

# 公共数据开放能否促进 区域协调发展？<sup>\*</sup>

——来自政府数据平台上线的准自然实验

方锦程 刘颖 高昊宇 董纪昌 吕本富

**摘要：**公共数据是公共服务的重要基础资源，评估其对区域协调发展的经济影响是科学施政的前提。本文以我国地方政府推动公共数据开放平台上线为准自然实验，基于2009~2021年295个城市级面板数据开展实证分析，检验公共数据开放是否促进了区域协调发展。研究发现，公共数据开放显著降低了城市内区域发展水平差距，促进了区域协调发展。机制分析表明，该影响通过破除区域信息壁垒、弥合区域资源禀赋差距的作用机制予以实现。此外，公共数据开放不仅促进落后区域后发追赶，也提升区域整体发展水平，实现发展与共富的有机统一。本文量化评估了数据要素开放共享的经济效应，对于深化数据资源体系建设，激发数据要素潜能具有重要启示意义。

**关键词：**公共数据开放 数据要素 区域协调发展

DOI:10.19744/j.cnki.11-1235/f.2023.0112

## 一、引言

数据作为新型生产要素，已成为数字中国建设和经济高质量发展的战略性资源。习近平总书记强调要“发挥数据的基础资源作用和创新引擎作用”<sup>①</sup>，实现数据的高效流通和使用，为实体经济赋能。如何评估和释放数据要素的经济价值、发挥数据要素在资源配置中发挥引领优化和协调作用，是学术研究、社会实践和政策制定等领域共同关注的重要话题。我国具有丰富的数据资源基础，据统计，2021年全国数据资源总存储量达到598.4EB，同比增长27.4%，占全球数据总存量的14.1%<sup>②</sup>。公共数据是占比最大的数据资源，在数据要素开发利用中具有基础性、引领性和示范性作用。对此，《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》明确提出要“加强公共数据开放共享”，目标是将公共数据纳入公共服务体系，以增强公共数据服务的均衡性和可及性。2022年12月，《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》指出“推进实施公共数据确权授权机制”，为激活数据要素价值提供了基础性制度保障。2023年2月，《数字中国建设整体布局规划》进一步指出，推动公共数据汇聚利用是畅通数据要素大循环的重要环节。可见，公共数据已然成为数据要素供给的重点任务，推动公共数据开放对释放数据要素红利、驱动经济高质量发展具有重要意义。

从概念上看，公共数据是指具有公共属性的数据资源，不仅包括政府内部的政务数据，而且还涵盖了企业等市场主体在商业活动中产生与社会公共利益相关的数据。公共数据开放则是指政府整合公共数据资源，开放给社会主体平等使用的过程，不仅为公众提供了数据的知情权与查询权，更重要的是保障了公众对数据要素的使用权，社会主体能够直接下载或借助开放API提取使用公共数据，更好地进行企业选址、资源发现和产业升级，创造商业和社会价值（纳加拉吉，2022）。我国地方政府自2012年开始陆续上线的公共数据开放平台，汇聚了海量的政府及社会主体的数据资源，并以此推动数据要素开放共享，已成为激活数据要素潜能、促进区域经济高质量发展的重要举措。国内外学者普遍认为公共数据开放有助于促进数据要素融

<sup>\*</sup> 本项研究得到国家社科基金重大项目“健全互联网领导和管理体制研究”（基金号：22ZDA079）的资助。感谢匿名评审专家的宝贵意见，感谢第四届互联网与数字经济论坛王榆靖博士及其他参会学者的宝贵建议，当然文责自负。刘颖为本文通讯作者。

入生产体系,是驱动经济增长的重要因素(戈德法布、塔克,2019;陈晓红等,2022),但是关于公共数据开放如何改变经济发展的空间格局,能否促进区域协调发展,相关研究仍缺乏微观实证证据,未形成一致性结论。

公共数据开放作为一种数字技术创新应用,其技术扩散属性可能导致经济发展不均衡,但作为一种公共数据资源的供给服务,其公共资源属性则可能有助于促进区域协调发展。一方面,从技术扩散的视角来看,创新技术通常在发达地区出现、集聚并扩散。发达地区不仅具有远超落后地区的技术实力,而且在对技术的使用上同样具有显著优势(罗杰斯等,2014),数字技术的技能偏向特征使其对经济增长的促进作用通常在高收入和高技能的地区更为明显。因此,数字技术创新将造成更为严重的区域经济发展不均衡和不平等(约特、保尔森,2019;查克费等,2023)。一个典型例证是区域发展中广泛存在的“数字鸿沟”现象,其深层原因是发达地区具有更大的互联网接入机会和更强的数字技术使用能力(邱泽奇等,2016),由此造成了更为显著的区域发展水平差距(福尔曼等,2012;张勋等,2021)。另一方面,从公共资源的视角来看,公共数据开放提高数据资源的普惠化水平,有助于降低数据要素的使用门槛,一定程度上起到破除信息壁垒、弥合资源禀赋差距的作用(纳加拉吉,2022),可能促进区域协调发展。一个典型案例是上海市大数据普惠金融应用,金融机构将上海市公共数据开放平台提供的高质量公共数据与内部数据有机结合,为中小微企业进行精准画像,提升了区域内金融资源供给的均衡性。综上所述,公共数据开放的“双重属性”为理论上思考其对区域协调发展的经济后果带来新的挑战,而且现有的国内外文献尚未对该问题展开深入系统地分析,因此分析公共数据开放能否以及如何影响区域协调发展是一个重要的实证问题。本文基于我国地方政府推动公共数据开放平台上线的准自然实验,从微观实证角度检验公共数据开放能否促进区域协调发展,并探讨相关经济机制。

本文采用地方政府推动公共数据开放平台上线作为公共数据开放的外生政策冲击,利用多时点双重差分法开展实证分析。其代表性在于:目前我国信息数据资源80%以上掌握在各级政府部门手里<sup>①</sup>,而公共数据开放平台是连接公共数据供给端和使用端的载体,面向社会提供可再利用的、具有公共属性的原始数据。本文结合DMSP/VIIRS地球夜间灯光数据的空间分布刻画区域经济发展的协调性,基于2009~2021年295个城市一年度层面的3835个面板数据,实证考察公共数据开放能否促进区域协调发展,并探讨相关经济机制。研究发现,公共数据开放平台上线后,区域发展水平差距平均下降4.23%,起到了促进区域协调发展的作用,且该结论通过了平行趋势检验、安慰剂检验和工具变量回归等一系列稳健性检验。机制分析表明破除区域信息壁垒、弥合区域资源禀赋差距是公共数据开放促进区域协调发展的经济机制。具体而言,首先,公共数据开放具有破除交通地理、行政壁垒、文化差异等多种因素形成的区域信息壁垒的作用,因此促进区域协调发展。其次,公共数据开放能够增强对稀缺资源发掘利用的能力,并降低对传统资源投入发展经济模式的依赖,通过弥合区域资源禀赋差距的作用促进区域协调发展。进一步地,研究提供直接证据表明公共数据开放不仅促进落后区域后发追赶,也提升了区域整体发展水平,实现发展与共富的有机统一。

本文的边际贡献有以下几点。

第一,本文对纳入数据生产要素的经济增长理论提供了微观证据。现有研究主要从理论上探讨数据要素的经济价值(加塞马加海、卡利克,2019;陈晓红等,2022),鉴于公共数据是中国情景下数据要素的最大组成部分,本文借助地方政府推动数据开放的独特政策实验,直接评估了数据生产要素非稀缺性和非排他性的经济价值。同时,本文还为数据要素赋能传统要素的配置优化和效能提升理论提供了微观证据,实证发现了公共数据开放有助于市场主体利用数据赋能经济决策,通过破除区域信息壁垒、弥合区域资源禀赋差距两种方式发挥其经济效应。研究结论揭示了数据要素与传统要素协同联动的经济作用。

第二,本文揭示了公共数据开放对区域协调发展的效应。现有研究大多关注公共数据开放促进发展的整体性作用(布林约尔松、麦克埃尔赫兰,2016;纳加拉吉等,2020),而本文则聚焦公共数据开放能否促进区域协调发展的问題。借助城市层面的微观数据,本文从创新技术与公共资源的不同视角辨析了公共数据开放的经济影响,并验证了公共数据开放促进区域协调发展的公共资源效应更为显著。研究结论表明了数据要素兼具

创造财富与共享财富的属性,拓展了评估公共数据开放经济效应的相关文献,并为推动数据要素开放共享、促进区域协调发展提供了新的研究视角与思路。

第三,与以往研究中国区域经济发展不平衡的文献聚焦于省域或城市群层面的研究对象不同(刘修岩等,2017;李彦龙、沈艳,2022),本文使用微观城市层面更精细化的数据,综合利用夜间卫星灯光、工商企业注册、全国路网矢量、银行分支机构等细粒度数据刻画了城市要素资源的空间分布结构,更清晰地探讨了公共数据开放促进区域协调发展的传导机制及经济影响。研究结论为数据要素开放共享提供了更加直接、更加丰富的实证证据。

## 二、制度背景与研究假设

### (一)中国公共数据开放的发展历史及现状

我国政府高度重视公共数据资源的制度建设。2015年9月,《国务院关于印发促进大数据发展行动纲要的通知》强调政府推动公共数据开放共享是缓解数据要素供给不足的有力手段,以政务数据公开共享带动企业等市场主体的数据采集和开放,能够使沉睡的数据被应用起来。2022年1月,国务院印发了《“十四五”数字经济发展规划》,提出要“建立健全国家公共数据资源体系,统筹公共数据资源开发利用”,“提升公共数据开放水平,释放数据红利”,其目的在于充分发挥数据要素对市场融通和资源配置的促进作用,打破数据资源有限供给对经济发展的制约。2022年6月,国务院印发了《关于加强数字政府建设的指导意见》,将“构建开放共享的数据资源体系”作为重点工作任务之一,提出“构建统一规范、互联互通、安全可控的国家公共数据开放平台”的要求。2022年12月,《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》进一步明确了“推进实施公共数据确权授权机制”,要求全面增强各地区、各部门的公共数据统筹管理能力,推动公共数据的汇聚融合、共享交换和开发应用,发挥公共数据的流通效能和经济社会价值。

从现有实践经验上看,公共数据开放的两点特性与其核心价值密不可分。其一是数据性,即开放的数据未经过加工或解读,这有助于使用者对原始数据进行整合、提炼与分析,以数据优化经济决策的形式,发掘数据的潜在价值。其二是公共性,即政府推动的数据要素开放不仅涵盖公共管理和服务机构的大部分数据,在数据获取上也存在非竞争性与非排他性。公共数据不能仅封闭在政府和少数大企业内部进行开发利用,而是要面向全部社会主体,依靠市场的力量激发数据要素活力。因此,公共数据开放不仅为社会主体参与数据开发利用开辟了全新渠道,也为革新区域经济发展模式创造了难得机遇,对区域经济的发展 and 共富具有促进作用。

