DOI:10.16339/j.cnki.hdxbskb.2025.02.008

数据要素开发与经济增长新动能
——基于国家级大数据综合试验区设立的经验证据*

陈泳昌[1],林迎蕾[2]

(1.湖南大学 经济与贸易学院,湖南 长沙　410079;2.南开大学 经济学院,天津　300110)

[摘　要]通过匹配合并2013—2020年国家级大数据综合试验区的试点数据、《中国城市统计年鉴》数据以及夜间卫星灯光数据等,实证研究了数据要素开发对地区经济增长的影响效应和作用机制。研究发现:以大数据综试区设立为表现形式的数据要素开发会显著促进各城市的经济增长,且这一结论在经过一系列稳健性检验后依然显著成立。影响机制分析发现,数据要素开发主要是通过创新驱动和数字基建双重路径来为地区经济增长构筑新动能。异质性分析发现,数字服务供给能力越强、政府对数字经济关注度越高的地区越能实现经济增长,在新冠疫情冲击后数据要素开发带来的经济增长效应更突出。拓展性分析表明,相对于“先富”地区而言,数据要素开发主要促进了“后富”地区的经济增长,且数据要素开发还将有效缩小城乡之间的收入差距。

[关键词]数据要素开发;经济增长;创新驱动;数字基建;收入差距;大数据

[中图分类号]F061.2　　[文献标识码]A　　[文章编号]1008-1763(2025)02-0068-11

Development of Data Factor and New Driver of Economic Growth: Empirical Evidence Based on the Establishment of National Level Big Data Comprehensive Pilot Zones

CHEN Yongchang[1],LIN Yinglei[2]

(1. School of Economics and Trade,Hunan University,Changsha　410079,China;
2. School of Economics,Nankai University,Tianjin　300110,China)

Abstract:By matching and merging the pilot data of national level big data pilot zones from 2013 to 2020, *the China Urban Statistical Yearbook*, and nighttime satellite lights, we empirically investigated the impact effect and mechanism of data factor development on regional economic growth. This paper finds that data factor development in the form of setting up comprehensive pilot zones significantly promotes the economic growth of cities, and this conclusion still holds significantly after a series of robustness tests. Mechanism analysis reveals that the development of data factor mainly constructs new momentum for regional economic growth through the dual paths of innovation-driven and digital infrastructure. Heterogeneity analyses find that the stronger the supply capacity of digital services and the higher the government's attention to the digital economy, as well as the more prominent economic growth effect from data factor development after the shock of the COVID-19 pandemic. Expansion analyses show that compared with the “first-rich” regions, data factor development mainly promotes the economic growth of the “second-rich” regions, and data factor development will also effectively reduce the income gap between urban and

* [收稿日期]2024-07-05
[基金项目]国家社会科学基金重大项目:新发展格局下中国产业链供应链安全稳定战略研究(21&ZD098);国家自然科学基金面上项目:全球疫情大流行下国际国内价值链重构对中国的影响和应对策略(72073025);中央高校基本科研业务费项目(531118010985)
[作者简介]陈泳昌(1997—),男,河南周口人,湖南大学经济与贸易学院助理教授,博士,研究方向:平台经济与数字贸易。

rural areas.

Key words：data factor development；economic growth；innovation drive；digital infrastructure；income gap；big data

一　引　言

在当前全球经济形势复杂多变与我国经济增长进入新常态的双重背景下，党的二十届三中全会明确提出，塑造发展新动能新优势。那么，如何深入挖掘新型生产要素的增长潜能进而构筑支撑经济高质量发展的长效驱动便成为社会各界需要深入思考的重大现实议题。近年来，我国加快推进实体经济和数字经济的深度融合，数据要素已经成为与土地、劳动、资本和技术等传统生产要素并列的全新要素。根据全国数据工作会议的最新统计，2023 年我国数据生产总量预计超过 32ZB，我国成为全球重要的数据大国之一①。那么，蓬勃发展的数据要素是否能够成为驱动经济持续增长的新型动能？为回答这一问题，本文将以近年来我国加快数据要素开发的重大试点举措——国家级大数据综合试验区（以下简称大数据综试区）的设立为准自然实验，从理论和实证两个层面深入考察数据要素开发对经济增长的影响。这一研究不仅能为我国加快构建和完善数据基础制度提供有益的理论支持，同时也能从数据要素视角探索推动我国经济持续高质量攀升的现实路径。

与本文研究较为相关的一支文献考察了我国在数字经济领域的各类试点政策带来的经济效应。一方面，基于宽带中国示范城市试点政策，方福前等通过匹配“宽带中国”试点政策信息与中国家庭追踪调查数据发现，数字基础设施能够通过破解教育资源以及高水平社会资本对农村户籍子代的约束，实现对劳动力供给侧的数字化赋能[1]。另一方面，从跨境电商综合试验区的视角，马述忠和郭继文基于国内领先跨境电商物流企业 2016 年 9 月至 2019 年 8 月的运单数据发现，大数据综试区的设立使城市内各行业的跨境电商出口平均增长了 21%[2]。此外，也有研究详细考察了电子商务进农村综合示范政策[3]及自贸试验区建设[4]等带来的影响。而聚焦到大数据综试区设立的影响来看，学者们发现该举措有助于提高生产效率[5]，加快创新[6]以及促进企业数字化转型[7]。但这些文献并未深入考察大数据综试区与区域经济增长的关系，尤其缺乏从增长动能构筑的视角探讨其中的具体作用机制的研究，本文将对此做出探讨。

国内外研究也围绕各类数字技术的经济效益展开了深入分析。其中，国外的研究发现，无论是互联网[8]、宽带[9]等数字基础设施建设，还是移动手机的普及和应用[10]，都会显著带动地区内的经济增长。国内学者同样利用互联网[11]、电信基础设施[12]以及人工智能[13]等数据做出了广泛分析，总体上认为数字技术的发展有利于地区经济增长。如郭峰等利用夜间卫星灯光数据来衡量地区经济发展情况，并发现数字经济的渗透可以通过克服地理障碍和缓解行政分割来促进行政边界地区的经济发展[14]。与本文研究最相关的文献初步从理论层面分析了数据要素投入及大数据技术应用对经济增长的影响。徐翔和赵墨非采用产业组织的创新模型为数据资本构建了微观基础，并将其引入内生增长模型，发现模型存在非平衡的稳态增长路径，其中数据资本的稳态增速高于其他类型资本及总产出的稳态增速[15]。杨俊等则认为大数据与其他生产要素存在“融合成本”，导致“研发模式转型”抑制短期经济产出，但长期来看，大数据会通过“乘数作用”提升中间品质量并促进技术进步，持续推动经济增长[16]。但综合这些研究来看，其大多局限在理论模型的推导层面，缺乏从实证维度做出详细的政策评估，尤其是缺少对我国代表性的大数据试点政策进行聚焦分析，这为本文留下了进一步拓展的空间。

