)	0.0299*** (0.0024)	0.3687*** (0.0477)	0.2636*** (0.0394)

表 1(续)

变量	绿色专利水平		绿色专利比例	
	(1) 绿色专利 申请量	(2) 绿色专利 获得量	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比
净资产收益率	0.001 2 (0.001 1)	-0.000 1 (0.000 9)	0.012 8 (0.017 3)	-0.000 4 (0.014 2)
资产回报率	0.002 2 ^{***} (0.000 5)	0.001 7 ^{***} (0.000 4)	0.020 6 ^{***} (0.007 9)	0.015 3 ^{**} (0.006 5)
换手率	-0.006 1 (0.008 2)	-0.009 3 (0.006 4)	-0.654 7 ^{***} (0.130 1)	-0.749 3 ^{***} (0.107 3)
第一大股东持股比例	0.000 3 (0.000 3)	0.000 2 (0.000 2)	-0.002 1 (0.004 2)	0.001 9 (0.003 5)
前十大股东持股比例	-0.000 5 [*] (0.000 3)	-0.000 2 (0.000 2)	-0.000 9 (0.004 2)	-0.001 2 (0.003 4)
机构投资者持股比例	0.001 3 ^{***} (0.000 1)	0.001 0 ^{***} (0.000 1)	0.007 5 ^{***} (0.002 2)	0.003 1 [*] (0.001 8)
资产负债率	0.001 9 ^{***} (0.000 2)	0.001 5 ^{***} (0.000 1)	0.006 5 ^{**} (0.002 9)	0.004 4 [*] (0.002 4)
独立董事比例	-0.000 5 (0.000 6)	-0.000 1 (0.000 4)	-0.036 4 ^{***} (0.008 9)	-0.022 7 ^{***} (0.007 3)
管理费用比率	-0.001 5 ^{***} (0.000 5)	-0.001 6 ^{***} (0.000 4)	-0.029 0 ^{***} (0.008 1)	-0.035 6 ^{***} (0.006 7)
年份固定效应	是	是	是	是
行业固定效应	是	是	是	是
观测值	31 037	31 037	31 037	31 037
调整的 R ²	0.132 5	0.123 7	0.117 4	0.118 9

注: ***、**和* 分别表示在 1%、5%和 10% 的显著性水平下显著,括号内为标准误。限于篇幅,以下各表格仅汇报绿色专利占比的回归结果。

表 2 和表 3 汇报了大数据发展的二级指标和三级指标的回归结果,回归系数均通过了 1% 水平下的显著性检验,说明大数据技术和大数据应用都对绿色技术选择具有显著的优化作用。比较两者的回归系数可知,大数据应用对绿色技术选择的促进作用更强。可能是因为大数据技术作为底层技术,对于绿色技术发展主要提供的是支撑和保障作用。大数据应用是在大数据技术基础上进行二次创新,构建出应用型产品或商业模式,更有利于释放大数据潜能,并实现绿色技术成果转化。表 2 的列(3)至列(6)回归结果显示,相较于数据挖掘,算法开发对绿色技术选择产生了更强的正向影响。从表 3 的列(3)至列(6)可以看出,相较于移动服务,智能制造对绿色技术选择的优化作用更为明显。

表 2 大数据技术对企业绿色技术选择的影响效应

变量	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比	(5) 绿色专利 申请占比	(6) 绿色专利 获得占比
大数据技术	0.133 2 ^{***} (0.022 3)	0.116 6 ^{***} (0.018 4)				
数据挖掘			0.036 7 ^{***} (0.008 8)	0.032 0 ^{***} (0.007 3)		
算法开发					0.049 2 ^{***} (0.010 9)	0.052 7 ^{***} (0.009 0)

表2(续)

变量	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比	(5) 绿色专利 申请占比	(6) 绿色专利 获得占比
控制变量	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
行业固定效应	是	是	是	是	是	是
观测值	31 037	31 037	31 037	31 037	31 037	31 037
调整的 R ²	0.116 7	0.118 8	0.116 2	0.118 2	0.116 3	0.118 6

