

张自然 何竞：数字经济发展对城市碳排放的影响 ——基于国家大数据综合试验区的准实验

张自然 中国社会科学院经济研究所经济增长研究室主任、研究员，
中国社会科学院大学教授

何竞 中国社会科学院大学经济学院博士研究生

来源：《经济问题探索》2024 年第 6 期 发表时间：2024 年 06
月 28 日

摘要：数字经济作为具有强劲增长动力的新型经济形态，厘清其对碳排放的影响机制是实现中国绿色转型和经济高质量发展的关键。本文以国家大数据综合试验区的设立表征数字经济的发展，基于 2011—2019 年 286 个地级市数据，构建交错双重差分模型和考虑空间溢出效应的修正两阶段双重差分模型，探究数字经济发展对城市碳排放的直接效应和其对于其他城市碳排放的溢出效应。研究结论表明，数字经济发展会降低城市的碳排放总量和电能消耗产生的碳排放，对其他方式产生的碳排放无显著影响，且工业用电碳排放的减少是碳排放水平降低的直接原因；数字经济发展会对周围城市产生碳减排的溢出效应，该效应会随着地理距离的增加而递减；异质性分析表明，数字经济发展在不同试验区类型、不同创新环境 and 经济水平下的降碳效果存在差异；机制分析表明，数字经济发展通过促进消费业态和模式创新、推动产业数字化转型降低城市碳排

放水平。关键词：碳排放;数字经济;国家级大数据综合试验区;交错
双重差分模型

一、引言

为积极应对全球气候变化，中国提出将以新发展理念为引领，在推动高质量发展中促进经济社会发展全面绿色转型，落实 2030 年前二氧化碳排放达到峰值，努力争取 2060 年前实现碳中和目标，提高国家自主贡献力度，充分体现了主动承担责任的大国担当。然而在我国人均 GDP 较低、尚未基本实现现代化的情况下，同时实现“双碳”目标和经济高质量发展意味着极大的挑战。与此同时，伴随着新一轮科技革命和产业变革，数字经济与经济社会各领域融合的广度和深度不断拓展，并逐渐成为中国国民经济的重要组成部分和增长动力(赵涛等，2020)。大数据是数字经济的关键生产要素，依托其产生了大量的新兴数字技术和相关产业，并与传统产业进行深度融合。因此，大数据和数字经济对于实现双碳目标的作用备受关注。一方面，大数据的应用能够通过提高能源监测和规划的效率、引导产业结构的升级、促进全要素生产率的提升等途径减少碳排放。另一方面，数据存储平台、算力中心和 5G 基站本身会产生大量的碳排放，计算机、通信和电子设备制造业等数字产业对电力的较高需求也会导致碳排放的增加。因此，厘清现阶段大数据政策对碳排放的影响方向和作用机制，不仅能为大数据进一步的发展提供思路和决策依据，还对中国实现绿色转型和经济高质量发展具有十分重要的研究意义。

为了充分发挥大数据的基础战略性资源作用，统筹推进大数据产业发展，国务院于 2015 年 8 月印发的《促进大数据发展行动纲要》中，明确提出“开展区域试点，推进大数据综合试验区建设”，旨在促进区域内及区域间数据要素的整合和大数据基础设施的共享。通过先行先试的方式，总结并把握大数据发展的规律和特征，为其他地区提供值得借鉴的实践经验，最终实现协同发展和大数据红利共享。同年 9 月，贵州大数据综合试验区建设正式启动。2016 年 8 月，包括京津冀、珠江三角洲、上海、河南、重庆、沈阳和内蒙古在内的第二批大数据综合试验区建设方案正式获批。试验区政策的核心在于形成大规模、高集聚度的大数据产业，全面促进大数据在各行业各领域的深度应用，推动相关制度及技术创新，培育大数据产业人才。可以看出，该政策的落脚点在于从产业、民生、制度、技术和人力资本等多个维度全面推动数字经济发展。因此，本文以大数据综合试验区的建立表征数字经济的发展。

在“双碳”目标的背景下，学术界对可能影响碳排放的因素进行了大量的理论与实证研究，包括经济增长、产业结构(吴振信等，2012)、城市化(林伯强和刘希颖，2010)、能源强度、能源结构(鲁万波等，2013)和科技进步(张颂心等，2021)等。而数字经济作为一种新的经济、新的动能、新的业态，引发了社会和经济各领域的深刻变革，其作用于碳排放的路径十分复杂多样。因此，在关于数字经济发展对碳排放影响的研究中，无论是研究视角还是研究结论都不尽相同。根据主要研究结论，具体可以划分为促进论、抑制论

和非线性影响三种。支持促进论的学者认为数字经济正在成为中国主要的碳排放来源(渠慎宁等, 2022), 且发展初期会挤占生产部门的资金投入, 进而降低碳排放效率(王山和余东华, 2023);支持抑制论的学者认为数字经济会通过促进产业升级(谢文倩等, 2022)、提高技术创新水平(郭风等, 2022)、优化资源配置(向宇等, 2023)、降低能耗强度(王芳和董战峰, 2023)、改善能源结构(谢云飞, 2022)等机制减少碳排放;支持非线性影响的学者在考虑空间效应的框架下, 认为数字经济发展具有“先促增、后抑制”的倒 U 型本地碳减排效应, 以及“先抑制、后促增”的 U 型空间溢出减排效应(李治国和王杰, 2022), 该影响在不同的空间范围内有所差异, 其空间溢出具有边界效应和地理衰减特征(王亮等, 2023)。

另外, 自国家级大数据综合试验区的政策实施以来, 其政策效应也受到了广泛的关注。绝大部分学者将大数据试验区的建立作为一项准实验, 利用广义双重差分模型、渐近双重差分模型、空间双重差分模型以及倾向得分匹配-双重差分模型, 探究其对全要素生产率、数字经济发展、要素配置和流通、绿色创新绩效和城市创新能力等的影响, 肯定了该政策的积极效应。具体而言, 邱子讯和周亚虹(2021)研究认为大数据试验区的建立可以通过促进纯技术进步显著提高地区全要素生产率;李桥兴和杜可(2021)提出试验区的建立会显著促进区域数字经济发展, 且其政策效应与区域发达程度有关;李潇和韦晓慧(2022)认为以大数据综合试验区建设所表征的数字经济发展能够通过市场一体化和劳动力要素优化配置等途径实现

流通业绩效优化；陈文和常琦(2022)提出大数据可以有效赋能企业提高绿色创新绩效；徐林等(2022)研究认为国家级大数据综合试验区政策对城市创新能力提升具有显著促进作用。

综上，在考察数字经济对碳排放的影响研究中，学者们通常采用熵值法和主成分分析法计算得到的数字经济指数表征数字经济的发展程度，并使用相应的工具变量处理其模型内生性。而以国家级大数据试验区的建立表征数字经济发展的研究中，更多探讨的是政策的经济效应和对其他宏观经济变量的影响。同时，也有少部分学者探讨了大数据试验区对环境福利的影响，如通过促进产业结构升级和提高机器人的应用，降低雾霾污染和碳排放水平(程云洁和段鑫，2023)；通过技术创新和数字普惠金融发展，降低电力消费碳排放水平(常皓亮等，2023)。因此，本文以大数据综合试验区的建设表征数字经济的发展，探究其对碳排放总量以及电能、热能、交通运输和直接能源消耗产生碳排放的影响，以期明晰数字经济增加或抑制碳减排的传导路径。

本文可能提供的边际贡献包括三点：其一是在碳排放核算方面。区别于依据化石能源消耗量或能源活动水平的总量核算方法，本文基于各城市的面板数据核算了不同方式产生的碳排放数量，并加总得到城市层面的碳排放总量。数据结构的细化使本文能够更准确地识别数字经济影响碳排放的路径，细分至对工业用电和居民用电产生的碳排放；其二是在大数据综合试验区虚拟变量的选取范围方面。在使用地级市数据分析的过程中，已有研究通常将获批试验

区的省份内所有城市都作为处理组，而未考虑到省内部分地级市在发展规划和相关政策方面的参与度都极低。本文综合考量各试验区的相关政策、发展思路、空间布局和相关产业发展，手工整理识别出城市层面的 35 个有效处理组，提高了双重差分模型估计的准确性；其三是在模型设定方面。本文在使用交错双重差分模型进行分析的基础上，进一步采用包含政策空间溢出效应的修正两阶段双重差分模型。即在分析试验区政策对本城市碳排放影响的同时，还全面考察了政策对处于不同距离范围内其他城市碳排放的影响。

二、理论分析与研究假设

国家大数据综合试验区的设立对于促进大数据基础设施的建设和整合、加快数据资源的应用和创新、形成大数据产业集聚和完善大数据产业生态都具有十分重要的作用。整体而言，试验区的建立会对城市碳排放产生两个不同方向的影响：一是大数据本身的能耗特性和相关产业的集聚会引致碳排放规模的扩大；二是大数据的应用所带来的运营效率提升、产业结构优化和技术进步会使得碳排放降低。具体而言，大数据的基础设施包括工业互联网、物联网和 5G 基站等网络基础设施，人工智能、云计算和区块链等新技术基础设施，数据中心和智能计算中心等算力基础设施。这些信息基础设施承担着大规模的数据存储、数据生产、数据传输和数据处理的任务，其过程会消耗大量的电力资源并导致碳排放的增加。同时，各试验区相继建立了一批大数据产业示范园区、大数据创业创新生态区和大数

据综合应用引领区，并培育了一批大数据骨干企业。该举措使得大数据产业在试验区内大规模集聚，增加碳排放的数量。另外，数字媒体设备、智能设备和其他数字产品通常具有较短的创新周期和较高的设备运转率，因此具有较快的更新换代速度，从而加剧能源的消耗。然而，试验区政策同时也促进了数字技术与传统产业的深度融合，赋能千行百业进行数字化转型，推动智能制造工程深入开展。一方面，在此过程当中数字化新技术、新产品、新业态不断渗透和升级，降低对传统化石能源的依赖程度，推动能源消费结构向更清洁的模式转变。另一方面，大数据技术和机器学习可以实现数据的高频采集和分析，有助于传统行业实施全流程的能源数据监测和成本核算等精细化管理，为进一步优化生产方案提供了保证。同时，数字化过程加强了企业间的信息共享和协同运作，提升了产业链和供应链的运作效率，降低了碳排放的数量。根据上述分析可知，大数据技术的自身发展和引致的产业集聚会导致城市碳排放的增多，但其与传统产业的深度融合以及其对技术创新、产业链、政务服务和居民生活的辐射效应会带来效率的提升，并由此释放各领域的碳减排潜力。据此，本文提出以下假设：