综合以上文献梳理，本文将以大数据综试区的设立为例，通过匹配合并 2013—2020 年大数据综试区的试点数据、《中国城市统计年鉴》以及夜间卫星灯光等数据，实证研究数据要素开发对地区经济增长的影响效应和作用机制。相较于已有文献，本文可能存在以下边际贡献。第一，研究视角上，在我国经济处在新旧动能转换的关键节点下，如何充分发挥数字经济优势，构筑支持我国经济持续健康增长的全新动能至关重要。为此，本文从数据要素开发视角展开分析，探索以创新驱动和数字基建为关键路径的增长新思路，从而丰富已有考察经济增长决定因素的前期研究。第二，研究数据上，以往文献主要使用地区人均生产总值来测度经济增长水平[17]，本文则以卫星灯光来反映地区的经济发展情况[18]。该方法能够较好地规避由于地区间价格差异对估计

① 2023 年，我国数据生产总量或将超过 32ZB，https://www.gov.cn/lianbo/bumen/202404/content_6943184.htm.

结果带来的潜在干扰,同时也能更加充分地反映市场提供的商品和服务价值。第三,研究内容上,文章不仅讨论了数据要素开发是否会影响地区经济增长,更是将研究从"是什么"拓展到了"为什么"的维度,从创新驱动和数字基建双重视角深入考察了其中的内在作用机制。此外,本文还从地区数字服务商发展情况、政府数字关注度以及新冠疫情影响这三个视角开展了异质性分析。最后,文章进一步从共同富裕的角度进行探讨,通过区分"先富"地区和"后富"地区以及关注城乡收入差距,从两个方面分析了大数据综试区对地区收入差距的影响。

二　政策背景与理论分析

(一)政策背景

数字经济正在成为重组全球要素资源、重塑全球经济结构、改变全球竞争格局的关键力量。为抢占发展先机,许多发达经济体已将大数据技术开发与利用上升为国家战略。例如,美国在2011年发布了《大数据研究与发展计划》,澳大利亚在2013年启动了《公共服务大数据战略》,法国和英国也在同年出台了大数据相关政策。为此,我国在2014年将"大数据"一词写入《政府工作报告》,并于2015年8月出台《促进大数据发展行动纲要》(以下简称"《纲要》"),这些措施对我国大数据产业进行了顶层设计和总体布局,并将大数据提升到国家战略的高度。

为贯彻落实《纲要》规划,贵州省于2015年9月启动了全国首个大数据综试区建设工作。随后,2016年10月,第二批大数据综试区名单发布,包括京津冀、珠三角两个跨区域类大数据综试区,上海市、河南省、重庆市、沈阳市四个区域示范类大数据综试区,以及内蒙古自治区一个大数据基础设施统筹发展类大数据综试区。目前,国内共计建设了八个大数据综试区,这将共同引领东部、中部、西部和东北地区的大数据产业协同发展,释放数据要素活力,为促进我国经济增长提供了扎实支持。

(二)理论分析

1.数据要素开发对经济增长的影响

大数据综试区在大数据创新应用、大数据产业集聚发展、大数据资源流通与交易、大数据国内外交流合作等诸多方面先行先试,极大地促进了数据要素的开发与共享。数据要素目前已成为与土地、劳动力、资本和技术并列的五大基础性生产要素,广泛渗透到生产、分配、流通、消费各种经济活动中,为经济增长注入新活力。从数据要素本身特征来看,相较于其他传统生产要素,数据具有非竞争性、非排他性/部分排他性、不易复制性、外部性和即时性等特征。这些特征使得数据要素对经济增长的提升效应不再受到边际报酬递减规律的限制,即数据要素在初期投入经济系统会存在高昂的投入成本,但是在后期会呈现低边际成本特征,即一旦突破门槛就会显示其巨大经济效益[19]。因此,数据要素对经济增长的长期效应更为深远。

将数据要素特征纳入宏微观经济环境来看,在微观层面,数据要素可以通过其本身承载的高价值信息来提高劳动、资本等其他要素之间的协同性,进而使微观运行效率得以提升,加速经济增长。具体而言,将数据要素中的有效信息提取出来可以加强企业分析能力,引导企业精准投资,降低企业运行的不确定性,使得微观主体的运行兼备效率与稳健。进一步,数据要素的外部性特别是网络外部性,能够通过规模经济和范围经济实现更多的数据积累和有价值信息提取,在一定阶段内加速微观主体运行效率的提升。特别是在数据驱动的双边或多边服务平台中,基于网络外部性形成的规模经济和范围经济尤为明显,特定阶段数据要素在效率提升和价值创造方面也会出现边际报酬递增的情形[20]。

在宏观层面,数据要素的独有特征将会放大,成为挖掘经济增长潜力的重要途径。例如,数据要素的非竞争性、非排他性同低成本复制特征相叠加,使得公共产品经常出现的"公地悲剧"转换为有利于经济增长的"公地喜剧"。具体而言,数据要素可以低成本无限复制,从而实现同时在多个微观应用场景中与其他各种要素进行组合、重复使用,使其在宏观层面的价值创造能力实现倍增。此外,数据要素的这些特征,同时以数据为基础的新模式、新业态、新服务等在全社会范围内推广并得到涌现,进而提高市场整体运行效率,并作用于经济增长[21]。