在国家政策指引下,我国各级政府不断建设公共数据开放平台作为主要的公共数据供给机制,通过打通各级部门数据流通渠道的方式,依照统一数据开放目录向市场开放公共数据平台。公共数据开放平台上线后,一方面,依照不同数据的开放等级,社会主体可以零成本下载或低成本申请数据,这降低了数据资源获取成本。另一方面,公共数据开放平台也为社会主体提供了数据提交入口,社会主体可将其采集、整理的数据提交给平台,经过审核后向社会开放,这丰富了公共数据的来源范围。更多数据资源的流通提高了公共数据利用效率,降低了市场中的信息搜寻、获取与甄别成本,促进市场经济的健康稳定发展。

图1展示了中国城市级(含直辖市与地级市)公共数据开放平台数量的增长趋势。从时间顺序看,中国公共数据开放平台的建

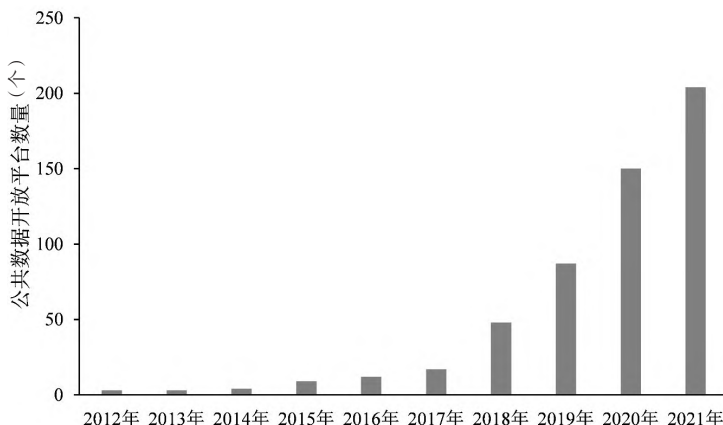


图1 历年城市级公共数据开放平台的数量增长趋势



设始于2012年北京与上海等地率先上线公共数据开放平台,之后平台数量呈现逐年加速增长的趋势。截止到2021年底,中国城市级公共数据开放平台数量已达到204个,占比达69.15%<sup>④</sup>。可以预见,公共数据开放面向社会供给丰富的数据要素,将引发传统经济社会生产体系变革,对我国未来的发展产生重大而深远的影响。

## (二)理论分析与研究假设

公共数据开放作为我国首个大规模推进数据要素开放共享、促进数据要素价值释放的政策试验,对区域经济发展模式的革新具有重要影响。既有文献认为,公共数据开放主要具有两方面的经济特性。第一,公共数据开放面向社会供给高质量的数据要素,有助于市场主体利用数据赋能经济决策(纳加拉吉,2022),促进生产效率的提升与经济高质量发展。公共数据开放平台不仅降低市场主体数据搜寻、获取和验证成本(戈德法布、塔克,2019),而且有助于企业从公共数据中挖掘信息进行选址、创新、择时销售等一系列需要准确把握商机的经济活动(弗霍尔斯特、安德鲁,2016;弗尔曼等,2021)。同样公共数据开放也为数据市场的规范发展树立标杆,有助于促进全社会数字技术与数字经济的发展,具有一定的价值溢出效应。第二,与传统公共基础资源不同,数据要素具有可复制性、高共享性及非稀缺性(陈晓红等,2022),面向社会提供数据公共品(纳加拉吉,2022)。因此,公共数据开放具有共享性与普惠性,这有助于破解传统数据要素融入生产体系中存在的高获取门槛、大企业垄断等现象,促使数据要素的开放共享惠及到所有经济群体。从公共数据开放的两点经济特性来看,其有助于促进数据要素融入生产体系,变革传统经济增长模式。

公共数据开放促进经济发展的作用较为明确。然而,现有文献对公共数据开放如何重塑区域经济发展模式尚不明确,且多数关注公共数据开放的经济增长价值(弗霍尔斯特、安德鲁,2016;纳加拉吉等,2020),而对其促进协调发展的价值尚未形成统一的结论。从技术创新视角和公共资源视角两方面对公共数据开放的经济影响进行分析,可以得出公共数据开放对区域协调发展的影响存在阻碍与促进两种对立的逻辑观点。下面将分别阐述两类观点并给出详实的实证证据进行探究。

第一种逻辑观点将公共数据开放视为一种数字创新技术。第一,发达地区与落后地区存在技术发展的差异。发展创新技术是一项具有高度不确定性的风险行为,通常需要参与主体进行较高的前期投入并承担难以预测的未知风险,因此会优先在具有雄厚经济实力与风险承担能力的发达地区出现并集聚,优先对发达地区的经济发展产生促进作用(罗杰斯等,2014)。随着技术知识的传播与扩散,发达地区会对落后地区产生技术溢出效应,促使技术要素向落后地区转移(铃村,1992),带动落后地区的经济发展。在技术发展水平上发达地区与落后地区存在时间上的先后差异,因此可能形成区域发展水平差距。第二,发达地区与落后地区存在技术使用能力的不平等。创新技术的经济价值取决于其使用场景与方式,技术使用能力越强的地区在经济上受益越大(邱泽奇等,2016),能够发掘利用公共数据的更多经济价值。除此以外,数据要素具有规模报酬递增的效应,其叠加融合将使其更易于分析,有助于放大数据要素本身的经济价值(陈晓红等,2022)。数据要素积累较多的区域具有更强的数据要素使用能力,能够更好的释放其经济价值(布埃拉等,2022)。因此,公共数据开放可能起到阻碍区域协调发展的作用。基于上述分析,本文提出如下假设。

假设H1a:公共数据开放强化发达地区的技术发展优势和使用能力优势,阻碍区域协调发展。

第二种逻辑观点从公共基础资源的视角切入,将公共数据开放视为向市场提供一种非稀缺的数据要素公共品(纳加拉吉,2022)。要素资源空间分布不均形成区域经济发展差距的重要因素(杰内奥利等,2013;莱斯曼,2014;阿莱西纳等,2016),而公共数据开放平台实现了数据要素的均等化分配,有助于缓解区域要素资源的分布不均,促进区域协调发展。首先,公共数据开放能够破除区域信息壁垒,提高区域要素资源配置效率。信息壁垒是制约要素资源配置的重要因素,特别是在市场缺乏透明性的落后区域,其信息不对称情况较为严重,因此其投资和发展风险相对较大,要素资源的市场化配置受到了严重的阻碍(石军伟、付海艳,2010)。在市场机制的调节下,大量要素资源会向信息较为透明的发达地区集中,形成了区域发展水平差距(刘修岩等,2017)。而公共数据开放为市场参与者提供了更多的公共信息,这不仅有助于减少决策风险,且政府背书的高质量数据降低了市场参与者信息搜寻、获取与验证的高昂成本(戈德法布、塔克,2019),提高了商

业项目的经济价值。因此,公共数据开放有助于打通制约要素资源配置的信息壁垒,提升区域市场可及性,促进区域协调发展。

其次,公共数据开放能够弥合区域资源禀赋差距,有助于克服区域要素资源配置不均对经济发展的制约,实现区域协调发展。第一,数据要素具有非稀缺性的特质,在一些生产活动中能够替代相对稀缺的传统要素投入并提升传统要素的利用效率(戈德法布、塔克,2019),这一效应对传统资源禀赋相对稀缺的落后区域尤为重要。例如,公共数据开放为市场主体提供了政府独有的数据信息,使得市场参与者无需为市场经营信息的获取投入大量人力物力,降低相应要素投入成本。第二,数据要素还能与传统要素互补赋能,能够促进要素的合理配置,提高稀缺资源的使用效率(陈晓红等,2022),这种效应在传统要素相对稀缺的后发地区尤为明显。第三,公共数据开放促进数据要素融入生产体系,与传统要素重新组合带动新业态产生(南比桑等,2017),为缺乏数据要素与传统要素投入发展经济的落后区域开拓全新经济增长模式,提供后发追赶的机遇。因此公共数据开放面向社会提供数据要素公共品,保障公众对数据要素的平等使用,可能对落后地区经济发展的促进效应较强。基于上述分析,本文提出如下假设。

假设 H1b:公共数据开放通过破除区域信息壁垒和弥合区域资源禀赋差距,促进区域协调发展。

## 三、研究设计

### (一)模型设定

本文从降低区域发展水平差距的角度考察公共数据开放对城市内区域协调发展的影响。为了探究政策影响的净效应,本文借助双重差分模型对研究对象在公共数据开放平台上线前后的时间差异及个体差异进行控制。因此,本文将上线公共数据开放平台的城市设定为处理组,其余城市设定为控制组,采用多期双重差分模型进行实证检验,具体的回归模型设定如下:

$$INEQ_{it} = \beta_0 + \beta_1 DATA_{it} + \beta_2 CONTROLS_{it} + \delta_i + \delta_t + \delta_{it} + \varepsilon_{it} \quad (1)$$

式中,下标  $i$  表示城市,  $t$  表示年份。被解释变量  $INEQ_{it}$  为城市  $i$  在  $t$  年的内部区域发展水平差距。解释变量  $DATA_{it}$  为公共数据开放平台上线的虚拟变量,当城市  $i$  在  $t$  年上线公共数据开放平台取 1,否则取 0。本文在回归模型中增加了控制变量及固定效应,用以吸收城市自身资源禀赋等对区域发展水平差距的影响。其中  $CONTROLS_{it}$  表示其他控制变量,代表除公共数据开放平台上线以外其他影响城市内区域发展水平差距的因素。 $\delta_i$  为城市固定效应,  $\delta_t$  为时间固定效应,用以吸收回归中不随时间变化与不随城市变化的混淆因素,在一定程度上缓解遗漏变量偏误。为控制城市层面因素在线性维度对估计结果的影响,本文进一步控制了城市固定效应与时间趋势交互项  $\delta_{it}$ 。 $\varepsilon_{it}$  为随机扰动项。考虑到同一城市内样本随机扰动项可能潜在相关,本文将标准误聚类到城市层面。 $\beta_1$  为本文解释变量的回归系数,表示城市公共数据开放对区域发展水平差距的影响,预期该系数显著为负,即公共数据开放降低区域发展水平差距,促进区域协调发展。

### (二)数据来源

本文的数据来源主要分为四大类:(1)城市层面的控制变量数据来自 2009~2021 年《中国城市统计年鉴》,部分缺失值以各省统计局数据进行补齐。(2)中国行政区划信息来源于中华人民共和国民政部,本文手工整理 2009~2021 年各省级、地级、县级、乡级<sup>⑤</sup>的行政区划变更情况。(3)城市内区域发展水平差距度量数据来自于 NOAA 提供的全球夜间灯光成像栅格数据,本文将其与中国城市行政区划信息进行对应后与城市面板进行交集匹配。(4)中国各城市公共数据开放平台上线的数据来源于本文根据公开报道进行手动搜集,并利用复旦大学 DMG 数字与移动治理实验室发布的《中国地方政府数据开放报告(城市)》、华中师范大学信息管理学院发布的《中国政府开放数据利用研究报告》对数据进行交叉验证。在获取原始数据的基础上,本文进行如下处理:(1)考虑到控制变量的数据可得性,删除非城市(地区、自治州、盟)的地级行政单元样本。(2)对主要机制变量与控制变量缺失的样本进行线性插值补充<sup>⑥</sup>。(3)为排除数据中存在的极端值,本文对所有连续变量进行了上下 1%的 Winsorize 处理。最终得到了中国 295 个城市层面 2009~2021 年度的 3835