假设 1:大数据试验区建设能够降低城市碳排放水平。

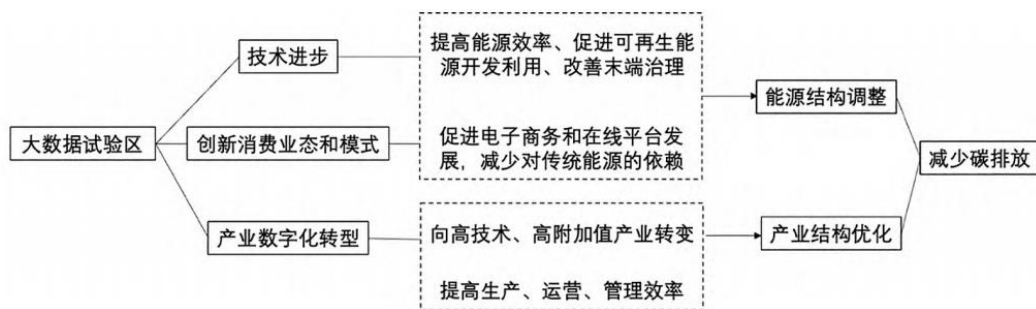


图 1 大数据试验区降低碳排放的理论机制

大数据试验区降低碳排放的理论机制如图 1 所示。首先，试验区政策可以通过促进技术进步降低城市碳排放水平。大数据综合试验区的发展重点包括数据资源共享开放、大数据人才培养和大数据国际合作等，良好的创新环境和专业人才的集聚都为技术进步奠定了基础。而技术进步主要可以从三个方面降低碳排放水平：一是生产工艺的改进可以提高能源的利用效率，降低单位产出的碳排放水平；二是绿色技术进步可以加快清洁能源、可再生能源和新能源的开发和使用，优化能源消费结构；三是处理技术的提升可以改善碳排放的末端治理，即以更低成本捕集、封存和利用二氧化碳排放。因此，本文提出以下假设：

假设 2:大数据试验区建设能够通过加快技术进步降低城市碳排放水平。

其次，试验区政策还可以通过改变经营和消费模式降低城市碳排放水平。大数据综合试验区内新型基础设施和系统平台的建立，为商业模式的创新提供了有利条件。依托于网络和数字平台的电子商务、即时零售等新型商业模式迅速发展，其相

较于传统商业模式而言具有更低的碳排放水平。一方面，通过信息挖掘和需求预测，企业可以在满足消费者需求的同时合理规划库存，降低由仓储、运输和损耗带来的额外能源消耗；另一方面，对实体店铺投入的减少降低了对煤炭、钢筋水泥等重工业的依赖程度，改善了能源消费结构，进而降低了碳排放水平。据此，本文提出以下假设：

假设 3:大数据试验区建设能够通过推动消费业态和模式的创新降低城市碳排放水平。

最后，试验区政策可以通过促进产业数字化转型降低城市碳排放水平。大数据综合试验区相继出台一系列产业政策，完成大数据产业的初步布局，有力推动数实融合和产业数字化转型。一方面，产业的数字化转型可以实现生产数据全面、精准、实时传输，提高生产、物流、仓储和设备维护的管理效率，使得企业自身逐步向高技术、高附加值的产业进行转变。另一方面，产业的数字化转型有助于上下游企业在质量管理、计划采购、仓储物流等方面进行信息互通，提升产业链的协作效率，将资源从低效率部门向高效率部门转移。由此带来的产业结构高级化、合理化和生产性服务业比重的加大，可以降低碳排放水平(王文举和向其凤，2014)。因此，本文提出以下假设：

假设 4:大数据试验区建设能够通过促进产业数字化转型降低城市碳排放水平。

三、研究设计与模型设定

(一)基本模型设定 1.交错双重差分模型。

为了探究国家级大数据综合试验区政策的实施对于城市碳排放的影响，本文分别将第一批和第二批国家级大数据综合试验区作为外生的政策冲击，构建交错双重差分模型。具体模型如下：

$$y_{it} = \alpha + \beta NBDEZ_{it} + \delta X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (1)$$

其中 y_{it} 为被解释变量，包括各城市的碳排放总量、电能消耗产生的碳排放量、热能消耗产生的碳排放量、直接能源消耗产生的碳排放量和交通运输产生的碳排放量。 i 表示城市， t 表示年份。 $NBDEZ_{it}$ 为核心解释变量，是代表国家级大数据综合试验区的政策虚拟变量，根据其成为试验区的时间进行赋值。若该城市当年成为国家级大数据综合试验区，则对其当年和以后年份均赋值为 1，对之前年份和非试验区城市均赋值为 0。 X_{it} 代表一系列控制变量，包括各城市层面的经济总量和结构特征、人口特征等。 λ_t 和 μ_i 分别表示时间固定效应和城市固定效应， ϵ_{ij} 代表随机误差项， β 是本文的关键估计系数。

2.考虑空间溢出效应的双重差分模型。考虑到国家级大数据试验区属于基于地理边界实施的政策，其政策效应很可能溢出到周边城市。在此情况下，若直接使用交错双重差分模型可能会导致以下问题，进而产生估计偏误：一是非试验区城市可能不是合格的控制组，其反事实趋势实际上也受到了邻近处理组

的影响；二是试验区城市的碳排放变化可能不仅反映了该城市自身接受处理的效应，还包括邻近其他处理组的影响。因此，本文借鉴 Butts(2023)提出的潜在因果框架，在双重差分模型中考虑空间溢出效应的影响，以区分政策处理效应和溢出效应。

$$y_{it} = \tau NBDEZ_{it} + \sum_{j=1}^J \gamma_j (1 - NBDEZ_{it}) Ring_{ij} + \mu_i + \lambda_t + \epsilon_{it} \quad (2)$$

其中， y_{it} 、 $NBDEZ_{it}$ 、 λ_t 、 μ_i 和 ϵ_{it} 的含义不变， $Ring_{ij}$ 是一组同心环状的虚拟变量，当距离 i 城市 0-100 公里、100-200 公里和 200-300 公里范围内存在其他处理组城市时，该变量取值为 1，否则为 0。在该潜在因果框架下， τ 是政策总效应的估计系数， γ_j 是各距离环内溢出效应的估计系数。

在满足无预期效应和平行趋势假定下，参照 Gardner(2022)提出修正的两阶段双重差分模型。第一阶段使用未被处理且未被溢出效应影响的城市样本估计时间固定效应和城市固定效应，即：

$$y_{it} = \mu_i + \lambda_t + u_{it} \quad (3)$$

然后基于估计的时间固定效应 $\hat{\lambda}_t$ 和城市固定 $\hat{\mu}_i$ 效应计算所有样本的残差，即：

$$\tilde{y}_{it} = y_{it} - \hat{\mu}_i - \hat{\lambda}_t \quad (4)$$

第二阶段，将残差 $\tilde{y}_{it}$ 回归到表示处理和溢出效应的虚拟变量上，即：

$$\tilde{y}_{it} = \tau NBDEZ_{it} + \sum_{j=1}^J \gamma_j (1 - NBDEZ_{it}) Ring_{ij} + \epsilon_{it} \quad (5)$$

(二)变量选取

1.城市碳排放水平。已有学者提供了不同的碳排放测算方式，主要包括依据化石能源消耗量和《2006 年 IPCC 国家温室气体清单指南》(简称 IPCC2006)提供的核算方法以及 CEADs 中国碳核算数据库提供的依据能源活动水平的核算方法。而针对地级市层面碳排放的核算方法主要包括：一是利用我国各城市每年原煤、焦炭、汽油、煤油、柴油、燃料油、天然气、电力等代表性能源的消费量乘以 IPCC 提供的二氧化碳排放系数，加总得到每个城市的年度碳排放总量；二是通过夜光灯数据法，即首先利用 IPCC 提供的排放系数计算全国范围内总的二氧化碳排放量，再用粒子群优化-反向传播法处理 DMSP/OLS 和 NPP/VIIRS 各城市的卫星影像数据，进而将全国层面的碳排放数据分解至每个城市(Chen et al., 2020; 张传兵和居来提·色依提, 2023)。

上述核算方法均提供了城市层面的碳排放加总数据，可以在分析某项政策对碳排放的影响时提供整体证据。而无法获取该项政策对不同方式产生碳排放的影响，以识别政策发挥作用的准确路径。因此，借鉴 Glaeser & Kahn (2010)、Li et al.(2013)、吴建新和郭

智勇(2016)提供的碳排放核算方法,本文分别核算了城市层面的直接能源消耗、电能消耗、热能消耗和交通运输消耗产生的碳排放。

具体而言,对于直接能源消耗产生的碳排放,本文利用各城市煤气和液化石油气等能源的直接消耗量与 IPCC2006 提供的碳排放系数计算其产生的碳排放。对于电能消耗产生的碳排放,收集整理得到国家发展和改革委员会应对气候变化司发布的 2011—2019 年我国华北、东北、华东、华中、西北和南方 6 大区域电网的基准线排放因子。通过各城市的全年用电量、工业用电量和城乡居民生活用电量与其对应年份和区域的电网基准线排放因子计算得到各城市电能消耗产生的碳排放。对于热能消耗产生的碳排放,本文从《中国城市建设统计年鉴》获取了各城市的锅炉房供热和热电厂供热总量,并参考《燃煤工业锅炉节能监测》(GB/T15317—2009)中燃煤工业热效率值的区间,确定采用 70%的热效率值。通过供热总量、热效率值和原煤的平均发热量系数可以计算出供热消耗的原煤数量,将其折算为标准煤后利用 IPCC2006 的相关转化因子计算得到各城市热能消耗产生的碳排放。对于交通运输消耗产生的碳排放,根据《中国交通运输统计年鉴》和《中国统计年鉴》计算出旅客周转量(亿人公里)和货物周转量(亿吨公里)所消耗的能源数量,再利用各城市的旅客周转量和货物周转量计算得到各城市交通运输消耗的能源数量和碳排放量。将上述电能消耗、热能消耗、交通运输消耗和直接能源消耗产生的碳排放加总即得到各

城市的碳排放总量。平均而言，电能消耗产生的碳排放占比为 61%，是产生碳排放的主要来源，热能、交通运输和直接能源消耗产生的碳排放占比分别为 7%、28%和 4%。

2.国家级大数据综合试验区。该变量是本文的核心解释变量，本文依据各地级市成为试验区的时间对该虚拟变量进行赋值。为引领大数据产业发展、实现数据共享，国家发展改革委员会、工业和信息化部 and 中央网信办于 2016 年 2 月批复同意贵州省建设首个国家大数据综合试验区，并于 2016 年 10 月公布第二批试验区省份名单。已有文献考虑到二者获批时间的差异和政策的滞后性，通常对第一批试验区从 2016 起开始赋值为 1，对第二批试验区从 2017 年开始赋值为 1。值得注意的是，尽管贵州省于 2016 年 2 月正式被批复同意建设大数据试验区，但在国务院于 2015 年 8 月印发的《促进大数据发展行动纲要》中早已明确提出“开展区域试点，推进贵州等大数据综合试验区建设”，且同年 9 月 18 日贵州大数据综合试验区建设正式启动。因此，考虑到大数据试验区建设的实际进程而非局限于批复时间，本文对第一批试验区从 2015 年起开始赋值为 1，对第二批试验区从 2017 年开始赋值为 1。