数据要素对经济增长的作用也得到现有研究的广泛认可。许宪春等认为数据要素相对于传统要素对经济增长的直接影响和推动作用更为明显[22]。Basu和Fernald指出数据要素开放和共享可以优化资源配置、减少信息不对称性,从而提高市场的运行效率[23]。将数据要素纳入内生增长模型来看,杨俊等[16]认为数据要素可以通过"乘数作用"提升中间品质量并促进技术进步,长期加速经济增长。袁健和段巍也采用类似的方法进一步验证了数据要素的经济效益[24]。综合以上分析,我们提出本文的假说1。

假说1:在其他条件不变的情况下,大数据综试区的设立将会促进地区经济增长。

2.数据要素开发影响经济增长的机制路径——

创新驱动与数字基建

大数据综试区的设立将从三个方面促进地区的创新水平。第一，大数据综试区设立的目标之一就是开放共享数据资源，加速大数据资源流通，便利大数据资源交易。伴随数据要素的流通加快和开放透明，创新活动中面临的信息不对称问题将得到缓解，这将通过优化资源配置来实现区域创新能力的提升。第二，与大数据综试区设立相伴出台的一系列财税、人才等政策也为区域创新注入了动力。就财税政策而言，如贵州大数据综试区在《关于加快大数据产业发展应用基于政策的意见》中提到，“被认定为高新技术企业的，减按 15% 征收企业所得税，并给予一次性 30 万元的补助”①。就人才政策而言，大数据综试区内的鼓励人才引进政策，为创新活动提供关键主体。如，贵州省自 2015 年以来陆续发布《关于加快大数据产业人才队伍建设的实施意见》等一系列人才培养与引进政策。第三，大数据综试区内的产学研合作也有利于推动区域创新。产学研合作过程中高校和科研院所以基础研究为重点，企业以产品开发为重点，三者之间具有较强的互补性和异质性，这有利于主体之间资源优势互补，增加多样化的学习和交流，不同领域的知识融合与知识互动，有利于创新灵感的迸发，从而加大创新产出的可能性。

创新是实现经济增长的关键驱动，是促进经济增长的直接原因[25]。创新可以通过两个方面来驱动经济增长。第一，企业对于大数据综试区内创新成果的应用，一方面可以提高企业的生产能力，提高企业产出与产品多样性，从而带来企业收入的增加；另一方面可以助力企业改造升级，带来更加先进的生产设备，有效地降低生产过程中的资源损耗和额外成本，扩大企业盈利空间。企业又是国民经济的细胞，其经营状况与国民经济紧密相关，企业经济效益的提升可以为我国经济增长注入活力。第二，大数据综试区内的创新活动将会带来技术进步，技术的突破则会极大地提升企业生产率，这是决定企业盈利能力的重要因素，生产率越高的企业往往利润水平越高，其支付的工资水平也就越高[26]。综合以上分析，我们提出本文的假说 2。

假说 2：在其他条件不变的情况下，大数据综试区的设立可以通过提高区域创新能力来促进经济增长。

大数据综试区的设立有效加快了地区的数字基建水平，截至 2021 年，贵州省全年新建 5G 基站 2.7 万个，累计建成超 5 万个；全年新增互联网出省宽带 1.1 万 G，累计达 2.8 万 G②。与传统基建不同，数字基建是指以 5G、大数据中心、人工智能、工业互联网等信息技术为代表的新型基础设施[27]，被视为数字经济发展的基石与底座。地区的数字基建会为重塑经济结构，推动经济转型升级，并为实现高水平增长搭建新机制、创造新契机，进一步促进经济增长。第一，数字基建本身就是大量的固定资产投资，将通过投资带来的乘数效应推动地区经济增长。数字基建伴随电信光缆、互联网基站等大范围的基础设施投资，将直接为经济增长带来新动能；此外，数字基建的不断完善会吸引更多的社会资本，带来配套产业投资的增加，促进社会资本的跨部门跨区域流动，进一步扩大投资的辐射范围，并在整个经济运转中发挥乘数效应，带来更高水平的经济增长。第二，数字基建有助于激发数据要素活力，通过数据要素赋能为地区经济探索增长新空间。数字基建不仅为地区发展数字经济建立了互联网、宽带等最基础的技术底座，同时也打造了数据交易平台等新型数字基础设施。例如，2022 年 4 月到 2023 年 5 月，贵阳大数据交易所内已经聚集了 577 家数据商、上架了 852 个数据产品、累计交易额超 10 亿元③。数据交易平台的搭建，一方面能提供集中式、规范化的数据交易场所和服务，为数据要素的使用提供良好的外部环境。另一方面，数据交易平台涵盖了包括大数据整合、交换、共享等多方面的数据服务，有助于打破数据孤岛，促进数据要素流通，这对于激发数据要素带来的经济增长活力至关重要。第三，数字基建还能通过优化产业结构来促进经济增长。数字基建能够为相关产业发展提供数字化、信息化及智能化的服务。基于数字基建的各类数字技术与传统产业的融合与集成应用，能有效提高传统产业的研发、生产、交易效率，加速传统产业的改造升级，甚至颠覆传统产业现有生产模式。传统产业的改造升级及新兴产业的培育壮大对经济增长的重要性已得到现有研究的广泛证实[28]。

假说 3：在其他条件不变的情况下，大数据综试区的设立可以通过完善数字基建来促进经济增长。

---

① 资料来源：《贵州省人民政府印发〈关于加快大数据产业发展应用若干政策的意见〉〈贵州省大数据产业发展应用规划纲要（2014—2020 年）〉的通知》，http://www.guizhou.gov.cn/zwgk/zcfg/szfwj/qff/201709/t20170925_70477119.html.

② 数据来源：《国家大数据（贵州）综合试验区发展报告 2021》，http://xxzx.guizhou.gov.cn/zxzk/zkcs/202205/t20220525_74213016.html.

③ 资料来源：“数实相融 算启未来”贵阳大数据交易所：勇闯数据新蓝海 构建数据交易新体系，https://k.sina.com.cn/article_5904554912_15ff05ba002001l0mg.html.

