

大数据赋能了企业绿色创新吗

——基于国家级大数据综合试验区的准自然实验

陈文 常琦¹

【摘要】：在数字经济蓬勃发展的当下，如何让大数据赋能企业绿色创新是“双碳”建设的重要议题。本文根据企业风险管理理论和实物期权理论，从企业的内、外部视角分析了大数据对企业绿色创新的影响路径，并基于国务院批准建设大数据试验区试点的准自然实验，以2013–2019年我国沪深两市A股上市公司为样本，采用多期双重差分法（DID）和倾向得分匹配双重差分法（PSM-DID）检验其影响及作用机制。实证结果显示，大数据可以有效赋能企业提高绿色创新绩效；在进行平行趋势检验、安慰剂检验以及更换被解释变量后，该结论依然成立。机制检验发现，数据要素市场建设具有提高企业内部控制水平和降低外部环境不确定性的“双重功效”，进而形成企业创新能力提高、创新意愿增强的正向反馈机制。此外，基于异质性检验，本文发现大数据试验区建设赋能企业绿色创新的政策效应在重污染行业和高污染地区并不明显；通过强化绿色信贷政策和加大地方政府环境规制，可以有效破解重污染行业和高污染地区对大数据赋能企业绿色创新的掣肘与梗阻。本研究旨在为如何协同推进经济体的数字化与绿色化发展提供启示与借鉴，使大数据切实成为经济绿色发展和提质增效的“新蓝海”。

【关键词】：大数据 绿色创新 绿色信贷 环境规制

【中图分类号】：F27 **【文献标识码】：**A **【文章编号】：**1000-8306(2022)08-0076-17

一、引言

随着社交网络、电子商务、搜索引擎等数字技术与经济生活各领域的深度融合，大数据对企业及行业的数字化赋能彰显了智力资本价值，支撑实体生产部门全面从传统的要素驱动、投资驱动向创新驱动转变。^[1]在此背景下，我国加快培育数据要素市场建设，建立健全要素市场化机制，确保数据要素在相关制度保障下的有序流通和高效利用。早在2015年8月31日，国务院就印发了《促进大数据发展行动纲要》，将大数据作为国家重要战略进行全面部署，并提出开展国家大数据综合试验区建设。大数据综合试验区在数据资源管理与共享开放、数据中心整合、数据资源应用等多方面做了大量重要探索，对我国数据要素市场的整体建设发挥了辐射带动和示范引领效应。^[2]

在国家致力于“双碳”目标和加快培育数据要素市场的当下，企业依托大数据资源，推动研发力量的整合，促进模式创新、业态创新和场景创新，对提升经济效率、全要素生产率至关重要。^{[2][3]}然而，大数据作为数字经济发展下高技术“负荷”的结晶，能否赋能企业绿色创新？这种影响主要通过何种路径发挥效用？是否存在异质性以及如何进一步强化？过往研究尚未解答，这为本文提供了边际贡献的机会。

本文可能的边际贡献与创新在于：一是基于中国宏观大数据政策与微观企业的绿色研发活动的互动关系，探讨如何进一步实现数字化与绿色化的畅通与协同，实现数据要素对可持续发展的全方位赋能；二是本文在统一的框架下探讨了大数据通过何

¹**作者简介：**陈文（1988—），西南财经大学金融学院，讲师。电子邮箱：chenwenpku2008@163.com。常琦（1993—），西南财经大学法学院，博士生（通讯作者）。电子邮箱：changqi@smail.swufe.edu.cn。

基金项目：国家自然科学基金“央行数字货币的可控匿名及其经济影响研究——基于隐私经济学的分析范式”（72003157）的资助

种路径驱动企业绿色创新水平这一关键问题，并支持外部环境不确定性和企业内部控制水平在其中发挥的中介作用机制，深化了已有文献；三是在借鉴既有实证研究的基础上，本文考察了大数据赋能企业绿色创新的异质性影响。同时，通过引入绿色信贷政策和地方政府环境规制两个思考维度，为化解重污染行业和高污染地区的绿色转型梗阻提供参考，进而夯实数字化与绿色化“双化”融合、协同并进的发展道路。

二、文献综述与研究假设

研究多使用绿色创新（或生态创新）来表征企业在提高竞争力的同时，关注并实现环境可持续发展的创新模式。^{[4][5]}我们认为，数字化转型对企业绿色创新的助推作用主要体现在以下三个方面：一是大数据在微观企业层面的进化效应。大数据时代，企业从传统的要素驱动、投资驱动向创新驱动转变，^[1]“信息孤岛”“数据壁垒”逐步被打破，技术迭代升级日益加快以及可持续性发展要求共同推动企业不断提升绿色研发能力以应对外部竞争。^{[6][7]}二是大数据在中观产业层面的渗透作用。伴随信息通信技术的持续扩散与广泛渗透，高新技术产业的高渗透性与强外部性特质得以进一步巩固和强化，在与传统产业的交叉、渗透与融合中实现产业联动、协同创新与绿色发展。^[6,7,8]三是大数据在宏观区域层面的整合效应。在“互联网+”的乘数效应和数字化转型强大的“链接”整合功能之下，可以实现区域间经济结构的重构以及跨区域创新资源要素的重组，形成协调商业、制造、研发生态系统发展的“正反馈循环”式创新模式。^{[2][6]}

与本文研究高度相关的文献仅有 El-Kassar and Singh 的研究，他们通过对 215 名在中东、北非地区和 Golf-Cooperation Countries(GCC) 工作的受访者的问卷调查，搜集了其所在企业绿色创新的驱动因素、大数据在其组织结构中的使用情况等相关数据，进而在企业绿色创新和组织绩效的框架下实证检验了大数据的重要作用。^[4]相较于此研究，本文基于大数据宏观政策进行探讨，可以规避问卷法下样本量较少且缺乏代表性等弊端，更科学、严谨地衡量大数据对企业绿色创新的重要影响。

大数据试验区建设是我国发展大数据产业的重大战略，在区域数字化转型中发挥着辐射带动和示范引领效应。^[2]同时，大数据已经成为激励企业创新的重要助推力量，更是赋能企业绿色研发与低碳发展的强大引擎。在此基础上，本文提出假设 1。

假设 1：大数据能够激励企业绿色创新能力的提升。

鉴于企业绿色创新活动具有长期性、高风险性和不可逆性，更易受到其内部控制水平和外部环境不确定性的影响。^[9,10,11,12]因此，基于企业风险管理理论和实物期权理论，本文从企业的内、外部视角分析了大数据如何推动企业绿色创新。