个面板数据作为研究样本<sup>⑦</sup>。

### (三)变量定义

#### 1. 区域发展水平差距(INEQ)

度量区域发展水平差距的前提是获取区域内部的经济发展水平分布。考虑到城市内细粒度数据可得性,并为克服传统以GDP度量区域经济发展水平的测量误差,本文采用2009~2021年NOAA提供的全球夜间灯光DMSP与VIIRS栅格数据刻画区域经济发展水平。由于DMSP和VIIRS的夜间灯光数据区间分别为1992~2013年和2012~2021年,且两套数据存在空间分辨率、卫星过桥时间及数据连续性的差异(吉布森等,2021)。本文进行如下处理:(1)对原始数据进行基期校准、空间分辨率匹配与连续性矫正处理(吴等,2022)。(2)采用孤立森林方法进一步过滤燃烧、极光及背景噪音等无关特征。(3)将合成DMSP-VIIRS数据与历年中国城市行政区划边界矢量数据进行匹配。(4)参考吴等(2022),利用sigmoid函数<sup>®</sup>对对数化的VIIRS数据进行分城市拟合。经过上述处理,本文最终得到了2009~2021年的合成DMSP-VIIRS夜间灯光栅格数据。

已有大量研究证明,夜间灯光亮度与人类经济活动密切相关(陈、诺德豪斯,2011;吉布森等,2021),且其不受人为造成的统计误差影响,数据的准确性相对较高。考虑到城市内部行政区划变动较多,传统基于行政区划单元度量经济活动空间分布的方式存在较大测量误差,本文借鉴莱斯曼(2014)和吉布森等(2021)的做法,采用城市内夜间灯光栅格的标准差作为城市内区域发展水平差距的度量。细化到栅格层面的度量能够在更小的地理单位上探究城市内部区域发展水平差距,这对于研究如何促进城市内部的协调发展尤为重要。

#### 2. 公共数据开放(DATA)

地方政府推动公共数据开放平台的上线为本文研究提供了理想的准自然实验场景。根据中国公共数据开放的政策实践,本文将公共数据界定为原始性、可机器读取、可供社会化利用、具有公共属性的数据,因此公共数据开放平台是指满足如下3个特征的平台:(1)提供未经过加工或解读的原始数据信息。(2)提供能够直接机器读取,不需要高成本技术手段转化就可使用的开放数据集。(3)所开放的数据集可被所有人直接或申请后利用。

本文主要采用3个步骤进行数据搜集:(1)限定地方政府推动的公共数据开放平台的域名必须包含gov.cn,且该平台为汇聚政府各部门数据的集中式统一数据开放平台,地方政府各个其他部门独立建设的非集中式公共数据开放平台被排除在外。(2)在百度等搜索引擎对城市名称+“数据开放”、“数据资源”、“公共数据”等关键词手动搜索,获取平台的首次上线时间。(3)根据前文定义对公共数据开放平台进行人工筛选后,参考复旦大学DMG数字与移动治理实验室发布的《中国地方政府数据开放报告(城市)》、华中师范大学信息管理学院发布的《中国政府开放数据利用研究报告》对数据进行交叉验证,获得了截止到2021年底共计204个城市公共数据平台的首次上线时间,将其作为城市一年度是否上线公共数据开放平台的虚拟变量,并借助双重差分法进行回归分析。

#### 3. 机制变量

为探究公共数据开放促进区域协调发展的传导机制。首先,区域信息壁垒的形成主要受到交通地理(李兰冰、张聪聪,2022;文雁兵等,2022)、行政壁垒(丁从明等,2018;马光荣、赵耀红,2022)、文化差异(丁从明等,2018;刘倩等,2020)等因素的影响。因此本文采用交通便捷程度(吴群锋等,2021)、行政分割程度(丁从明等,2018)、文化分割程度(丁从明等,2018)度量形成信息壁垒的因素。其次,考虑到区域发展不平衡的主要影响因素是资源禀赋的空间分布不均(莱斯曼,2014),本文采用土地资源禀赋分布(贾俊雪、梁煊,2020)、人口资源禀赋分布(杰内奥利等,2013)、金融资源禀赋分布(蔡庆丰等,2020)度量区域资源禀赋差距,用以验证公共数据开放弥合区域资源禀赋差距的作用。

#### 4. 控制变量

为排除影响城市内区域发展水平差距的资源禀赋因素,在结合相关文献的做法及本文研究需要的基础上,纳入城市金融资源(李成友等,2021)、发展水平(杨继东、罗路宝,2018)、发展能力(马红旗等,2017)、产业



结构(李成友等,2021)、人力资本(庄毓敏等,2020)及外商投资(刘修岩等,2017)作为控制变量。本文详细的变量定义列示于表1。

(四)描述性统计

表2报告了本文主要变量的描述性统计。*INEQ*的平均值为3.7044且中位数为3.4123,表明中国各城市内仍存在发展不平衡的问题,且中国各城市间的区域发展水平差距不存在显著的偏态分布。*DATA*的平均值为0.14,表明本文有14%的样本为处理组。*DATA*的中位数为0,表明本文近一半的样本上线公共数据开放平台的时间相对靠后,我国的公共数据开放仍处于新兴发展状态。其余控制变量的描述性统计特征与相关研究近似(贾俊雪、梁煊,2020;李成友等,2021)。

四、实证结果分析

(一)基准回归结果

表3报告了公共数据开放促进区域协调发展的实证检验结果。在基准回归结果中,本文采用递进式的回归策略。列(1)报告了仅加入解释变量与固定效应的回归结果,公共数据开放(*DATA*)的系数为-0.157且在1%的统计水平上显著,表明公共数据开放显著的降低了区域发展水平差距。列(2)在列(1)的基础上加入了其他控制变量,*DATA*的系数为-0.157且仍在在1%的统计水平上显著。较小的系数变动表明本文选取的政策实验外生性较强,受到资源禀赋因素及其他不可观测因素的潜在影响较小。从经济显著性来看,公共数据开放平台上线后,区域发展水平差距平均下降4.23%<sup>⑨</sup>。不同回归模型设定下公共数据开放的系数均在1%的显著性水平上为负,表明公共数据开放降低区域发展水平差距,促进区域协调发展。本文的假说H1b得到了实证支持。

尽管公共数据开放可能存在促进或阻碍区域协调发展两种不同理论可能,就本文研究的中国地方政府推动的公共数据开放平台上线这一特定情境而言,其经济特性主要偏向于促进区域协调发展的模式。首先,公

共数据开放平台所提供的数据要素是非稀缺的公共品,这与技术要素的稀缺性形成了鲜明的对比。从本

| 表1 主要变量定义  |          |                  |                        |
|------------|----------|------------------|------------------------|
| 变量类别       | 变量名称     | 变量名称             | 测算方式                   |
| 因变量        | 区域协调发展   | <i>INEQ</i>      | 城市内夜间灯光栅格的标准差          |
| 自变量        | 公共数据开放   | <i>DATA</i>      | 城市是否上线公共数据开放平台         |
| 破除区域信息壁垒   | 交通便捷程度   | <i>ROAD</i>      | 城市内道路长度与人口数的比值         |
|            | 行政分割程度   | <i>GOV</i>       | 城市单位面积乡级行政区划数          |
|            | 文化分割程度   | <i>CULTURE</i>   | 城市方言小类的人口加权比重          |
| 弥合区域资源禀赋差距 | 土地资源禀赋分布 | <i>LANDDIST</i>  | 城市内坡度栅格的标准差            |
|            | 土地资源禀赋分布 | <i>LANDPRICE</i> | 城市内土地价格的标准差            |
|            | 人口资源禀赋分布 | <i>POPDIST</i>   | 城市内人口栅格的标准差            |
|            | 金融资源禀赋分布 | <i>BANKDIST</i>  | 城市内各乡级行政区银行数量的标准差      |
| 控制变量       | 城市金融资源   | <i>FINANCE</i>   | 城市年末金融机构贷款余额与国民生产总值的比值 |
|            | 城市发展水平   | <i>PGDP</i>      | 城市人均生产总值的自然对数          |
|            | 城市发展能力   | <i>GDPR</i>      | 城市生产总值同比增长率            |
|            | 城市产业结构   | <i>SERVICE</i>   | 城市第三产业产值占比             |
|            | 城市人力资本   | <i>EDU</i>       | 城市高校学生在校人数与总人口的比值      |
|            | 外商投资水平   | <i>FDI</i>       | 城市外商直接投资项目数的自然对数       |

| 表2 主要变量描述性统计     |      |         |         |         |         |          |
|------------------|------|---------|---------|---------|---------|----------|
| 变量名称             | 观测数  | 平均值     | 标准差     | 最小值     | 中位值     | 最大值      |
| <i>INEQ</i>      | 3835 | 3.7044  | 1.8965  | 0.3098  | 3.4123  | 8.1462   |
| <i>DATA</i>      | 3835 | 0.1400  | 0.3471  | 0.0000  | 0.0000  | 1.0000   |
| <i>ROAD</i>      | 3835 | 5.4706  | 10.8935 | 0.1312  | 2.5707  | 111.8933 |
| <i>GOV</i>       | 3835 | 1.1077  | 0.6554  | 0.0313  | 1.0051  | 4.1179   |
| <i>CULTURE</i>   | 3835 | 0.0019  | 0.0049  | 0.0000  | 0.0001  | 0.0888   |
| <i>LANDDIST</i>  | 3835 | 0.8602  | 0.4431  | 0.3421  | 0.7716  | 1.5493   |
| <i>LANDPRICE</i> | 3835 | 0.2193  | 0.2163  | 0.0346  | 0.1439  | 0.8941   |
| <i>POPDIST</i>   | 3835 | 5.0233  | 4.2337  | 0.0551  | 3.9795  | 38.3907  |
| <i>BANKDIST</i>  | 3835 | 0.6064  | 0.4025  | 0.0000  | 0.5197  | 4.2350   |
| <i>FINANCE</i>   | 3835 | 1.0051  | 0.5682  | 0.3008  | 0.8338  | 3.2484   |
| <i>PGDP</i>      | 3835 | 10.6451 | 0.6026  | 9.1735  | 10.6247 | 12.0090  |
| <i>GDPR</i>      | 3835 | 8.5447  | 4.4357  | -4.7000 | 8.3000  | 19.2900  |
| <i>SERVICE</i>   | 3835 | 54.7483 | 13.8790 | 22.0200 | 54.4500 | 85.7600  |
| <i>EDU</i>       | 3835 | 0.1758  | 0.0391  | 0.0838  | 0.1747  | 0.2733   |
| <i>FDI</i>       | 3835 | 3.0353  | 1.5755  | 0.6931  | 2.7726  | 7.5653   |

| 表3 公共数据开放促进区域协调发展的作用 |                      |                      |
|----------------------|----------------------|----------------------|
|                      | (1)                  | (2)                  |
|                      | <i>INEQ</i>          | <i>INEQ</i>          |
| <i>DATA</i>          | -0.157***<br>(-3.97) | -0.157***<br>(-3.90) |
| <i>FINANCE</i>       |                      | 0.186***<br>(3.28)   |
| <i>PGDP</i>          |                      | -0.016<br>(-0.16)    |
| <i>GDPR</i>          |                      | 0.011***<br>(2.83)   |
| <i>SERVICE</i>       |                      | -0.001<br>(-0.53)    |
| <i>EDU</i>           |                      | -0.072<br>(-0.15)    |
| <i>FDI</i>           |                      | 0.009<br>(0.74)      |
| 城市固定效应               | 是                    | 是                    |
| 年份固定效应               | 是                    | 是                    |
| 城市时间趋势固定效应           | 是                    | 是                    |
| 观测值                  | 3835                 | 3835                 |
| 调整 R <sup>2</sup>    | 0.98                 | 0.98                 |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数t值,采用城市层面的聚类异方差稳健标准误进行估计;*INEQ*表示区域发展水平差距;列(1)(2)的经济显著性为-4.23%,相应的计算方式为自变量的回归系数/因变量的平均值。