三 实证方案设计

(一)计量模型设定

1. 基准模型构建

为从实证层面有效识别大数据综试区的设立与地区经济增长的因果关系,本文使用双重差分法(difference-in-differences,DID)进行回归分析。考虑到各城市被设立为大数据综试区的时间不同,本文具体构造如下多时点DID模型进行分析:

$$Econ_{it} = \alpha_0 + \alpha_1 Policy_{it} + \alpha_2 CX_{it} + \eta_i + v_t + \varepsilon_{it} \tag{1}$$

式中,i、t分别表示城市和年份。$Econ_{it}$表示i城市在第t年的经济增长水平,本文以各城市的夜间卫星灯光指数进行测度,下文将具体阐述该指标的构建思路。$Policy_{it}$表示i城市在第t年是否被设立为大数据综试区。CX_{it}表示一系列可能影响估计结果的城市层面控制变量。η_i和v_t分别表示城市固定效应和年份固定效应,ε_{it}为残差项。回归将标准差聚类到城市—年份层面,以控制残差在城市—年份层面的相关性。

2. 机制检验模型构建

本文参考江艇的研究[29],采用如下模型检验核心解释变量大数据综试区对机制变量的影响,并基于已有理论或者文献来论述机制变量对被解释变量经济增长的影响。

$$Mech_{it} = \alpha_0 + \alpha_1 Policy_{it} + \alpha_2 CX_{it} + \eta_i + \nu_t + \varepsilon_{it} \tag{2}$$

式中,$Mech_{it}$表示本文的机制变量,分别指创新驱动和数字基建这两类指标,具体构建方式将在机制分析中进行详细介绍。其余变量与基准模型保持一致。

(二)指标构建

1. 被解释变量($Ecno$)

本文采用夜间卫星灯光数据来反映地区的经济增长水平。相较于已有研究常采用的GDP数据,夜间卫星灯光数据的优势在于:第一,剔除了价格因素的干扰。直接采用国家统计局公布的GDP数据研究经济增长问题,需要对地区间的价格偏差做出有效处理,但由于部分地区的价格指数缺少或未统计,往往无法有效地剔除价格因素带来的干扰,而夜间卫星灯光数据避免了这一误差[18]。第二,GDP数据仅包括能够被直接统计的商品和服务交易,但现实中并非所有的经济活动都采用物物交换的方式进行,也有部分无法以价格形式表现的经济活动。夜间卫星灯光数据则能较为全面地反映地区总体的经济活动,因此能更加真实地反映地区的经济发展状况。

本文所使用的夜间卫星灯光数据来自美国军事气象卫星,其能探测到城市灯光甚至小规模居民地、车流等发出的低强度灯光,且该数据消除了云及火光等偶然噪声影响。将历年全球灯光影像数据重投影转换为兰伯特等角圆锥投影,并用中国行政边界进行裁剪,得到年空间分辨率为1千米,可用于本文研究的中国行政区内的非辐射定标夜间卫星灯光数据。具体来说,本文采用各地级市的平均夜间灯光亮度来作为被解释变量。

2. 核心解释变量($Policy$)

本文选取大数据综试区这一虚拟变量作为核心解释变量。倘若某年某一城市被设立为大数据综试区,那么该城市在当年及随后年份赋值为1,其他年份赋值为0,该变量实际上是大数据综试区分组变量与政策时间虚拟变量的交互项。

3. 其余控制变量

参考郭峰等对地区经济增长决定因素的研究,本文分别从企业层面和城市层面选取控制变量[14]。(1)经济发展水平($Agdp$):使用城市层面人均GDP取对数来表示;(2)科学支出水平(Sci):采用各城市的科学支出费用与GDP比值来测度;(3)教育支出水平(Edu):采用各城市的教育支出费用与GDP比值来测度;(4)地区金融发展水平($Finance$):以城市金融机构的存款余额与贷款余额比值来衡量;(5)市场化水平($Market$):以《中国分省份市场化指数报告》中的市场化指数来衡量;(6)高铁开通情况($Trans$):开通高铁的城市当年及以后赋值为1,其他城市赋值为0。

4. 数据处理及描述性统计

本文研究数据包括以下三个方面:第一,美国军事气象卫星公布的夜间卫星灯光数据;第二,大数据综试区名单数据;第三,《中国城市统计年鉴》数据。在将上述三项数据逐一匹配并清洗后,本文得到了2013—2020年全国288个城市的研究数据。数据的描述性统计如表1所示。

四 实证结果与分析

(一)基准回归结果

基于以上模型设定,本文使用双重差分法来实证考察大数据综试区设立对地区经济增长的影响,并将估计结果呈现在表2中。其中,第(1)列不加入任何控制变量,仅控制城市固定效应,以处理城市层面的不可预测因素干扰,结果表明,大数据综试区会

显著促进城市经济增长。为进一步控制时间维度的潜在干扰因素，我们在第(2)列中加入年份固定效应，双向固定效应的估计结果依然支持大数据综试区带来的增长效应。为处理遗漏变量导致的内生性问题，本文在第(3—4)列同时加入经济发展水平、科学支出水平、教育支出水平、地区金融发展水平、市场化水平及高铁开通情况六项控制变量，结果同样显示，核心解释变量始终在 1%的水平上显著为正，即大数据综试区的设立的确有助于促进地区的经济增长。综上所述，本文假说 1 得以验证。

表 1　变量描述性统计

变量	样本量	均值	标准差	最小值	最大值
Ecno	2257	1.157	2.155	0.001	22.180
Policy	2257	0.111	0.314	0.000	1.000
Agdp	2257	10.770	0.545	9.532	12.070
Sci	2257	0.029	0.025	0.003	0.141
Edu	2257	0.037	0.020	0.014	0.127
Finance	2257	1.521	0.406	0.790	2.951
Market	2257	8.327	1.652	3.580	11.620
Trans	2257	0.640	0.480	0.000	1.000

表 2　基准回归结果

变量	(1)	(2)	(3)	(4)
Policy	0.509***	0.219***	0.316***	0.193***
	(0.06)	(0.06)	(0.06)	(0.06)
Agdp			0.406***	−0.096
			(0.08)	(0.09)
Sci			4.275***	3.484***
			(1.11)	(1.07)
Edu			6.521***	−3.316*
			(1.57)	(1.83)
Finance			−0.237***	−0.075
			(0.06)	(0.05)
Market			0.144***	0.012
			(0.03)	(0.02)
Trans			−0.106***	−0.181***
			(0.03)	(0.03)
城市固定效应	YES	YES	YES	YES
年份固定效应	NO	YES	NO	YES
N	2257	2257	2257	2257
R^2	0.961	0.970	0.967	0.971