第一，根据企业风险管理理论，内部控制是企业创新战略实施的重要保障。^[10]大数据可以通过提升内部控制质量赋能企业绿色创新。一是大数据等数字技术的应用强化了管理层对绿色创新活动各环节的整体把控与指挥监督，能够提升企业内部的风险评估水平和应对能力，有助于防范和化解其在绿色研发中的风险隐患；^[9]二是数据资源的整合与集聚降低了企业的信息搜寻成本，^[10]便于企业内部各部门之间的沟通与交流，减少甚至规避由于信息不对称导致的创新投资失误，提高企业创新成功的可能性。^[13]

第二，根据实物期权理论，较高的环境不确定性会增加企业的边际投资成本，企业倾向于延迟绿色研发的投资时机，等待明朗的投资环境。^[12]而大数据通过对市场需求和技术发展的评估与预测，降低了企业外部环境的不确定性，1 增强了其从事绿色研发活动的信心与决心。一方面，在当前以激发数据要素活力为导向的制度背景下，企业对市场需求预测的准确率大幅提高。机器学习和人工智能等大数据技术都是预测算法。企业可以通过这些技术的应用，预测绿色研发产品的市场需求及销售和收益情况，进而科学制定研发战略。^[11]另一方面，大数据分析降低了技术发展的不确定性风险，提高了企业对创新成果被模仿和替代的可能性及绿色研发成果转换率的预判能力。^[12]这对于缓解企业绿色创新活动易面临的融资困境大有裨益。在此基础上，本文提出假设 2。

假设 2：大数据可以通过增强内部控制和降低外部环境不确定性赋能企业绿色创新。

历史研究表明,清洁型行业和污染型行业在污染排放与技术构成等方面存在显著差异,其对于大数据发展的绿色溢出效应响应程度同样具有异质性。^[14]相较于清洁型行业,重污染行业技术升级、改造的难度大、成本高,^[14]数字技术的渗透与扩散较为缓慢。这就需要从污染源头把控,发挥绿色信贷的重要功效。绿色信贷主要通过投融资渠道内生企业环境成本,在信贷可得性以及贷款融资成本上压缩污染企业融资空间并抬高其债务融资成本,从而倒逼其淘汰污染型生产技术。^[15]此外,绿色信贷为企业的研发方向增加了约束条件,缩小技术创新方向的选择域。^[15]这不仅有助于坚定企业开展绿色技术研发的决心,更有助于破解污染行业所面临的“创新活跃但并不绿色”(effective innovation but not green)发展难题。^[5]基于此,我们认为,大数据在重污染行业发挥的绿色创新激励作用相对较弱,而绿色信贷则可以通过影响污染企业技术创新收益—成本比为其清洁技术研发和绿色转型升级提供内生动力。在此基础上,本文提出假设3。

假设3:在重污染行业,大数据对企业绿色创新的激励效应不足;而绿色信贷水平的提升,能够强化大数据对于重污染行业企业的绿色创新激励。

污染密集型产业(PII)往往对劳动力、运输成本、监管政策等特定的地点属性有着相似的需求,产业集聚在形成规模效应的同时,也导致了严重的环境恶化,产生污染地区。^[17]研究表明,PII越来越多地在我国经济欠发达地区形成空间集聚。^[16]随着财政分权的深入,地方保护主义扩大,^[18]一些欠发达地区仍然以牺牲环境为代价片面地追求产业发展,而数字化转型所带来的绿色低碳发展红利便会被轻视和消弭。这些地区被视为“污染天堂”的一个重要原因在于环境规制较弱、存在监管漏洞。^[17]根据“波特假说”,环境规制能倒逼高耗能、高污染的企业革新生产技术,激发创新补偿效应,不仅可以弥补企业的合规成本,而且可以提高企业的生产力和竞争力。^[19]鉴于此,我们认为,大数据同样无法在高污染地区充分发挥其对绿色创新的激励作用。这就需要释放环境规制的效用,增大PII的规制压力和沉没成本,诱发其开展绿色技术创新,从而遏制在污染地区形成的“污染避风港效应”。^[20]在此基础上,本文提出假设4。

假设4:在高污染地区,大数据对企业绿色创新的激励效应不足;而环境规制力度的加大,能够强化大数据对于高污染地区企业的绿色创新激励。

三、研究设计

(一) 模型设计

在样本期内,我国先后有10个省份(城市)分批次获准建设大数据试验区。因此,我们将2015年作为贵州省的政策时间节点,将2016年作为北京、天津、河北、内蒙古、辽宁、河南、上海、重庆和广东的政策时间节点。^[2]本研究的实验组为其所在省份获准建设大数据试验区的上市公司,控制组则是其所在省份未获准建设大数据试验区的上市公司。

鉴于传统的DID模型仅适用于政策发生于同一时间点的情况,而本研究中的大数据试验区试点则属于分批次获准建设,故这种“一刀切式”的传统DID模型并不适用于本文的政策效用评估。^[21]因此,我们借鉴罗知等(2015)的研究,^[21]采用多期DID模型来考察我国大数据试验区的建设对上市公司绿色创新的影响。本文的多期DID基准模型设定如下:

$$\text{EnvrPat}_{it} = \beta_0 + \beta_1 \text{Treatit} \times \text{Timeit} + \text{Zit} + \sum \text{Year} + \sum \text{Industry} + \varepsilon_{it} \quad (1)$$

其中,虚拟变量Treatit表示企业是否属于实验组:若企业i所在的省份在2015年或2016年获准建设大数据试验区,则Treatit取值均为1,否则取值为0。虚拟变量Timeit表示获批建设大数据试验区的时间前后:若企业i所在的省份在t年获准建设,则从t年到2019年的Timeit取值均为1,否则取值为0。Zit为一系列控制变量, $\sum \text{Year}$ 为年份固定效应, $\sum \text{Industry}$ 为行业固定效应, ε_{it} 为随机扰动项。其中, β_1 是此模型中的待估参数,若 $\beta_1 > 0$,表明建设大数据试验区具有激发企业绿色创

