

# 大数据要素集聚、技术能力缺口 与生产率区域差距

张国胜, 严鹏, 李欣珏, 杜鹏飞

**[摘要]** 作为知识密集型生产要素,大数据的集聚效应并不是古典经济学中“投入—产出”的必然结果,而是建立在与要素禀赋相匹配的技术能力基础之上。本文基于适宜性技术进步理论,构建了大数据要素集聚、技术能力缺口及空间放大效应对生产率区域差距影响的理论框架,并使用2011—2021年县级面板数据进行实证检验。研究发现,大数据要素集聚扩大了集聚地与前沿地区的生产率差距,上述结论在多种稳健性检验中均成立。结合大数据要素的非稀缺性、知识密集性和场景依赖性等异质性特征,本文把技术能力这一“黑箱”打开,并将其分解为数据获取能力和数据应用能力,发现两类能力缺口均抑制了大数据要素集聚对全要素生产率提升的赋能效应,这是大数据要素集聚扩大生产率区域差距的重要原因。进一步研究发现,大数据要素存储与使用的空间分离特性强化并放大了技术能力对生产率区域差距的影响,即存在空间放大效应。本文揭示了大数据要素集聚激发生产率赋能效应的技术能力条件,并为中国发挥大数据要素规模报酬递增优势以及促进区域协调发展提供了理论借鉴。

**[关键词]** 大数据要素; 技术能力缺口; 全要素生产率; 集聚; 区域差距

**[中图分类号]** F424 **[文献标识码]** A **[文章编号]** 1006-480X(2024)10-0118-19

## 一、引言

党的二十大报告指出,经济结构性体制性矛盾突出,发展不平衡、不协调、不可持续,传统发展模式难以为继,一些深层次体制机制问题和利益固化藩篱日益显现。经历了改革开放四十多年的快速增长,依靠传统要素投入的增长逐渐放缓,中国经济正面临着增长缺乏新动力、区域发展不平衡等结构性矛盾(杨汝岱等,2023)。与此同时,伴随着数字技术快速发展,大数据作为数字经济的基础性、关键性、决定性新型生产要素,对全要素生产率的赋能日趋明显,为突破现阶段传统要素投入增长空间有限的约束提供了可能性。实际上,伴随着新兴技术革命和产业转型升级深入推进,大数据作为重要生产要素的经济价值愈加显现。2024年1月,国家数据局等17部门联合印发《“数据

**[收稿日期]** 2024-05-29

**[基金项目]** 国家社会科学基金重大项目“推进以农业转移人口市民化为首要任务的新型城镇化研究”(批准号21ZDA068);国家社会科学基金重点项目“户籍制度城乡双向改革与新型城镇化研究”(批准号20AJL012)。

**[作者简介]** 张国胜,云南大学经济学院、云南数字经济研究院教授,经济学博士;严鹏,云南大学经济学院博士研究生;李欣珏,云南大学经济学院、云南数字经济研究院副教授,统计学博士;杜鹏飞,国家信息中心博士后,经济学博士。通讯作者:严鹏,电子邮箱:whutyp@163.com。感谢匿名评审专家和编辑部的宝贵意见,文责自负。

要素×”三年行动计划(2024—2026年)》,指出“充分发挥数据要素的放大、叠加、倍增作用,构建以数据为关键要素的数字经济,是推动高质量发展的必然要求”“发挥数据要素乘数效应,赋能经济社会发展”。从“互联网+”到“数据要素×”,体现了数字经济条件下发挥大数据作为新型生产要素推动经济社会发展的愿望。尤其是在经济持续复苏以及国际环境复杂多变的情形下,大数据要素的价值创造和经济赋能潜力愈发受到关注,有关部门正积极建立健全数据要素相关的体制机制,推动国家大数据综合试验区设立。在此背景下,大数据要素规模呈指数级增长并在特定地区形成集聚,即同一类型或不同类型的大数据资源在一定地域范围内集中与聚合。因此,发挥大数据要素的集聚效应,有望成为推动地区经济高质量发展和促进区域协调发展的新引擎。然而,从中国的特征事实看,区域之间的全要素生产率差距却在逐渐扩大(孙志燕和侯永志,2019)。可见,政策意图与特征事实之间出现了偏离,对其原因的探讨是一个值得深入研究的命题。

目前,现有文献主要集中于大数据的要素特征及其对经济增长和生产率的影响上。从要素特征看,大数据要素具有典型的非竞争性、复制成本低和规模报酬递增特征(Jones and Tonetti, 2020)。除此之外,大数据要素还具有明显的知识密集、场景依赖等异质性特征。从大数据要素对经济增长的影响看,大数据要素能够促进技术交叉融合、驱动技术创新(杨俊等, 2022),加速资源流通速度、优化资源配置,提高经济增长速度并促进长期经济增长(徐翔和赵墨非, 2020)。从大数据要素对生产率的影响看,大数据要素投入能够降低交易成本、扩大市场范围和提高效率(许宪春等, 2019; 谢康等, 2020)。推动大数据要素与劳动力、资本等要素协同,有助于突破传统资源要素约束,提高全要素生产率。理论上,大数据要素集聚的规模经济和外部性能够让集聚地的生产率赋能效应更加明显,从而缩小集聚地与前沿地区的生产率差距。显然,这些理论研究的逻辑推导完全与政策意图相吻合,但与中国的现实情况相悖。<sup>①</sup>

实际上,生产率赋能效应是建立在大数据要素集聚具有显而易见的规模经济和强大的范围经济基础之上(王超贤等, 2022),也就是说大多数研究者倾向于认为,数据作为一种高级的新型生产要素,是报酬递增的。但是,这忽视了大数据要素集聚发挥作用所需要的技术条件,因为技术进步是一切集聚效应的根源。对于大数据要素集聚而言,大数据是知识密集型要素,并且具有极强的场景依赖特性,随着数据存量的指数级增加和循环累积,对大数据挖掘、存储、计算和分析等数据获取能力的要求越来越高。本文参考吴江和陶成熙(2024),将大数据要素的生产率赋能效应定义为大数据要素通过融入土地、劳动、资本、技术等要素,并在多要素协同下实现生产效率倍增、实现生产力的几何式增长。因此,赋能效应的发挥离不开协同优化、复用增效、融合创新等数据应用能力的支撑。综上所述,大数据要素集聚能否发挥生产率赋能效应取决于集聚地相应的数据获取能力和数据应用能力,本文将两者统一定义为技术能力。也就是说,大数据作为新型生产要素,其集聚效应并非古典经济学分析框架中“投入—产出”的必然结果,而是需要建立在一定的技术能力基础之上。

根据适宜性技术进步理论,只有当技术进步偏向与要素禀赋相匹配时,全要素生产率才会显著提高,而当技术进步偏离要素禀赋时,生产部门的效率提升则会受到限制甚至有所损失,并且失衡程度越大,损失幅度越大(孔宪丽等, 2015)。因此,集聚地要想发挥大数据要素集聚的生产率赋能效应就必须具备相应的技术能力。然而,基于大数据要素禀赋的差异,不同地区的技术能力匹配度不同,表现为集聚地的技术能力与发挥大数据要素集聚的生产率赋能效应所要求的技术能力之间的匹配程度不同。如果技术能力不匹配,就会产生相应的缺口,而本文将其定义为技术能力缺口。在技术

<sup>①</sup> 特征事实参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

能力缺口的作用下,大数据要素集聚的生产率赋能效应可能会大打折扣,并最终加剧生产率区域差距。从这个角度讲,尽管大数据要素蕴含巨大的潜力与价值,但如果缺乏与之匹配的技术能力,大数据要素集聚带来的经济增长与生产率提升将远远小于理论逻辑中的增长效应。所以,对于落后地区,在加快大数据要素集聚的同时,更为重要的是提高自身使用大数据要素的技术能力。

基于此,本文构建县域层面的生产率区域差距和技术能力缺口等指标,试图回答上述问题,并剖析大数据要素集聚、技术能力缺口与全要素生产率区域差距三者之间的经济学关系;技术能力在大数据要素的集聚效应中发挥的重要角色,以及技术能力缺口在大数据要素集聚对生产率区域差距影响中所起的关键作用。概言之,在数字经济时代,大数据要素集聚促进地区全要素生产率增长的效应能否发挥,将取决于集聚地的技术能力与大数据要素禀赋是否匹配。因此,理论层面,本文首先以适宜性技术进步理论为基础,构建包含大数据要素集聚、技术能力缺口和生产率区域差距的理论框架。实证方面,本文利用国家大数据综合试验区设立的准自然试验,考察大数据要素集聚对生产率区域差距的影响并检验相关机制,最后结合大数据要素的空间分离特性进一步分析大数据要素集聚与技术能力缺口扩大生产率区域差距的空间放大效应。研究发现,大数据要素集聚显著扩大了集聚地与前沿地区的生产率差距,其关键原因是,在数据获取能力缺口和数据应用能力缺口的作用下,大数据要素集聚的生产率赋能效应未能有效发挥,从而加剧了生产率区域差距,即技术能力缺口抑制了大数据要素集聚的生产率赋能效应。并且,这一过程受到大数据要素存储与使用的空间分离特性影响,生产率区域差距扩大的结果被空间放大效应进一步强化。上述结论也验证了本文提出的“大数据的集聚效应并不是古典经济学中‘投入—产出’的必然结果,而是需要建立在必要的技术能力基础之上”这一观点。