注：括号内为城市一年份层面的聚类稳健标准误，***、** 和 * 分别表示 1%、5%和 10%的显著性水平，下表同。

(二)有效性评估

1. 平行趋势检验与动态效应分析

采用双重差分法对设立大数据综试区进行效应评估的重要前提是要满足平行趋势假设。因此，本文借鉴 Beck 等的做法[30]，采用事件研究法进行平行趋势检验，设定如下计量模型：

$$Econ_{it} = \alpha_0 + \sum_{k=-3,k\neq-1}^{4} \alpha_{1k} Policy_{i,t_{0+k}} + \alpha_2 CX_{it} + \eta_i + \nu_t + \varepsilon_{it} \tag{3}$$

式(3)中，t_0 代表设立大数据综试区当年时点，$Policy_{i,t_{0+k}}$ 代表设立大数据综试区前后第 k 年的虚拟变量。当处理组城市 i 在第 t_{0+k} 年被设立为大数据综试区时 $Policy_{i,t_{0+k}}$ 取值为 1；否则为 0。其余设定与基准模型保持一致。检验结果如图 1 所示，我们按照基准回归的四列模型设定依次开展有效性评估。以 Model 4 为例，可以发现，在大数据综试区设立之前，处理组和对照组中的城市经济发展水平不存在显著差别，满足双重差分模型所要求的平行趋势假设。这表明大数据综试区的设立的确显著促进了地区的经济增长。此外，从政策实施后的动态效应来看，该政策效应具有较好的持续性，在政策实施后的四年时间内 *Policy* 系数均显著为正。

2. 随机抽样安慰剂检验

为进一步确保基准回归结果的可靠性，本文同时构造虚假的处理组和政策冲击时点开展随机抽样安慰剂检验。具体而言，我们根据被设立为大数据综试区的城市数量，对所有城市进行不重复随机抽样，将抽中的城市设定为虚假的处理组；同时根据大数据综试区实施的时间，随机生成虚假的政策冲击年份，最后根据所构建的虚拟处理组进行估计。在此基础上，重复上述过程 500 次。根据处理结果①，排除了伪回归的可能性，佐证了本文研究结论的可靠性。

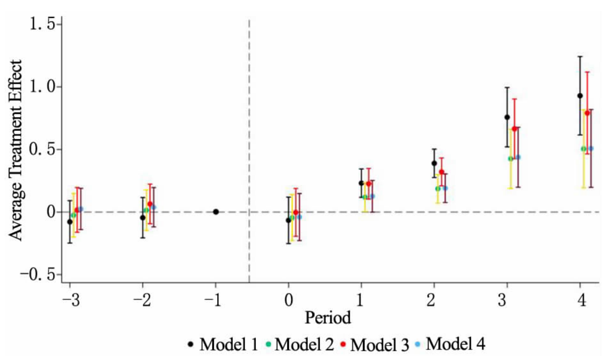

图 1　平行趋势检验结果

① 随机抽样安慰剂检验结果，因篇幅所限未汇报，感兴趣读者可向作者索取。

3. PSM_DID 检验

由于国家在设立大数据综试区的过程中,可能会出于经济发展水平、产业结构以及科学技术等因素的考虑,而优先将某些城市设立为大数据综试区,这或将导致文章的估计结果存在自选择偏差。因此,本文进一步采用倾向得分匹配法(propensity score matching,PSM),为每一个被设立为大数据综试区的城市匹配一个除了获批大数据综试区之外其他特征高度相似的匹配城市。在此基础上,进一步对匹配样本采用双重差分模型进行检验。首先,选取本文的控制变量作为协变量,采用 Logit 模型估计出倾向得分值,即各城市成为大数据综试区的概率,并根据倾向得分值使用近邻匹配法进行匹配,使得控制组与处理组两组样本具有相似特征,达到类似于随机实验的效果;其次,对匹配后的样本进行平衡性检验,确保匹配后的两组样本除了在经济发展水平方面存在差异,其他方面不存在系统性偏差;最后,在匹配后的样本中重新估计大数据综试区设立对地区经济增长的影响。估计结果进一步证明了大数据综试区设立有利于地区经济增长的结论是稳健的①。

4. 处理霍桑效应

DID 估计成立的前提条件之一,是在政策实施前处理组不存在政策预期,否则将导致估计结果存在偏误。因此,本文参考方福前等的做法[1],在基准回归的基础上加入大数据综试区分组变量与政策实施前一年虚拟变量的交互项,回归结果如表 3 所示。其中,*Policy* 系数仍然显著为正且交互项的系数不显著,可以排除霍桑效应的存在。此外,借鉴 Carballo 的研究[31],本文还分别对 2015 年和 2016 年设立的大数据综试区构建两项虚拟变量来控制霍桑效应,如贵州于 2015 年设立大数据综试区,那么虚拟变量 *Policyb*1 仅对 2014 年的贵州省赋值 1,其余情况赋值 0。类似的,2016 年设立的大数据综试区以虚拟变量 *Policyb*2 来进行控制。由于旨在考察设立大数据综试区对没有设立大数据综试区的地区经济增长的影响,因此,我们在回归中剔除 *Policy*=1 的结果,估计结果如表 3 第(2)列所示,两项虚拟变量均不显著,即并未表现出霍桑效应。

5. Goodman-Bacon 分解检验

在应用多时点 DID 时,研究样本可以根据是否处理和处理早晚划分为四种类型:早处理、晚处理、总是处理以及从未处理。而一项回归在早处理样本和总是处理样本作为控制组时,可能会由于估计结果中已包含了处理效应而导致双向固定效应估计量存在偏误。对此,本文借鉴 Stevenson 和 Wolfers 的做法[32],对双向固定效应估计量进行分解,结果如表 4 所示。根据结果可知,“晚处理 VS 早处理”这一类 2×2 DID 所占权重为 0.008 即 0.8%,这意味着对最终双向固定效应估计量的影响不大。与此同时,“时变处理 VS 从未处理”这一类所占权重最大(98.9%),对双向固定效应估计起主导作用。综上所述,本文的研究结论较为稳健。