新的效用，佐证了假设 1；反之，则假设 1 不成立。

为探讨大数据试验区建设之于企业绿色创新的作用机制，本文引入内部控制水平（InControl）和外部环境不确定（EU）两个指标，并采用余东升等的研究方法对它们是否为两者之间的中介变量分别进行验证。^[22]式（1）至式（3）共同构成了本研究采用的结构方程组，变量 M 代指 EU 或 InControl。

$$M_{it} = \gamma_0 + \gamma_1 Treat_{it} \times Time_{it} + Z_{it} + \sum Year + \sum Industry + \varepsilon_{it} \tag{2}$$

$$EnvrPat_{it} = \varphi_0 + \varphi_1 Treat_{it} \times Time_{it} + \varphi_2 M_{it} + Z_{it} + \sum Year + \sum Industry + \varepsilon_{it} \tag{3}$$

（二）研究样本与数据来源

2013 年被认为是我国的“大数据元年”，²故我们借鉴邱子迅和周亚虹的做法，将 2013 年设定为本研究的起始年份，^[2]并以 2013—2019 年我国 A 股上市公司为样本展开研究。鉴于大数据试验区政策于 2015 和 2016 年分批发布，我们设定：若企业 i 所在的省份于 2015 年获准建设大数据实验区，则 2015—2019 年为实验期，2013—2014 年为非实验期。同样地，若企业 i 所在的省份于 2016 年获准建设，则 2016—2019 年为实验期，2013—2015 年为非实验对比期。

此外，上市公司专利数据来自国家知识产权局；其经济特征数据来自 CSMAR；企业内部控制水平则是在参考申慧慧等研究的基础上，^[23]使用深圳市迪博企业风险管理技术有限公司发布的“迪博·中国上市公司内部控制指数”来予以表征，该公司构建的 DIB 内部控制库是我国首个专业的内部控制信息数据库；环境规制强度和绿色信贷水平测算中所使用的数据来自我国各级政府工作报告、《中国工业统计年鉴》以及国家统计局公布的经济普查数据。同时，本文借鉴了相关研究成果，^{[5][19][24]}选取了如下可能影响企业绿色创新的指标作为控制变量：企业规模（Size）、资产负债率（Lev）、资产收益率（ROA）、企业托宾 Q 值（TQ）、无形资产比率（Inasset）、股权集中指标（3%）（Concent）和股票回报率（Return）。此外，在合并上市公司样本数据时，我们剔除了金融行业公司、ST 或 PT 公司，以及变量缺失的样本。并且，为了消除极端值对本文结果的影响，我们对连续型变量在 1%和 99%的水平上进行 Winsorize 处理。本文被解释变量、核心解释变量及控制变量的指标定义详见表 1。

表 1 变量定义

变量符号	变量名称	变量测度
EnvrPat	绿色创新水平	上市公司绿色专利申请量
InControl	企业内部控制水平	DIB 迪博内部控制指数
EU	外部环境不确定性	未经行业调整的环境不确定性/行业环境不确定性
EnRegulat	环境规制强度	政府工作报告中的环境词频占比
Gcredit	绿色信贷水平	六大高耗能工业产业利息支出/工业利息总支出
Size	企业规模	总资产的对数值

Lev	资产负债率	总负债/总资产
ROA	资产收益率	营业利润/总资产
TQ	企业托宾 Q 值	企业托宾 Q 值
Inasset	无形资产比率	无形资产/总资产
Concent	股权集中指标 (3%)	公司前三位大股东持股比例之和
Return	股票回报率	当年 5 月全次年 4 月买入并持有收益率

本文被解释变量为上市公司绿色创新水平，以绿色专利申请量（EnvrPat）来测度。鉴于专利申请流程耗时较长，采用专利申请数据可以更具时效性地考察试点政策对企业绿色技术创新活动产生的影响。对于上市公司绿色专利的筛选，本文依据世界知识产权组织（WIPO）于 2010 年推出的“国际专利分类绿色清单”，结合国际专利分类号甄别并提取出样本上市公司的绿色专利数据。该绿色清单根据《联合国气候变化框架公约》中对绿色专利的划分标准生成。

内部控制水平（InControl）和外部环境不确定性（EU）是本文的中介变量。其中，我们用深圳市迪博企业风险管理技术有限公司公布的内部控制指数来衡量企业的内部控制水平（InControl）。而外部环境不确定性（EU）则在参考申慧慧等（2012）的研究成果基础上，用公司经营业绩波动来予以表征。^[23]具体测算过程如下：首先，使用每个样本公司过去五年的销售收入数据，建立回归模型（4）。从过去第四年开始至当年分别将年份赋值为 1~5，利用模型（4）分别估计 5 年的非正常销售收入。其中，SALE 代表销售收入，YEAR 指代年份，回归得到的残差结果即为非正常销售收入。其次，计算过去 5 年非正常销售收入的标准差，除以过去五年销售收入的平均值，得到未经行业调整的环境不确定性。最后，各公司未经行业调整的环境不确定性除以行业环境不确定性（同年度同行业内所有公司未经行业调整的环境不确定性的中位数），即为公司经业调整后的环境不确定性，也就是本文所运用的外部环境不确定性（EU）。

$$SALE = \alpha_0 + \alpha_1 YEAR + \varepsilon$$

(4)

四、实证结果及分析

（一）描述性统计

表 2 报告了本文主要变量的描述性统计的结果。从表 2 中可以看出，企业绿色创新申请量（EnvrPat）的均值为 1.15，而最大值为 31，说明我国上市公司的绿色创新水平差异较大，这一定程度反映了我国的实际情况。通过对上市公司数据的匹配和处理，在本研究的时间范围内，我们共获得 19122 个有效观测值。但受部分上市公司成立时间较晚的影响，企业内部控制和外部环境不确定这两个测算指标在该时间范围存在部分缺失值，故观测值数量减少。此外，我们未能获取我国全部地级市的 PM_{2.5} 浓度数据，受制于其可得性的问题，本研究共获得了 286 个地级市的 PM_{2.5} 浓度数据，故对应的观测值数量也有所减少。