文对截至2021年底的公共数据开放平台的统计来看,无需相关部门审批、获取门槛较低的数据集占比达99.87%<sup>⑨</sup>,绝大部分的公共数据申请门槛较低。而且,公共数据在使用上具有可复制性和非竞争性,因此公共数据开放所公开的数据要素能够同时惠及所有经济主体的发展。其次,公共数据开放起到的是激发沉睡数据资源活力、降低数据要素获取门槛的作用,能够有效的降低经济主体使用数据要素的成本,以其普惠性破解数据要素融入生产体系中存在的高获取门槛、大企业垄断、“数字鸿沟”等现象。已有文献证明,许多先发的经济主体在政府大规模推动数据开放之前,即可基于自身雄厚的经济或技术实力自行收集或向有关部门获取数据,探索数据要素支撑的发展模式(纳加拉吉等,2020;贝拉哈等,2022;纳加拉吉,2022)。因此,公共数据开放的边际效益可能对落后的经济主体更强。这一实证发现与其他公共基础资源设施资源文献的结论类似,即公共资源对落后经济主体更有价值(阿格拉沃尔等,2017;伊文斯等,2018)。

## (二)平行趋势检验

使用双重差分方法进行因果推断的前定假设是处理组与控制组存在相同的时间变化趋势,即平行趋势假设。然而公共数据开放政策的选择可能是非随机的,这产生了内生性问题。首先,在区域发展水平差距较高的城市,政府可能推动公共数据开放的动机更高,通过数据要素的均等化分配促进区域协调发展,产生反向因果问题。其次,较为富裕的区域数据基础设施建设较好,政府数据治理能力较强,更有可能推动公共数据开放。为验证本文的结果是由公共数据开放平台上线这一政策驱动而非城市特征驱动,本文借助事件研究法检验处理组与控制组的平行趋势假设。具体而言,本文采用样本内时间窗口 $(-n, +m)$ 内的各独立年份 $D_k$ 虚拟变量与处理组虚拟变量( $TREAT_i$ )的交互项作为虚拟变量,替换基准回归模型中的双重差分项进行回归检验,以时间窗口的第一期 $(-n)$ 期作为基准,检验公共数据开放的动态时间趋势,具体回归模型设定如下:

$$INEQ_{it} = \beta_0 + \sum_{k=-n+1, k \neq t-n}^{k=t+n} \gamma_k D_k \times TREAT_i + \beta_2 CONTROLS_{it} + \delta_i + \delta_t + \delta_{it} + \varepsilon_{it} \quad (2)$$

式中, $D_k$ 表示时间窗口 $(-n, +m)$ 内第 $k$ 年公共数据开放平台上线的时间虚拟变量,该模型设定以公共数据平台实际开放的前 $n$ 年 $(-n)$ 期为基期, $\gamma_k$ 为 $D_k \times TREAT_i$ 的回归系数。本文选取 $(-6, 5)$ 的时间窗口,并以 $-6$ 期为基期进行平行趋势检验。图2报告了平行趋势检验结果<sup>⑩</sup>,其中 $\gamma_k (k < 0)$ 估计系数均不显著,表明在公共数据开放平台上线之前,处理组与控制组的区域发展水平差距没有显著差异。而 $\gamma_k (k \geq 0)$ 估计系数均显著为负,表明在公共数据开放平台上线之后,处理组区域发展水平差距显著低于控制组,验证了本文基准回归的平行趋势假设。 $\gamma_k (k \geq 0)$ 估计系数的绝对值随着 $k$ 的增加而增加,说明公共数据开放促进区域协调发展的作用具有长期时间动态效应,且该效应随时间增加呈现递增趋势。

## (三)安慰剂检验

为进一步排除其他不可观测因素对区域协调发展的影响,以确保本文的研究结论由公共数据开放平台上线驱动而非随机因素驱动,本文进一步通过安慰剂检验的方式加

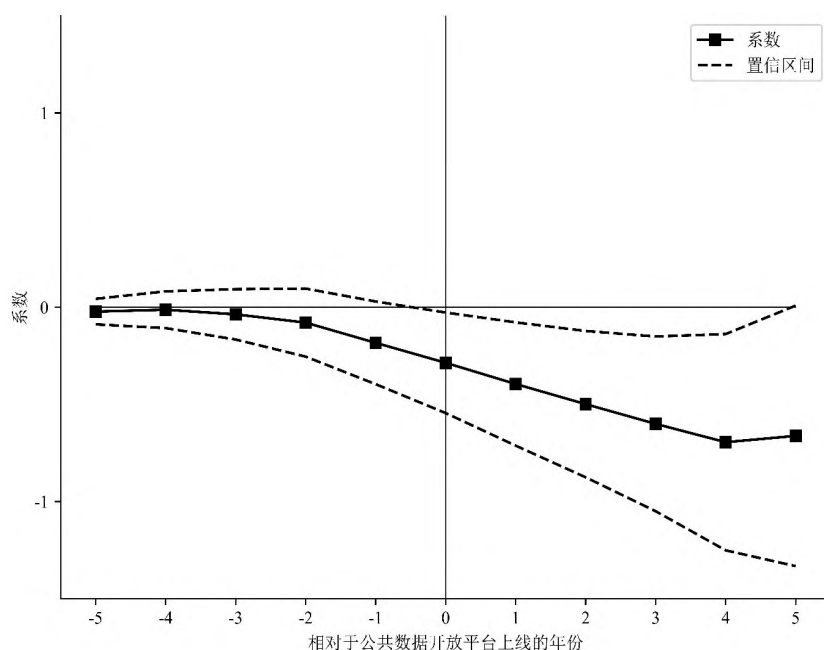


图2 平行趋势检验

注:上图绘制了(2)式中系数 $\gamma_k$ 的回归估计值。图中纵坐标表示 $\gamma_k$ 的数值大小,横坐标表示公共数据开放平台上线的年份时间线,竖直实线对应公共数据开放平台上线的年份。图中虚线表示90%统计水平上的置信区间。



以验证。具体地,本文在城市与公共数据平台上线时间的双重维度上进行500次随机抽样回归,每次抽样随机选取204个城市与相应的公共数据平台上线时间作为虚拟处理组,其余91个城市作为虚拟控制组按基准回归模型进行回归。图3报告了安慰剂检验中解释变量估计系数及P值的核密度分布图,抽样回归的估计系数核密度近似均值为0的正态分布,绝大多数估计系数的P值大于0.1的10%统计显著性水平,且所有估计系数均未超过左侧黑色实线处的基准回归系数-0.157。上述结果表明公共数据开放促进区域协调发展的结论并非随机因素驱动,本文的结论通过了安慰剂检验。

(四)工具变量回归

尽管上述实证分析发现了公共数据开放促进区域协调发展的影响,但回归结果仍可能受到内生性问题的困扰。区域协调发展水平可能会受到区域内随时间变化的资源禀赋因素或长期生产力冲击的影响,这使得本文评估公共数据开放的净效应可能受到混淆因素的影响。为缓解研究结论的潜在内生性问题,本文采取工具变量法进行如下内生性处理。

参考徐业坤和马光源(2019)及袁淳等(2021)的研究,本文手工收集了1985~2009年(城市公共数据开放平台上线前)各城市历任市委书记的籍贯特征,以各城市所有市委书记籍贯所在城市距离海岸线的距离平均值的倒数作为公共数据开放平台上线的工具变量。城市距离海岸线的距离通常与城市化、对外开放与经济发展相关(余华义,2015),因此市委书记籍贯所在城市距离海岸线的距离可以衡量市委书记对于开放型政策的接受度。2009年前城市历任市委书记特征会从政策延续等方面影响样本期内城市对开放型政策的应用和接受程度,满足相关性条件;同时2009年前历任市委书记的籍贯特征与当期区域协调发展直接影响近乎消失,满足外生性条件。考虑到各城市2009年前历任市委书记的籍贯特征为截面数据,难以直接作为面板数据的工具变量,本文参考(黄群慧等,2019)的做法,采用全国互联网宽带接入端口数与2009年前历任市委书记籍贯特征的交乘项作为工具变量IV,借助两阶段最小二乘法(2SLS)缓解本文基准回归的潜在内生性问题。

表4报告了工具变量回归结果。列(1)为第一阶段回归结果,工具变量IV的系数在1%显著性水平上为正,符合理论预期。列(2)展示了工具变量IV的第二阶段回归结果,DATA\_IV的回归系数在1%的显著性水平上为负,本文的基准回归结果保持稳健。二阶段DATA\_IV系数大小为-0.332,相较于基准回归

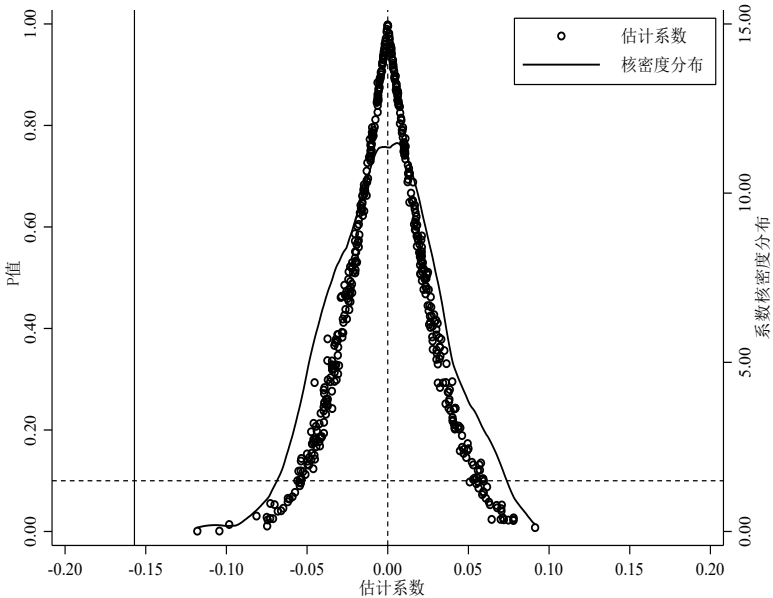


图3 安慰剂检验

注:上图展示了随机生成的公共数据开放平台上线事件进行安慰剂检验的结果。竖直线表示式(1)系数 $\beta_1$ 的回归估计值,水平虚线与竖直虚线分别表示P值为0.1和估计系数为0。黑色空心点表示每次安慰剂检验的回归系数估计值,相应的黑色曲线为系数核密度估计曲线。