与既有研究相比,本文可能的边际贡献是:①与已有关注大数据的生产要素属性研究大数据要素促进全要素生产率增长的文献不同,本文从技术能力的视角剖析大数据要素发挥生产率赋能效应的技术条件,并深入挖掘大数据要素集聚导致生产率区域差距扩大的关键原因,这是对已有研究非常重要的补充。②基于适宜性技术进步理论,构建了大数据要素集聚、技术能力缺口及其空间放大效应对生产率区域差距影响的理论框架和传导机制。特别地,本文将技术能力聚焦于大数据要素的非稀缺性、知识密集性和场景依赖性等相较于传统要素的异质性特征上,进一步将技术能力这一“黑箱”打开并分解为数据获取能力和数据应用能力。在此基础上,基于空间分离特性探究大数据要素集聚扩大生产率区域差距的空间放大效应,拓展了现有理论。③测算了2011—2021年中国县域层面的全要素生产率,以及利用全国工商注册企业信息数据构建技术能力指标并与县域数据进行匹配,为宏观层面的相关研究提供了借鉴和参考。④借助双重差分模型、双重稳健估计和多种全要素生产率测算方法等,进行更加稳健的因果效应推断和机制检验,得到一系列关于大数据要素集聚、生产率区域差距的新结论,为数字经济时代生产率赶超和区域协调发展提供新的经验证据,并为充分发挥大数据要素的经济倍增效应、生产率赋能效应和实现高质量发展提供决策参考依据。

## 二、理论分析

### 1. 大数据要素集聚与生产率区域差距

从要素集聚与生产率区域差距的影响看,新经济地理理论认为,在集聚效应的作用下,要素集聚地会吸引更多高生产率企业,从而缩小与前沿地区的差距。大量实证研究表明,集聚经济是影响区域生产率差距的主要因素之一。基于上述逻辑,大数据要素集聚必然会导致生产率区域差距的变化,但是差距究竟是缩小还是扩大并不明确。

按照传统集聚理论,集聚的外部性是生产率提升的重要途径。因此,理论上大数据要素集聚能够为集聚地带来更大的规模经济,生产率赋能效应也会更加明显,这有助于缩小集聚地与前沿地区之间的生产率差距。与其他生产要素的集聚一样,大数据要素集聚具有马歇尔外部性(Marshall's Externalities),即能够带来专业化分工,具体表现为大数据要素集聚相应带来了大数据企业的集聚,从而产生共享经济,因此,集聚地的生产率增长可能更快。大数据要素集聚还具有雅各布斯外部性(Jacobs' Externalities),即能够促进产业间的密切往来,地区内围绕大数据要素会催生多样化产业,能够实现规模报酬递增,最终将促进集聚地的全要素生产率提升。根据新经济地理理论,大数据要素的集聚效应带来规模经济,推动上下游生产中间品的厂商集中于同一地区,从而减少冰山运输成本、扩大产品需求,并为厂商带来规模收益(Krugman, 1991)。在此过程中,这一效应会不断加强上下游产业的关联性,形成循环累积机制,使与大数据相关的配套产业长期锁定在某个地区,与之一起锁定的还有生产率赋能效应。从这些角度讲,大数据要素集聚通过共享经济、专业化分工和循环累积等机制为大数据相关产业集聚提供了重要的向心力,在路径依赖和自我预期的作用下,这种向心力会不断累积和强化,最终形成以大数据要素集聚地为核心的“中心—外围”空间结构,并进一步加剧区域间的不均衡发展。其结果就是中心地区的全要素生产率得到更快增长,因此,集聚地与前沿地区的生产率差距就有可能缩小。

然而,上述生产率区域差距缩小的发生是有条件的。集聚是一种空间概念,是指经济活动在特定地域内的集中过程,而要素则是经济空间的载体。经济活动的集中与分散实质上反映了要素的集聚或分散。因此,由要素集聚带来的生产率赋能效应实际上是一种要素禀赋与技术能力在经济增长中的动态耦合和相互作用的过程(郝大江和张荣, 2018),尤其对于大数据这种新型要素更是如此。因为大数据相较于传统要素有三个重要的异质性特征:①非稀缺性。大数据具有数据产生数据的特性,并且按照指数模式增长,总量趋于无限。②知识密集性。一方面,从众多繁杂的碎片化信息到可用的数据,这一过程必须被计算机以数字化、可视化的形式呈现出来,是一个知识附加的过程;另一方面,大数据要素之所以有价值,是因为其承载着大量的客观事实,这是数据的内在实质。③场景依赖性。大数据发挥作用必须依赖于具体的场景,随着应用场景的不同,大数据要素的价值也会不同。正如适宜性技术进步理论所强调的那样,大数据要素的知识密集性等异质性特征意味着其集聚的生产率赋能效应的发挥会受到特定技术能力的制约,即当技术能力与大数据要素禀赋相适应时,全要素生产率才会得到大幅提升,否则,技术能力与大数据要素禀赋失衡会降低其生产率的提升效率。因此,大数据要素集聚对生产率区域差距的影响主要取决于大数据要素禀赋与技术能力的匹配程度,换言之,大数据要素集聚并不必然带来集聚地生产率的提升。若集聚地具备相应的技术能力,则可以最大程度发挥生产率赋能效应,进而缩小与前沿地区的差距。否则,大数据要素集聚发挥生产率赋能效应的过程就会受到制约,导致全要素生产率无法达到理论预期的增长,基于传统集聚理论的“中心—外围”空间结构可能也不会产生,集聚地与前沿地区的生产率差距反而会扩大。基于上述分析可以发现,如果具备相应的技术能力,集聚经济的正外部性有助于集聚地缩小与前沿地区的生产率差距,否则差距反而会扩大。因此,大数据要素集聚对生产率区域差距的影响并不明确。据此,本文提出:

假说 1a: 大数据要素集聚会缩小集聚地与前沿地区的生产率差距。

假说 1b: 大数据要素集聚会扩大集聚地与前沿地区的生产率差距。

## 2. 技术能力缺口与生产率区域差距

技术能力与大数据要素禀赋的匹配与否是决定大数据要素集聚对生产率区域差距影响的关



键所在。适宜性技术进步理论认为,发展中国家如果选择与其要素禀赋结构相一致的技术结构,那么这个国家和发达国家之间在全要素生产率以及每个劳动力的人均产出上的差异就会变得最小(林毅夫和张鹏飞,2006)。根据这一理论,一个国家内部的落后地区与前沿地区也是如此,对于大数据要素而言,地区的发展同样需要与其要素禀赋相匹配的技术能力。然而,这种技术能力并非与生俱来,落后地区往往需要通过自身学习或者吸收前沿地区的技术扩散来获取。但是,落后地区在学习和吸收前沿地区技术扩散的过程中存在一定的约束(Acemoglu and Zilibotti,2001),致使落后地区与要素禀赋所要求的技术能力之间产生缺口。①在技术扩散的过程中,前沿地区往往会根据自身的要素禀赋条件进行有针对性的研发和学习。对于大数据要素而言,不同区域间要素禀赋的异质性导致技术需求的差异性,这进一步影响了各地区所能获取和积累的技术能力,并最终反映在对大数据要素的获取和应用的差距上。因此,落后地区会出现与所需技术能力不匹配的情况,导致全要素生产率水平远远落后于前沿地区。②由于区域间技术发展水平和外部环境的异质性,特别是在技术认知、应用及创新等方面,落后地区与前沿地区存在显著差距(龚斌磊,2022),落后地区在引进和转化大数据相关技术时,必须针对本地大数据要素的具体生产和使用条件进行适应性投资。然而,当要素禀赋差异过大时,所需的适应性投资规模较大,而投资的边际回报率相对较低。适应性投资与回报的不对称性降低了地方政府和科技企业进行此类投资的积极性。这种情况可能会进一步削弱落后地区的技术能力,加剧其与前沿地区在技术能力上的差距并产生较大的技术能力缺口,从而形成一个不利于落后地区技术发展的恶性循环。③大数据要素的生产率赋能效应体现了其提升效率和推动技术创新的核心特征,而这些特征的基础在于海量数据所蕴含的巨大潜力与价值(杨俊等,2022)。但是,这些潜力和价值的挖掘依赖于较为复杂的技术能力,从实践看,大数据要素价值的实现主要是依靠大数据收集、处理、共享和应用等技术能力。因此,倘若落后地区存在较大的技术能力缺口,会使该地区无法充分、有效利用集聚在此的大数据资源,甚至会造成资源错配等问题,其全要素生产率很可能会落后于其他地区。