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

表3 大数据应用对企业绿色技术选择的影响效应

变量	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比	(5) 绿色专利 申请占比	(6) 绿色专利 获得占比
大数据应用	0.206 1*** (0.026 5)	0.146 3*** (0.021 8)				
移动服务			0.051 6*** (0.009 4)	0.036 0*** (0.007 7)		
智能创造					0.077 9*** (0.009 7)	0.064 6*** (0.008 0)
控制变量	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
行业固定效应	是	是	是	是	是	是
观测值	31 037	31 037	31 037	31 037	31 037	31 037
调整的 R ²	0.117 5	0.118 9	0.116 6	0.118 3	0.117 6	0.119 5

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

(二) 异质性研究

1. 区域异质性

表4汇报了我国东部、中部和西部的异质性回归结果,无论被解释变量是绿色专利申请占比还是绿色专利获得占比,大数据发展的拟合系数均在1%的统计水平下显著为正,但该拟合系数存在一些差异,中部地区系数远远大于东部和西部地区。可能的解释为:中部地区借助大数据发展便利,充分发挥其区位和基础建设优势,在承接沿海地区产业转移等方面发展迅猛,大数据发展对绿色技术选择的影响效应不断增强。东部地区创新优势比较明显,通过建立研发和制造基地以及加强区域合作,部分企业的绿色技术发展已趋于饱和,在推动绿色技术选择方面,大数据发展水平的优势略低于中部地区。西部地区受其地理位置和资源状况的限制,与东中部地区相比处于一定的劣势,数字技术与数字基础设施建设发展尚不完善,在充分发挥大数据发展优势优化企业绿色技术选择偏向上有进一步的发展空间。

表4 大数据发展影响绿色技术选择的区域异质性效应

变量	东部地区		中部地区		西部地区	
	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比	(5) 绿色专利 申请占比	(6) 绿色专利 获得占比
大数据发展	0.514 3*** (0.050 6)	0.261 2*** (0.041 6)	0.767 6*** (0.098 5)	0.574 9*** (0.081 9)	0.480 0*** (0.105 5)	0.386 3*** (0.089 3)
控制变量	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是

表4(续)

变量	东部地区		中部地区		西部地区	
	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比	(5) 绿色专利 申请占比	(6) 绿色专利 获得占比
行业固定效应	是	是	是	是	是	是
观测值	21 364	21 364	5 319	5 319	4 354	4 354
调整的 R ²	0.051	0.049	0.051	0.048	0.038	0.036

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

2. 产权异质性

Laffont and Tirole^[26]研究表明,国有企业因存在严重的委托代理问题导致技术创新的积极性和效率低下,难以对经营者进行有效的监督与激励。此外,由于政策性任务压力,国企的目标趋向于政治化或非经济化,从而影响企业的生产和运营决策^[27]。而非国有企业的约束与限制在这些方面较小,具有较高的创新效率,进行绿色技术创新的意愿更加强烈。因此,本文假设大数据发展将对非国有企业绿色创新产生比国有企业更为显著的推动效应。基于这一假设,本文将样本根据所有权性质划分为国有企业和非国有企业两组进行回归分析。表5结果显示,大数据发展的回归系数均显著为正数,非国有企业组中的回归系数都高于国有企业组。由此可见,大数据发展水平提升对国有企业和非国有企业的绿色技术选择都产生了促进作用,对非国有企业的促进作用强于国有企业,支持了上述论点。

3. 管理者短视主义异质性

本文以管理者短视主义指数是否高于样本平均水平为划分依据,分成两组分别进行回归,表6的回归结果显示,大数据发展的估计系数均显著为正,对绿色技术选择优化效应在弱管理者短视主义的企业样本中更强。究其原因,公司管理层短视主义会使其决策偏向于期限较短、风险较低的投资项目,追求短期目标,而减少期限较长的绿色技术创新活动。此外,管理层的短视行为可能会削减公司的研发和资本投入,进而对创新活动造成影响,且对探索式创新具有边际递增的抑制效应。

五、影响机制与稳健性检验

(一) 影响机制分析

表7中的列(1)至列(3)、列(4)至列(6)、列(7)至列(9)分别汇报了绿色投资、智力资本、研发投入

表5 大数据发展影响绿色技术选择的产权异质性效应

变量	国有企业		非国有企业	
	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比
大数据发展	0.383 6*** (0.061 8)	0.287 3*** (0.051 1)	0.326 3*** (0.075 0)	0.187 0*** (0.061 1)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
行业固定效应	是	是	是	是
观测值	19 595	19 595	11 438	11 438
调整的 R ²	0.127	0.134	0.118	0.110