表 3 处理霍桑效应的检验结果

变量	(1)	(2)
Policy	0.201*** (0.07)	
分组变量×前一年虚拟变量	0.367 (0.07)	
*Policyb*1		−0.026 (0.03)
*Policyb*2		−0.019 (0.03)
控制变量	YES	YES
城市固定效应	YES	YES
年份固定效应	YES	YES
N	2257	2006
R^2	0.971	0.968

表 4 Goodman-Bacon 分解

组别	权重	平均 DID 估计系数
早处理 VS 晚处理	0.003	0.017
晚处理 VS 早处理	0.008	0.355
时变处理 VS 从未处理	0.989	0.218

(三)其他稳健性检验

第一,为了控制影响地区经济增长的不可预测因素,同时避免主要控制变量带来的“维度诅咒”以及估计有偏问题,并考虑到在地区经济发展的过程中,变量间可能存在非线性关系,本文通过更换双重机器学习模型来进行稳健性检验。第二,通过控制同时期发布的“宽带中国”战略以及跨境电商综合试验区设立政策来排除其他数字经济政策对经济增长效应的干扰。第三,考虑到房地产行业在过去 20 年间对我国经济发展的特殊影响,我们将城市房地产行业的增加值取对数作为额外控制变量加入回归,以此进行控制。第四,相对其他城市而言,直辖市的经济发展基础较好,资源人才集聚水平较高,这些城

① PSM_DID 检验结果,因篇幅所限未汇报,感兴趣读者可向作者索取。

市可能会存在较强的线性增长趋势。为处理这一问题，我们将北京、上海、天津和重庆四个直辖市内的企业样本剔除，采用余下数据开展稳健性检验。第五，政策出台与市场的实际感知往往存在一定的时间滞后性，即大数据综试区的设立或许并非会在当期便对地区的经济增长产生影响，而是随着时间的推移，政策效果才逐渐显现。因此，我们将核心解释变量分别滞后一期和两期处理，以此考虑政策的滞后效应。以上稳健性检验结果如表 5 所示，均表明本文的基准回归结果是可靠的。

表 5　其他稳健性检验结果

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	双重机器学习	控制同期数字经济政策		控制地产业	剔除直辖市	考虑政策滞后效应	
		宽带中国	跨境电商综试区			滞后一期	滞后两期
Policy	0.230***	0.190***	0.140**	0.197***	0.121**	0.326**	0.378**
	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.13)	(0.15)
控制变量	YES	YES	YES	YES	YES	YES	YES
城市固定效应	YES	YES	YES	YES	YES	YES	YES
年份固定效应	YES	YES	YES	YES	YES	YES	YES
N	2257	2257	2257	2257	2225	1892	1595
R^2	—	0.971	0.973	0.971	0.970	0.979	0.991

（四）异质性分析

尽管上文从多角度验证了大数据综试区的设立对地区经济增长的积极影响，但依然缺少分析这一影响所表现出的异质性特征。因此，本节将依次从城市数字服务商发展情况、政府对数字经济关注度以及新冠疫情冲击的影响这三个方面来进行探讨。

1. 数字服务商

地方数字经济的全面发展及经济效益释放离不开各类数字服务商的协同发力，即这些服务商可能会作为数字技术主导企业的上下游配套，从而支撑其更好地释放经济效力。因此，我们基于工商登记数据获取了各城市归属于信息技术和软件服务业的初创企业信息，并根据企业的成立年份和所在城市进行加总，得到全国各个城市的数字服务商数据，将其作为回归指标，结果如表 6 第(1)列所示。结果表明，在数字服务商聚集多的城市，大数据综试区这一政策更能促进区域经济增长。

2. 数字经济关注度

政策支持是产业发展的关键支撑，数据要素的开发利用同样离不开政府关注和支持。因此，我们参考现有研究[33−35]，借助 Python 软件对政府工作报告进行分词处理，统计了省级政府工作报告中与数字经济相关的关键词词频，采用该词频占政府工作报告词频总数的比重来衡量政府对数字经济的关注程度。回归结果如表 6 第(2)列所示，政府对数字经济关注度更高的情况下，大数据综试区带来的经济增长效果会更显著。

3. 新冠疫情冲击

新冠疫情暴发期间，数字技术为克服疫情冲击、推动复工复产做出了重要贡献。因此，我们进一步考察大数据综试区带来的增长效应是否会在疫情冲击发生后更加突出。本文采用指标 *Epidemic*，对 2019 年及之后的样本取值为 1，在此之前的样本取值为 0，并以交互项的方式进行实证检验。估计结果如表 6 第(3)列所示，交互项系数显著为正，表明在新冠疫情暴发后，大数据综试区的设立更能促进城市经济增长。

表 6　异质性分析结果

变量	(1) 数字服务商	(2) 数字经济关注度	(3) 新冠疫情冲击
Policy	0.039	0.041	0.290***
	(0.06)	(0.08)	(0.06)
Information	0.675**		
	(0.29)		
Policy×*Information*	0.626*		
	(0.36)		
Goverment		−0.139	
		(0.10)	
Policy×*Goverment*		0.362*	
		(0.20)	
Epidemic			0.205***
			(0.06)
Policy×*Epidemic*			0.283*
			(0.17)
控制变量	YES	YES	YES
城市固定效应	YES	YES	YES
年份固定效应	YES	YES	YES
N	2257	2257	2257
R^2	0.974	0.971	0.971

五 机制分析

结合前文大数据综试区设立背景以及该政策实施对地区经济增长影响的理论分析,这里进一步通过构建相应的机制变量来从实证层面识别大数据综试区设立对地区经济增长影响的内在作用机制,具体包括创新驱动和数字基建两个方面。

(一)创新驱动

对于创新驱动这一机制而言,我们参考张杰和郑文平[36]的研究,分别使用城市层面发明专利申请和授权的数量加一取对数来构建测度指标。回归结果如表7第(1)列和第(2)列所示,结果均表明大数据综试区的设立可以驱动城市创新从而进一步推动地区经济增长。此外,本文还参考《中国城市和产业创新力报告》中城市层面的创新指数来衡量地区创新水平,回归结果如表7第(3)列所示,结果同样表明大数据综试区的设立可以提升地区创新水平。综上所述,本文假说2得以验证。