另外，第二批试验区包括京津冀、珠三角、上海、河南、重庆、沈阳和内蒙古大数据综合试验区，已有文献通常直接将获批试验区的省份内所有城市都作为处理组，而未考虑到省内部分地级市在发展规划和相关政策方面的参与度都极低。因

此，本文梳理了各试验区的发展思路，结合相关政策、发展思路、空间布局和相关产业发展识别出有效处理的 35 个地级市。具体内容见表 1。

表 1 国家级大数据综合试验区主要内容

	试验区	发展思路	主要相关政策	部分大数据相关产业发展	有效处理组
第一批	贵州	黔中引领、两极带动、协同发展。两地三区、三大体系、七大平台、七大任务以及十大工程。	《贵阳大数据产业行动计划》《贵阳国家高新区促进大数据技术创新十条政策措施》《六盘水市人民政府关于大力发展电子商务的实施意见》《安顺市政府关于进一步加快黎阳高新技术产业园区建设促进高新技术产业发展的意见》《毕节市大数据信息产业发展规划（2016—2025）》	黔南州超算中心、贵安新区电子信息产业园大数据基地、六盘水水月产业园区大数据终端生产物流园、遵义云卫通大数据中心、安顺西秀智能终端产业园、毕节金沙电子信息产业园、铜仁高新技术产业园区	贵阳市、六盘水市、遵义市、安顺市、毕节市、铜仁市
第二批	京津冀	一心一地两区，形成北京中关村+天津滨海新区、武清+河北张家口、廊坊、承德和秦皇岛“1+2+4”协同发展功能格局，建设张北、廊坊、承德、秦皇岛、石家庄 5 大数据产业基地	《北京市大数据和云计算发展行动计划（2016—2020 年）》《天津市关于加快推进智能科技产业发展若干政策》《秦皇岛市推动科技服务业高质量发展实施方案》《中国数坝·张家口市大数据产业发展规划》《承德大数据产业发展实施意见》《承德大数据产业发展规划》《廊坊加快发展“大智移云”引领产业升级行动计划（2016—2018）》《廊坊市大数据产业发展规划纲要》	中关村大数据产业园、天津西青大数据产业园、秦皇岛健康大数据产业基地、张家口市张北县云基地阿里数据中心、张家口怀来县东花园大数据产业基地、承德德鸣大数据产业园区、廊坊·中关村软件园	北京市、天津市、秦皇岛市、张家口市、承德市、廊坊市

第二批	珠三角	以广州、深圳为核心区，重点打造佛山、珠海、中山、肇庆、江门等珠江西岸大数据产业带；惠州、东莞等珠江东岸大数据产业带；打造汕头、汕尾、阳江、湛江等沿海大数据产业带	《广州市关于促进大数据发展的实施意见》《深圳市促进大数据发展行动计划》《佛山市智能制造发展行动计划》《珠海市创建珠江三角洲国家大数据综合试验区实施方案》《中山市促进大数据发展行动计划》《肇庆市大数据发展规划（2017—2020）》《惠州市发展大数据产业总体设计方案》《东莞市大数据发展实施方案》《汕头市促进大数据发展行动计划》《汕尾市“明珠数谷”大数据产业园发展规划》《阳江市大力发展电子商务加快培育经济新动力的实施方案》《清远市大数据发展“十三五”规划》	广州开发区大数据产业园、广州天河大数据产业园、深汕特别合作区大数据产业园、深圳腾讯大数据众创空间项目、佛山福能大数据产业园、佛山市南海区大数据产业园、珠海欧比特地理信息产业大数据创业创新孵化园、中山美居智能制造大数据产业园、肇庆大数据云服务产业园、江门市“珠西数谷”省级大数据产业园、惠州潼湖生态智慧区大数据产业园、东莞中科智大数据创业创新孵化园、汕头世纪互联大数据协同创新产业园、汕尾市“明珠数谷”大数据产业园、阳江涉农电子商务平台、“粤西数谷”湛江国家高新区大数据产业园、清远万方大数据产业园	广州市、深圳市、佛山市、珠海市、中山市、肇庆市、江门市、惠州市、东莞市、汕头市、汕尾市、阳江市、湛江市、清远市
	上海	以静安为核心的“1中心+4基地”的发展格局	《上海市大数据发展实施意见》《上海市关于促进云计算创新发展培育信息产业新业态的实施意见》	上海静安大数据产业基地、上海杨浦大数据产业基地	上海市
	河南	引导规范重点行业、大型企业数据中心建设，打造以郑州、洛阳为中心的区域性数据中心	《郑州市人民政府关于促进大数据产业发展的若干意见》《关于促进大数据发展的实施意见》	郑东新区智慧岛、郑州高新区大数据产业园、洛阳大数据产业园	郑州市、洛阳市
	重庆	积极引领东部、中部、西部、东北四大板块发展	《重庆市健康医疗大数据应用发展行动方案》《重庆市关于运用大数据加强对市场主体服务和监管的实施意见》	重庆仙桃国际大数据谷、重庆云谷永川大数据产业园	重庆市
	沈阳	立足沈阳，辐射辽宁，带动东北的市场布局	《沈阳市促进大数据发展三年行动计划（2015—2018）》	沈阳数字经济产业园、法库·贵澳大数据农业科技产业园	沈阳市
	内蒙古	重点建设和林格尔新区大数据产业核心区，加快建设包头大数据创新产业园区、鄂尔多斯高新技术产业园、乌海大数据产业园等重点园区	《呼和浩特市关于促进大数据产业发展的若干政策》《呼和浩特市大数据人才引进暂行办法》《呼和浩特市扶持电子商务产业发展若干政策》《内蒙古国家大数据综合试验区建设实施方案》	呼和浩特和林格尔县盛乐云计算产业园区、呼和浩特新城区鸿盛云计算产业园、包头云计算中心、乌海大数据产业园、鄂尔多斯绿色数据港	呼和浩特市、包头市、乌海市、鄂尔多斯市

数据来源：各省(区、市)人民政府官网、公开资料整理。

3.控制变量。为了控制其他变量对城市碳排放的影响，本文选取了一系列城市层面的控制变量。包括地区生产总值、年末

人口规模、第三产业占总产值的比重、当年实际使用的外资金额、社会消费品零售总额、公共图书馆图书总藏量、固定资产总额、教育支出、科学支出和年末金融机构各项贷款余额。其中，除第三产业产值占总产值的比重外，对其他变量均进行对数化处理。

(三)数据来源与描述性统计

鉴于部分地级市的数据缺失较为严重，在剔除该部分城市后，本文最终采用 2011—2019 年我国 286 个地级市的样本数据。对于仅有个别年份缺失的部分变量，使用插值法进行补齐。另外，由于西藏自治区不属于我国六大区域电网覆盖的范围，无法通过用电量和电网基准线排放因子计算其电能消耗产生的碳排放，因此本文样本中不包含西藏地区。

除特别说明的整理和计算外，本文所使用的其余数据均来自《中国统计年鉴》《中国交通运输统计年鉴》《中国能源统计年鉴》《中国城市统计年鉴》《中国城市建设统计年鉴》《中国区域统计年鉴》。本文主要变量的描述性统计如表 2 所示，变量的单位为对数化处理之前原始变量的单位。

表 2 主要变量的符号和描述性统计

变量名称（单位）	变量符号	平均值	最小值	最大值	标准差
碳排放总量（kg）	lnce	23.01	13.99	26.02	1.114
电能消耗产生的碳排放（kg）	lnele_ce	22.58	13.75	25.54	1.198
热能消耗产生的碳排放（kg）	lnheat_ce	20.87	14.37	24.56	1.359
交通运输消耗产生的碳排放（kg）	lntrans_ce	21.45	12.38	25.14	1.062
直接能源消耗产生的碳排放（kg）	ln_dir_ce	19.29	9.947	24.50	1.532
国家级大数据综合试验区	NBDEZ	0.043	0	1	0.203
创新环境虚拟变量	innov	0.511	0	1	0.499
经济水平虚拟变量	eco	0.401	0	1	0.490
发明专利授权数目（个）	patent	50.17	0.341	100	28.87
人均邮政业务量（万元）	postal	0.204	0.000	0.710	0.039
计算机和软件从业人员占比（%）	indus_digi	1.332	0.054	10.85	1.009
地区生产总值（万元）	lngdp	16.54	14.10	19.75	0.934
年末人口规模（万人）	lnpop	5.845	-3.218	8.136	0.803
第三产业占总产值的比重（%）	third_ratio	41.12	10.15	83.52	9.939
当年实际使用的外资金额（万美元）	lnfdi	10.03	1.098	14.94	1.893
社会消费品零售总额（亿元）	lncons	15.52	5.472	18.88	1.089
公共图书馆图书总藏量（万册）	lnlib	7.314	2.302	11.27	1.034
固定资产总额（亿元）	lnasset	15.76	10.89	18.60	0.967
教育支出（亿元）	lnedu	13.10	0.773	9.905	16.24
科学支出（亿元）	lnsci	10.33	6.251	15.52	1.394
年末金融机构各项贷款余额（亿元）	lnfin	16.38	6.369	20.41	1.194

四、实证结果分析

(一)基准回归结果

在交错双重差分模型当中若存在异质性的处理效应，包括不同时点接受处理的组别之间，和接受处理后不同的时长之间，则使用传统的双向固定效应模型(TWFE)会导致潜在的偏误。且在动态情形下，双向固定效应模型还可能产生“坏的控制组”问题，即较晚接受处理的样本会以较早接受处理的样本作为控制组，使估计系数受到跨期交叉污染的影响。Callaway & Sant’ Anna(2021)通过计算每个组别-时期的平均处理效应并进行加权估计，提供了政策效应的异质性-稳健估计量以解决上述问题。为了评估异质性和“坏的控制组”问题的严重程度和导致的偏误大小，本文在表 3 当中汇报了双

向固定效应模型的回归结果作为对照，在表 4 中汇报了 Callaway & Sant’ Anna 的异质性-稳健估计结果。

表 3 双向固定效应回归结果

被解释变量	(1)	(2)	(3)	(4)	(5)
	碳排放总量	电能消耗 产生的碳排放	热能消耗 产生的碳排放	交通运输 产生的碳排放	直接能源消耗 产生的碳排放
<i>NBDEZ</i>	-0.1801** (0.0913)	-0.3630*** (0.1134)	-0.0276 (0.1155)	-0.0288 (0.1416)	0.0220 (0.1025)
控制变量	是	是	是	是	是
年份固定效应	是	是	是	是	是
城市固定效应	是	是	是	是	是
N	2391	2367	1050	2441	2385
R ²	0.8898	0.8977	0.8756	0.7334	0.9037

注：*、**和***分别表示在 10%、5%和 1%的水平上显著。下同。

表 4 Callaway & Sant’ Anna 异质性-稳健估计结果

被解释变量	(1)	(2)	(3)	(4)	(5)
	碳排放总量	电能消耗 产生的碳排放	热能消耗 产生的碳排放	交通运输 产生的碳排放	直接能源消耗 产生的碳排放
<i>NBDEZ</i>	-0.1656* (0.0932)	-0.2488* (0.1325)	-0.2747 (0.1856)	0.0764 (0.0859)	-0.1156 (0.1466)
控制变量	是	是	是	是	是
年份固定效应	是	是	是	是	是
城市固定效应	是	是	是	是	是
N	2381	2346	1050	2433	2375