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

表6 大数据发展影响绿色技术选择的管理者短视主义异质性效应

变量	弱短视主义		强短视主义	
	(1) 绿色专利 申请占比	(2) 绿色专利 获得占比	(3) 绿色专利 申请占比	(4) 绿色专利 获得占比
大数据发展	0.411 9*** (0.072 9)	0.289 3*** (0.059 7)	0.259 0*** (0.060 1)	0.169 3*** (0.050 2)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
行业固定效应	是	是	是	是
观测值	16 465	16 465	14 563	14 563
调整的 R ²	0.110 6	0.115 8	0.109 7	0.111 4

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

入的中介效应回归结果,回归结果通过中介效应检验,关键变量系数在1%的统计显著性水平下显著,实证结果支持了三条影响机制均存在部分中介效应,充分验证了本文的假说2至假说4。首先,在绿色投资传导渠道中,大数据发展能够将企业经营活动进行数据化,减少绿色产业交易成本和传递过程中的信息损失,并为优化企业的投资决策提供帮助,使得更多资金被用于绿色技术水平的提升,从而提升企业绿色技术选择。其次,在智力资本传导渠道中,依托大数据可以有效降低公司的外部交易成本^[28],同时也能降低其信息搜索及知识获取成本,缓解信息不对称,更好地对生产经营活动进行监测,降低研发成本和创新风险,促进智力资本增加,进而显著优化企业的绿色技术选择。最后,在研发投入传导渠道中,大数据发展和应用明显拓宽了企业外源融资渠道,降低企业研发的不确定性和风险,从而充分提高了企业的绿色技术创新能力并优化了企业的绿色技术选择。

表7 大数据发展影响绿色技术选择的中介效应

变量	(1) 绿色投资	(2) 绿色专利申请占比	(3) 绿色专利获得占比	(4) 智力资本	(5) 绿色专利申请占比	(6) 绿色专利获得占比	(7) 研发投入	(8) 绿色专利申请占比	(9) 绿色专利获得占比
大数据发展	0.1280*** (0.0249)	0.2141** (0.1035)	0.1273 (0.0874)	0.4769*** (0.0113)	0.3176*** (0.0627)	0.1752*** (0.0516)	0.2385*** (0.0084)	0.1744*** (0.0650)	0.0914* (0.0538)
绿色投资		0.3698*** (0.0437)	0.2976*** (0.0369)						
智力资本					0.1124*** (0.0361)	0.1225*** (0.0297)			
研发投入								0.5653*** (0.0510)	0.4824*** (0.0422)
控制变量	是	是	是	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是	是	是	是
行业固定效应	是	是	是	是	是	是	是	是	是
观测值	9201	9201	9201	21865	21865	21865	22563	22563	22563
调整的R ²	0.2492	0.1361	0.1337	0.3834	0.1149	0.1189	0.3790	0.1096	0.1123

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

(二) 稳健性检验

1. 替换解释变量

本文通过更换关键词各维度的赋值方式,采用熵权法构建大数据发展指标代替变异系数法得到的总指标,结果见表8的列(1)和列(2)。回归结果显示,大数据发展对企业绿色技术选择的影响系数在1%的统计水平下显著为正。

2. 改变样本范围

国务院在2015年颁布了《促进大数据发展行动纲要》,我国的大数据发展与应用得到了全面的推动。本文选择2015年之后的上市公司数据进行实证研究,表8的列(3)和列(4)显示,2015年之后大数据发展对企业绿色技术选择仍然表现为显著的正向促进作用。

3. 工具变量法

启动于2009年的“基于2021年第12期学科拔尖学生培养试验计划”(又称“珠峰计划”)首轮

表8 大数据发展影响绿色技术选择的稳健性检验

变量	熵权法		2015年之后	
	(1) 绿色专利申请占比	(2) 绿色专利获得占比	(3) 绿色专利申请占比	(4) 绿色专利获得占比
大数据发展	0.3833*** (0.0487)	0.2738*** (0.0401)	0.3673*** (0.0640)	0.2525*** (0.0525)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
行业固定效应	是	是	是	是
观测值	31037	31037	20111	20111
调整的R ²	0.1175	0.1190	0.1136	0.1174