表4 工具变量回归

|                       | (1)                | (2)                  | (3)                  |
|-----------------------|--------------------|----------------------|----------------------|
|                       | DATA               | INEQ                 | INEQ                 |
| IV                    | 0.162***<br>(9.54) |                      | -0.029<br>(-1.40)    |
| DATA_IV               |                    | -0.332***<br>(-2.71) |                      |
| DATA                  |                    |                      | -0.152***<br>(-3.73) |
| FINANCE               | 0.024<br>(1.10)    | 0.189***<br>(3.31)   | 0.185***<br>(3.25)   |
| PGDP                  | -0.057<br>(-1.33)  | -0.025<br>(-0.24)    | -0.014<br>(-0.14)    |
| GDPR                  | -0.003<br>(-1.22)  | 0.011***<br>(2.70)   | 0.011***<br>(2.82)   |
| SERVICE               | 0.000<br>(0.22)    | -0.001<br>(-0.52)    | -0.001<br>(-0.54)    |
| EDU                   | -0.253<br>(-1.00)  | -0.130<br>(-0.26)    | -0.084<br>(-0.17)    |
| FDI                   | -0.006<br>(-0.73)  | 0.009<br>(0.71)      | 0.010<br>(0.82)      |
| 城市固定效应                | 是                  | 是                    | 是                    |
| 年份固定效应                | 是                  | 是                    | 是                    |
| 城市时间趋势固定效应            | 是                  | 是                    | 是                    |
| 观测值                   | 3835               | 3835                 | 3835                 |
| Cragg-Donald Wald F   | 96.186             |                      |                      |
| Kleibergen-Paap rk LM | 30.464             |                      |                      |
| 调整R <sup>2</sup>      | 0.68               |                      | 0.98                 |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数t值,采用城市层面的聚类异方差稳健标准误进行估计;INEQ表示区域发展水平差距。

系数-0.157仅扩大2.11倍,远低于2003~2014年间国际顶级金融期刊JF、JFE、RFS<sup>®</sup>文献中平均扩大9倍的工具变量估计系数(姜,2017),且Cragg-Donald Wald F值为96.186,大于经验法则10,表明本文的弱工具变量问题较小。为进一步验证工具变量的外生性,即仅通过公共数据开放这一渠道对区域协调发展产生影响,本文参考唐要家等(2022)的做法,将工具变量与自变量同时纳入回归,列(3)汇报了相关回归结果。工具变量IV的系数不显著且接近于0,而DATA的系数仍显著为负,进一步验证了工具变量的外生性。上述结果表明,在缓解了内生性问题之后,本文的基本结论保持不变。

### (五)排除混淆政策冲击

本文的样本期内存在众多与公共数据开放平台上线相似的政策冲击,可能混淆本文的回归结果。因此,本文手动搜集了6个与公共数据开放较为相似的政策进行检验。

第一,由于公共数据开放平台与互联网+政务服务(*INTERNET*)同属数字建设的重要环节,因此本文参考曲永义和王可(2022)将互联网+政务服务试点冲击时间设定为2016年并与公共数据开放平台同时纳入回归。表5列(1)中DATA的系数仍然显著而INTERNET的回归系数不显著,表明本文公共数据开放平台促进区域协调发展的效应不受互联网+政务服务政策冲击的影响。

第二,由于公共数据开放平台与智慧城市建设(*SMART*)同属提供新型城市数字公共服务的重要抓手,因此本文参考石大千等(2018)将智慧城市试点中的县级试点剔除,与公共数据开放平台同时纳入模型进行回归检验。表5列(2)中DATA的系数仍然显著而SMART的系数不显著,表明本文公共数据开放平台促进区域协调发展的效应不受智慧城市政策冲击的影响。

第三,由于公共数据开放平台与宽带中国试点(*BROADBAND*)同属建设城市数字基础设施的重要环节,因此本文将宽带中国与公共数据开放平台同时纳入模型进行回归检验。表5列(3)中DATA的系数仍然显著而BROADBAND的系数不显著,表明本文公共数据开放平台促进区域协调发展的效应不受宽带中国试点政策冲击的影响。

第四,由于公共数据开放平台与信息惠民国家试点(*INFO*)同样具有依托信息化建设优化公共资源配置的目的,因此本文将信息惠民国家试点与公共数据开放平台同时纳入模型进行回归检验。表5列(4)中DATA的系数仍然显著而INFO的系数不显著,表明本文公共数据开放平台促进区域协调发展的效应不受信息惠民国家试点政策冲击的影响。

第五,由于公共数据开放平台与地方数据交易平台建设(*DATATRADE*)同样具有优化数据要素资源配置的目的,因此本文将地方数据交易平台建设与公共数据开放平台同时纳入模型进行回归检验。

表5 排除混淆政策冲击

|                  | (1)                  | (2)                  | (3)                  | (4)                  | (5)                  | (6)                  | (7)                  |
|------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|                  | INEQ                 | INEQ                 | INEQ                 | INEQ                 | INEQ                 | INEQ                 | INEQ                 |
| DATA             | -0.156***<br>(-3.88) | -0.154***<br>(-3.85) | -0.157***<br>(-3.90) | -0.157***<br>(-3.93) | -0.157***<br>(-3.87) | -0.157***<br>(-3.89) | -0.154***<br>(-3.83) |
| INTERNET         | -0.017<br>(-0.38)    |                      |                      |                      |                      |                      | -0.012<br>(-0.27)    |
| SMART            |                      | 0.098<br>(1.61)      |                      |                      |                      |                      | 0.098<br>(1.59)      |
| BROADBAND        |                      |                      | -0.006<br>(-0.11)    |                      |                      |                      | -0.003<br>(-0.07)    |
| INFO             |                      |                      |                      | 0.010<br>(0.15)      |                      |                      | 0.007<br>(0.11)      |
| DATATRADE        |                      |                      |                      |                      | 0.003<br>(0.09)      |                      | 0.003<br>(0.08)      |
| BIGDATA          |                      |                      |                      |                      |                      | 0.014<br>(0.31)      | 0.011<br>(0.25)      |
| FINANCE          | 0.185***<br>(3.27)   | 0.183***<br>(3.22)   | 0.186***<br>(3.26)   | 0.186***<br>(3.27)   | 0.186***<br>(3.28)   | 0.186***<br>(3.28)   | 0.182***<br>(3.20)   |
| PGDP             | -0.017<br>(-0.16)    | -0.023<br>(-0.22)    | -0.016<br>(-0.16)    | -0.016<br>(-0.15)    | -0.016<br>(-0.16)    | -0.016<br>(-0.15)    | -0.023<br>(-0.22)    |
| GDPR             | 0.011***<br>(2.83)   | 0.011***<br>(2.82)   | 0.011***<br>(2.84)   | 0.011***<br>(2.83)   | 0.011***<br>(2.83)   | 0.011***<br>(2.83)   | 0.011***<br>(2.82)   |
| SERVICE          | -0.001<br>(-0.54)    | -0.001<br>(-0.53)    | -0.001<br>(-0.53)    | -0.001<br>(-0.54)    | -0.001<br>(-0.54)    | -0.001<br>(-0.55)    | -0.001<br>(-0.55)    |
| EDU              | -0.075<br>(-0.15)    | -0.072<br>(-0.15)    | -0.073<br>(-0.15)    | -0.071<br>(-0.14)    | -0.070<br>(-0.14)    | -0.074<br>(-0.15)    | -0.074<br>(-0.15)    |
| FDI              | 0.009<br>(0.75)      | 0.009<br>(0.73)      | 0.009<br>(0.74)      | 0.009<br>(0.74)      | 0.009<br>(0.74)      | 0.009<br>(0.74)      | 0.009<br>(0.73)      |
| 城市固定效应           | 是                    | 是                    | 是                    | 是                    | 是                    | 是                    | 是                    |
| 年份固定效应           | 是                    | 是                    | 是                    | 是                    | 是                    | 是                    | 是                    |
| 城市时间趋势<br>固定效应   | 是                    | 是                    | 是                    | 是                    | 是                    | 是                    | 是                    |
| 观测值              | 3835                 | 3835                 | 3835                 | 3835                 | 3835                 | 3835                 | 3835                 |
| 调整R <sup>2</sup> | 0.98                 | 0.98                 | 0.98                 | 0.98                 | 0.98                 | 0.98                 | 0.98                 |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数t值,采用城市层面的聚类异方差稳健标准误进行估计;DATA、INTERNET、SMART、BROADBAND、INFO、DATATRADE、BIGDATA分别表示公共数据开放平台、互联网+政务服务、智慧城市试点、宽带中国试点、信息惠民国家试点、地方数据交易平台建设、国家级大数据综合试验区的双重差分项。

表5列(5)中 *DATA* 的系数仍然显著而 *DATA-TRADE* 的系数不显著,表明本文公共数据开放平台促进区域协调发展的效应不受地方数据交易平台建设冲击的影响。

第六,由于公共数据开放平台与国家级大数据综合试验区(*BIGDATA*)同属落实国务院《促进大数据发展行动纲要》的后续行动,因此本文将国家级大数据综合试验区与公共数据开放平台同时纳入模型进行回归检验。表5列(6)中 *DATA* 的系数仍然显著而 *BIGDATA* 的系数不显著,表明本文公共数据开放平台促进区域协调发展的效应不受国家级大数据综合试验区政策冲击的影响。

第七,本文将全部政策与公共数据开放平台上线同时纳入回归进行检验,表5列(7)中除 *DATA* 外,其余政策估计系数均不显著,且 *DATA* 的估计系数所受影响较小,仅从基准回归结果的-0.157降低到-0.154,因此说明本文选取的公共数据开放平台上线这一试点冲击具有独特性且相对外生,所得结论不受其余混淆政策因素的影响。