表 2 描述性统计

变量	观测值	平均值	标准差	最小值	最大值
EnvrPat	19122	1.15	4.20	0.00	31.00
Lev	19122	0.42	0.20	0.05	0.94
Size	19122	22.19	1.31	18.94	26.59
Return	19122	0.10	0.50	-0.54	2.16
TQ	19122	2.99	2.32	0.89	14.53
ROA	19122	0.04	0.06	-0.34	0.19
Inasset	19122	0.05	0.05	0.00	0.33
Concent	19122	53.99	15.09	19.77	88.19
Pollution	19122	0.30	0.46	0.00	1.00
PM2.5	18881	0.52	0.50	0.00	1.00
InControl	18003	633.07	135.91	0.00	941.31
EU	14024	1.36	1.28	0.13	7.73
EnRegulat	19122	0.01	0.00	0.00	0.01
Gcredit	19122	0.40	0.12	0.19	0.80
EnvrInvPat	19122	0.61	2.21	0.00	15.00
EnvrNewPat	19122	0.43	1.67	0.00	12.00
EnauPat	19122	1.23	4.12	0.00	30.00
EnauInvPat	19122	0.36	1.38	0.00	10.00
EnauNewPat	19122	0.78	2.70	0.00	19.00

（二）基础回归和 PSM 结果分析

本文借鉴余东升等（2021）的研究方法，采用一对一又放回的近邻 PSM 匹配了实验组和控制组样本，^[22] 具体实证结果如表 3 所示。

表 3 基准回归和 PSM 结果

变量	匹配前		匹配后	
	EnvrPat	EnvrPat	EnvrPat	EnvrPat
	(1)	(2)	(3)	(4)
Treat×Time	0.3650*** (0.0747)	0.2030*** (0.0725)	0.4069*** (0.1038)	0.3472*** (0.1014)
Lev		0.1317 (0.1887)		0.0176 (0.2985)
Size		0.8967*** (0.0312)		0.9436*** (0.0736)
Return		-0.0721 (0.0737)		0.0150 (0.1443)
TQ		0.0368** (0.0169)		0.0200 (0.0219)
ROA		2.5741*** (0.5168)		1.8224*** (0.6758)
Inasset		-1.3856** (0.6114)		-1.7105** (0.8567)
Concent		-0.0075*** (0.0020)		-0.0016 (0.0033)
常数项	0.3489 (0.2731)	-18.7168*** (0.7010)	0.4186 (0.2793)	-20.2780*** (1.5729)

行业固定效应	Yes	Yes	Yes	Yes
时间固定效应	Yes	Yes	Yes	Yes
R ²	0.0730	0.1337	0.0885	0.1504
N	19122	19122	9300	9300

注：*P<0.10、**P<0.05、***P<0.01；括号内为标准误。

其中，第（1）（2）列表示匹配前实验组和控制组样本的 DID 估计结果，第（3）（4）列表示匹配后实验组和控制组样本的 PSM-DID 估计结果。回归结果显示，匹配前后 $Treat \times Time$ 的系数均显著为正。在加入一系列控制变量后，这一结果依然保持稳健。由此表明，我国大数据试验区试点的建设有助于激发上市公司绿色创新的效能，该实证结果支持了假设 1。

为进一步检验 PSM-DID 的匹配效果，我们还绘制了倾向得分值密度函数图。图 1 显示，相较于匹配前的样本，匹配后实验组和控制组倾向得分值的概率密度分布更为接近，说明本文的匹配效果较好。此外，图 2 表明大多数变量的标准化偏差在匹配后缩小；而图 3 的条形图则显示大多数观测值均在共同取值范围，故在进行匹配时仅会损失少量样本。综上，在共同支撑假设成立的基础上，我们进一步证明了采用 PSM-DID 方法分析本文研究样本的合理性与可靠性。因此，本文后续研究也在 PSM-DID 匹配结果的基础上进一步展开。

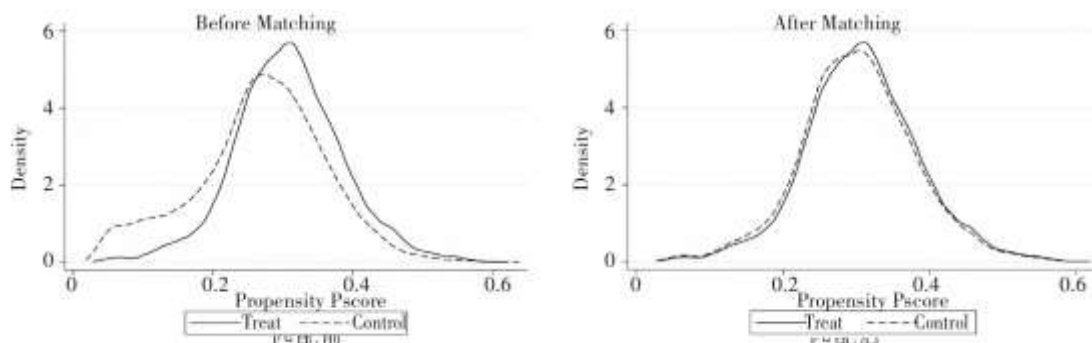


图 1 倾向得分值概率分布密度函数图

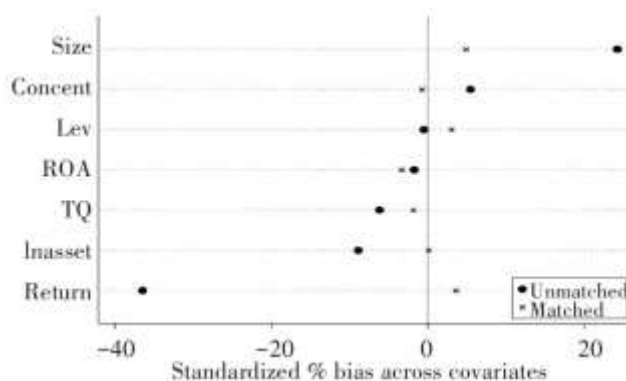


图 2 各变量的标准化偏差图

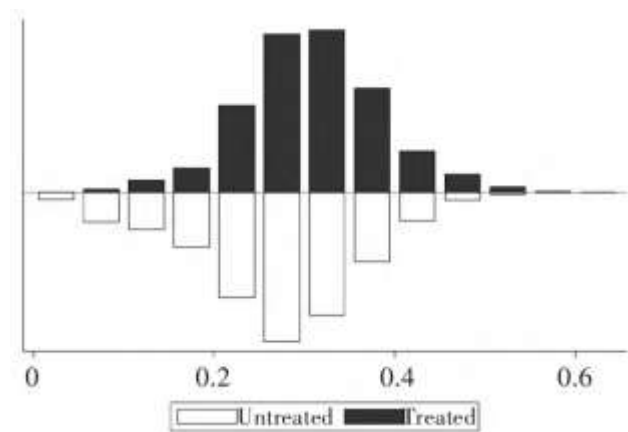


图 3 倾向得分的共同取值范围

（三）基于 PSM-DID 估计的有效性检验

1. 平行趋势检验

双重差分法可以部分缓解政策评估中的内生性问题，但其潜在假设是在政策实施前，实验组与控制组要满足平行趋势。为验证这一假设，借鉴曹越等的做法，本文设置了政策实施前的年份哑变量，并将其与 Treat 的交互项纳入模型，^[19]回归结果详见表 4 中的第（1）列。其中，Treat×Before1、Treat×Before2、Treat×Before3 取值为 1 的条件分别为“政策实施前 1 年”“政策实施前 2 年”“政策实施前 3 年”。实证结果显示，政策实施前不同年份与 Treat 的交互项估计系数均不显著。由此表明，在大数据试验区试点建设获批前，实验组和控制组的变化趋势并没有显著差异，即平行趋势假说成立，符合 DID 模型使用的前提条件。