注:***、**和* 分别表示在1%、5%和10%的显著性水平下显著,括号内为标准误。

选取了 17 所院校的理工科专业进行试点。研究发现^[24 29-30]，该计划增加了毕业生进入理工行业的倾向，从而扩充了理工科人才的供应，而理工科教育强调与大数据紧密相关的数学、编程和工程设计技能。此外，基础研究处于“研究-应用-生产”的科研链条起始端，是整个科学体系的源头，也是解决企业技术问题的根本，“珠峰计划”对企业借助大数据实现绿色技术的创新与应用具有重要意义。

参考张叶青等^[24]的研究方法，“珠峰计划”的影响始于 2010 年入学的本科生，而这些学生大多在 2014 年或之后毕业。鉴于 2014 年毕业生的时间节点可能跨越两个年度，在后续工具变量的回归分析中，2014 年的数据被排除在外。具体工具变量的构造公式如下：

$$IV_{it} = \frac{\ln\left(\sum_{k=1}^n \frac{1}{distance_{ik}}\right)}{N_c} \times post_t \quad (2)$$

其中 i 代表各个公司， k 为首批入选的试点高校， c 表示公司的注册地所在城市； $distance_{ik}$ 表示从高校所在的地级市至公司所在地级市的直线距离； n 为参与试点的高校数量； N_c 表示 2014 年位于公司注册城市的上市公司总数； $post_t$ 代表时间虚拟变量，其中 2014 年以后被标记为 1，2014 年之前标记为 0。按照工具变量的构建，上市公司与试点高校的接近程度决定了其受政策影响的程度。接近度越高，公司提升大数据使用水平的可能性也越大。当一家上市公司所在城市拥有更多的上市公司时，其对周边的影响可能会相对减弱，因此，从理论角度来看，符合工具变量中关于相关性的条件要求^[24]。同时，这一工具变量难以通过其他途径影响公司的绿色技术决策，满足工具变量的排他性标准^[24]。

表 9 报告了工具变量的回归结果和相关检验，列(1)至列(2)描述了工具变量的回归结果，列(3)至列(4)是对工具变量取对数处理后的回归结果，两者的结果基本一致，工具变量和大数据发展均在 1% 的水平下显著为正。进一步对工具变量进行有效性检验，结果显示工具变量满足了相关性假设并通过了识别不足检验。

表 9 大数据发展影响绿色技术选择的工具变量回归结果

变量	(1)	(2)	(3)	(4)
	第一阶段结果 大数据发展	第二阶段结果 托宾 Q	第一阶段结果 大数据发展	第二阶段结果 托宾 Q
工具变量	2.1857*** (0.084)		0.1984*** (0.008)	
大数据发展		2.044*** (0.281)		2.077*** (0.289)
控制变量	是	是	是	是
行业固定效应	是	是	是	是
年份固定效应	是	是	是	是
观测值	24 712	24 712	24 712	24 712
弱工具变量检验	127.182		93.454	
F 统计量 (Cragg-Donald Wald)				
识别不足检验	127.371		93.722	
LM 统计量 (Anderson canon. corr.)				

注：***、**和* 分别表示在 1%、5%和 10% 的显著性水平下显著，括号内为标准误。

六、主要结论与启示
在当前宏观环境下，全球经济发展具有较大不确定性，大数据发展作用

不断凸显，逐渐转变为推动经济高质量增长的新驱动力，助力中国经济向绿色、可持续发展方向转型。本文实证研究了大数据发展进程对企业绿色技术选择的影响效应，得到如下研究结论：第一，大数据发展能够显著增强企业绿色技术选择，推动企业技术创新，实现高效率、低能耗，促进经济可持续发展。在进行多种稳健性检验后，该结论仍然成立。第二，绿色投资、智力资本和研发支出是大数据发展影响绿色技术选择的重要影响机制，均存在部分中介效应。第三，大数据发展水平对绿色技术选择的优化影响在中部地区、非国有企业样本和管理者弱短视主义样本中更为显著。