从这些角度讲,集聚地尤其是落后地区使用大数据要素的技术能力与要素禀赋所要求的技术能力之间存在缺口,就可能会造成大数据在发挥其生产率赋能效应上存在差距,最终导致地区生产率差距扩大。进一步,本文聚焦大数据要素的异质性特征并将技术能力分解为数据获取能力和数据应用能力。其中,数据获取能力是指数据挖掘、清洗、处理等依靠自身学习、经验和技能基础的技术能力;数据应用能力是指依靠外部需求以及产业发展、研发创新等实现数据价值的技术能力。基于此,本文提出:

假说 2a:数据获取能力缺口会抑制大数据要素集聚对全要素生产率的赋能效应,从而扩大生产率区域差距。

假说 2b:数据应用能力缺口会抑制大数据要素集聚对全要素生产率的赋能效应,从而扩大生产率区域差距。

### 3. 生产率区域差距的空间放大效应

大数据要素还具有另外一个重要的特性:大数据要素的使用与存储在空间上是可分的,即空间分离特性。这意味着,只要具备相应的技术能力,集聚在某个地区的大数据要素不仅可供“近在咫尺”的本地企业使用,也可以被“远在天边”的地区使用。从这个角度讲,空间分离特性进一步凸显了技术能力在大数据要素发挥生产率赋能效应时的重要性和必要性,也正是在技术能力和空间分离特性的作用下,生产率区域差距被进一步放大。由于技术能力的存在,前沿地区较落后地区在数据挖掘、存储、分析处理和应用等方面具有绝对优势,能够充分使用大数据要素赋能全要素生产率增长,因此,对于那些大数据要素禀赋较弱但具备较强技术能力的地区,能够远程使用各地区包括落后地

区的大数据资源,这会形成“赢者通吃”的局面并进一步放大技术能力缺口对生产率差距的影响。大数据要素价值的重要体现之一就是其具有较强的时效性。因此,一旦集聚地没有相应的技术能力使用集聚在此的大数据资源,数据的价值将随着时间的推移不断降低甚至沉淀为“无用数据”。而这些数据在其“过期失效”之前则可以源源不断地供给技术能力强的地区进行挖掘使用,即使那些具备技术能力的地区“远在天边”,由此扩大集聚地与前沿地区的生产率差距。随着互联网技术的快速发展,信息、知识交流的手段和效率得到极大改善。缄默知识的不可编码性降低,不可编码信息的空间局限性降低,集聚的空间溢出不再依赖于地理邻近性。“远在天边”也可以通过各种科技手段如同“近在咫尺”地使用大数据资源。概言之,大数据要素集聚对生产率的赋能效应存在空间溢出。而这种溢出在技术能力和空间分离特性的双重作用下会形成“强者更强、弱者更弱”的分化局面,进而让落后地区与前沿地区的生产率差距进一步放大。需要强调的是,空间放大效应的产生依赖于两个重要条件:一是大数据要素的空间分离特性;二是发挥大数据要素生产率赋能效应所需要的技术能力。在此条件下,无论大数据要素集聚在前沿地区还是落后地区,生产率区域差距的空间放大效应都会存在。一方面,如果大数据要素集聚在落后地区,由于落后地区可能存在较大的技术能力缺口,因此,无法充分有效地使用大数据要素赋能自身生产率增长。同时,前沿地区的技术能力缺口可能相对较小甚至为零,因此,能够远程使用包括落后地区在内所有地区的大数据要素实现生产率快速增长,由此扩大与落后地区的差距。另一方面,如果大数据要素集聚在前沿地区,则前沿地区可以更为方便快捷地使用本地的大数据要素,而落后地区可能不会或较小概率享受大数据要素集聚带来的生产率赋能效应,因而生产率区域差距可能会更大。基于上述分析,本文提出:

假说3:大数据要素集聚对生产率区域差距的影响会被大数据要素的空间分离特性进一步强化和放大,即存在空间放大效应。

基于上述理论逻辑,本文首先以大数据要素的非稀缺性、知识密集性和场景依赖性三个异质性特征为起点,挖掘出大数据要素集聚发挥生产率赋能效应所需要的技术能力,即数据获取能力和数据应用能力。然后,聚焦技术能力缺口,分别从数据获取能力缺口和数据应用能力缺口两个方面检验大数据要素集聚扩大生产率区域差距的影响机制,并进一步考察大数据要素的空间分离特性对生产率区域差距产生的空间放大效应。本文的理论框架如图1所示。

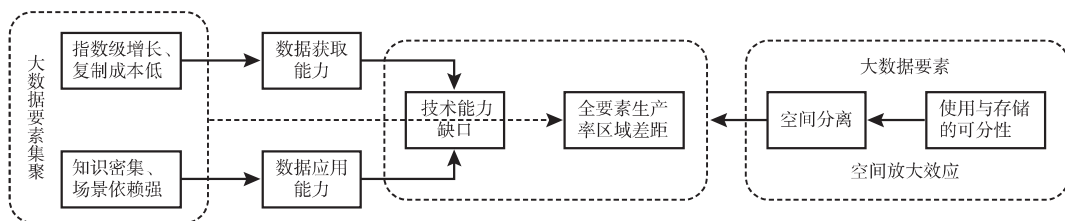


图1 大数据要素集聚与生产率区域差距的理论框架

### 三、研究设计

#### 1. 政策背景

2015年8月31日,国务院发布了《促进大数据发展行动纲要》(简称《纲要》)。为了全面推进大数据部署和深化大数据应用,实现数字发展、区域协调发展,《纲要》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。2015年9月,贵州省启动了首个大数据综合试验区建设;2016年,第二批国家级大数据综合试验区名单发布。从试验区的分布情况看,这些试验区既有前沿地区也有落后地区,并且在资源禀赋、经济实力、信息社会水平、生态环境等方面存在一定差异,其发展定位与目标也各不相同。但从《纲要》的目标任务看,多次提及并强调要推动整合分散的数据中心资源、实现数据资源的汇聚应用。这表明,国家大数据综合试验区的设立一定程度上能够促进大数据要素的集聚。以国家大数据(贵州)综合试验区为例,截至2023年底,贵州省在建和运营重点数据中心47个,其中大型以上25个,已建成“贵州·中国南方数据中心示范基地”,聚集了三大运营商及华为、腾讯、富士康和苹果等数据中心和灾备中心。此外,贵阳大数据交易所集聚的数据商和第三方服务机构已超过1000家,累计交易额超过49亿元。<sup>①</sup>由此可见,国家大数据综合试验区设立这一准自然实验为本文研究大数据要素集聚效应提供了较为理想的外生冲击。

## 2. 模型设定

基于上述政策背景介绍,本文借鉴余明桂等(2016)的研究设计,利用国家大数据综合试验区设立这一准自然实验提供的外生冲击,采用双重差分(DID)模型,分析大数据要素集聚对县域层面全要素生产率区域差距的影响。具体模型设定如下:

$$GapTFP10_{it} = \alpha + \beta bigdata_i \times post_t + X'_{it}\gamma + \eta_t + \lambda_i + \varepsilon_{it} \quad (1)$$

其中, $i$ 表示县域, $t$ 表示年份。 $GapTFP10_{it}$ 为县域 $i$ 与前沿地区的生产率差距; $bigdata_i$ 为县域 $i$ 是否属于国家大数据综合试验区的虚拟变量; $post_t$ 为国家大数据综合试验区政策实施前后的虚拟变量; $X'_{it}$ 为县域层面可能影响全要素生产率区域差距的一系列控制变量, $\eta_t$ 为年份固定效应, $\lambda_i$ 为县域固定效应, $\varepsilon_{it}$ 为随机扰动项。本文主要关注估计系数 $\beta$ 。需要注意的是,作为省级层面的政策变量,国家大数据综合试验区的设立可能无法全面反映县域层面的大数据要素集聚差异,为此,本文将在稳健性检验中尝试替换核心解释变量,以缓解这一问题。