表 4 平行趋势检验与安慰剂检验

变量	平行趋势检验	安慰剂：虚构政策时间		安慰剂：虚构实验组	
	EnvrPat	EnvrPat	EnvrPat	EnvrPat	EnvrPat
	(1)	(2)	(3)	(4)	(5)
Treat×Time		0.1693 (0.1250)	0.1285 (0.1245)	-0.2237 (0.3037)	-0.0889 (0.0996)
Treat×Before1	0.3524 (0.3867)				

Treat×Before2	0.0989 (0.5121)				
Treat×Before3	0.7150 (0.5187)				
Lev	0.0363 (0.2963)		0.0240 (0.2992)	0.0246 (0.2991)	0.0241 (0.2992)
Size	0.9365*** (0.0733)		0.9464*** (0.0738)	0.9465*** (0.0738)	0.9491*** (0.0737)
Return	0.0321 (0.1457)		0.0408 (0.1451)	0.0502 (0.1436)	0.0490 (0.1436)
TQ	0.0165 (0.0219)		0.0224 (0.0220)	0.0226 (0.0220)	0.0229 (0.0220)
ROA	1.8369*** (0.6773)		1.8194*** (0.6757)	1.8133*** (0.6748)	1.7877*** (0.6774)
Inasset	-1.6437* (0.8554)		-1.6955** (0.8565)	-1.6756* (0.8574)	-1.7055** (0.8556)
Concent	-0.0018 (0.0033)		-0.0016 (0.0033)	-0.0016 (0.0033)	-0.0017 (0.0033)
常数项	-20.3951*** (1.5828)	0.3674 (0.2794)	-20.3962*** (1.5760)	-20.4049*** (1.5758)	-20.4612*** (1.5754)
行业固定效应	Yes	Yes	Yes	Yes	Yes
时间固定效应	Yes	Yes	Yes	Yes	Yes
R ²	0.1516	0.0872	0.1494	0.1494	0.1494
N	9300	9300	9300	9300	9300

注：* $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$ ；括号内为标准误。

2. 安慰剂检验：虚构政策时间

在参考黄凯等研究的基础上，为剔除其他政策以及不可观测的缺失变量对本研究的干扰，从而进一步支持本文基准回归的因果效应，我们采用虚构事件时间的方式进行了第一个安慰剂检验，即假想大数据试点政策滞后至 2018 年发生。^[25]其原因如下：一是选择滞后事件可以更好地避免因部分企业上市时间较短，而产生绿色专利数据缺失样本较多的问题；二是鉴于试点是在 2015 年和 2016 年分别获批，故为了避免窗口期重叠，我们选择滞后政策时间的方式以保证上文结论的可靠性。^[25]表 4 中的第 (2) (3) 列显示，无论是否包含控制变量，双重差分模型估计出来的交互项 $Treat \times Time$ 系数均不显著，符合本文的预期。

3. 安慰剂检验：虚构实验组

本文借鉴余东升等的方法，采用虚构实验组的方式进行了第二个安慰剂检验。^[22]鉴于我国大数据试点是于 2015 年和 2016 年分别获批建设，因此我们分别以此两年实验组中上市公司的样本数量（2015 年实验组的样本量为 18, 2016 年实验组的样本量为 1155）为基准，在全样本中进行随机抽样，并将其设定为“伪”实验组，进而将该虚构实验组与控制组进行双重差分模型回归，回归结果见表 4 中的第 (4) (5) 列，其分别为 2015 年和 2016 年的回归结果。实证结果显示，交互项 $Treat \times Time$ 系数均不显著。此外，为保证研究结果的可靠性与准确性，我们按照以上方式进行了 500 次虚拟实验组的安慰剂试验。图 4 报告了 500 次“伪”实验组估计系数核密度及相应 P 值分布。其中，X 轴表示“伪政策虚拟变量”估计系数的大小，Y 轴表示密度值和 P 值大小，曲线是估计系数的核密度分布，圆点是估计系数对应的 P 值，垂直虚线是 PSM-DID 模型[见上文表 3 中的第 (4) 列]，基准回归交互项的估计系数 0.35，水平虚线为其显著性水平 0.1。从图中可以看出，一方面，基于随机抽样的交互项估计系数均值分布在 0 附近，且与基准回归交互项的估计系数存在显著差异；另一方面，500 次随机处理过程中，大多数估计值的 P 值都大于 0.1（在 10% 的水平上不显著），这表明我们的估计结果不太可能是偶然得到的。因此，我国大数据试验区的绿色创新效应并未明显受到潜在非观测随机因素的影响。