#### (六)稳健性检验

前述实证结果借助丰富的统计检验确立了本文的主要结论,然而回归结果仍可能受到变量计算方式、统计口径、样本选择、遗漏变量等偏误的影响,本文从5个方面对结论的稳健性进一步验证。第一,本文更换被解释变量的计算方式,采用夜间灯光计算的基尼系数(*GINI*)与泰尔指数(*THEIL*)度量区域发展水平差距。第二,考虑到本文所采用的数据为合成DMSP-VIIRS,在数据重叠年份(2012~2013年)处的拼接可能存在统计口径的偏误,本文仅采用2013年及之后的VIIRS数据样本进行回归。第三,尽管本文根据中国城市历年的行政区划对样本变量的统计口径进行调整,行政区划变动仍可能对本文的回归结果造成影响,因此删去行政区划发生变动的样本或城市进行回归。第四,由于双重差分模型中处理组的个体特征可能在处理后发生了改变,与相关对照组的个体特征并不一致,存在样本选择偏误,本文采用PSM-DID方法予以排除。第五,为排除样本选择偏误,参考曹和陈(2022)加入了各城市政策冲击是否发生的虚拟变量与年份的交互固定效应(城市冲击前年份固定效应),用以吸收冲击发生前处理组与样本组的系统性差异<sup>③</sup>。

表6报告了本文的稳健性检验结果。列(1)(2)为纳入基尼系数与泰尔指数为被解释变量的回归,*DATA* 系数均显著为负,表明本文的结论在更改被解释变量计算方式后保持稳健。列(3)~(7)分别为采用VIIRS数据样本的回归、删去行政区划发生变动的样本的回归、删去行政区划发生变动的城市的回归、采用PSM-DID方法的回归、增加城市冲击前年份固定效应的回归,*DATA* 的系数均显著为负,表明在更换被解释变量计算方式、统计口径、样本选择、遗漏变量等偏误后,本文的回归结论保持稳健。

## 五、进一步研究

#### (一)机制分析

前述实证结果通过丰富的识别策略与严谨的实证设计,证明了公共数据开放促进区域协调发展的作用。本节则在此基础上,进一步考察公共数据开放促进区域协调发展的具体经济机制,拟从破除区域信息壁垒、弥合区域资源禀赋差距两个方面进行验证。

表6 稳健性检验

|                   | (1)                 | (2)                  | (3)                  | (4)                  | (5)                 | (6)                  | (7)                  |
|-------------------|---------------------|----------------------|----------------------|----------------------|---------------------|----------------------|----------------------|
|                   | <i>GINI</i>         | <i>THEIL</i>         | <i>INEQ</i>          | <i>INEQ</i>          | <i>INEQ</i>         | <i>INEQ</i>          | <i>INEQ</i>          |
| <i>DATA</i>       | -0.011**<br>(-2.59) | -0.009**<br>(-2.29)  | -0.074***<br>(-3.50) | -0.145***<br>(-3.21) | -0.080**<br>(-2.06) | -0.140***<br>(-3.07) | -0.108***<br>(-3.62) |
| <i>FINANCE</i>    | -0.003<br>(-0.47)   | -0.017***<br>(-2.81) | 0.107**<br>(2.39)    | 0.189***<br>(3.30)   | 0.206***<br>(3.39)  | 0.074<br>(1.15)      | 0.188***<br>(3.31)   |
| <i>PGDP</i>       | -0.023**<br>(-2.04) | -0.035***<br>(-3.64) | 0.030<br>(0.30)      | -0.027<br>(-0.26)    | -0.147<br>(-1.26)   | 0.187<br>(1.50)      | -0.014<br>(-0.14)    |
| <i>GDPGR</i>      | 0.001<br>(1.26)     | -0.000<br>(-0.19)    | 0.009***<br>(3.18)   | 0.011***<br>(2.62)   | 0.012***<br>(3.12)  | 0.008*<br>(1.71)     | 0.011***<br>(2.90)   |
| <i>SERVICE</i>    | -0.001**<br>(-2.12) | -0.001**<br>(-1.98)  | -0.003**<br>(-2.47)  | -0.001<br>(-0.38)    | 0.001<br>(0.34)     | 0.001<br>(0.43)      | -0.001<br>(-0.45)    |
| <i>EDU</i>        | -0.018<br>(-0.25)   | -0.040<br>(-0.70)    | -0.152<br>(-0.40)    | -0.043<br>(-0.09)    | -0.204<br>(-0.41)   | -0.940*<br>(-1.92)   | -0.052<br>(-0.11)    |
| <i>FDI</i>        | -0.001<br>(-0.77)   | 0.001<br>(0.60)      | -0.005<br>(-0.48)    | 0.010<br>(0.87)      | 0.002<br>(0.16)     | 0.027<br>(1.41)      | 0.011<br>(0.93)      |
| 城市固定效应            | 是                   | 是                    | 是                    | 是                    | 是                   | 是                    | 是                    |
| 年份固定效应            | 是                   | 是                    | 是                    | 是                    | 是                   | 是                    | 是                    |
| 城市时间趋势<br>固定效应    | 是                   | 是                    | 是                    | 是                    | 是                   | 是                    | 是                    |
| 城市冲击前年<br>份固定效应   | 否                   | 否                    | 否                    | 否                    | 否                   | 否                    | 是                    |
| 观测值               | 3835                | 3835                 | 2655                 | 3600                 | 3201                | 2088                 | 3835                 |
| 调整 R <sup>2</sup> | 0.92                | 0.97                 | 0.99                 | 0.98                 | 0.98                | 0.98                 | 0.98                 |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数t值,采用城市层面的聚类异方差稳健标准误进行估计;*GINI*、*THEIL*分别表示基于夜间灯光计算的基尼系数、泰尔指数等度量区域发展水平差距的变量;列(6)为满足共同支撑假设的PSM-DID回归。



## 1. 破除区域信息壁垒

根据 H1b, 如果公共数据开放具有破除区域信息壁垒的作用, 那么公共数据开放促进区域协调发展的作用在区域信息壁垒较高的情境下较强。为检验这一机制, 该部分从交通便捷程度、行政分割程度、文化分割程度 3 个方面寻找代理变量, 表征区域信息壁垒的强弱, 即某区域的交通便捷程度越低、行政分割程度越高、文化分割程度越高, 表明该区域的信息壁垒越强。

表 7 报告了公共数据开放破除区域信息壁垒的机制检验。首先, 对于交通便捷程度, 本文从开放街道地图 (OSM)、全球公路开放获取数据集 (gROADS) 等获取了历年全国道路网络矢量数据, 并参照吴群锋等 (2021) 的做法, 分城市计算人均道路长度作为交通便捷程度 (ROAD) 的度量。交通基础设施的不完善形成了信息壁垒, 降低了区域内要素资源配置效率, 不利于区域协调发展的实现 (文雁兵等, 2022)。表 7 的列 (1) 纳入交通便捷程度与公共数据开放进行交互回归, 交互项系数显著为正, 表明公共数据开放促进区域协调发展的效应在交通便捷程度较低, 即区域信息壁垒较强的情况下更加明显。

其次, 对于行政分割程度, 本文从民政部获取了样本期内历年全国乡级行政区划代码, 并参考丁从明等 (2018) 的做法, 采用城市单位面积包含的乡级行政区划数度量区域行政分割程度 (GOV)。由于地方政府的权责被局限于一定的行政辖区内, 不同行政区划内的政策存在一定的差异与竞争, 这阻碍了资源与信息在地区间的交流沟通 (马光荣、赵耀红, 2022), 形成了区域发展水平差距。单位面积的行政区划数越多, 区域内行政分割程度越高, 各行政区划间信息壁垒越强, 不利于区域协调发展的实现。表 7 的列 (2) 纳入行政分割程度与公共数据开放进行交互回归, 交互项系数显著为负, 表明公共数据开放促进区域协调发展的效应在行政分割程度较高, 即区域信息壁垒较强的情况下更加明显。

最后, 对于文化分割程度, 本文参考丁从明等 (2018), 从中国研究数据服务平台 (CNRDS) 获取了中国方言片区数据, 计算方言分散度指标度量区域文化分割程度 (CULTURE)。方言作为区域交流媒介, 尽管存在一定时期的历史作用, 也为人际之间的沟通带来了一定的障碍。每种方言代表了不同群体的情感认同且蕴含着特定的思维模式, 形成了各种极具凝聚力的社会圈子 (丁从明等, 2018)。区域文化分割程度越高, 不同圈子之间的信息不对称程度越高, 可能对区域协调发展产生阻碍。表 7 的列 (3) 纳入文化分割程度与公共数据开放进行交互回归, 交互项系数显著为负, 表明公共数据开放促进区域协调发展的效应在文化分割程度较高, 即区域信息壁垒较强的情况下更加明显。

综上所述, 表 7 结果证实了公共数据开放促进区域协调发展的效应在区域信息壁垒较强的情况下更加明显。公共数据开放具有破除交通便捷程度、行政分割程度、文化分割程度等多种因素形成的区域信息壁垒的作用, 因此可以有力促进区域协调发展。

## 2. 弥合区域资源禀赋差距

根据 H1b, 公共数据开放在弥合区域资源禀赋差距方面起到重要作用。数据要素不仅具有促进发展的经济价值, 同时对传统要素也具有赋能的作用, 能够促

表 7 破除区域信息壁垒

|                   | (1)                  | (2)                  | (3)                  |
|-------------------|----------------------|----------------------|----------------------|
|                   | INEQ                 | INEQ                 | INEQ                 |
| DATA              | -0.204***<br>(-4.69) | 0.256***<br>(3.29)   | -0.123***<br>(-2.80) |
| DATA×ROAD         | 0.008***<br>(3.26)   |                      |                      |
| DATA×GOV          |                      | -0.356***<br>(-4.85) |                      |
| DATA×CULTURE      |                      |                      | -18.803*<br>(-1.75)  |
| ROAD              | -0.007**<br>(-2.36)  |                      |                      |
| GOV               |                      | -0.212***<br>(-3.46) |                      |
| CULTURE           |                      |                      | 28.079***<br>(2.69)  |
| FINANCE           | 0.191***<br>(3.42)   | 0.198***<br>(3.47)   | 0.182***<br>(3.26)   |
| PGDP              | -0.007<br>(-0.07)    | 0.011<br>(0.11)      | -0.048<br>(-0.50)    |
| GDPR              | 0.011***<br>(2.85)   | 0.010***<br>(2.70)   | 0.011***<br>(2.92)   |
| SERVICE           | -0.001<br>(-0.47)    | 0.000<br>(0.10)      | -0.001<br>(-0.47)    |
| EDU               | -0.088<br>(-0.18)    | -0.145<br>(-0.30)    | -0.077<br>(-0.16)    |
| FDI               | 0.008<br>(0.71)      | 0.007<br>(0.59)      | 0.010<br>(0.85)      |
| 城市固定效应            | 是                    | 是                    | 是                    |
| 年份固定效应            | 是                    | 是                    | 是                    |
| 城市时间趋势固定效应        | 是                    | 是                    | 是                    |
| 观测值               | 3835                 | 3835                 | 3835                 |
| 调整 R <sup>2</sup> | 0.98                 | 0.98                 | 0.98                 |

注: \*\*\*、\*\*、\* 分别表示 1%、5% 和 10% 的显著水平; 括号内数值为回归系数 t 值, 采用城市层面的聚类异方差稳健标准误进行估计; INEQ 表示区域发展水平差距。