(二)数字基建

对于数字基础设施建设而言,我们首先构建几项较为经典的测度指标。具体是用地区的互联网宽带接入端口数量、域名数量及网页数量与当地行政面积的比值来构建三项测度指标。回归结果如表7第(4—6)列所示。可以看出,大数据综试区的设立可以显著促进地区数字基础设施的建设情况,助力经济发展。此外,我们还进一步从数字平台视角刻画了更为新颖的数字基础设施建设指标。本文搜集了由中国信息通信研究院于2021年12月发布的《大数据白皮书》中的大数据交易平台建设信息,倘若某一城市在第t年设立了大数据交易平台,那么对该城市在当年及随后年份赋值1,否则赋值0,基于该指标的回归结果呈现在表7第(7)列,回归结果表明,大数据综试区的设立可以推动当地数字平台的建设。综上所述,本文假说3得以验证。

表7 机制检验结果

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	创新驱动			数字基建			
	申请数量	授权数量	城市创新指数	宽带接入端口	域名数量	网页数量	数字平台
Policy	0.116***	0.072**	0.414***	0.362***	0.515**	0.102***	0.102***
	(0.04)	(0.03)	(0.12)	(0.10)	(0.22)	(0.03)	(0.02)
控制变量	YES	YES	YES	YES	YES	YES	YES
城市固定效应	YES	YES	YES	YES	YES	YES	YES
年份固定效应	YES	YES	YES	YES	YES	YES	YES
N	2257	2257	2257	2257	2257	2257	2257
R^2	0.962	0.971	0.840	0.950	0.887	0.921	0.662

六 拓展性分析

尽管文章发现大数据综试区的设立的确有助于提升地区的收入水平,但在我国扎实推进共同富裕的当下,如何在实现总体收入增长的条件下,弥合地区间的收入差距需要进一步关注。因此,我们从以下两个角度进行分析。

第一,从地区间的经济发展水平出发,考察大数据综试区对"先富"地区和"后富"地区的影响是否存在显著差异。借鉴覃成林和杨霞的做法[37],区分出62个"先富"城市以及其他"后富"城市,以此来反映城市间富裕水平的差异,并对"先富"城市内的企业赋值1,对"后富"城市内的企业赋值0。分组回归结果如表8第(1—2)列所示,结果表明,相对于"先富"城市,"后富"城市更能通过大数据综试区这一政策来促进地区经济增长,该结果为从制度创新视角扎实推进共同富裕提供了重要证据。

第二,从城乡收入差距出发,农村地区由于经济发展阶段相对而言落后于城市地区,因此在数字基础设施建设以及对大数据技术的应用方面也多处于劣势。由大数据综试区设立带来的数字基建水平提升或许能够为农村地区利用大数据技术来提升收入水平创造更多机会。为此,我们分别进行如下检验以回答这一问题。首先,分别使用城镇居民可支配收入和农村居民可支配收入来验证大数据综试区是否会对城市和农村地区的经济增长带来差异性影响。估计结果呈现在表8第(3—4)列,大数据综试

区的设立不仅会显著促进城镇地区经济水平的提升，对农村经济增长同样具有重要意义，且从估计系数初步来看，大数据综试区对农村人均收入增长的促进作用更强。进一步，我们参考刘贯春的思路[38]，使用农村人均可支配收入与城镇人均可支配收入的比值来衡量城乡收入差距情况，回归结果如表 8 第(5)列所示。能够发现，大数据综试区的设立显著缩小了城乡收入差距，这一发现进一步丰富了文章的研究结论。

表 8　地区经济收入差距视角的拓展性分析

变量	(1) “先富”地区	(2) “后富”地区	(3) 城镇	(4) 农村	(5) 城乡收入差距
Policy	0.212 (0.22)	0.088*** (0.03)	0.277* (0.16)	0.684*** (0.16)	0.160*** (0.06)
控制变量	YES	YES	YES	YES	YES
城市固定效应	YES	YES	YES	YES	YES
年份固定效应	YES	YES	YES	YES	YES
N	477	1780	2257	2257	2257
R^2	0.974	0.954	0.997	0.998	0.990

七　结论与建议

文章通过匹配合并 2013—2020 年大数据综试区的试点数据、《中国城市统计年鉴》数据以及夜间卫星灯光数据等，实证研究了数据要素开发对地区经济增长的影响效应和作用机制。本文发现：(1)以大数据综试区设立为表现形式的数据要素开发会显著促进各城市的经济增长，且这一结论在利用处理霍桑效应、Goodman-Bacon 分解等稳健性检验后依然显著成立。(2)影响机制分析发现，数据要素开发会通过创新驱动和数字基建双重视角来构建地区经济增长的新动能，从而推动经济水平提升。(3)异质性分析发现，从地区数字服务商的供给来看，数字服务供给能力越强的城市，越能通过大数据综合试验区来实现经济增长；从地方政府对数字经济的关注度来看，关注度越高的地区越能依托数据要素开发来提升经济发展水平；从新冠疫情的冲击来看，各地区在冲击后更能利用数据要素开发来促进经济增长。(4)拓展性分析表明，相对于“先富”地区而言，数据要素开发主要促进了“后富”地区的经济增长，且数据要素开发还将有效缩小城乡之间的收入差距。

结合本文研究结论，我们给出如下政策建议。

第一，深入实施大数据战略行动，纵深推进国家大数据综试区建设。以大数据制度创新为发力点之一加快建设数字中国，推进大数据基础设施建设与大数据技术创新应用，构筑我国经济增长新动能。第二，加快大数据综试区内的数字基建步伐，以数字基建推动地区经济持续增长。具体来说，在充分结合地区经济发展需要的条件下，推动 5G 基站、光纤光缆、数据交易中心和边缘计算中心等数字基建的逐步完善。同时，加快数字基建对传统产业的赋能升级，将大数据运算技术充分嵌入到工业生产和产品销售过程中，为传统产业增长构建新优势。第三，广泛推动大数据技术在各行各业的创新应用，进一步夯实创新驱动下地区经济增长的长效机制。其中包括，鼓励企业提高对大数据技术创新成果的生产应用，助力突破企业生产制造难题。此外，依托大数据综试区建设，加快打造一批大数据高新技术产业，在优化地区产业结构的同时进一步带动经济增长。第四，通过优化大数据综试区的空间布局，不断缩小地区间的收入差距。尤其是，考虑在“后富”地区进一步探索符合设立大数据综试区条件的优良试点，加大大数据综试区的受益范围，为“后富”地区充分利用数据要素资源，进而挖掘经济增长新动能提供保障和支持。此外，支持农村地区积极利用大数据技术赋能生产经营活动，以大数据综试区助力数字乡村建设，加快推动城乡共同富裕。

[参　考　文　献]

[1]　方福前，田鸽，张勋. 数字基础设施与代际收入向上流动性——基于“宽带中国”战略的准自然实验[J]. 经济研究，2023(5)：79－97.