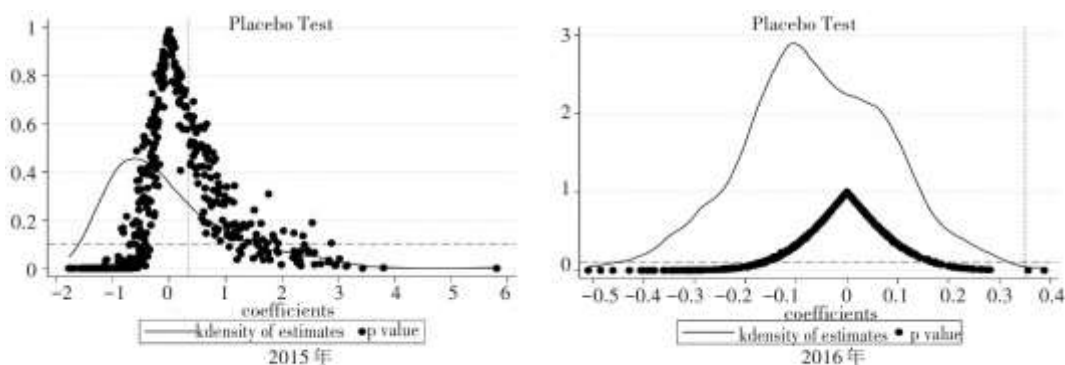


图 4 “虚构实验组”估计系数核密度及相应 P 值分布图

4. 更换被解释变量

鉴于绿色专利授权量也是表征上市公司绿色创新水平的重要测度指标。^[26]因此，本文在参考既有研究的基础上，将绿色专利授权量（EnauPat）作为绿色专利申请数量（EnvrPat）的替代指标，来进一步观测我国大数据试验区的绿色创新效应。此外，上市公司的绿色专利可以分为绿色发明型专利和绿色实用新型专利。借鉴许可和张亚峰（2021）的研究，为了进一步考察专利类型的异质性影响，我们将绿色发明型专利的申请量（EnvrInvPat）与授权量（EnauInvPat），^[26]以及绿色实用新型专利的申请量

(EnvrNewPat) 与授权量 (EnauNewPat) 也直接作为因变量, 以确保研究结果的准确性与可靠性, 实证结果见表 5。实证结果显示, $Treat \times Time$ 的系数均在 1% 的水平下显著为正, 再次表明了本文基准回归结果的稳健性。

表 5 更换被解释变量

变量	EnvrInvPat	EnvrNewPat	EnauPat	EnauInvPat	EnauNewPat
	(1)	(2)	(3)	(4)	(5)
Treat $\times$ Time	0.1546*** (0.0524)	0.1432*** (0.0404)	0.3728*** (0.1109)	0.0949** (0.0386)	0.1927*** (0.0721)
Lev	-0.0271 (0.1556)	0.1011 (0.1155)	-0.0646 (0.2941)	-0.2418** (0.0979)	0.2912 (0.1900)
Size	0.5090*** (0.0369)	0.2907*** (0.0277)	1.0208*** (0.0723)	0.3683*** (0.0243)	0.5533*** (0.0463)
Return	-0.0088 (0.0792)	0.0343 (0.0564)	0.0777 (0.1400)	0.0333 (0.0523)	0.0360 (0.0944)
TQ	0.0184	-0.0004	0.0673***	0.0294***	0.0303**
	(0.0124)	(0.0084)	(0.0221)	(0.0088)	(0.0136)
ROA	1.1203*** (0.3275)	0.4132 (0.3059)	0.2812 (0.7433)	-0.1704 (0.2199)	0.3778 (0.5062)
Inasset	-0.7532* (0.4451)	-0.8558** (0.3634)	-2.2828** (0.9175)	-0.4047 (0.2834)	-1.7518*** (0.6308)
Concent	-0.0018 (0.0018)	0.0009 (0.0013)	0.0033 (0.0033)	-0.0010 (0.0012)	0.0036* (0.0021)
常数项	-10.8747*** (0.7891)	-6.3455*** (0.5947)	-22.2532*** (1.5715)	-7.8097*** (0.5374)	-12.2932*** (0.9958)
行业固定效应	Yes	Yes	Yes	Yes	Yes

时间固定效应	Yes	Yes	Yes	Yes	Yes
R ²	0.1455	0.1402	0.1755	0.1436	0.1612
N	9300	9300	9300	9300	9300

注：*P<0.10、**P<0.05、***P<0.01；括号内为标准误。

（四）机制检验

前文分别从内部控制水平（InControl）和外部环境不确定性（EU）的视角，理论分析了建设大数据试验区对上市公司绿色创新影响的影响机制。为验证假设 2，本文选用中介效应模型进行实证检验³

第（1）（2）列汇报了企业外部环境不确定性（EU）对大数据试验区绿色创新效应的影响机制结果。实证结果显示，第（3）列中的 Treat×Time 的系数和第（4）列中的 EU 的系数均显著为负，这表明大数据试验区试点的建设有助于通过降低企业经营外部环境的不确定性推动绿色创新效应的提升。根据中介机制效应检验程序，本文进一步对样本进行了 Sobel 检验，发现 Sobel 检验的 P 值为 0.076，即中介效应存在。第（3）（4）列的实证结果同样证实企业内部控制水平（InControl）在大数据实验区的绿色创新驱动效应中的同样发挥着中介作用。

五、进一步分析与检验

为检验行业和区域的异质性影响，本文引入分组变量：污染行业和污染地区的测度指标。重污染行业的标准是在参考 Tanaka 的研究基础上，^[27]根据《上市公司行业分类指引》《上市公司环保核查行业分类管理名录》以及《上市公司环境信息披露指南》的相关规定进行认定。重污染行业主要包括煤炭、采矿、纺织、制革、造纸、石化、制药、化工、冶金、火电等 16 类行业。高污染地区则是以 2013—2019 年度 286 个地级市 PM_{2.5} 浓度数据的均值为阈值，若上市公司注册地在该时间段内的 PM_{2.5} 浓度均值高于此阈值水平，则视为高污染地区。借鉴 ZHAO 等的研究，我们采用的是达尔豪斯大学大气成分分析小组 4 发布的 PM_{2.5} 浓度数据，并以该数据年度平均估计值（在 1 公里空间分辨率下）来刻画城市污染程度。^[28]

第（1）列至第（4）列汇报了相关实证结果。其中，第（1）（2）列以污染行业为界分标准，第（3）（4）列以污染地区为界分标准。第（1）（3）列报告了重污染行业和高污染地区的回归结果，实证结果显示 Treat×Time 的回归系数不显著，说明大数据试验区试点的建设对企业的绿色创新激励效应在重污染行业和高污染地区受到了抑制。第（2）（4）列则报告了轻污染行业和低污染地区的回归结果，实证结果显示 Treat×Time 的回归系数均显著为正，说明大数据试验区试点的建设对企业的绿色创新激励效应在轻污染行业和低污染地区成效显著。