进相对稀缺的土地、人口、金融等要素的合理配置,主要体现在优化城市发展规划和促进就业市场与金融市场对资源的合理配置。因此,数据要素通过弥合区域资源禀赋差距的方式,促进了区域协调发展。为验证这一假设,该部分从土地、人口、金融等传统资源禀赋空间分配不均的3个方面寻找代理变量,表征区域资源禀赋差距的大小。如果某区域的土地、人口、金融资源禀赋的空间分布越不均等,表明该区域的资源禀赋差距越大。

首先,对于土地资源禀赋的空间分布,本文参考贾俊雪、梁煊(2020)采用城市内坡度栅格数据计算区域土地资源禀赋的空间分布(*LANDDIST*),其值越大,说明城市内地形崎岖程度越高,城市内开发价值较高的土地资源在空间分布上相对不均,阻碍了区域经济协调发展。由于土地资源禀赋空间分布的度量指标尚没有明确统一的代理变量,因此本文进一步从中国土地市场网获取了历年各城市的土地交易数据,并利用历年城市内各地块单位面积价格计算区域土地资源禀赋的空间分布(*LANDPRICE*)。土地价格的高低通常与土地的价值强相关,因此*LANDPRICE*较大则相应表明土地资源在空间分布上相对不均。本文借助上述两个指标验证公共数据开放弥合区域资源禀赋差距的经济机制。表8的列(1)、(2)分别纳入*LANDDIST*与*LANDPRICE*与公共数据开放进行交互回归,交互项系数均显著为负,表明公共数据开放促进区域协调发展的效应在土地资源禀赋的空间分布较为不均,即区域资源禀赋差距较大的情况下更加明显。这表明公共数据开放在革新城市经济增长模式、降低经济发展对土地资源的依赖方面发挥着积极作用。而且,通过提供更充分的市场信息,公共数据开放有助于城市发展规划的合理制定,从而促进区域协调发展。

其次,对于人口资源禀赋的空间分布,参考杰内奥利等(2013),本文从LandScan获取了历年全球人口栅格数据,将其与中国行政区划合并后,借助城市内人口栅格数据计算区域人口资源禀赋的空间分布(*POPDIST*),其值越大,表明区域内人口空间分布相对不均。

表8的列(3)纳入人口空间分布与公共数据开放进行交互回归,交互项系数显著为负,表明公共数据开放促进区域协调发展的效应在人口资源禀赋的空间分布较为不均,即区域资源禀赋差距较大的情况下更加明显。人口资源分布不均是区域发展差距形成的重要原因之一(杰内奥利等,2013),公共数据开放为城市内部的潜在发展区域提供了丰富的市场信息,这有助于促进就业市场对人口资源的合理配置,从而对区域协调发展起到积极推动作用。

最后,对于金融资源禀赋的空间分布,参考蔡庆丰等(2020),本文结合金融许可证信息与工商企业数据获取了全量的商业银行分支机构点位数据。由于分支机构点位数据并非栅格,因此本文将其聚合到各城市的乡级行政区层面计算区域金融资源禀赋的空间分布(*BANKDIST*),其值越大,表明区域内金融资源禀赋的空间分布越相对不均。银行分支机构作为获取区域软信息,进行金融资源配置的重要金融机构,有利于提升区域金融资源可得性,扩大金融资源的覆盖范围(李志生、金陵,2021)。银行分支机构相对较少的区域通常面临着银行与市场间的信息不对称,阻碍了信贷资源的合理配置(王京滨、李博,2021)。表8

表8 弥合区域资源禀赋差距

|                                | (1)                | (2)                 | (3)                  | (4)                 |
|--------------------------------|--------------------|---------------------|----------------------|---------------------|
|                                | <i>INEQ</i>        | <i>INEQ</i>         | <i>INEQ</i>          | <i>INEQ</i>         |
| <i>DATA</i>                    | -0.048<br>(-0.58)  | -0.068<br>(-1.11)   | 0.031<br>(0.53)      | 0.013<br>(0.20)     |
| <i>DATA</i> × <i>LANDDIST</i>  | -0.122*<br>(-1.67) |                     |                      |                     |
| <i>DATA</i> × <i>LANDPRICE</i> |                    | -0.305**<br>(-2.16) |                      |                     |
| <i>DATA</i> × <i>POPDIST</i>   |                    |                     | -0.032***<br>(-2.83) |                     |
| <i>DATA</i> × <i>BANKDIST</i>  |                    |                     |                      | -0.229**<br>(-2.45) |
| <i>LANDDIST</i>                | -0.240<br>(-0.55)  |                     |                      |                     |
| <i>LANDPRICE</i>               |                    | -0.046<br>(-0.89)   |                      |                     |
| <i>POPDIST</i>                 |                    |                     | 0.021<br>(0.90)      |                     |
| <i>BANKDIST</i>                |                    |                     |                      | -0.198<br>(-0.65)   |
| <i>FINANCE</i>                 | 0.187***<br>(3.30) | 0.184***<br>(3.29)  | 0.176***<br>(3.22)   | 0.183***<br>(3.25)  |
| <i>PGDP</i>                    | -0.017<br>(-0.16)  | -0.018<br>(-0.18)   | -0.030<br>(-0.32)    | -0.024<br>(-0.23)   |
| <i>GDPR</i>                    | 0.011***<br>(2.85) | 0.011***<br>(2.76)  | 0.011***<br>(2.84)   | 0.011***<br>(2.75)  |
| <i>SERVICE</i>                 | -0.001<br>(-0.52)  | -0.001<br>(-0.47)   | -0.001<br>(-0.47)    | -0.001<br>(-0.43)   |
| <i>EDU</i>                     | -0.090<br>(-0.18)  | -0.089<br>(-0.18)   | -0.049<br>(-0.10)    | -0.048<br>(-0.10)   |
| <i>FDI</i>                     | 0.008<br>(0.70)    | 0.008<br>(0.68)     | 0.007<br>(0.64)      | 0.009<br>(0.76)     |
| 城市固定效应                         | 是                  | 是                   | 是                    | 是                   |
| 年份固定效应                         | 是                  | 是                   | 是                    | 是                   |
| 城市时间趋势固定效应                     | 是                  | 是                   | 是                    | 是                   |
| 观测值                            | 3835               | 3835                | 3835                 | 3835                |
| 调整 R <sup>2</sup>              | 0.98               | 0.98                | 0.98                 | 0.98                |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数*t*值,采用城市层面的聚类异方差稳健标准误进行估计;*INEQ*表示区域发展水平差距。

的列(4)纳入金融资源空间分布的指标与公共数据开放进行交互回归,交互项系数显著为负,表明公共数据开放促进区域协调发展的效应在金融资源禀赋的空间分布较为不均,即区域资源禀赋差距较大的情况下更加明显。这表明公共数据开放有助于提升信贷市场的透明度,为市场主体提供了以数据支撑的信用标识。银行可以借助数据要素降低开拓下沉市场的成本,通过减少区域金融资源分配差距的方式促进区域协调发展。

综上所述,表8结果证实公共数据开放促进区域协调发展的作用在区域资源禀赋差距较大的情况下更加明显。公共数据开放能够增强对稀缺资源发掘利用的能力且降低对传统资源投入发展经济的依赖,通过弥合区域资源禀赋差距的作用促进区域协调发展。

## (二)拓展分析

### 1. 公共数据开放的要素流通作用

前文结果证实了公共数据开放促进区域协调发展的效应。由于要素的空间分布不均是区域发展水平差距形成的重要因素,一个合理的推断是,公共数据开放也应具有促进要素流通的作用。本文首先参考刘修岩等(2017)的做法,并结合数据可得性的考虑,借助《中国城市统计年鉴》的数据计算了历年城市人均公路货运量与人均公路客运量代表商品流动与人口流动,进而代理区域要素流通。其次,本文参考蔡庆丰等(2021)的做法,采用北京大学数字普惠金融指数的信贷分指数度量小微信贷的覆盖程度,其指数主要由小微经营贷与个人消费贷的广度和深度组成,一定程度上代表了区域小微信贷的流通配置。本文借助上述3个变量检验公共数据开放的要素流通作用。表9报告了公共数据开放的要素流通作用的实证检验。列(1)、(2)、(3)分别将商品流动(*GOODS*)、人口流动(*PASSENGERS*)与小微信贷流通(*PKUCRE*)作为因变量进行回归,系数均在5%的统计水平上显著为正,说明公共数据开放有助于促进区域要素流通。

### 2. 公共数据开放的普惠发展效应

前文的分析证实了公共数据开放降低区域发展水平差距的经济后果,那么公共数据开放促进区域协调发展的效应究竟是促进落后区域的追赶,还是对先发区域的抑制,从而缩小区域经济发展差距?也就是说,公共数据开放是否具有普惠发展效应?本节将对此进行深入探究。

公共数据开放为市场提供了有价值的数据资源,这有利于激发市场主体经营活力,引领市场主体借助数据挖掘优化经济决策,促进地区的创业活力与经济发展。由于我国最大的发展差距仍是城乡差距(温家宝,2012),本文借助中国的城乡二元结构现象区分城市内不同区域的发展程度,进而对公共数据开放的普惠发展效应进行探究。为规避城市内行政区划频繁变动对估计结果造成的偏误,本文以国家统计局2009年的城乡划分代码作为城乡的划分依据,参考万海远(2021)的做法,将主城区视为发展程度相对较高的发达区域,非主城区视为发展程度相对较低的落后区域<sup>①</sup>。国家统计局的城乡划分代码在村级行政单元上界定了城乡的类别,然而村级行政区划的数据获取极为困难,因此本文以乡级行政单元为基本研究对象,将主城区数占比大于0.5的乡级行政单元界定为主城区,其余界定为非主城区,并依此进行后续计算。

根据上述定义,本文以2009年乡级行政区划矢量数据进行城乡不同区域发展程度的刻画。从区域创业活力与经济发展两个角度进行深入探究。首先,企业是市场经济中最基本、最重要的主体,因此创业活跃度能够有效的代表一个区域的创新与发展活力(格拉泽尔等,2010)。公共数据开放激发了区域的创业活力,其不仅能够满足创业者的信息获取需求,以公共数据辅助创业者进行企业选址、

表9 公共数据开放的要素流通作用

|                   | (1)               | (2)                 | (3)               |
|-------------------|-------------------|---------------------|-------------------|
|                   | <i>GOODS</i>      | <i>PASSENGERS</i>   | <i>PKUCRE</i>     |
| <i>DATA</i>       | 0.026**<br>(1.98) | 0.013**<br>(2.04)   | 1.011**<br>(2.25) |
| <i>FINANCE</i>    | -0.017<br>(-1.50) | -0.001<br>(-0.10)   | -0.006<br>(-0.00) |
| <i>PGDP</i>       | -0.005<br>(-0.20) | -0.037**<br>(-2.22) | 2.448<br>(1.24)   |
| <i>GDPR</i>       | -0.001<br>(-0.60) | -0.000<br>(-0.01)   | -0.098<br>(-1.19) |
| <i>SERVICE</i>    | 0.002**<br>(2.03) | 0.001***<br>(3.31)  | 0.027<br>(0.66)   |
| <i>EDU</i>        | -0.250<br>(-1.13) | -0.070<br>(0.80)    | -0.872<br>(-0.08) |
| <i>FDI</i>        | 0.004<br>(0.79)   | 0.007***<br>(3.39)  | -0.076<br>(-0.31) |
| 城市固定效应            | 是                 | 是                   | 是                 |
| 年份固定效应            | 是                 | 是                   | 是                 |
| 城市时间趋势固定效应        | 是                 | 是                   | 是                 |
| 观测值               | 3397              | 3408                | 3230              |
| 调整 R <sup>2</sup> | 0.82              | 0.84                | 0.99              |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数t值,采用城市层面的聚类异方差稳健标准误进行估计。