[2]　马述忠，郭继文. 制度创新如何影响我国跨境电商出口？——来自综试区设立的经验证据[J]. 管理世界，2022(8)：83－102.

[3]　涂勤，曹增栋. 电子商务进农村能促进农户创业吗？——基于

电子商务进农村综合示范政策的准自然实验[J].中国农村观察,2022(6):163-180.

[4] 罗舟.自贸试验区对产业结构升级的影响——基于SDID的验证[J].湖南大学学报(社会科学版),2024(4):75-82.

[5] 邱子迅,周亚虹.数字经济发展与地区全要素生产率——基于大数据综合试验区的分析[J].财经研究,2021(7):4-17.

[6] 任英华,刘宇钊,罗涵.数据基础制度建设对制造业企业数字技术创新的影响研究[J].湖南大学学报(社会科学版),2024(6):63-72.

[7] 石玉堂,王晓丹.大数据综合试验区设立能否驱动企业数字化转型——基于准自然实验的实证研究[J].科学学研究,2023(6):1-18.

[8] Choi C,Yi M H. The effect of the Internet on economic growth:evidence from cross-country panel data[J]. Economics Letters,2009(1):39-41.

[9] Czernich N, Falck O, Kretschmer T, et al. Broadband infrastructure and economic growth[J]. The Economic Journal,2011,552:505-532.

[10] Lee S H,Levendis J, Gutierrez L. Telecommunications and economic growth:an empirical analysis of sub-Saharan Africa [J]. Applied Economics, 2012(4):461-469.

[11] 韩宝国,朱平芳.宽带对中国经济增长影响的实证分析[J].统计研究,2014(10):49-54.

[12] 郑世林,周黎安,何维达.电信基础设施与中国经济增长[J].经济研究,2014(5):77-90.

[13] 陈彦斌,林晨,陈小亮.人工智能、老龄化与经济增长[J].经济研究,2019(7):47-63.

[14] 郭峰,熊云军,石庆玲,等.数字经济与行政边界地区经济发展再考察——来自卫星灯光数据的证据[J].管理世界,2023(4):16-34.

[15] 徐翔,赵墨非.数据资本与经济增长路径[J].经济研究,2020(10):38-54.

[16] 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(4):103-119.

[17] 郭炳南,王宇,张浩.大数据试验区设立能否驱动中国城市经济增长质量提升?[J].南京财经大学学报,2022(4):98-108.

[18] 徐康宁,陈丰龙,刘修岩.中国经济增长的真实性:基于全球夜间灯光数据的检验[J].经济研究,2015(9):17-29+57.

[19] 郑健壮,钱元旻,陈立峰.地区数据要素、数字技术水平与其经济增长[J].科学学研究,2024(11):2318-2329.

[20] Gordon R J. Why has economic growth slowed when innovation appears to be accelerating? [R]. NBER Working Paper, No. 24554,2018.

[21] Carriere-Swallow M Y, Haksar M V. The economics and implications of data: an integrated perspective[M]. Washington: International Monetary Fund,2019.

[22] 许宪春,张钟文,胡亚茹.数据资产统计与核算问题研究[J].管理世界,2022(2):16-30+2.

[23] Basu S, Fernald J. Information and communications technology as a general-purpose technology: evidence from US industry data[J]. German Economic Review, 2007(2):146-173.

[24] 袁健,段巍.数据交易、专业化人工智能与经济增长[J].管理评论,2024(6):3-18.

[25] 李苗苗,肖洪钧,赵爽.金融发展、技术创新与经济增长的关系研究——基于中国的省市面板数据[J].中国管理科学,2015(2):162-169.

[26] Macis M, Schivardi F. Exports and wages: rent sharing, workforce composition, or returns to skills? [J]. Journal of Labor Economics,2016(4):945-978.

[27] 郭凯明,潘珊,颜色.新型基础设施投资与产业结构转型升级[J].中国工业经济,2020(3):63-80.

[28] 郭克莎,彭继宗.二三产业结构变动与经济发展质量——上中等收入阶段向高收入阶段演进的国际经验[J].财贸经济,2022(8):5-26.

[29] 江艇.因果推断经验研究中的中介效应与调节效应[J].中国工业经济,2022(5):100-120.

[30] Beck T, Levine R, Levkov A. Big bad banks? The winners and losers from bank deregulation in the United States[J]. The Journal of Finance,2010(5):1637-1667.

[31] Carballo J, Chatruc M R, Santa C S. Online business platforms and international trade[J]. Journal of International Economics, 2022,137:103599.

[32] Stevenson B, Wolfers J. Bargaining in the shadow of the law: divorce laws and family distress[J]. The Quarterly Journal of Economics, 2006(1):267-288.

[33] 陶长琪,丁煜.数字经济政策如何影响制造业企业创新——基于适宜性供给的视角[J].当代财经,2022(3):16-27.

[34] 金灿阳,徐蔼婷,邱可阳.中国省域数字经济发展水平测度及其空间关联研究[J].统计与信息论坛,2022(6):11-21.

[35] 刘胜,徐榕鑫,陈秀英.数字化转型与企业风险承担——兼论我国产业链供应链"稳中求进"的启示[J].经济与管理,2023(3):74-81.

[36] 张杰,郑文平.创新追赶战略抑制了中国专利质量么?[J].经济研究,2018(5):28-41.

[37] 覃成林,杨霞.先富地区带动了其他地区共同富裕吗——基于空间外溢效应的分析[J].中国工业经济,2017(10):44-61.

[38] 刘贯春.金融结构影响城乡收入差距的传导机制——基于经济增长和城市化双重视角的研究[J].财贸经济,2017(6):98-114.