那么，如何提高大数据试验区试点对于重污染行业和高污染地区企业绿色创新的促进作用？为此，本文以绿色信贷水平（Gcredit）和环境规制强度（EnRegulat）指标为调节变量。在参考 Hu 等（2020）研究的基础上，本研究以六大高耗能工业产业利息支出在工业利息总支出中的占比来衡量绿色信贷（Gcredit），该指标为绿色信贷发展水平的反向指标。^[29]此外，作为经济发展和环境治理工作的指导性政策文件，地方政府工作报告可以较为真实、全面地彰显地方政府在环境保护和整治中的水平和力度。^[30]因此，我们借鉴上官绪明和葛斌华（2010）的做法，^[30]通过手工整理、计算 2013—2019 年我国各省级政府工作报告中的环境词频占比来表征环境规制强度（EnRegulat）。由此，构建以下两个回归方程：

$$\text{EnvrPat}_{it} = \chi_0 + \chi_1 \text{Treat}_{it} \times \text{Time}_{it} \times \text{Gcredit}_{it} + \chi_2 \text{Treat}_{it} \times \text{Time}_{it} + \chi_3 \text{Gcredit}_{it} + Z_{it} + \sum \text{Year} + \sum \text{Industry} + \varepsilon_{it} \quad (5)$$

$$\text{EnvrPat}_{it} = \delta_0 + \delta_1 \text{Treat}_{it} \times \text{Time}_{it} \times \text{EnRegulat}_{it} + \delta_2 \text{Treat}_{it} \times \text{Time}_{it} + \delta_3 \text{EnRegulat}_{it} + Z_{it} + \sum \text{Year} + \sum \text{Industry} + \varepsilon_{it} \quad (6)$$

第（5）（6）列汇报了相关实证结果。其中，第（5）列检验了绿色信贷（Gcredit）的调节作用。实证结果显示 $\text{Treat} \times \text{Time} \times \text{Gcredit}$ 的回归系数显著为负，说明绿色信贷政策的推进，可以强化大数据试验区试点对于重污染行业企业绿色创新的激励效应。第（6）列则报告了对环境规制（EnRegulat）调节作用的检验结果。实证结果显示， $\text{Treat} \times \text{Time} \times \text{EnRegulat}$ 的回归系数显著为正，说明环境规制水平的提升，可以强化大数据试验区试点对于高污染地区企业绿色创新的激励效应。此外，我们还基于环境规制和绿色信贷两个变量的调节效应进行了边际分析，详见图 5。借鉴 Zhao 等（2020）的做法，将每个调节变量的低、高水平分别设置为低于和高于均值一个标准差。^[28]如图 5 所示，绿色信贷规模的扩大可以强化大数据实验区对企业绿色创新的激励效应，这一强化作用在重污染行业尤为突出。同时，无论是高污染地区还是低污染地区，环境规制强度的提高均会对大数据实验区的绿色溢出效应产生显著的正向激励作用。

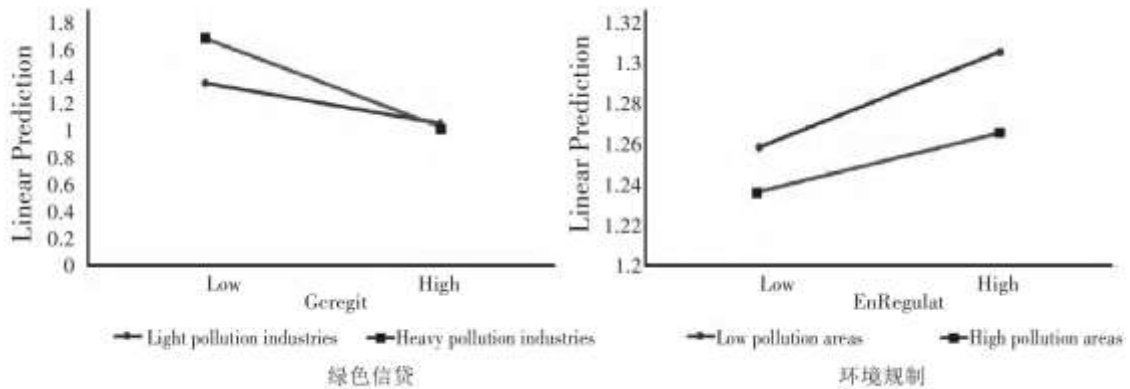


图 5 调节效应的边际分析

综上，图 5 的检验结果支持假设 3 和假设 4 的成立。因此，若进一步释放大数据试验区对企业绿色创新的激励作用，尤其是对重污染行业和高污染地区企业的绿色创新激励，需要充分发挥绿色信贷政策和地方环境规制的“双轮驱动”作用，形成对大数据试验区绿色溢出效应的助推合力。

六、研究结论与政策建议

文章以沪深两市 A 股上市公司为研究样本，基于大数据试验区试点的准自然实验，实证检验了大数据对企业绿色创新的影响。研究发现：一是大数据可以有效驱动企业提高绿色创新绩效，引入平行趋势检验等稳健性检验后，依然支持该结论；二是作用机制显示，大数据试验区试点的建设能够通过提高企业内部控制水平和降低外部环境不确定性，充分释放其对企业绿色创新的赋能和乘数效应；三是加快推进绿色信贷政策和加大地方政府的环境规制力度，不仅可以为重污染行业和高污染地区在数字化与绿色化协同转型中“纾困解难”，而且可以协同数据要素市场建设，共同形成推动经济转型与高质量发展的叠加效应。

基于上述结论，本文的政策建议如下：

第一,我国在引领大数据产业发展的进路中,应当坚持以科技为核心的全面创新与以可持续发展为导向的基本原则相融合,全方位深化大数据通过提高企业内部控制水平和降低环境不确定性驱动企业创新绿色高质量发展的红利优势。同时,监管部门既要在宏观层面重视环境绩效的整体表现,也应在微观层面精准识别环境治理的痛点和难点,关注重污染行业和高污染地区企业的绿色转型状况,多维度破解数字化与绿色化“双化”融合中的掣肘难题。

第二,充分发挥绿色信贷政策在大数据赋能企业绿色创新中的融资激励效用。在绿色信贷框架下,通过投融资渠道动态调整环境污染的机会成本,引导高能耗、高污染、资源性的粗放、低端产业通过数字化转型和研发创新逐步实现核心竞争力增强与生产技术绿色化的“双重目标”,从而准确把握、科学应对数字经济和绿色技术创新带来的机遇与挑战,实现经济与环境的双赢。

第三,充分释放环境规制对大数据赋能企业绿色创新的贡献能力。数字经济时代,合规有效的数据要素市场、高质量的政府治理能力是中国经济实现良性循环的关键。在国家统一的环保制度框架下,地方政府应当加快完善与绿色化转型适配的环境标准与制度体系,筑牢数字化与绿色化协同发展的“安全底座”。此外,中央层面也应强化顶层设计,建立健全对地方政府环境规制实施情况的监督与考核机制,严防“污染天堂”的形成。