基于上述研究结论，本文提出如下政策建议：(1) 增强相关政策配适度，为绿色技术发展赋能。助力绿色技术发展需要各级政府部门继续强化大数据发展的建设力度，推进数字技术与数字基础设施

建设,筑牢提升企业大数据发展水平的根基。此外,政府应当以鼓励发行绿色债券和提供低息贷款或风险投资资金的形式对绿色创新项目提供融资支持,并制定税收优惠的激励政策减轻企业前期面临的资金压力。在积极推动大数据发展和实施绿色激励政策的同时,应当激励大数据与传统经济的融合,鼓励企业进行绿色创新,以实现可持续的正向发展循环。(2)提高绿色投资能力、增加智力资本,助力经济高质量发展。一方面,提高绿色财政政策的扶持力度,优化政策工具,进一步吸引社会资本流向绿色产业发展和绿色技术创新领域。对于不同类别的绿色投资制定针对性策略以提高绿色投资效能,助力绿色产业链升级和发展,推动企业绿色技术进步。另一方面,实现企业的绿色技术进步亟须推动专利、商标及品牌声誉等无形资产的建设,以企业发展战略为导向,大力推动绿色技术发展,助力经济可持续发展。(3)改善企业治理结构,降低管理者短视主义。一方面,优化企业治理结构,建立完善的激励机制和创新评估标准,缓解绿色创新动力不足的问题;在企业内部建立绿色创新文化,并与初创企业、研究机构等外部创新生态系统建立联系,提升创新思维和技术。另一方面,应当以提升人力资本为抓手,健全完善人才培育体系。既要培养具有高综合素质的大数据发展领域的专业化人才,加快大数据发展进程;也要提升管理层的认知降低其短视主义,更好地认知社会责任并权衡短期决策和长期可持续性之间的关系。

参考文献:

- [1]金芳,齐志豪,梁益琳.大数据、金融集聚与绿色技术创新[J].经济与管理评论,2021,37(4):97-112.
- [2]郭丰,杨上广,任毅.数字经济、绿色技术创新与碳排放——来自中国城市层面的经验证据[J].陕西师范大学学报(哲学社会科学版),2022,51(3):45-60.
- [3]刘志彪,徐天舒.我国制造业数字化改造的障碍、决定因素及政策建议[J].浙江工商大学学报,2023(2):92-105.
- [4]BRAUN E, WIELD D. Regulation as a means for the social control of technology[J]. Technology analysis & strategic management, 1994, 6(3): 259-272.
- [5]FENG S, ZHANG R, LI G. Environmental decentralization, digital finance and green technology innovation[J]. Structural change and economic dynamics, 2022, 61: 70-83.
- [6]XUE L, ZHANG Q, ZHANG X, et al. Can digital transformation promote green technology innovation[J]. Sustainability, 2022, 14(12): 7497.
- [7]CHEN J, WANG J. Digital input and manufacturing innovation efficiency[J]. Economic survey, 2022, 39(3): 78-88.
- [8]ZHANG D, RONG Z, JI Q. Green innovation and firm performance: evidence from listed companies in China[J]. Resources, conservation and recycling, 2019, 144: 48-55.
- [9]胡汉辉,申杰.数字经济、绿色创新与“双碳”目标——“减排”和“增效”视角[J].南京财经大学学报,2023(4):79-88.
- [10]EL-KASSAR A N, SINGH S K. Green innovation and organizational performance: the influence of big data and the moderating role of management commitment and HR practices[J]. Technological forecasting and social change, 2019, 144: 483-498.
- [11]SONG M, FISHER R, KWOH Y. Technological challenges of green innovation and sustainable resource management with large scale data[J]. Technological forecasting and social change, 2019, 144: 361-368.
- [12]庞瑞芝,张帅,王群勇.数字化能提升环境治理绩效吗?——来自省际面板数据的经验证据[J].西安交通大学学报(社会科学版),2021,41(5):1-10.
- [13]REN S, HAO Y, WU H. How does green investment affect environmental pollution? Evidence from China[J]. Environmental and resource economics, 2022, 81: 25-51.
- [14]石英华,刘帅.“双碳”目标下绿色投资政策实践、问题及建议[J].地方财政研究,2022(10):11-17.
- [15]刘海生.嵌入无形资产的会计恒等式研究[J].商业经济与管理,2008(9):61-68.
- [16]赵宸宇,王文春,李雪松.数字化转型如何影响企业全要素生产率[J].财贸经济,2021,42(7):114-129.
- [17]苑泽明,金宇,王天培.上市公司无形资产评价指数研究——基于创业板上市公司的实证检验[J].会计研究,2015

(5): 72 – 79 + 95.