## 3. 变量选取

(1)被解释变量:全要素生产率区域差距。对于全要素生产率(TFP),本文采用随机前沿模型(SFA)进行测算,其中,生产函数选择超越对数形式生产函数,这是因为,超越对数形式生产函数放松了对要素替代弹性的限制,能够更好地拟合各类投入产出关系(龚斌磊,2022)。随机前沿模型针对无效率项的不同假设有不同的形式,根据现有文献并结合数据特征,本文参考Wang and Ho(2010)选择考虑个体异质性的无效率项时变面板随机前沿模型(TFP10),原因主要有两点:①相较于无效率项非时变的SFA模型,时变的无效率项更适合时间跨度长的面板模型;②传统的面板SFA模型把非时变的个体异质性视为无效率项的组成部分,这会导致估计偏误,因此,需要考虑个体异质性。其中,产出指标为县域 $i$ 的实际地区生产总值(GDP),本文将基期设定为2000年,然后以各省份GDP指数对县域名义GDP进行平减,得到实际GDP。投入指标包括县域层面的物质资本和全社会从业人员数,由于县域并未直接统计这两个数据,参考金飞和张琦(2013),分别使用人口数据和全社会固定资产投资数据构建比例系数,估算出县域层面的劳动力投入和物质资本数据。需要特别说明的是,资本存量的估计参考张军等(2004)的研究。对于区域差距而言,本文参考龚斌磊(2022)的研究,使用前沿地区的全要素生产率与县域 $i$ 的全要素生产率的比值度量,其中,前沿地区为当年全要素生产率最高的县。该变量为正向指标,即比值越大,说明县域 $i$ 的全要素生产率与前沿地区

<sup>①</sup> 资料来源:贵州省人民政府网站([https://www.guizhou.gov.cn/ztl/2024nsbh/jjsb/202408/t20240829\\_85505998.html](https://www.guizhou.gov.cn/ztl/2024nsbh/jjsb/202408/t20240829_85505998.html))。

沿地区的差距越大。为了验证结论的可靠性,本文替换SFA模型重新计算全要素生产率,主要包括 Battese-Coelli-95 (TFP95)、Slacks-Based Measure (TFPsbm)、Belotti-Ilardi-18 (TFP18) 三种模型 (Battese and Coelli, 1995; Tone, 2002; Belotti and Ilardi, 2018);另外,使用作差的方法 (*Diff*) 替换差距的衡量方式进行稳健性检验。此外,由于前沿地区在资源禀赋、宏观政策、人口流动等方面可能具有优势,并且还存在着诸多影响大数据要素集聚和生产率区域差距的不可观测因素,使用前沿地区作为参照系可能会导致估计结果偏误,为此本文在稳健性检验部分加强了这一方面的讨论。

(2)核心解释变量:大数据要素集聚。本文采用时间虚拟变量 ( $post_t$ ) 与组间虚拟变量 ( $bigdata_i$ ) 的交乘项来捕捉大数据要素集聚的外生冲击,即若县域  $i$  所在的省/市属于国家大数据综合试验区,  $bigdata_i$  取值 1, 否则取值 0;若县域  $i$  所在的省/市在第  $t$  年成为国家级大数据综合试验区,则  $post_t$  在当年及之后年份赋值 1, 否则为 0。需要说明的是,虽然贵州省在 2016 年才正式获批建设国家级大数据综合试验区,但贵州省的实际建设启动时间是 2015 年,因此,将 2015 年作为贵州省的政策实施时点,2016 年作为其他地区的政策实施时点。另外值得注意的是,国家大数据综合试验区的设立并不完全等价于大数据要素集聚,但从试验区实际的建设情况和中国大数据中心的地理分布看,政策的实施确实推动了大数据中心的建设,大数据要素呈现出向试验区集聚的态势。此外,孙伟增等 (2023) 发现,国家大数据综合试验区的设立一定程度上使大数据要素在区域范围内聚合,并且两者之间存在着较强的因果关系,<sup>①</sup>因此,使用国家大数据综合试验区作为县域大数据要素集聚的外生冲击,具有代表性且合理。

(3)控制变量。本文的控制变量包括:①劳动力投入,使用县域层面全社会从业人员数的自然对数度量;②固定资本投入,使用物质资本存量的自然对数度量;③政府干预水平,使用地方财政支出占 GDP 的比值度量;④产业结构高级化程度,使用第三产业与第二产业产值之比度量;⑤人力资本水平,鉴于县域层面在校大学生数据有限,使用普通中学在校学生数占总人口的比重度量;⑥金融发展水平,使用年末金融机构各项贷款余额的自然对数度量;⑦医疗卫生情况,使用医院和卫生院床位数的自然对数度量。

#### 4. 数据来源

本文使用的数据为 2011—2021 年间全国县级行政区的面板数据,数据来自历年《中国县域统计年鉴》《中国区域经济统计年鉴》、中国研究数据服务平台 (CNRDS)、国泰安数据库 (CSMAR)。为了保证数据质量,本文借鉴贾俊雪等 (2011) 的研究,按如下步骤对原始数据进行处理:①以 2019 年的行政区划为基准,剔除行政区划有更改名字等较大变更的县和县级市,并统一各县级行政区划名称和代码;另外,查阅并参考各地区行政区划沿革的历史资料,对各个县和县级市的行政隶属关系、区划面积和人口等数据进行调整。②由于县级数据质量较差,本文剔除了数据缺失严重的省、市样本,如西藏自治区、三沙市、儋州市等。③对于部分缺失数据,本文首先从各省、市和区县的官方网站、统计局、国民经济和社会发展统计公报逐年手动搜集匹配,然后在此基础上采用插值法和趋势外推法进行插补。经过上述处理后,本文最终确定的样本为 2011—2021 年全国 2556 个县域的非平衡面板数据。从变量的描述性统计看,<sup>②</sup>除参考 Belotti and Ilardi (2018) 计算的结果 (TFP18) 略偏大以外,其余方法计算结果的平均值、标准差等统计量均相差不大,表明本文计算的县域 TFP 是稳健

① 根据匿名评审专家的建议,本文对大数据要素集聚与国家大数据综合试验区设立做因果分析,详细结果参见《中国工业经济》网站 (ciejournal.ajcass.com) 附件。

② 详细描述性统计结果参见《中国工业经济》网站 (ciejournal.ajcass.com) 附件。



的。此外,从各年县域层面的前沿地区分布及其TFP水平看,前沿地区基本固定分布在东部沿海地区,并且主要集中于江苏省和广东省,总体符合中国的现实情况,表明本文测算的县级层面的全要素生产率也是合理的。<sup>①</sup>其余变量无异常。

#### 四、实证分析

##### 1. 基准回归

表1汇报了国家大数据综合试验区设立对生产率区域差距影响的估计结果。其中,第(1)列只控制了县域和年份的固定效应,第(2)列在此基础上增加了县域层面可能影响大数据要素集聚和生产率区域差距的控制变量。结果发现,不论是仅加入固定效应还是在此基础上加入控制变量,国家大数据综合试验区设立的系数均在1%的水平上显著为正,说明相对于非试点地区,大数据要素集聚扩大了试点地区与前沿地区的生产率差距,即大数据要素集聚扩大了生产率区域差距,假说1b得以验证。这说明,大数据要素作为知识密集型的新型要素,其集聚效应并非古典经济学分析框架中“投入—产出”的必然结果,而是需要建立在与要素禀赋相匹配的技术能力基础之上。此外,回归系数和调整R<sup>2</sup>都随着控制变量的增加逐渐趋于稳定,<sup>②</sup>说明本文控制变量的选取是比较合理的。其中,固定资本投入、政府干预都在样本期内显著扩大了生产率区域差距,这可能是由于新增投资向固定资本转化的效率低下和市场活力不足导致的;劳动力投入、人力资本水平和医疗卫生情况显著缩小了生产率区域差距;产业高级化系数为正、金融发展水平系数为负,但两者在统计上均不显著。总体而言,控制变量的回归结果基本符合预期。

表1 大数据要素集聚与生产率区域差距:基准回归结果

	(1)	(2)
	<i>GapTFP10</i>	<i>GapTFP10</i>
<i>bigdata</i> × <i>post</i>	0.0016*** (0.0002)	0.0018*** (0.0002)
常数项	1.0431*** (0.0000)	1.0239*** (0.0091)
控制变量	否	是
县域固定效应	是	是
年份固定效应	是	是
观测值	28116	23100
调整R <sup>2</sup>	0.7240	0.8276

注: \*、\*\*、\*\*\*分别表示在10%、5%、1%水平上显著;括号内为聚类到县域层面的稳健标准误。以下各表同。

##### 2. 平行趋势检验<sup>③</sup>

为保证双重差分结果的有效性,本文的回归分析需要满足平行趋势条件。本文利用事件研究法(EventStudy)进行平行趋势检验,即分别考察国家大数据综合试验区设立前后处理组与对照组的