表 4 中列(1)-(5)的估计结果表明国家级大数据试验区政策对城市碳排放总量和电能消耗产生碳排放影响的估计系数分别为-0.16 和-0.24，且均在 10%的显著性水平下显著。而对热能产生的碳排放、交通运输产生的碳排放和直接能源消耗产生的碳排放无显著影响，表明大数据试验区政策主要是通过影响电能消耗进而影响城市碳排放总量。该结论与表 3 中双向固定效应模型的估计结果大致相同，表 4 中关键估计量的绝对值更小

且显著性水平更低。表明若忽略处理效应异质性和“坏的控制组”会导致一定程度的估计偏误，高估国家级大数据试验区的降碳效应。

同时，利用 Goodman- Bacon (2021)的方法将多时点处理效应进行分解，分解结果如表 5 和表 6 所示。可以看出，在国家级大数据试验区对碳排放总量的影响当中，将第一批试验区(早期)作为实验组而将第二批试验区(后期)作为控制组时，DID 估计量的值为-0.06;反过来，将第二批试验区作为实验组而将第一批试验区作为控制组时，DID 的估计量为-0.36。因此，第二批试验区的设立相较于第一批而言，其降碳作用更为明显。在对电能消耗产生碳排放的影响当中，将第一批试验区作为实验组而将第二批试验区作为控制组时，DID 的估计量为 0.13;反过来将第二批试验区作为实验组而将第一批试验区作为控制组时，DID 的估计量为-0.33。该结果表明第一批试验区的设立相较于第二批而言增加了电能消耗产生的碳排放，而第二批试验区的设立相较于第一批而言减少了电能消耗产生的碳排放。

表 5 碳排放总量的 Bacon 分解结果

	权重	平均 DID 估计量
早期处理 vs 后期处理	0.005	-0.067
后期处理 vs 早期处理	0.004	-0.363
所有处理 vs 从未处理	0.990	-0.184

表 6 电能消耗产生碳排放的 Bacon 分解结果

	权重	平均 DID 估计量
早期处理 vs 后期处理	0.004	0.138
后期处理 vs 早期处理	0.003	-0.331
所有处理 vs 从未处理	0.993	-0.412

将两个批次的试验区作为实验组，将所有非试验区作为控制组时，可以看出国家级大数据试验区政策从整体上降低了城市的碳排放总量和电能消耗产生的碳排放。而上述分解情况表明，第一批试验区和第二批试验区无论在降碳总效应还是对电能消耗的影响上均存在异质性。这可能是由于不同批次试验区的发展定位和路径不同，贵州作为首个国家级大数据综合试验区，其发展的总体定位是从七大主要任务着手，开展系统性试验；而第二批试验区的定位主要是围绕数字要素的流通、数字资源的统筹和大数据产业的集聚。即相较于第二批试验区而言，第一批试验区的发展目标更为多元，且担负着系统性试验的任务。

(二)平行趋势检验

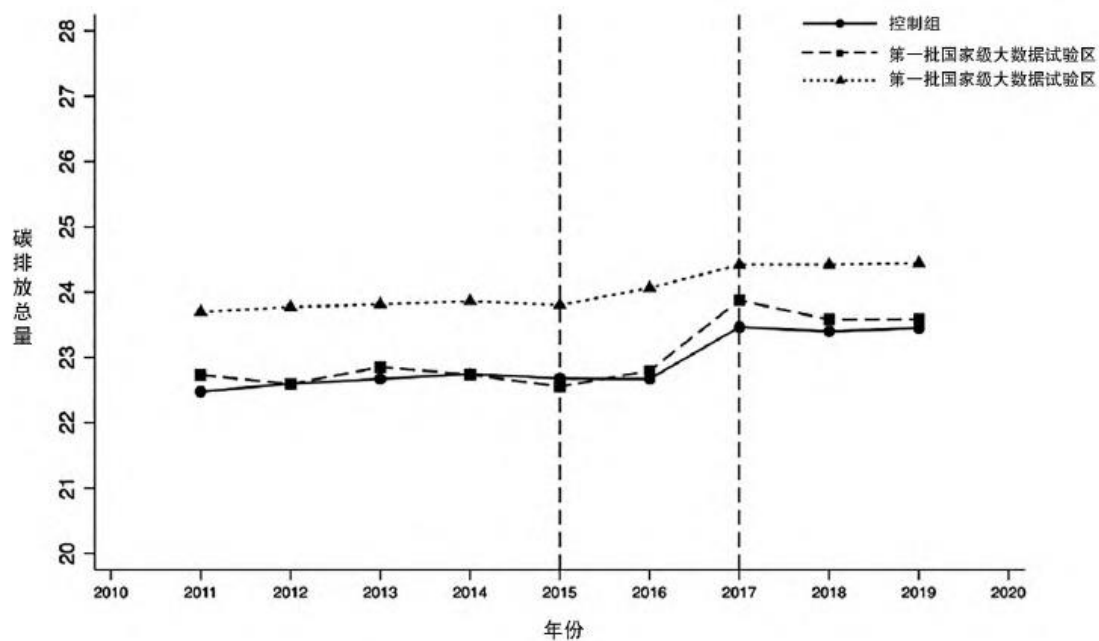


图 2 碳排放总量的时间趋势

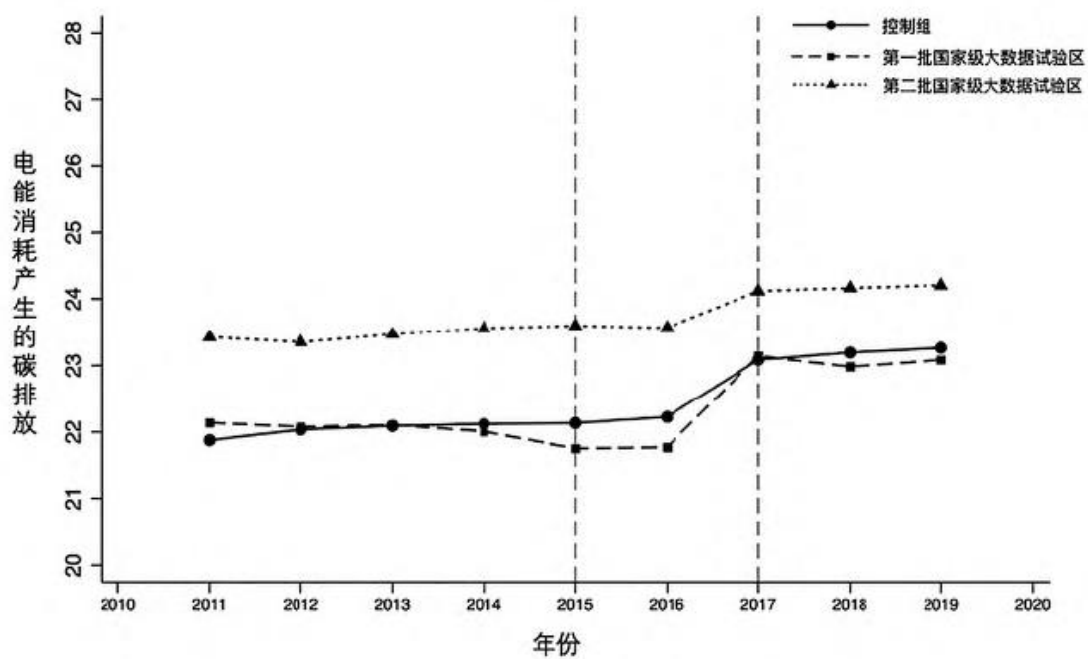


图 3 电能消耗产生碳排放的时间趋势

平行趋势假设是双重差分模型得以成立的核心假设，要求在未受到政策干扰时，试验区和非试验区城市的碳排放总量和电能消耗产生碳排放的变化趋势相同。图 2 和图 3 分别展示了

碳排放总量和电能消耗产生碳排放的绝对时间趋势，以初步观测其是否满足平行趋势假设。从图 2 中可以看出在第一批试验区受到处理(2015 年)之前，其与控制组的碳排放总量绝对值和变动趋势大致相同，有时略高于控制组的碳排放水平，但在 2015 年出现向下的偏离趋势，随后又有所增加。该现象与上文中 Bacon 分解的估计结果相一致，即第一批试验区设立对碳排放总量的影响较小。同时，第二批试验区受到处理(2017 年)之前，其与控制组的碳排放总量变动趋势高度一致，但 2017 年其上升幅度明显小于控制组，随后二者之间的差异水平逐渐缩小。图 3 中显示的情况大致相同，即第一批试验区电能消耗产生碳排放的变动趋势在受到处理当年开始与控制组产生偏离，第二批试验区电能消耗产生碳排放的变动趋势也在受到处理当年开始与控制组产生偏离。

为了进一步综合考虑本文使用的交错双重差分模型是否满足平行趋势假设，借鉴 Beck et al.(2010)利用事件分析法进行平行趋势检验，模型设置如下：

$$y_{it} = \alpha + \sum_{k=-3}^{k=2} \beta_k event_{ik} + \delta X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (6)$$

其中， $event_{ik}$ 表示相对政策发生时间的虚拟变量，当城市 i 距离其接受处理时间的为 k 年时， $event_{ik}$ 对赋值为 1，否则赋值为 0。在本文中 k 可能的取值包括 -6、-5、-4、-3、-2、-

1、0、1、2、3、4，但部分样本的 k 取值无法取到-6、-5、1、2，因此在模型中设定 k 的取值范围为-4 到 2 的整数，并以-4 作为基期进行估计。 β_k 表示相对政策发生时间对被解释变量的影响。