- [18]解维敏,方红星. 金融发展、融资约束与企业研发投入[J]. 金融研究 2011(5): 171 – 183.
- [19]申明浩,谭伟杰. 数字化与企业绿色创新表现——基于增量与提质的双重效应识别[J]. 南方经济 2022(9): 118 – 138.
- [20]刘意,谢康,邓弘林. 数据驱动的产品研发转型: 组织惯例适应性变革视角的案例研究[J]. 管理世界 2020, 36(3): 164 – 183.
- [21]杨俊,李小明,黄守军. 大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论[J]. 经济研究 2022, 57(4): 103 – 119.
- [22]HE Y, FU F, LIAO N. Exploring the path of carbon emissions reduction in China's industrial sector through energy efficiency enhancement induced by R&D investment[J]. Energy, 2021, 225: 207 – 217.
- [23]BAKER S R, BLOOM N, DAVIS S J. Measuring economic policy uncertainty[J]. Quarterly journal of economics, 2016, 131(4): 1593 – 1636.
- [24]张叶青,陆瑶,李乐芸. 大数据应用对中国企业市场价值的影响——来自中国上市公司年报文本分析的证据[J]. 经济研究 2021, 56(12): 42 – 59.
- [25]王林辉,王辉,董直庆. 经济增长和空气质量相容性政策条件——环境技术进步方向视角下的政策偏向效应检验[J]. 管理世界 2020, 36(3): 39 – 60.
- [26]LAFFONT J J, TIROLE J. A theory of incentives in procurement and regulation[M]. Cambridge: MIT Press, 1993.
- [27]杨继生,阳建辉. 行政垄断、政治庇佑与国有企业的超额成本[J]. 经济研究 2015, 50(4): 50 – 61 + 106.
- [28]杜传忠,管海锋. 数字经济与我国制造业出口技术复杂度——基于中介效应与门槛效应的检验[J]. 南方经济, 2021(12): 1 – 20.
- [29]宋弘,陆毅. 如何有效增加理工科领域人才供给? ——来自拔尖学生培养计划的实证研究[J]. 经济研究 2020, 55(2): 52 – 67.
- [30]TAMBE P. Big data investment, skills, and firm value[J]. Management science, 2014, 60(6): 1452 – 1469.

(责任编辑: 刘淑浩; 英文校对: 谈书墨)

The Impact of Big Data Development on Corporate Green Technology Selection: Text Analysis Based on Annual Reports of Listed Companies in China

WANG Li¹, ZHANG Cong¹, WANG Yanan²

(1. School of Economics and Management, East China Normal University, Shanghai 200062, China;

2. School of Economics and Finance, South China University of Technology; Guangzhou 510006, China)

Abstract: Green technological progress is the core driving force of green economic development, and the effect of Big Data development on the selection of green technologies is being continuously enhanced. This paper uses the panel data of Chinese listed companies from 2011 to 2020, matches and captures the annual reports of listed companies in China through Python software, uses the coefficient of variation method to process the keyword frequency of annual corporate reports to construct a comprehensive index of Big Data development, and adopts the two indexes of green patent scale and green patent ratio to characterize the green technology selection to empirically test the impact of Big Data development on green technology selection. The research finds that Big Data development can significantly optimize green technology selection, and green investment, intellectual capital, and R&D investment are important mechanisms. Big Data's effect has significant heterogeneity in terms of enterprise property rights, manager myopia, and economic policy uncertainty. The research conclusion is still valid in various robustness tests such as replacing the core explanatory variable, changing the sample time range, and using instrumental regression. The research conclusion can provide important guidance for tapping the potential of Big Data in making green production decisions.

Key words: Big Data development; green technology selection; green investment; intellectual capital; R&D input