① 前沿地区及其TFP参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

② 依次加入控制变量的回归结果参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

③ 平行趋势检验结果、其他稳健性检验结果参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

生产率区域差距变化,基期选择事件窗口的第一期。平行趋势检验结果显示,政策实施之前,政策虚拟变量各期估计系数均不显著,即处理组与对照组在政策实施之前生产率区域差距无显著性差异,而在政策实施之后差异显著。此外,就试点政策的动态效应而言,国家大数据综合试验区设立当年对生产率区域差距的影响并不明显,随着时间的推移,处理组全要素生产率与前沿地区的差距整体呈扩大态势,说明政策发挥具有迟滞性与动态可持续性。综上所述,研究样本满足平行趋势假设。

### 3. 内生性问题处理

本文可能存在潜在的内生性问题:一方面,由于无法穷尽所有可能影响大数据要素集聚和生产率区域差距的因素,本文可能遗漏掉重要的控制变量;另一方面,大数据要素可能集聚在全要素生产率较高的地区,即存在反向因果关系。基于以上考量,本文采取工具变量法以期缓解潜在的内生性问题。出于对大数据中心建设的安全考虑,本文收集整理县域所在省份过去5年内(不含当年)地震次数的平均值(*earthquake\_m*)作为大数据要素集聚的工具变量。从相关性角度看,大数据要素作为重要的基础性资源和战略性资源,大数据中心的选址布局必然需要考虑所在地的安全性。因此,一个地方历史上的地震次数很可能会成为政府考虑的因素,从而影响大数据要素的集聚。从外生性角度看,宏观层面的历史均值变量一定程度上外生于县域层面的混杂因素,并且宏观层面的历史均值变量并不会直接影响县域当期的全要素生产率,此外县域层面不可观测的时变混杂因素也难以直接影响宏观层面的历史变量。表2汇报了工具变量的相关结果。从一阶段回归结果看,*earthquake\_m*的回归系数在1%的水平上显著为负,表明历史地震次数确实会减少当地的大数据要素集聚,与预期相符。从二阶段回归结果看,大数据要素集聚的回归系数依然在1%的水平上显著为正,并且系数较基准回归结果有所扩大,说明由于潜在的内生性问题,基准回归可能低估了大数据要素集聚对生产率区域差距的影响。另外,识别不足检验的KP-LM统计量对应的P值为0,表明不存在识别不足问题;弱工具变量检验的KP-Wald F统计量约为555.02,远大于16.38(Stock-Yogo 10%显著性水平)这一临界值,通过了弱工具变量检验,表明本文选取历史地震次数作为大数据要素集聚的工具变量是合理可行的。综上所述,在缓解了可能的内生性问题后,本文的基本结论依然成立。

表2 工具变量的回归结果

	(1)	(2)
	一阶段回归	二阶段回归
	<i>bigdata</i> × <i>post</i>	<i>GapTFP10</i>
<i>earthquake_m</i>	-0.0718*** (0.0044)	
<i>bigdata</i> × <i>post</i>		0.0149*** (0.0018)
控制变量	是	是
省份固定效应	是	是
年份固定效应	是	是
KP-LM P-value		0.0000
KP-Wald F-value		555.0202
观测值	25200	23100
调整 R <sup>2</sup>	0.5554	0.7363

## 五、影响机制分析

基于前文的理论分析,本文参考江艇(2022)的研究,构建如下影响机制分析模型:

$$GapTFP10_{it} = \alpha + \beta_1 bigdata_i \times post_t \times gapability_{it} + \beta_2 bigdata_i \times post_t + \beta_3 gapability_{it} + X'_{it} \gamma + \eta_t + \lambda_i + \varepsilon_{it} \quad (2)$$

其中, $gapability_{it}$ 为机制变量,其余变量与基准回归模型一致,核心解释变量为 $bigdata_i \times post_t \times gapability_{it}$ , $\beta_1$ 是本文主要关注的估计系数。

对于机制变量,数据获取能力的度量难度较大,目前还没有统一指标。现有关于数据能力的研究多为企业层面,基本上是通过企业问卷调查的方式来测度企业数据能力(李唐等,2020)。部分研究虽然探讨了省级层面的大数据发展水平,但也是通过每百家企业拥有的网站数、每百人使用计算机数等有关数字化的指标间接度量,鲜有直接从宏观层面相对全面测度区域数据能力的指标。不过,得益于机器学习等技术的发展,文本分析方法为本文县域数据获取能力的测度提供了新的工具。从现有文献看,文本分析方法能够借助机器学习深度挖掘文本主体的相关特征(杨金玉等,2022),是指标构造的较好方法。因此,本文运用文本挖掘方法处理企业工商注册信息数据,以捕捉数据获取能力,据此测度县域层面的数据获取能力。具体步骤如下:①基于历年政府工作报告、大数据相关政策文献及实际情况,采用人工阅读与关键词提取方式,整理出数据获取能力的关键词库。②鉴于地区层面的数据获取能力主要由从事数据收集、清洗、分析和处理等业务的企业提供,本文利用企业工商注册信息中的“经营范围”文本作为分析对象。③使用第一步生成的关键词库对获得的文本信息进行处理,首先对文本进行清洗,如剔除空格、乱码等,然后通过R进行文本挖掘和统计分析,捕捉并统计与数据获取能力相关的关键词词频。④根据企业所在地信息,将关键词词频汇总到县域层面,构建县域数据获取能力指标。在此基础上,为衡量县域大数据要素集聚程度与自身数据获取能力的匹配度,本文参考权小锋等(2010),使用线性模型预测县域*i*大数据要素集聚程度所需的数据获取能力,然后用预测值减去实际的数据获取能力即为数据获取能力缺口( $gapwords$ ),<sup>①</sup>数值越大表明缺口越大。为了验证结果的稳健性,本文还将“经营范围”里含有数据获取能力关键词的企业设定为具有数据获取能力的企业,然后根据注册地信息统计县域层面具有数据获取能力的企业数作为县域数据获取能力的替代指标,最后同样通过预测值减实际值的方式计算得到数据获取能力缺口的替代指标( $gapfirms$ )。对于数据应用能力,由于大数据是数字经济的基础性、关键性、决定性生产要素,大数据的应用自然与数字经济的创新密切相关,因此,本文使用县域层面的数字经济发明专利数作为数据应用能力的代理指标;同样地,用县域*i*数据应用能力的预测值减去实际值表征数据应用能力缺口( $gapdpat$ )。除此之外,本文还使用当年全国技术能力的平均值代表各县域大数据要素集聚所需的基本技术能力,然后将其与实际技术能力相减得到技术能力缺口的替代指标作为稳健性检验。<sup>②</sup>

### 1. 数据获取能力缺口的影响

表3第(1)列汇报了数据获取能力缺口的机制检验结果。可以看出,数据获取能力缺口与国家大数据综合试验区政策虚拟变量的交互项系数显著为正,说明在两种技术能力缺口的作用下,国家

① 详细的计算步骤参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

② 回归结果参见《中国工业经济》网站(ciejournal.ajcass.com)附件。

表3 机制分析结果

	(1)	(2)	(3)	(4)
	<i>GapTFP10</i>	<i>GapTFP10</i>	<i>GapTFP10</i>	<i>GapTFP10</i>
<i>bigdata</i> × <i>post</i>	0.0004 (0.0003)	0.0005 (0.0003)	0.0020*** (0.0003)	0.0018*** (0.0003)
<i>gapawords</i>	0.0003 (0.0002)			
<i>bigdata</i> × <i>post</i> × <i>gapawords</i>	0.0014*** (0.0003)			
<i>gapfirms</i>		0.0003 (0.0002)		
<i>bigdata</i> × <i>post</i> × <i>gapfirms</i>		0.0014*** (0.0003)		
<i>gapdpat</i>			-0.0020** (0.0008)	
<i>bigdata</i> × <i>post</i> × <i>gapdpat</i>			0.0024*** (0.0007)	
<i>gapctydpat</i>				-0.0010*** (0.0003)
<i>bigdata</i> × <i>post</i> × <i>gapctydpat</i>				0.0005* (0.0003)
常数项	1.0231*** (0.0092)	1.0232*** (0.0092)	1.0375*** (0.0114)	1.0284*** (0.0093)
控制变量	是	是	是	是
县域固定效应	是	是	是	是
年份固定效应	是	是	是	是
观测值	22694	22694	17335	22317
调整 R <sup>2</sup>	0.8277	0.8276	0.8204	0.8265

大数据综合试验区设立显著扩大了集聚地与前沿地区的生产率差距。数据获取能力缺口越大的地区,其利用大数据要素赋能全要素生产率增长的水平越有限,相比之下,前沿地区的数据获取能力缺口较小甚至为零,因此,大数据要素集聚的生产率赋能效应非常明显,这就使得数据获取能力低的地区与前沿地区的生产率差距变大。为了验证这一机制的可靠性,本文使用具有数据获取能力的企业数作为县域数据获取能力的替代指标并生成新的数据获取能力缺口变量(*gapfirms*)来进行稳健性检验。表3第(2)列汇报了这一回归结果,不难发现,回归结果与第(1)列基本一致,数据获取能力缺口与国家大数据综合试验区政策虚拟变量的交互项系数依然显著为正,说明数据获取能力缺口这一机制是稳健的,假说2a得以验证。