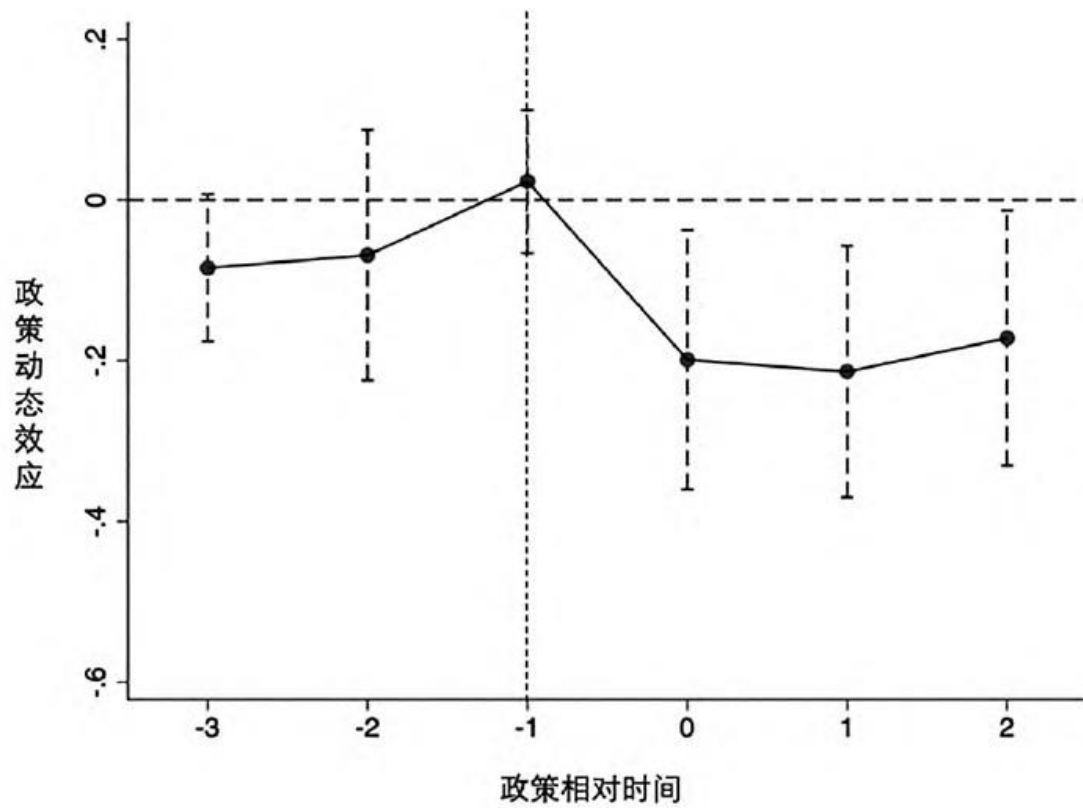


图 4 碳排放总量的平行趋势

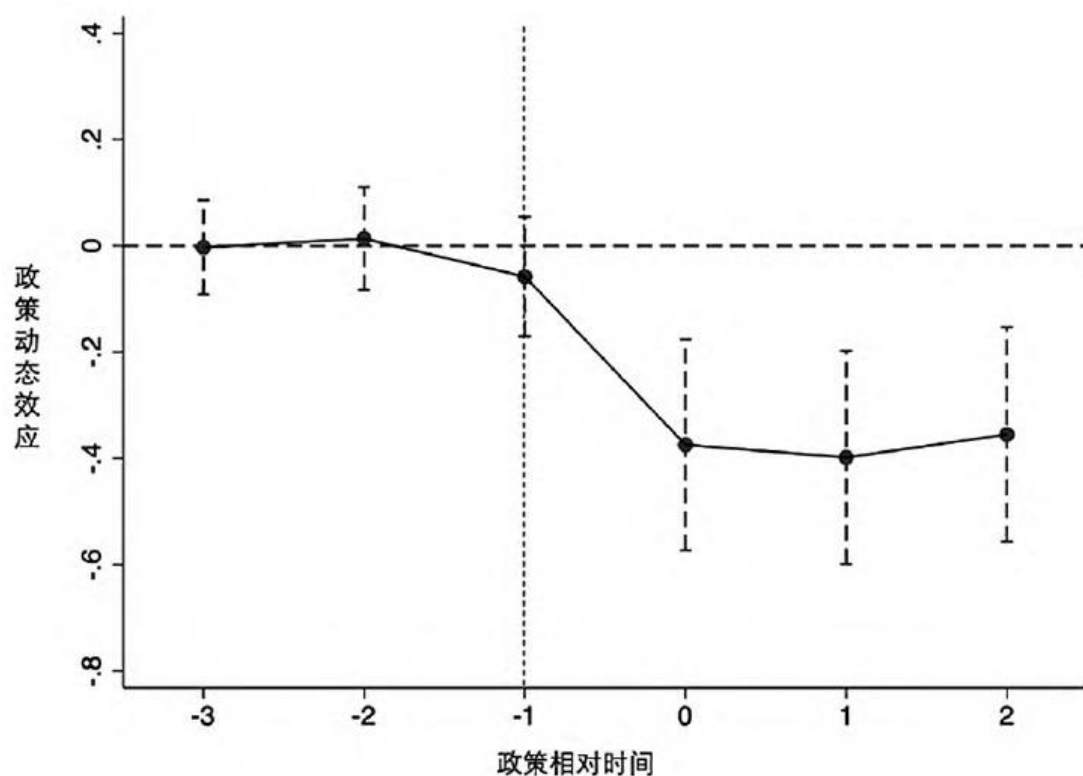


图 5 电能消耗产生碳排放的平行趋势检验

图 4 和图 5 分别展示了以碳排放总量和电能消耗产生碳排放为被解释变量时，式(6)中 β_k 的估计系数，虚线表示其估计量 95%的置信区间。可以看出，政策发生前各虚拟变量系数的置信区间均将 0 值包含在内，即未对被解释变量产生显著影响。表明在国家级大数据试验区设立前，试验组和控制组的碳排放总量、电能消耗产生碳排放水平均不存在趋势性变动，满足平行趋势假设。在国家级大数据试验区设立后，各虚拟变量的系数显著为负，表明该政策的实施对试验区城市的碳排放总量、电能消耗产生碳排放的水平产生负向且持续的影响。

(三)稳健性检验

1.安慰剂检验。

为了验证上述实证结果的稳健性，借鉴周茂等(2018)提供的方法，通过随机虚构政策试点的方式生成“伪实验组”进行安慰剂检验。本文共进行了 1000 次的随机抽样，并对原模型进行重新估计，得到的估计系数及其分布的概率密度如图 6、图 7 所示。由图中可知，随机处理状态下政策虚拟变量的估计系数基本服从 0 均值的正态分布，即不会对碳排放总量和电能消耗产生的碳排放产生显著作用。因而本文基准回归结果当中，国家级大数据试验区政策对碳排放产生的影响是真实存在的，而非由其他不可预测因素驱动的。因此，安慰剂检验进一步验证了本文实证结果的稳健性。

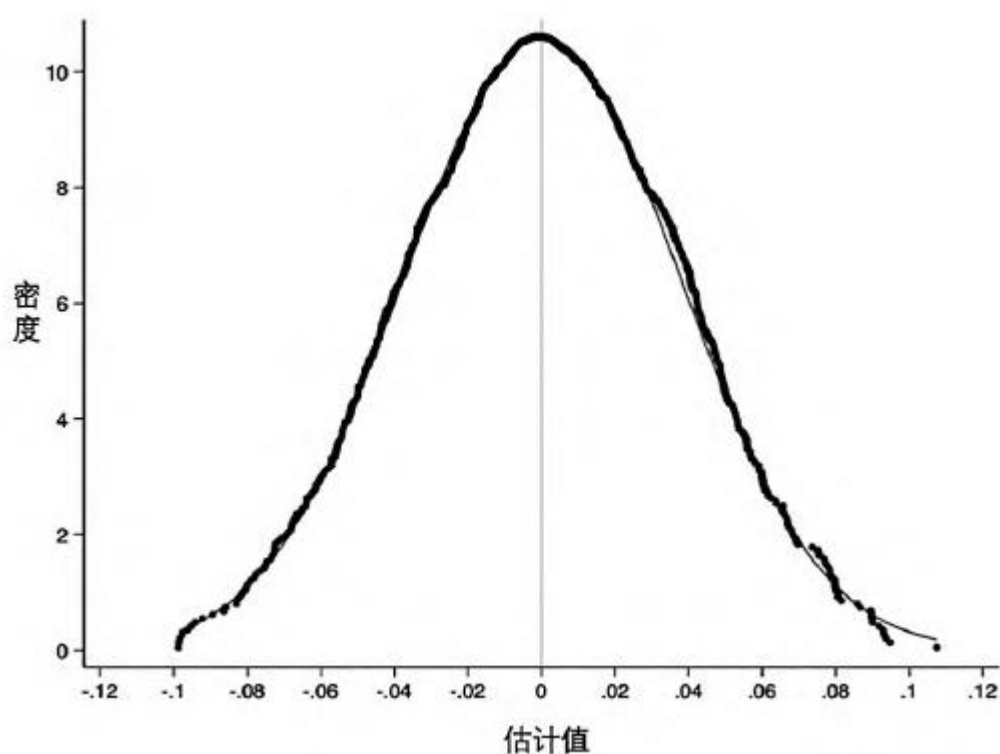


图 6 碳排放总量的安慰剂检验

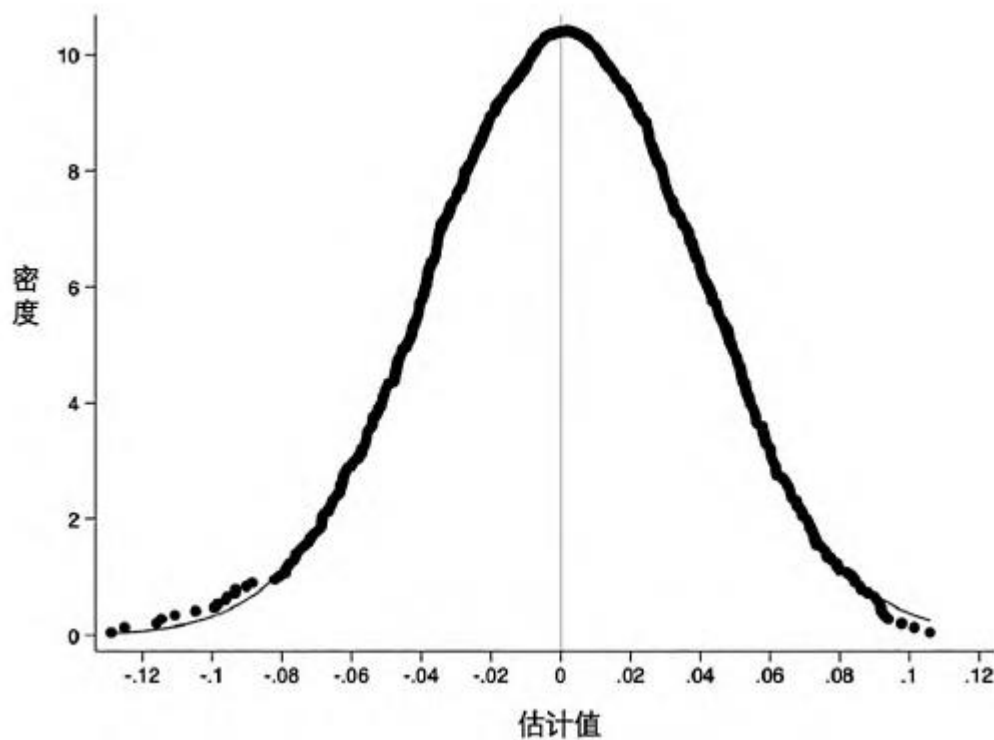


图 7 电能消耗产生碳排放的安慰剂检验

2.反事实检验。为了进一步证实基本实证结果的稳健性，本文通过更改国家级大数据试验区的实施时间进行反事实检验。即虚构政策时间统一提前或滞后，重新回归后观察核心政策变量的系数，若其与基本结果一样显著为负，则说明碳排放总量和电能消耗产生碳排放的下降不是由大数据试验区的设立导致的，而与其他因素有关。反之，若反事实检验的结果显示核心政策变量的系数不显著，则表明城市碳排放、电能消耗产生碳排放的下降主要是大数据试验区设立的结果，因此基本实证的结果是稳健的。本文将政策时间分别提前至实际发生时间的 2 年、3 年和 4 年，以及延后至实际发生时间的 1 年，其核心政策变量的估计系数如