参考文献:

- [1]Porter M E.Strategy and the Internet[J/OL].Harvard Business Review,2001.[https://hbr.org/2001/03/strategy-and-the-inter net](https://hbr.org/2001/03/strategy-and-the-inter-net).
- [2]邱子迅,周亚虹.数字经济发展与地区全要素生产率——基于国家级大数据综合试验区的分析[J].财经研究,2021,47(7):4-17.
- [3]李天宇,王晓娟.数字经济赋能中国“双循环”战略:内在逻辑与实现路径[J].经济学家,2021(5):102-109.
- [4]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.
- [5]Hu G,Wang X,Wang Y.Can the Green Credit Policy Stimulate Green Innovation in Heavily Polluting Enterprises? Evidence from a Quasi-natural Experiment in China[J].Energy Economics,2021,98:105134.
- [6]Li J,Chen L,Chen Y,et al.Digital Economy,Technological Innovation,and Green Economic Efficiency-Empirical Evidence f from 277 Cities in China[J/L].Managerial and Decision Economics,2021[2021-07-31].<https://onlinelibrary.wiley.com/doi/abs/10.1002/mde.3406>.
- [7]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.
- [8]Zhang H.Industrial Cluster Innovation Based on 5 G Network and Internet of Things[J/OL].Microprocessors and Microsystems,2021,83:103974.
- [9]Gordon L A,Wilford A L.An Analysis of Multiple Consecutive Years of Material Weaknesses in Internal Control[J].The Accounting Review,2012,87(6):2027-2060.

-
- [10]田丹,吕文栋,刘凯丽.内部控制对创新风险的作用机制——基于风险缓和模型的研究[J].财贸经济,2022,43(5):129-144.
- [11]Brynjolfsson E,Collis A.How Should We Measure the Digital Economy?[J].Harvard Business Review,2019.
- [12]朱德胜.不确定环境下股权激励对企业创新活动的影响[J].经济管理,2019,41(2):55-72.
- [13]Jensen M C,Meckling W H.Theory of the firm:Managerial Behavior,Agency Costs and Ownership Structure[J].Journal of Financial Economics,1976,3(4):305-360.
- [14]Du K,Cheng Y,Yao X.Environmental Regulation,Green Technology Innovation,and Industrial Structure Upgrading:The Road to the Green Transformation of Chinese Cities[J].Energy Economics,2021,98:105247.
- [15]Liu X,Wang E,Cai D.Green Credit Policy,Property Rights and Debt Financing:Quasi-natural Experimental Evidence from China[J].Finance Research Letters,2019,29:129-135.
- [16]Wu J,Wei Y D,Chen W,et al.Environmental Regulations and Redistribution of Polluting Industries in Transitional China:Understanding Regional and Industrial Differences[J].Journal of Cleaner Production,2019,206:142-155.
- [17]Zhao X,Zhao Y,Zeng S.Corporate Behavior and Competitiveness:Impact of Environmental Regulation on Chinese Firms[J].Journal of Cleaner Production,2015,86:311-322.
- [18]Rooij B V,Lo C W H.Fragile Convergence:Understanding Variation in the Enforcement of China's Industrial Pollution Law[J].Law&Policy,2010,32(1):14-37.
- [19]曹越,辛红霞,张卓然.新《环境保护法》实施对重污染行业投资效率的影响[J].中国软科学,2020(08):164-173.
- [20]He J.Pollution Haven Hypothesis and Environmental Impacts of Foreign Direct Investment:The Case of Industrial Emission of Sulfur Dioxide (SO₂) in Chinese Provinces[J].Ecological Economics,2006,60(1):228-245.
- [21]罗知,赵奇伟,严兵.约束机制和激励机制对国有企业长期投资的影响[J].中国工业经济,2015(10):69-84.
- [22]余东升,李小平,李慧.“一带一路”倡议能否降低城市环境污染?——来自准自然实验的证据[J].统计研究,2021,38(6):44-56.
- [23]申慧慧,于鹏,吴联生.国有股权、环境不确定性与投资效率[J].经济研究,2012,47(7):113-126.
- [24]李瑛玫,史琦.内部控制能够促进企业创新绩效的提高吗?[J].科研管理,2019,40(6):86-99.
- [25]黄凯,刘亚辉,尹玉刚.成分股调整改变企业权益资本成本了吗——兼论产权性质的影响[J].会计研究,2021(3):122-136.

[26]许可,张亚峰.绿色科技创新能带来绿水青山吗?——基于绿色专利视角的研究[J].中国人口·资源与环境,2021,31(5):141-151.

[27]Tanaka S.Environmental Regulations on Air Pollution in China and Their Impact on Infant Mortality[J].Journal of Health Economics,2015,42:90-103.

[28]Zhao N,Smargiassi A,Hudson M,et al. Investigating Associations Between Anti-nuclear Antibody Positivity and Combined Long-term Exposures to NO₂,O₃,and PM_{2.5} Using a Bayesian Kernel Machine Regression Approach[J].Environment International,2020,136:105472.

[29]Hu Y,Jiang H,Zhong Z. Impact of Green Credit on Industrial Structure in China:Theoretical Mechanism and Empirical Analysis[J].Environmental Science and Pollution Research,2020,27(10):10506-10519.

[30]上官绪明,葛斌华.科技创新、环境规制与经济高质量发展——来自中国 278 个地级及以上城市的经验证据[J].中国人口·资源与环境,2020,30(6):95-104.

注释:

1 市场不确定性和技术不确定性是企业创新行为的主要环境维度^[12]。

2 参见:人民网-人民日报,2013年12月25日,“2013大数据元年”(网址:<http://cpc.people.com.cn/n/2013/1225/c83083-23940252.html>)。

3 受部分上市公司成立时间较晚的影响,企业内部控制和外部环境不确定性两个测算指标在本研究的时间范围内存在部分缺失值。因此,表5的样本量与前文不同。

4 受制于数据可得性的问题,本研究共获得了我国286个地级市的PM_{2.5}浓度数据。因此,表5中的第(3)(4)(6)列的回归中存有缺失值。参加:“Dalhousie University Atmospheric Composition Analysis Group”(网址:http://fizz.phys.dal.ca/atmos/martin/?page_id=140)。