### 2. 数据应用能力缺口的影响

表3第(3)、(4)列为数据应用能力缺口的机制检验结果,其中,第(3)列为加入数据应用能力缺口(*gapdpat*)及其与政策虚拟变量交互项的回归结果。由于县域层面的数字经济发明专利是通过县域全部发明专利与城市层面的数据构造比例系数估算得到的,可能存在测量偏误,因



此,为了验证结论的可靠性,本文直接使用城市层面的数字经济发明专利生成数据应用能力缺口变量(*gapctydpat*)进行回归,结果见第(4)列。回归结果显示,在加入数据应用能力缺口及其与国家大数据综合试验区政策虚拟变量的交互项后,交互项的回归系数均显著为正,说明在数据应用能力缺口越大的地区,大数据要素集聚对生产率区域差距的扩大作用越大,即数据应用能力缺口抑制了大数据要素集聚的生产率赋能效应,因此,对于那些不具备数据应用能力或能力较低的地区而言,其生产率水平就与前沿地区“渐行渐远”,差距由此扩大。综上所述,本文提出的假说2b得以验证。

## 六、进一步分析

### 1. 空间放大效应

根据前文理论分析,大数据要素具有存储与使用的空间分离特性,因此,只要具备相应的技术能力,大数据要素不仅可以供本地使用,还可以被其他地区使用。换言之,无论是“近在咫尺”还是“远在天边”的具有较强技术能力的地区都能调用本地数据。如此一来,若本地缺乏与使用大数据要素相匹配的技术能力,大数据要素集聚对生产率区域差距的影响就会被空间分离特性进一步放大。需要说明的是,空间放大效应是建立在基准回归和机制分析的结果之上。一方面,由于技术能力缺口的存在,集聚地无法充分发挥大数据要素的生产率赋能效应,导致与前沿地区的生产率差距扩大,即直接效应;另一方面,由于前沿地区具有较强的技术能力,因而能够远程使用集聚地的大数据要素并且赋能自身全要素生产率增长,这必然会在直接效应的基础上加剧和放大生产率区域差距。本文分两步来进行验证:①验证大数据要素集聚具有空间溢出作用,即间接效应,对应模型为式(3);②进一步验证间接效应在技术能力的作用下会扩大生产率区域差距,对应模型为式(4)。模型的具体设置参考邵帅等(2019)采用空间杜宾模型(SDM):<sup>①</sup>

$$GapTFP10_{it} = \alpha + \rho \sum_{j=1}^N W_{ij} \times GapTFP10_{jt} + \beta bigdata_{it} \times post_{it} + \theta \sum_{j=1}^N W_{ij} \times bigdata_{jt} \times post_{jt} + X'_{it} \gamma + \eta_t + \lambda_i + \varepsilon_{it} \quad (3)$$

$$GapTFP10_{it} = \alpha + \rho \sum_{j=1}^N W_{ij} \times GapTFP10_{jt} + \beta_1 bigdata_{it} \times post_{it} \times ability_{it} + \beta_2 bigdata_{it} \times post_{it} + \beta_3 ability_{it} + \theta_1 \sum_{j=1}^N W_{ij} \times bigdata_{jt} \times post_{jt} \times ability_{it} + \theta_2 \sum_{j=1}^N W_{ij} \times bigdata_{jt} \times post_{jt} + \theta_3 \sum_{j=1}^N W_{ij} \times ability_{jt} + X'_{it} \gamma + \eta_t + \lambda_i + \varepsilon_{it} \quad (4)$$

其中, $W_{ij}$ 为空间权重矩阵,本文使用反地理距离平方矩阵,即 $W_{ij} = 1/d_{ij}^2$ , $d_{ij}$ 为县域*i*与县域*j*之间的球面距离。*ability*为技术能力,分为数据获取能力和数据应用能力,其余变量与基准回归模型保持一致。其中,数据获取能力为县域层面数据获取能力关键词词频的自然对数( $\ln words$ ),数据应用能力为县域层面数字经济发明专利数的自然对数( $\ln cdpat$ )。需要说明的是,选择反地理距离平方矩阵是基于如下考量:①理论依据。根据重力模型(Gravity Model)理论,地区间相互作用的强度与距离的平方成反比,因此,使用反地理距离平方矩阵可以较好反映地区间相互作用和影响的实际情况。②事实依据。为了验证大数据要素集聚的非线性空间溢出,本文先使用一般的空间权重矩阵,按照100千米的步进距离,依次对式(4)做回归,即每增加100千米做一次回归,一直到1500千米。结果发现,随着空间距离的增加,大数据要素集聚与数据获取能力交互项的空间溢出

<sup>①</sup> 空间效应分解参见《中国工业经济》网站([ciejournal.ajcass.com](http://ciejournal.ajcass.com))附件。

效应先变大后减小,符合二次函数的特征。<sup>①</sup>

表4第(1)列汇报了式(3)的回归结果,第(2)、(3)列为式(4)的回归结果,其中,第(2)列是数据获取能力(*lnwords*)的回归结果,第(3)列是数据应用能力(*lncdpat*)的回归结果。总的看,在回归中引入空间权重矩阵后,生产率区域差距的自回归系数 $\rho$ 均显著为正,表明生产率区域差距存在空间自相关,即存在所谓的“同群效应”。这种效应可能源于技术能力的“同群”。具体而言,技术能力低的县域,其周边地区的技术能力可能也较低,因此,该县及周边地区与前沿地区的生产率差距较大。此外,直接效应项 *bigdata* × *post* 的回归系数在1%的水平上显著为正,说明在考虑政策的空间溢出后,大数据要素集聚依然扩大了本地区与前沿地区的生产率差距。同时,政策虚拟变量的空间滞后项 (*W* × *bigdata* × *post*) 也显著为正,说明本县国家大数据综合试验区的设立会扩大周边县域的生产率区域差距,换言之,本县大数据要素集聚的影响会溢出到周边地区,即大数据要素集聚具有空间溢出作用(间接效应),空间放大效应的第一步得以验证。

表4 空间放大效应检验结果

	(1)	(2)	(3)
	<i>GapTFP10</i>	<i>GapTFP10</i>	<i>GapTFP10</i>
<i>bigdata</i> × <i>post</i>	0.0016*** (0.0002)	0.0015*** (0.0003)	0.0023*** (0.0002)
<i>W</i> × <i>bigdata</i> × <i>post</i>	0.0033*** (0.0011)	0.0033** (0.0014)	0.0007 (0.0015)
<i>bigdata</i> × <i>post</i> × <i>ability</i>		0.0019*** (0.0002)	0.0013*** (0.0002)
<i>W</i> × <i>bigdata</i> × <i>post</i> × <i>ability</i>		-0.0281*** (0.0097)	-0.1810*** (0.0206)
<i>ability</i>		0.0010*** (0.0003)	-0.0010*** (0.0002)
<i>W</i> × <i>ability</i>		-0.0491*** (0.0064)	-0.0091** (0.0041)
$\rho$	0.9633*** (0.0057)	0.8905*** (0.0121)	0.9395*** (0.0078)
控制变量	是	是	是
县域固定效应	是	是	是
年份固定效应	是	是	是
观测值	23100	16951	16951
R <sup>2</sup>	0.2290	0.3234	0.3125

注:由于SDM模型回归需要平衡面板数据,故第(2)、(3)列样本观测值有所减小。表中结果基于Quasi-Maximum Likelihood Estimation(QMLE)方法估计得到。

从表4第(2)、(3)列的回归结果看,大数据要素集聚与技术能力相互作用的直接效应(*bigdata* × *post* × *ability*)显著为正,即在技术能力的作用下,大数据要素集聚显著扩大了本地区与前沿地区的生产率区域差距,这实际上也进一步验证了前文的机制。但两者交互项的间接效应(*W* × *bigdata* ×

① 具体结果参见《中国工业经济》网站([ciejournal.ajcass.com](http://ciejournal.ajcass.com))附件。

$post \times ability$ )显著为负,表明在技术能力的作用下,大数据要素集聚的间接效应能够缩小其他地区与前沿地区的生产率差距。从系数的绝对值看,直接效应的影响远小于间接效应,这是导致生产率区域差距被进一步扩大的主要原因。具体而言,在技术能力较低的地区,大数据要素集聚不仅没有缩小集聚地的生产率区域差距,反而加剧了该地区与前沿地区之间的生产率差距。与此同时,对于其他地区尤其是技术能力较强的前沿地区,由于可以远程利用包括低技术能力地区在内所有地区的大数据要素并赋能自身全要素生产率增长,这会导致生产率的极化并加剧集聚地与前沿地区的生产率差距。概言之,在技术能力的作用下,集聚地与前沿地区的差距会随着前沿地区生产率的快速增长而进一步扩大,即在直接效应和间接效应的共同作用下,大数据要素集聚对生产率区域差距的扩大效应会被进一步放大,本文假说3提出的空间放大效应得到支持和回应。