表 7 所示。表 7 中结果显示，虚构政策变量的系数均不显著，据此可以说明本文结论的稳健性。

表 7 反事实检验

被解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	碳排放总量				电能消耗产生的碳排放			
	提前 2 年	提前 3 年	提前 4 年	滞后 1 年	提前 2 年	提前 3 年	提前 4 年	滞后 1 年
<i>NBDEZ</i>	0.2986 (0.2102)	0.0342 (0.0914)	0.0439 (0.1462)	-0.0248 (0.0247)	-0.0922 (0.1298)	-0.0279 (0.1198)	0.0230 (0.0609)	-0.0185 (0.0173)
控制变量	是	是	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是	是	是
城市固定效应	是	是	是	是	是	是	是	是
N	2342	2337	2330	2296	1708	2283	2254	2300

3.倾向评分匹配-双重差分模型。考虑到国家级大数据试验区的试点可能存在一定的选择偏差，即选取具有特定条件的地区进行试点以实现最好的政策效果。例如贵州省能源富集，其水能资源蕴藏和煤炭资源保有储量位居全国前列，且自然环境优良。同时，尽管其人均 GDP 和信息社会发展程度都较低，但其 GDP 增速位于八大试验区之首，增长潜力巨大，适宜开展全面的系统性试验；内蒙古风能储量居全国之首，太阳能辐射总量居全国第二，地域辽阔且劳动力价格水平较低，适宜建设成为基础设施统筹发展类试验区；京津冀和珠三角作为我国重要的城市经济带，其人口、经济、区位优势显著，同时拥有良好的信息化水平和创新环境，适宜从区域一体化的角度建设跨区域类试验区；上海地处长江、黄浦江入海汇合处，是我国最大的枢纽港之一；河南拥有“承东启西、连南贯北”的交通枢纽；重庆处于“两江、三线、三国道”的交汇点；沈阳是环渤海地区与东北地区的重要结合部。因

此上海、河南、重庆、沈阳作为各区域内的核心，适宜构建引领东、中、西和东北四大板块的区域示范类示范区。

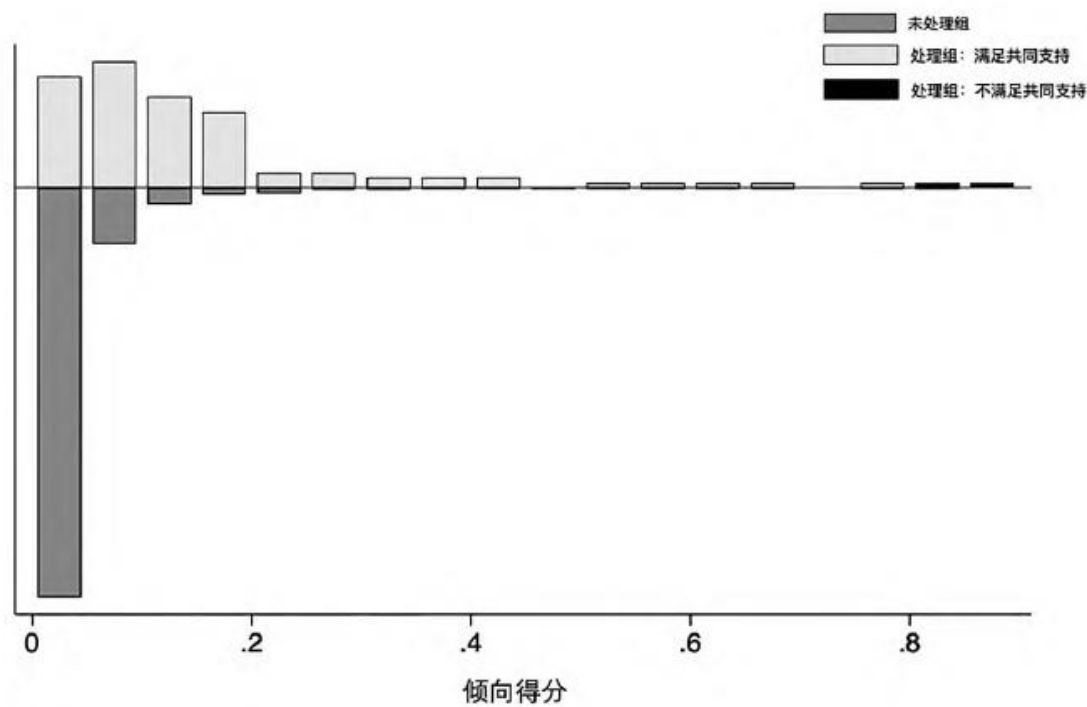


图 8 倾向得分匹配结果

因此，本文利用倾向评分匹配-双重差分(PSM-DID)模型解决上述选择偏差问题。主要思路是在控制组当中寻找与试验区可测变量接近的样本，基于倾向得分进行匹配后，对满足共同支持的样本进行双重差分。具体而言，本文利用 Logit 模型对所有控制变量进行一对一最近邻匹配，得到的倾向得分值及匹配结果如图 8 所示。图 8 结果显示匹配的平衡性较好，绝大部分样本满足共同支持假设。同时，平衡性检验结果的 t 统计量显示，样本匹配之后各控制变量在处理组与控制组之间的差异均不显著，表明支持采用 PSM-DID 模型进行分析。利用匹配成功的样本构建双重差分模型，其估计结果如表 8 中的列(1)和列(4)所示。列(1)和

列(4)中政策虚拟变量的系数分别为-0.49 和-0.65，且均在 1%的显著性水平下显著。对比基准回归结果可知，对更具有可比性的样本进行双重差分，大数据试验区的降碳效果更显著。即相对于具有相似特征的其他城市而言，试验区城市的碳排放总量和电能消耗产生的碳排放均有明显的下降，该结论进一步支持了基本实证结果的稳健性。

表 8 稳健性检验结果

被解释变量	(1)	(2)	(3)	(4)	(5)	(6)
	碳排放总量			电能消耗产生的碳排放		
	PSM-DID	处理时间 统一为 2016	剔除其他政策	PSM-DID	处理时间 统一为 2016	剔除其他政策
<i>NBDEZ</i>	-0.4959*** (0.0936)	-0.0867 (0.0688)	-0.1803** (0.0914)	-0.6533*** (0.1190)	-0.2728*** (0.0890)	-0.3672*** (0.1148)
<i>BC</i>			0.0552 (0.0508)			-0.0519 (0.0673)
<i>CE</i>			-0.0250 (0.0655)			-0.0400 (0.0678)
控制变量	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
城市固定效应	是	是	是	是	是	是
N	187	2391	2391	183	2367	2367
R ²	0.4546	-0.8892	0.8898	0.4262	0.8968	0.8978

4.改变模型时间设定。本文考虑到国家级大数据试验区启动建设的实际时间，以及政策对碳排放影响的时滞效应，在基本模型中分别将 2015 年和 2017 年作为第一批和第二批试验区的设立时间。若非考虑试验区的实质建设进程，而统一选择实际获得批复的年份作为政策实施的时间，即将第一批和第二批试验区的设立时间都统一为 2016 年构建静态双重差分模型。政策虚拟变量的估计结果如表 8 中列(2)和列(5)所示，大数据试验区的设立

对碳排放总量影响不显著，对电能消耗产生碳排放的影响为-0.27，且在1%的显著性水平下显著。对比基准回归结果可知，在不考虑政策时滞和建设进程的情况下，无法准确识别出大数据试验区对城市碳排放总量的影响。因此，本文基础模型中不以统一的获批年份作为政策发生时点的设定是更准确的。

5.控制其他政策的影响。在本文选取的样本区间内，还存在其他相关政策可能会影响城市的碳排放水平，因此应当识别并控制这些政策以验证大数据试验区政策的处理效应。清洁能源示范省的确立会通过改变能源的使用结构直接影响地区碳排放水平，“宽带中国”政策的实施会通过影响地区的数字基础设施建设、电能利用效率等影响碳排放总量和电能消耗产生的碳排放。自2014年开始，国家能源局先后将浙江、四川、西藏、甘肃、宁夏和青海确立为清洁能源示范省。本文据此构建虚拟变量，当城市*i*所在的省份在*t*年被确立为清洁能源示范省时，确立当年及之后的年份都对其赋值为1，否则赋值为0。中华人民共和国工业和信息化部、国家发展和改革委员会为贯彻落实《国务院关于印发“宽带中国”战略及实施方案的通知》，于2015年确定38个城市为2015年度“宽带中国”示范城市。本文据此构建虚拟变量，当城市*i*在*t*年被确立为“宽带中国”示范城市时，确立当年及之后的年份都对其赋值为1，否则赋值为0。将上述两项政策虚拟变量纳入模型后，对模型进行重新估计，估计结果如表8中列(3)和列(6)所示。可以看出，在控制了其他政策变量后，大数

据试验区政策的估计系数分别为-0.18 和-0.36, 与基本模型估计系数相近, 且均在 5%的显著性水平下显著。因此, 本文的基本实证结论不受其他相关政策的影响, 具有一定的稳健性。

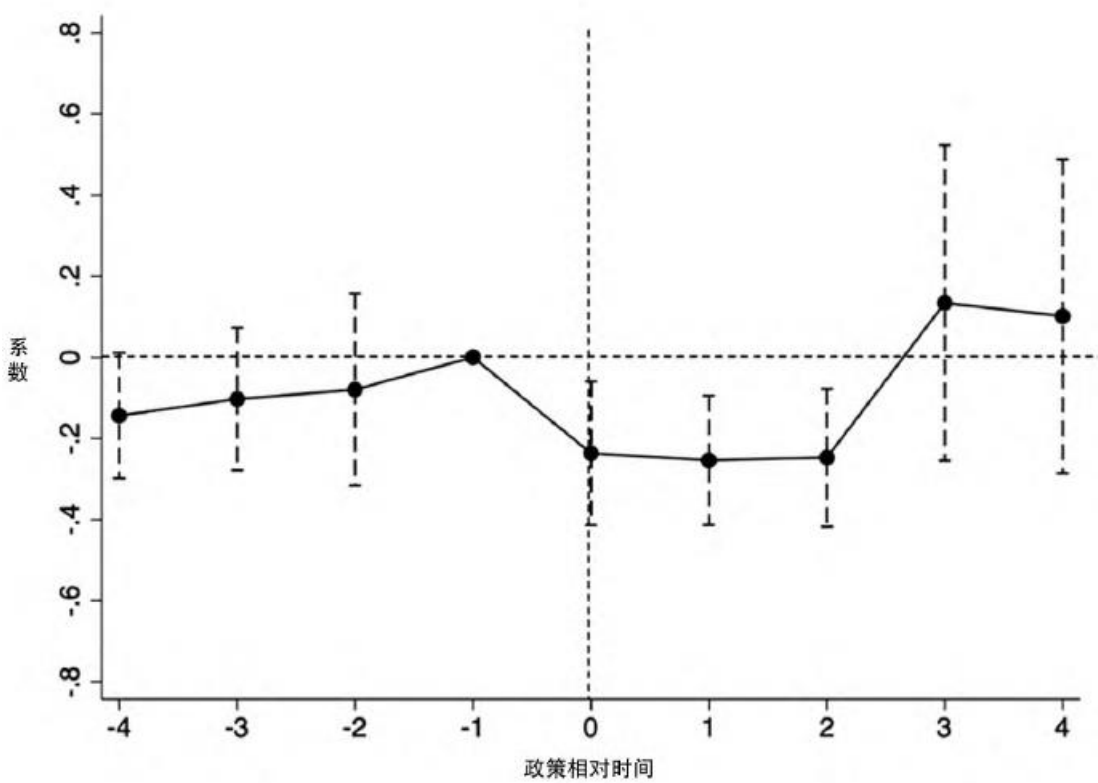


图 9 CD 估计量的动态效应(碳排放总量)