### 2. 空间放大效应的再检验

若“空间放大效应”成立,则大数据要素集聚对生产率区域差距的影响应该在小范围内最小,随着空间范围的扩大这一影响逐渐变大。这是因为,在其他条件不变的情况下,小范围内的县域技术能力可能相差不大,其对大数据要素的使用效果可能并无差异,因此,这些县域与所处区域内的前沿相比差距可能也较小。随着范围的扩大,县域之间的技术能力参差不齐,大数据要素的生产率赋能效应在不同县域之间就可能存在较大差异,并且更大的空间范围意味着具有较强技术能力地区的可用数据资源越多,其大数据要素的规模效应和生产率赋能效应也就越大,结果就是这些县域与所处区域内的前沿地区相比差距可能会变大,即生产率的分化会更为明显,所以大数据要素集聚对于生产率区域差距的影响也会越大。需要强调的是,这一差距是大于不考虑空间分离特性时的差距,也就是说,大数据要素集聚对生产率区域差距的影响被空间分离特性放大了。为了验证上述猜想,本文对基准回归模型中的被解释变量进行如下调整:依次将县域*i*所在城市、省份、地区内生产率最高的县作为前沿地区,然后重新计算市域内生产率差距( $Gapctfp$ )、省域内生产率差距( $Gapptfp$ )和地区内生产率差距( $Gapwetfp$ ),其中,地区划分为东部和西部两个地区,原因在于,从大数据要素集聚的分布情况看,东西分布差异最为明显。例如,对于第*t*年东部地区*p*省*c*市的*i*县而言,其生产率区域差距可以是3个维度上的:①城市*c*内部的差距  $Gapctfp_{it}$ ,即城市*c*中前沿县的生产率与县域*i*生产率的比值;②省份*p*内部的差距  $Gapptfp_{it}$ ,即省份*p*中前沿县的生产率与县域*i*生产率的比值;③东部地区内部的差距  $Gapwetfp_{it}$ ,即东部地区中前沿县的生产率与县域*i*生产

表 5 空间放大效应的再检验

变量	(1)	(2)	(3)
	$Gapctfp$	$Gapptfp$	$Gapwetfp$
$bigdata \times post$	0.0014*** (0.0002)	0.0019*** (0.0004)	0.0038*** (0.0004)
常数项	1.0025*** (0.0077)	1.0194*** (0.0106)	1.0562*** (0.0138)
控制变量	是	是	是
县域固定效应	是	是	是
年份固定效应	是	是	是
观测值	22694	22694	22694
R <sup>2</sup>	0.5174	0.4416	0.5566

率的比值。表5汇报了大数据要素集聚对各县域与不同范围内前沿的生产率差距的回归结果。可以发现,随着空间范围的扩大,政策虚拟变量的回归系数逐渐变大,<sup>①</sup>从系数大小看,对地区内生产率差距的回归系数是市域内生产率差距的3倍左右,是省域内生产率差距的2倍,说明空间放大效应确实存在,上述猜想得以验证,空间放大效应再次得到检验。

## 七、结论与启示

大数据作为新型生产要素,是数字化、网络化、智能化的基础。发挥中国数据规模庞大和数据应用场景丰富的双重优势,对于激发大数据要素集聚的生产率赋能效应、促进区域协调发展和发展新质生产力具有重要意义。本文聚焦于技术能力,在适宜性技术进步理论的框架下,首先以大数据要素的非稀缺性、知识密集性和场景依赖性三个异质性特征为起点,挖掘出大数据要素所需的两种重要技术能力:数据获取能力和数据应用能力。然后,围绕技术能力缺口,分别从数据获取能力缺口、数据应用能力缺口两个方面检验大数据要素集聚扩大生产率区域差距的影响机制。最后,基于大数据要素的空间分离特性,从空间的角度考察了生产率区域差距扩大的空间放大效应。主要结论如下:①实证结果显示,大数据要素集聚显著扩大了集聚地与前沿地区的生产率差距,表明作为知识密集型生产要素,大数据的集聚效应并不是古典经济学中“投入—产出”的必然结果,而是建立在与要素禀赋相匹配的技术能力基础之上。②机制分析表明,大数据要素集聚在技术能力缺口,具体是数据获取能力缺口和数据应用能力缺口的作用下加剧了生产率区域差距,即技术能力缺口抑制了大数据要素集聚的生产率赋能效应。③进一步研究发现,大数据要素存储与使用的空间分离特性强化并放大了技术能力对生产率区域差距的影响,即存在空间放大效应。上述结论为发挥大数据要素规模报酬递增优势、提高各类要素协同效率,优化资源配置,推动高质量发展极具理论与现实意义。据此,本文提出如下政策启示:

(1)提升落后地区的技术能力。机制分析表明,技术能力缺口会抑制大数据要素集聚的生产率赋能效应从而扩大生产率区域差距。因此,提升大数据要素集聚地的技术能力,不仅能缩小区域间生产率差距,还将为大数据要素的应用奠定坚实基础。建立专项财政支持机制,加大政府投入,为落后地区企业、科研机构等提供稳定的研发资金和技术开发补贴,降低技术创新的资金压力,鼓励大数据技术项目落地。通过科研资源下沉,鼓励一线科研机构、高校与落后地区的本地企业建立合作,联合开展技术攻关,增强当地的数据处理和应用能力。在此过程中,政府应针对性地提供税收优惠、融资担保等配套政策,为企业减负。加强落后地区的人才培养,设置专项培训计划和职业教育,提升当地技术人才的数据技能 and 创新能力。同时,政策上鼓励高端技术人才向落后地区流动,通过财政补贴和住房支持等措施吸引大数据领域的高级人才,建立大数据领域的人才支撑体系。通过这些措施,逐步弥补落后地区的技术能力缺口,为大数据要素禀赋和生产率赋能提供有力支持。

(2)强化空间协同效应。空间放大效应表明,空间分离特性叠加技术能力会放大大数据要素集

<sup>①</sup> 虽然与全国前沿差距(基准回归)相比,对地区内生产率差距的回归系数更大,但这进一步验证了时空分离特性的影响可能是非线性的,并且存在一个较优范围。从每百千米步进距离的空间溢出效果看,大数据要素集聚与技术能力交互项的间接效应随距离的增加先变大后减小,表明大数据要素集聚的间接效应或者时空分离特性确实存在一个较优范围,同时也说明本文选用反地理距离平方矩阵是合理的。详细结果参见《中国工业经济》网站(ciejournal.ajcass.com)附件。



聚对生产率区域差距的影响,因此,需要建立跨区域合作机制,推动技术和资源在更大空间内的优化配置。在有条件的区域内设立“区域大数据协同发展平台”试点,以鼓励各地共享大数据资源、研发成果并进行联合创新。具体而言,地方政府应率先破除行政壁垒,通过减少跨区域的审批流程和资源限制,推动大数据要素在多个地区的自由流动,营造合作共赢的创新氛围。通过建立区域大数据创新网络,将技术资源相对丰富的地区与周边落后地区进行对接,发挥溢出效应。例如,依托东部沿海发达地区的技术和人才优势,带动中西部地区共同提升技术能力。地方政府应鼓励跨区域的企业联盟和产学研合作,通过共建实验室、联合研发等方式促进技术合作。政府应出台针对大数据技术应用的专属法规,防止数据跨地区流动过程中产生的不正当竞争行为,从而确保区域协同能够真正落地实施,实现全要素生产率在各地区的共同提升。

(3)加强大数据要素集聚的引导与监管。鉴于目前大数据要素集聚扩大了生产率区域差距,短期之内要更加注重大数据要素的合理集聚,加强对大数据要素集聚的合理引导和市场监管。鼓励大数据要素资源丰富的地区充分发挥其要素禀赋和技术优势,通过政策引导将其打造为大数据要素集聚高地。例如,支持技术发达地区在数据存储、云计算等核心技术上不断优化,推动其形成辐射带动能力,增强对周边地区的溢出效应。针对技术能力较弱的地区,政府应加大扶持力度,为其提供技术转移支持与人才培养,帮助其弥补在数据获取和应用上的能力短板。考虑设立技术转移基金,扶持大数据企业在落后地区设立分支机构,实现技术外溢。加强对大数据市场的监管,避免大数据资源过度集中而形成垄断现象,导致市场秩序失衡。政府应设立透明的市场准入标准和要求,鼓励数据资源的合理竞争,确保大数据要素在各区域间的良性循环,实现生产率赋能的最大化。