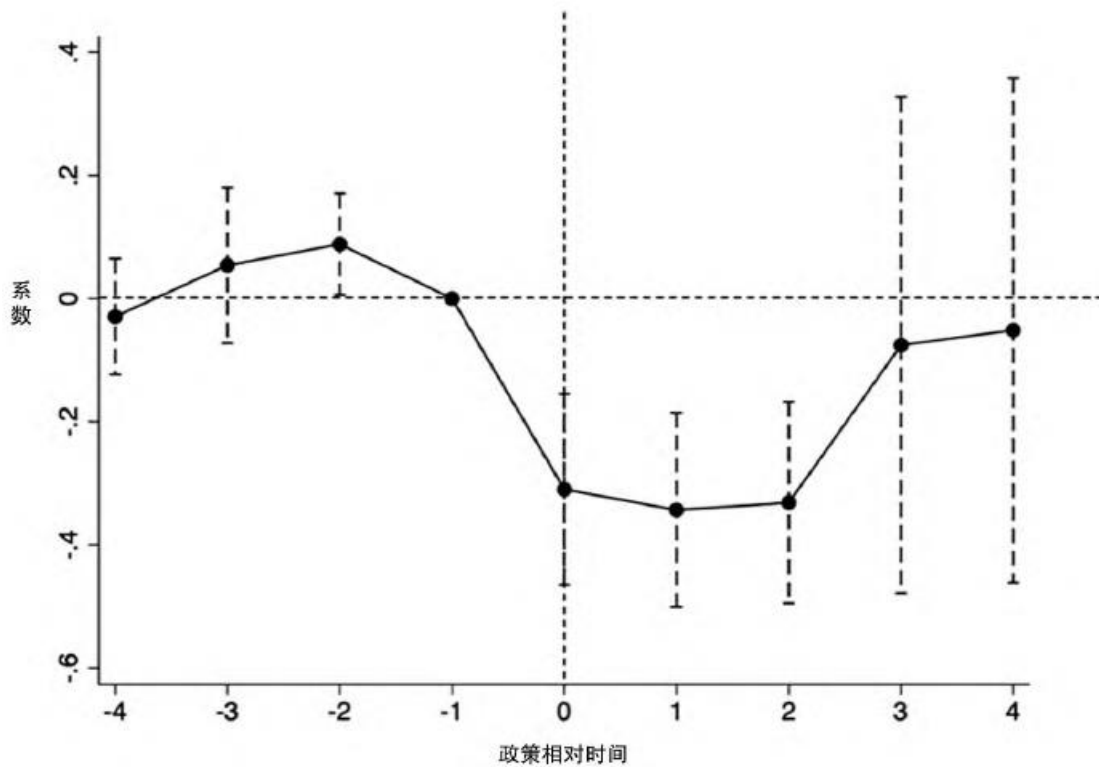


图 10 CD 估计量的动态效应(电能消耗产生的碳排放)

6.更换估计方法。针对交错双重差分模型当中潜在的异质性处理偏误，不同的学者提供了其他异质性-稳健估计量。本文使用了 Chaisemartin & D' Haultfoeuille (2020)提供的估计量，其基本思想也是基于每个组别-时期的处理效应，将其进行加权平均以获得平均处理效应的无偏估计，但有可能存在负的权重。估计结果显示，在以碳排放总量、电能消耗产生的碳排放为被解释变量时分别有 110 个、114 个处理效应，且其均获得正的权重。最终加权得到的平均处理效应如表 9 中的列(1)和列(3)所示，政策虚拟变量影响碳排放总量的估计系数为-0.23，影响电能消耗产生碳排放的估计系数为-0.30，且均在 1%的显著性水平下显著。另外，该估计量的动态效应如图 9 和图 10 所示，以政策发生的

前一期为基期，虚线表示 95%的置信区间。从图 9 和图 10 中可以看出，政策发生前被解释变量均满足平行趋势假设，政策的发生降低了碳排放总量和电能消耗产生的碳排放，且对电能消耗产生碳排放的影响幅度更大。另外，本文还采用了 Cengiz et al.(2019)提出的堆叠回归估计方法，其基本思想是为每一个处理组匹配尚未处理或从未处理的样本，形成一个数据集。进而将这些数据集堆叠在一起进行线性回归，该方法同样解决了“坏的控制组”问题，其估计结果见表 9 中的列(2)和列(4)。上述不同估计方法的结果均支持了基本模型的结论，因此，本文得到的实证结论具有稳健性。

表 9 不同方法的估计结果

被解释变量	(1)	(2)	(3)	(4)
	碳排放总量		电能消耗产生的碳排放	
	CD 稳健估计量	堆叠估计量	CD 稳健估计量	堆叠估计量
<i>NBDEZ</i>	-0.2370*** (0.0829)	-0.1849* (0.0965)	-0.3099*** (0.1077)	-0.3972*** (0.1170)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
城市固定效应	是	是	是	是
N	536	7078	516	20131
R ²	-	0.8860	-	0.8951

(四)考虑空间溢出效应的双重差分模型结果

为了在包含空间溢出效应的潜在因果模型框架下识别政策处理效应和溢出效应,本文采用 Gardner(2022)提出的修正两阶段双重差分模型进行估计，估计结果如表 10 所示。表 10 中列(1)显示，在考

考虑不同距离内其他试验区的溢出效应后，政策对于试验区碳排放总量的影响变得不显著，而其他试验区对碳排放总量的溢出效应均在 1% 的显著性水平下显著。可以看出，在降低碳排放总量方面，大数据试验区政策对其他城市的溢出效应随着地理距离的增加而递减。表 10 中列(2)显示，政策对于试验区电能消耗碳排放的影响依然在 1% 的显著性水平下显著，同时对处于 100-200 公里和 200-300 公里范围内的城市产生大致相同的负向溢出效应。而试验区政策对热能消耗产生的碳排放、交通运输产生的碳排放和直接能源消耗产生的碳排放几乎不存在显著的直接效应和溢出效应。

表 10 修正的两阶段双重差分模型

被解释变量	(1)	(2)	(3)	(4)	(5)
	碳排放总量	电能消耗 产生的碳排放	热能消耗 产生的碳排放	交通运输 产生的碳排放	直接能源消耗 产生的碳排放
<i>NBDEZ</i>	-0.1215 (0.0983)	-0.3039*** (0.1133)	-0.1312 (0.1802)	-0.0441 (0.1918)	0.0970 (0.1106)
0-100 公里溢出效应	-0.0178*** (0.0067)	-0.0109 (0.0091)	-0.0132 (0.0155)	0.0097 (0.0078)	0.0137 (0.0087)
100-200 公里溢出效应	-0.0164*** (0.0050)	-0.0190*** (0.0068)	-0.0218* (0.0118)	0.0032 (0.0055)	0.0042 (0.0064)
200-300 公里溢出效应	-0.0150*** (0.0051)	-0.0195** (0.0093)	0.0098 (0.0111)	0.0003 (0.0053)	0.0028 (0.0053)
控制变量	是	是	是	是	是
N	2392	2368	1052	2442	2387

五、进一步分析

(一)工业用电和居民用电碳排放

由基本模型的实证结果可知，大数据试验区政策对电能消耗产生的碳排放具有显著的负向影响。电力消费的主要组成部分包括工业用电和居民生活用电，二者平均占比分别为 71%和 13%。因此，为了进一步探究政策的效果侧重和潜在的影响机制，本文将电能消耗产生

的碳排放细分为工业用电产生的碳排放和居民用电产生的碳排放，表 11 报告了试验区政策对二者的影响。可以看出，大数据试验区政策仅对工业用电产生的碳排放具有显著的抑制作用，而对居民用电产生的碳排放没有明显的影响。该结果表明，大数据试验区的建立现阶段主要是通过产业政策引导工业部门的生产活动，尚未直接影响到城乡居民的生活方式，进而降低了工业电力消费总量和其产生的碳排放。

表 11 试验区政策对工业用电和居民用电产生碳排放的影响

被解释变量	(1)	(2)
	工业用电产生的碳排放	居民用电产生的碳排放
<i>NBDEZ</i>	-0.4434*** (0.1556)	-0.2315 (0.1577)
控制变量	是	是
年份固定效应	是	是
城市固定效应	是	是
N	2361	2350

(二)不同试验区类型

八大国家大数据综合试验区根据区域类型可以划分为贵州试验区、跨区域类试验区、区域示范类试验区和基础设施统筹发展类试验区四种，不同类型试验区的总体定位和发展侧重不同，其降低碳排放的效果也存在一定的异质性。具体而言，贵州作为第一批试验区，其建设围绕数据资源管理与共享开放、数据中心整合、数据资源应用、数据要素流通、大数据产业集聚、大数据国际合作和大数据制度创新七大任务，全面开展系统性试验。第二批当中的京津冀试验区和珠江三角洲试验区属于跨区域类试验区，其总体定位是围绕落实国家区域发展战略，更加注重数据

要素流通，促进区域一体化发展。第二批试验区当中的上海、河南、重庆和沈阳属于区域示范类试验区，其总体定位是积极引领东、中、西和东北四大板块发展，更加注重数据资源统筹，促进区域协同发展。第二批试验区当中内蒙古属于基础设施统筹发展类试验区，其总体定位是在充分发挥区域能源、气候、地质等条件的基础上，强化绿色集约发展。各类型试验区政策对碳排放总量和电能消耗产生碳排放的影响系数如表 12 所示。可以看出，跨区域类试验区和区域示范类试验区政策都会显著降低碳排放总量，且区域示范类试验区的降碳作用更明显。而贵州试验区和内蒙古基础设施统筹发展类试验区政策对碳排放总量并没有显著的影响。另外，除内蒙古基础设施统筹发展类试验区外，其余类型的试验区政策均会对电能消耗产生的碳排放产生一定程度的抑制作用，其中区域示范类试验区的降碳效果最为明显，跨区域类试验区的降碳效果次之。

表 12 不同试验区类型异质性

被解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	碳排放总量				电能消耗产生的碳排放			
	贵州	跨区域类	区域示范类	基础设施统筹发展类	贵州	跨区域类	区域示范类	基础设施统筹发展类
<i>NBDEZ</i>	-0.1066 (0.2252)	-0.1958** (0.0957)	-0.4374*** (0.0858)	0.2267 (0.4656)	-0.2524* (0.1424)	-0.4136*** (0.1186)	-0.7186*** (0.0925)	0.2381 (0.6502)
控制变量	是	是	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是	是	是
城市固定效应	是	是	是	是	是	是	是	是
N	2148	2267	2140	2130	2123	2242	2116	2108
R ²	0.8912	0.8991	0.9015	0.8920	0.9012	0.9095	0.9080	0.8987

(三)创新环境异质性

各试验区的基础创新环境差异较大,会直接影响到产业政策的实施效果,进而对城市碳排放产生不同的影响。因此,为了探究不同基础创新环境下大数据试验区政策降碳效果的异质性,本文选取北京大学企业大数据研究中心公布的中国区域创新创业指数作为创新环境的衡量指标。具体而言,选取大数据试验区政策实施前一年,即2014年的中国区域创新创业指数人均得分构建虚拟变量 $innov$,若该城市的人均得分高于中位数则赋值为1,定义为创新环境较好;否则赋值为0,定义为创新环境较差。由于该指数的涵盖范围不包括自治州、区(直辖市),因此将这部分样本剔除之后,本文在基准模型中加入大数据试验区政策虚拟变量($NBDEZ$)和创新环境虚拟变量($innov$)的交互项,检验在不同创新环境下政策效应的异质性。创新环境对碳排放总量和电能消耗产生碳排放的异质性影响如表13中列(1)和列(3)所示。结果显示,对于创新环境较好的城市而言,试验区的建立会对其电能消耗产生的碳排放具有更明显的抑制作用,而对处于不同创新环境下的城市碳排放总量影响不存在显著差异。

$$y_{it} = \alpha + \beta_1 NBDEZ_{it} + \beta_2 NBDEZ_{it} \times innov_i + \delta X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (7)$$

(四)经济水平异质性

数字经济是以数字技术为核心驱动力量的新型经济形态,其发展进程、特征以及对碳排放的影响与当地的经济水平有关。而各试验区的经济发展水平差异较大,既有贵州和内蒙古等经济水平较低但发展潜力较大的地区,也有珠三角和上海等经济水平较高的地区。因

此, 为了探究不同经济发展水平下大数据试验区政策降碳效果的异质性, 本文选取政策实施前一年(2014 年)各试验区的人均国内生产总值构建经济水平虚拟变量 eco 。若该城市的人均国内生产总值高于中位数则赋值为 1, 定义为经济发展水平较高; 否则赋值为 0, 定义为经济发展水平较低。将大数据试验区政策虚拟变量($NBDEZ$)和经济水平虚拟变量(eco)的交互项加入到基准模型当中, 检验结果如表 13 中列(2)和列(4)所示。结果显示, 相较于经济发展水平较低的城市而言, 试验区的建立对经济发展水平更高城市具有更显著的降碳作用, 包括碳排放总量和电能产生的碳排放。该结果可能的原因是, 试验区政策在经济发展水平较低的地区侧重于先构建大数据基础设施, 而在经济发展水平较高的地区侧重于推动数实产业深度融合, 导致现阶段试验区政策对前者的降碳作用没有充分发挥出来。

$$y_{it} = \alpha + \beta_1 NBDEZ_{it} + \beta_2 NBDEZ_{it} \times eco_i + \delta X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (8)$$

表 13 创新环境和经济水平异质性

被解释变量	(1)	(2)	(3)	(4)
	碳排放总量		电能消耗产生的碳排放	
$NBDEZ$	-0.0765 (0.1221)	-0.1530* (0.0873)	-0.1737** (0.0784)	-0.3277*** (0.1033)
$NBDEZ \times innov$	-0.1503 (0.1610)		-0.2696* (0.1579)	
$NBDEZ \times eco$		-0.3737** (0.1651)		-0.4353** (0.2056)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
城市固定效应	是	是	是	是
N	2391	2391	2367	2367
R ²	0.9038	0.9046	0.9111	0.9117

(五)机制检验

根据本文的理论分析部分可以得知，大数据试验区政策会通过加速技术进步、推动消费业态和模式的创新以及促进产业数字化转型降低城市碳排放水平。因此，本文利用中介效应模型检验上述影响机制。具体而言，使用该城市当年的发明专利授权数目代表技术进步，使用人均邮政业务量代表经营和消费模式的创新，使用计算机和软件从业人员比重代表产业数字化转型，构建模型如下：

$$medi_{it} = \alpha_0 + \alpha_1 NBDEZ_{it} + \alpha_2 X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (9)$$

$$y_{it} = \beta_0 + \beta_1 NBDEZ_{it} + \beta_2 medi_{it} + \beta_3 X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (10)$$

其中，medi 为中介变量，包括发明专利授权数目、人均邮政业务量和产业数字化转型。其余控制变量与基准模型相同，检验结果见表 14。列(1)至列(3)分别展示了大数据试验区政策变量对三个中介变量的影响系数，结果显示大数据试验区的建立对发明专利授权数目、人均邮政业务量和产业数字化转型均存在正向的影响，符合理论预期。但试验区政策对发明专利授权数量的影响不具有统计显著性，而对人均邮政业务量和产业数字化转型的影响在 5%的显著性水平下显著。表明大数据试验区的建立促进了消费业态和模式的创新，推动了产业数字化转型过程，而对发明专利数没有显著的影响。列(4)和列(6)分别展示了人均邮政业务量和产业数字化转型对城市碳排放水平的影响，二者在不同的显著性水平下均为负值，表明经营和消费模式的创新和产业数字化转型具有一定的降碳作用。列(5)和列(7)分别展示了人均邮政

业务量和产业数字化转型对电能消耗产生碳排放的影响，二者均在 1%的显著性水平下显著为负，表明经营和消费模式的创新和产业数字化转型有助于降低电能消耗产生的碳排放。综上，大数据试验区的设立通过促进消费业态和模式创新、推动产业数字化转型降低了城市的碳排放总量和电能消耗产生的碳排放，验证了本文的假设 3 和假设 4。

表 14 机制检验

被解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>patent</i>	<i>postal</i>	<i>indus_digi</i>	碳排放	电能消耗产生的碳排放	碳排放	电能消耗产生的碳排放
<i>NBDEZ</i>	1.621 (1.506)	0.1624** (0.0082)	0.2829** (0.1266)	-0.1629* (0.0912)	-0.3279*** (0.1110)	-0.1648* (0.0896)	-0.3412*** (0.1108)
<i>postal</i>				-0.0001* (0.0000)	-0.0002*** (0.0000)		
<i>indus_digi</i>						-0.0519*** (0.0149)	-0.0781*** (0.0196)
控制变量	是	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是	是
城市固定效应	是	是	是	是	是	是	是
N	2413	2442	2442	2386	2360	2386	2360
R ²	0.9354	0.7035	0.7190	0.9044	0.9122	0.9045	0.9119

六、结论与政策建议

国家大数据综合试验区的设立从产业、民生、制度、技术和人力资本等多个维度全面推动数字经济发展，因此本文以试验区的建立表征数字经济的发展。通过理论分析试验区政策对碳排放水平的影响方向和影响机制，提出了本文的研究假设。在实证分析部分，将大数据试验区的设立作为一项准实验，利用交错双重差分模型和修正的两阶段双重差分模型对研究假设进行了检验，

并得到以下结论：(1)数字经济的发展会降低城市的碳排放总量和电能消耗产生的碳排放，而对热能产生的碳排放、交通运输产生的碳排放和直接能源消耗产生的碳排放没有显著影响。且该结论在经过安慰剂检验、反事实检验、倾向评分匹配—双重差分模型、改变模型时间设定、控制其他政策的影响和更换估计方法等一系列稳健性检验后依然成立。(2)大数据试验区的设立可以降低周围城市的碳排放总量和电能消耗产生的碳排放水平，且该溢出效应会随着地理距离的增加而递减。而试验区的设立对周围城市其他方式产生的碳排放几乎不存在显著的溢出效应。(3)根据进一步分析可知，试验区设立对电能消耗产生碳排放的抑制作用，主要是通过减少工业用电产生的碳排放实现的。而试验区的设立现阶段对居民用电产生的碳排放没有显著的影响。(4)大数据试验区政策的降碳效应在不同试验区类型、不同创新环境和不同经济水平下存在一定程度的异质性。具体而言，区域示范类试验区的降碳效果最显著，跨区域类试验区的降碳效果次之，贵州试验区的降碳效果最低。内蒙古基础设施统筹发展类试验区的设立呈现出一定的增碳效应，但并不具有统计显著性。另外，试验区的设立在创新环境更好、经济发展水平更高的城市呈现出更大的降碳效果。(5)机制检验结果表明，试验区的设立会通过促进消费业态和模式创新、推动产业数字化转型降低城市的碳排放总量和电能消耗产生的碳排放水平。基于上述结论，本文提出以下政策建议：

首先，基于大数据试验区的实践经验，全面推动数字经济发展，实现经济社会发展绿色转型。根据本文的研究结论可知，数字经济作为一种具有巨大增长潜力的新型经济形态，其发展可以显著降低二氧化碳排放量。因此，为了深度挖掘数据价值，全面释放数字经济的降碳潜力，应当加快数据空间和信息基础设施建设、构建多样化多层次的数据要素市场、完善数据产权制度和数据治理体系。同时，充分发挥大数据试验区的辐射和带动作用，并将其可复制的实践经验应用、推广到其他地区，实现数据红利的共享和碳减排范围的扩大。

其次，大力推进产业数字化深层次转型，引导企业进行数字化智能改造。本文的研究结果显示，工业用电碳排放的减少是试验区碳排放水平降低的直接原因，其主要是通过消费模式和业态的创新、产业数字化转型实现的。由于大数据本身的能耗特性和新型基础设施的建设需要大量的电力支持，即数字产业自身的发展可能会带来碳排放的增加，但其数字技术赋能到其他产业会带来碳排放的降低。因此，为了更好地发挥大数据技术的效率提升作用，充分释放各领域各产业的降碳潜力，应当强化财政专项资金统筹、引导制造业企业数字化升级、不断提升企业的数字化能力。

最后，根据不同地区的资源禀赋和发展特点制订差异化的数字经济发展战略，在推动数字普惠的过程中最大化各区域的发展优势。本文研究结论显示，数字经济发展的降碳效应在不同试验区类型、不同创新环境和不同经济水平下呈现出一定的异质性。因此，各地区应当根据自身发展特征选取适宜的数字经济发展战略，而不应盲目复制降

碳效果显著地区的发展战略。特别是对于侧重新型基础设施统筹发展的区域，其现阶段的碳排放可能会有所增加，但以兼顾全局和长期的视角来看，其仍然具有较大的碳减排潜能。