(4)实施差异化政策。空间放大效应的再检验结果表明,大数据要素集聚对生产率区域差距的影响在不同区划范围内存在差异。因此,在政策设计上应因地制宜,实施差异化管理,以便充分发挥各地区大数据要素的生产率赋能效应。对于具备较强技术能力的地区,应更加注重区域内部的均衡发展,通过设立技术共享机制,进一步缩小区域内部的技术能力差距。例如,通过“数据资源共享计划”促使核心城市的数据资源向周边区域有序流动,提升区域内部生产率水平。针对其他技术能力较为薄弱的地区,应重点进行政策倾斜,尤其是在财政投入和资源配置上予以重点支持。例如,在这些地区设立大数据孵化园区,培育创新企业,提供低成本的融资渠道,逐步形成大数据创新的产业链。对于资源优势明显但经济发展相对滞后的区域,可以推动其建立大数据领域的特色产业集群,提供差异化的政策扶持,鼓励其集中发展适合当地特点的创新技术,最终实现区域间生产率的动态平衡和可持续发展。

目前,大数据要素集聚的测度还存在瑕疵,生产率区域差距的研究也不够全面,因此,本文的研究可以进一步扩展,深入探讨县域层面的大数据要素及其集聚的测度问题,并衍生出更加全面的区域发展研究,如大数据要素对于共同富裕的影响等。

#### 〔参考文献〕

- [1] 龚斌磊. 中国农业技术扩散与生产率区域差距[J]. 经济研究, 2022, (11): 102-120.
- [2] 郝大江, 张荣. 要素禀赋、集聚效应与经济增长动力转换[J]. 经济学家, 2018, (1): 41-49.
- [3] 贾俊雪, 郭庆旺, 宁静. 财政分权、政府治理结构与县级财政解困[J]. 管理世界, 2011, (1): 30-39.
- [4] 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.
- [5] 金飞, 张琦. 中国市区县级TFP变动问题的讨论: 2007—2010年[J]. 数量经济技术经济研究, 2013, (9): 55-71.

- [6]孔宪丽,米美玲,高铁梅.技术进步适宜性与创新驱动工业结构调整——基于技术进步偏向性视角的实证研究[J].中国工业经济,2015,(11):62-77.
- [7]李唐,李青,陈楚霞.数据管理能力对企业生产率的影响效应——来自中国企业—劳动力匹配调查的新发现[J].中国工业经济,2020,(6):174-192.
- [8]林毅夫,张鹏飞.适宜技术、技术选择和发展中国家的经济增长[J].经济学(季刊),2006,(3):985-1006.
- [9]权小锋,吴世农,文芳.管理层权力、私有收益与薪酬操纵[J].经济研究,2010,(11):73-87.
- [10]邵帅,张可,豆建民.经济集聚的节能减排效应:理论与中国经验[J].管理世界,2019,(1):36-60.
- [11]孙伟增,毛宁,兰峰,王立.政策赋能、数字生态与企业数字化转型——基于国家大数据综合试验区的准自然实验[J].中国工业经济,2023,(9):117-135.
- [12]孙志燕,侯永志.对我国区域不平衡发展的多视角观察和政策应对[J].管理世界,2019,(8):1-8.
- [13]王超贤,张伟东,颜蒙.数据越多越好吗——对数据要素报酬性质的跨学科分析[J].中国工业经济,2022,(7):44-64.
- [14]吴江,陶成煦.激活数据要素赋能千行万业——《“数据要素×”三年行动计划(2024—2026年)》政策解读[J].情报理论与实践,2024,(3):16-19.
- [15]谢康,夏正豪,肖静华.大数据成为现实生产要素的企业实现机制:产品创新视角[J].中国工业经济,2020,(5):42-60.
- [16]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.
- [17]许宪春,任雪,常子豪.大数据与绿色发展[J].中国工业经济,2019,(4):5-22.
- [18]杨金玉,彭秋萍,葛震霆.数字化转型的客户传染效应——供应商创新视角[J].中国工业经济,2022,(8):156-174.
- [19]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.
- [20]杨汝岱,李艳,孟珊珊.企业数字化发展、全要素生产率与产业链溢出效应[J].经济研究,2023,(11):44-61.
- [21]余明桂,范蕊,钟慧洁.中国产业政策与企业技术创新[J].中国工业经济,2016,(12):5-22.
- [22]张军,吴桂英,张吉鹏.中国省际物质资本存量估算:1952—2000[J].经济研究,2004,(10):35-44.
- [23]Acemoglu, D., and F. Zilibotti. Productivity Differences [J]. Quarterly Journal of Economics, 2001, 116(2): 563-606.
- [24] Battese, G. E., and T. J. Coelli. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data [J]. Empirical Economics, 1995, 20: 325-332.
- [25] Belotti, F., and G. Haddad. Consistent Inference in Fixed-Effects Stochastic Frontier Models [J]. Journal of Econometrics, 2018, 202(2): 161-177.
- [26] Jones, C. I., and C. Tonetti. Nonrivalry and the Economics of Data [J]. American Economic Review, 2020, 110(9): 2819-2858.
- [27] Krugman, P. Increasing Returns and Economic Geography [J]. Journal of Political Economy, 1991, 99(3): 483-499.
- [28] Tone, K. A Slacks-based Measure of Super-efficiency in Data Envelopment Analysis [J]. European Journal of Operational Research, 2002, 143(1), 32-41.
- [29] Wang, H. J., and C. W. Ho. Estimating Fixed-Effect Panel Stochastic Frontier Models by Model Transformation [J]. Journal of Econometrics, 2010, 157(2): 286-296.

## **Big Data Factor Agglomeration, Technological Capability Gaps and Regional Disparities in Productivity**

ZHANG Guo-sheng<sup>1,2</sup>, YAN Peng<sup>1</sup>, LI Xin-jue<sup>1,2</sup>, DU Peng-fei<sup>3</sup>

(1. School of Economics, Yunnan University;

2. Institute of Yunnan Digital Economy, Yunnan University;

3. The State Information Center)

**Abstract:** After more than four decades of rapid growth under reform and opening, the economic growth dependent on traditional factor inputs has gradually slowed, and the Chinese economy now faces structural challenges, including a lack of new growth drivers and unbalanced regional development. With the rapid advancement of digital technologies, big data, as a fundamental and decisive new factor in digital economy production, increasingly enhances total factor productivity (TFP), offering a potential solution to the growth constraints of traditional factor inputs. In this context, big data is expanding rapidly and clustering within specific regions. Consequently, the productivity-enhancing effects of big data agglomeration are anticipated to serve as a catalyst for fostering high-quality regional economic growth and coordinated regional development. However, the regional disparities in TFP within China are widening, indicating a clear deviation between policy intentions and regional outcomes, highlighting a need for an in-depth investigation into the underlying causes of this disparity.

Based on this, this paper constructs county-level indicators to assess regional productivity disparities and technological capability gaps, aiming to explore the economic relationships among big data factor agglomeration, technological capability gaps, and regional TFP disparity. This paper finds that big data factor agglomeration significantly intensifies the productivity disparity between agglomerated and frontier regions. This disparity arises from the unrealized productivity-enhancing potential of big data, which is constrained by gaps in data acquisition and application capabilities. In other words, technological capability gaps inhibit big data's productivity-enhancing effects. Additionally, spatial separation in big data storage and utilization further amplifies these regional productivity disparities through spatial effects. These findings validate this paper's proposition that "the agglomeration effect of big data is not an automatic outcome of classical 'input-output' dynamics but requires a foundation of essential technological capabilities".

The potential contributions of this paper are as follows. Firstly, it analyzes the technical prerequisites for the big data factor to achieve productivity-enhancing effects from the standpoint of technological capabilities, offering insights into why big data factor agglomeration may widen regional productivity disparity. Secondly, based on the theory of appropriate technological progress, it establishes a theoretical framework and transmission mechanism that illustrates how big data factor agglomeration, technological capability gaps, and spatial amplification effects shape regional productivity disparities. Thirdly, this paper calculates TFP at the county level across China from 2011 to 2021 and develops technological capability indicators using data from national business registrations, matching these with county-level data, thereby providing a useful reference for macro-level research. Lastly, presenting novel findings on big data factor agglomeration and regional productivity disparities, this paper contributes fresh empirical evidence for productivity convergence and coordinated regional development in the digital economy era. It offers a reference framework for policy decisions aimed at maximizing the economic multiplier and productivity-enabling effects of the big data factor to drive high-quality growth.

**Keywords:** big data factor; technological capability gaps; TFP; agglomeration; regional disparities

**JEL Classification:** O33 O40 R11

[责任编辑:崔志新]