交易对手搜寻、商机把握等行为,形成高质量的创业决策(郭家堂、骆品亮,2016)。而且,公共数据开放拓宽了区域的市场可达性,有利于创业者借助数字技术挖掘公开数据发现尚待开发的价值洼地,这有利于创业资源从发达区域向落后区域的转移(张勋等,2019),实现后发地区的高质量追赶。因此借助工商企业注册数据,分别计算了城市层面(*ENTRE*)、城市内非主城区层面(*ENTRE\_R*)、城市内主城区层面(*ENTRE\_C*)历年人均工商企业注册数量,对比探究公共数据开放培育城市内不同区域创业活力的异质性,进而探讨其促进区域协调发展的方式。其次,为进一步验证公共数据开放的普惠发展效应的存在,本文计算了城市层面(*DNMEAN*)、城市内非主城区层面(*DNMEAN\_R*)、城市内主城区层面(*DNMEAN\_C*)历年夜间灯光的均值,用以探究公共数据开放促进城市内不同区域经济增长的异质性。

表10报告了公共数据开放的普惠发展效应检验。列(1)、(2)、(3)分别采用城市企业注册数、非主城区企业注册数、主城区企业注册数作为被解释变量与公共数据开放进行回归,列(1)、(2)中 *DATA* 系数均显著为正,而列(3)的估计系数不显著。结合经济显著性来看,公共数据开放对城市总体人均企业注册数的促进效应约为3.69%,对非主城区人均企业注册数的促进效应约为3.99%。上述结果说明公共数据开放有助于激发区域创业活力,且对发展相对较差的非主城区的促进效应更强,存在一定的普惠发展效应。然而,公共数据开放对主城区创业活力的促进作用并未显现,这可能源于主城区的市场信息较为丰富,且相应各区域开发程度较高,创业者难以发掘尚未开发的商业机会,因此公共数据促进创业活力的边际效应相对较弱。因此公共数据开放实现数据共享,满足创业者的商机信息需求,有助于实现高质量经济增长与区域协调发展。列(4)、(5)、(6)分别采用城市夜间灯光均值、非主城区夜间灯光均值、主城区夜间灯光均值作为被解释变量与公共数据开放进行回归,列(4)、(5)中 *DATA* 系数均显著为正,而列(6)的估计系数不显著。结合经济显著性来看,公共数据开放对城市总体发展程度的促进效应约为4.45%,对非主城区发展程度的促进效应约为5.86%,表明公共数据开放促进了区域整体的经济发展,且对非主城区等落后区域的促进效应更强。结合创业活力与经济发展两方面实证结果来看,公共数据开放不仅促进了区域整体经济发展,同样具有促进落后区域实现高质量追赶的普惠发展效应,因此有助于区域协调发展。

六、结论与建议

本文基于2009~2021年中国295个城市面板数据,借助地方政府公共数据开放平台上线的政策冲击,利用双重差分模型实证检验公共数据开放促进区域协调发展的作用。研究结论表明,公共数据开放降低了区域发展水平差距,促进了区域协调发展,这一作用在考虑内生性问题和诸多稳健性检验后仍然成立。机制分析表明,公共数据开放通过破除区域信息壁垒、弥合区域资源禀赋差距的方式,实现区域协调发展。而且,公共数据开放不仅促进落后区域后发追赶,同样促进了区域整体经济发展,实现了发展与共富的有机统一。

本文的研究结论对支持地方政府公共数据开放,实现区域协调发展与共同富裕具有重要的启示,相关政

表 10 公共数据开放的普惠发展效应

|                   | (1)                | (2)                | (3)                 | (4)                | (5)                | (6)                |
|-------------------|--------------------|--------------------|---------------------|--------------------|--------------------|--------------------|
|                   | <i>ENTRE</i>       | <i>ENTRE_R</i>     | <i>ENTRE_C</i>      | <i>DNMEAN</i>      | <i>DNMEAN_R</i>    | <i>DNMEAN_C</i>    |
| <i>DATA</i>       | 1.132*<br>(1.79)   | 1.006**<br>(2.29)  | 1.873<br>(1.56)     | 0.336**<br>(2.30)  | 0.375***<br>(2.93) | -0.003<br>(-0.02)  |
| <i>FINANCE</i>    | 1.872***<br>(2.71) | 1.405***<br>(2.66) | 0.772<br>(0.67)     | 0.665***<br>(3.33) | 0.669***<br>(3.31) | 1.001***<br>(3.31) |
| <i>PGDP</i>       | 2.101*<br>(1.80)   | 0.480<br>(0.40)    | -3.793<br>(-1.55)   | 1.635***<br>(5.33) | 1.277***<br>(4.87) | 1.725***<br>(4.32) |
| <i>GDPR</i>       | 0.106*<br>(1.96)   | 0.090*<br>(1.88)   | 0.020<br>(0.28)     | 0.005<br>(0.41)    | 0.001<br>(0.11)    | 0.048**<br>(2.42)  |
| <i>SERVICE</i>    | 0.082***<br>(2.62) | 0.074***<br>(2.70) | 0.110**<br>(2.09)   | 0.024***<br>(3.35) | 0.030***<br>(4.09) | 0.011<br>(0.98)    |
| <i>EDU</i>        | -3.554<br>(-0.52)  | -8.997<br>(-1.43)  | -25.292*<br>(-1.94) | 1.257<br>(0.97)    | 1.510<br>(1.22)    | 0.939<br>(0.45)    |
| <i>FDI</i>        | 0.495**<br>(2.27)  | 0.137<br>(0.88)    | 0.487<br>(1.43)     | 0.030<br>(0.88)    | 0.043<br>(1.13)    | 0.100**<br>(2.02)  |
| 城市固定效应            | 是                  | 是                  | 是                   | 是                  | 是                  | 是                  |
| 年份固定效应            | 是                  | 是                  | 是                   | 是                  | 是                  | 是                  |
| 城市时间趋势<br>固定效应    | 是                  | 是                  | 是                   | 是                  | 是                  | 是                  |
| 观测值               | 3835               | 3835               | 3835                | 3835               | 3835               | 3835               |
| 调整 R <sup>2</sup> | 0.89               | 0.93               | 0.95                | 0.99               | 0.99               | 0.99               |

注:\*\*\*、\*\*、\*分别表示1%、5%和10%的显著水平;括号内数值为回归系数t值,采用城市层面的聚类异方差稳健标准误进行估计;列(1)、(2)、(4)、(5)的经济显著性分别为3.69%,3.99%,4.45%,5.86%,相应的计算方式为自变量的回归系数/因变量的平均值。

策建议如下。

第一,充分认识和发挥公共数据开放共享的经济价值,推进公共数据归集整合、有序流通和共享使用。一方面,各级政府要积极推动公共数据资源体系和数据共享机制的建设,借助公共数据开放平台的建立打通数据要素流通使用的堵点,以矫正数据要素市场失灵,缓解数据要素供给不足的问题。另一方面,各级政府要根据不同区域的经济发展特点,拓展规范化、多样化的公共数据开发利用机制,引导市场力量挖掘数据潜在价值,以发挥公共数据开放优化经济发展格局的重要作用。

第二,充分发挥数据要素赋能传统要素的配置优化和效能提升的作用,突破区域协调发展中要素循环不畅、资源禀赋约束等制约。一方面,需要建立以数据为中心的市场资源配置体系,以数据的信息效应加强对市场秩序的规范和管理,通过市场竞争引导区域资源整合与要素配置。另一方面,以产业升级为抓手,在区域主导产业基础上建立垂直领域数据资源库,引进云计算、大数据等新一代信息技术,促进数据要素融入生产体系,以发挥数据要素促进实体经济普惠发展的作用。

第三,强化公共数据要素的公益属性,多措并举提高社会公众的数字技能和获得感,大力发展公平普惠的数据服务。政府应推动建立健全适应新产业形态的数字治理体系,积极开展公益性数字技能培训,提升社会整体数字素养,保障公众对数据要素红利的平等获取。面对层出不穷的数字创新技术,政府还应积极引导数字技术领头企业和社会组织参与公共服务,加速推动新技术由商业产品向公共产品的融合转变,着力打造公共数字技术服务体系,为弥合数字鸿沟、促进区域协调发展提供有力支撑<sup>⑤</sup>。

(作者单位:方锦程、刘颖、董纪昌、吕本富,中国科学院大学经济与管理学院、中国科学院大学数字经济监测预测预警与政策仿真教育部哲学社会科学实验室;高昊宇,中国人民大学财政金融学院)

#### 注释

①参见《习近平主持中共中央政治局第二次集体学习》,新华社北京,2017年12月9日电, [https://www.gov.cn/guowuyuan/2017-12/09/content\\_5245520.htm](https://www.gov.cn/guowuyuan/2017-12/09/content_5245520.htm)。

②参见中国信息通讯研究院、中国网络空间研究院:《国家数据资源调查报告(2021)》,2022年7月。

③参见《李克强:信息数据“深藏闺中”是极大浪费》,《北京日报》,2016年5月13日电, [https://www.gov.cn/xinwen/2016-05/13/content\\_5073036.htm](https://www.gov.cn/xinwen/2016-05/13/content_5073036.htm)。

④统计时以2021年的行政区划信息为准,不考虑新设立的三沙市和儋州市后共计295个城市。

⑤包括街道、镇、乡、民族乡、苏木、民族苏木、县辖区,属于中国的四级行政区。

⑥截至2023年1月,2022年《中国城市统计年鉴》尚未公布,因此本文将2021年的控制变量数据进行线性外推。

⑦样本以2021年(样本最后一年)的行政区划为标准筛选城市样本(共297个城市,包括293个地级市与4个直辖市,不含地区、自治州、盟等地级行政单位),对于样本期内(2009~2021年)行政区划发生变动的城市样本,本文进行如下处理:(1)删去了莱芜市、巢湖市等被撤销的地级城市。(2)将那曲地区、山南地区、哈密地区、吐鲁番地区、日喀则地区、林芝地区、昌都地区、海东地区、铜仁地区、毕节地区等变革为地级市的城市纳入样本。(3)删去三沙市、儋州市等新设立的地级城市。经过上述处理,本文最终得到了295个城市级行政单元(291个地级市与4个直辖市)作为研究对象。

⑧拟合采用的 sigmoid 函数形式为  $DN = \begin{cases} 63 & \text{如果 } DN > 63 \\ a + b \left( \frac{1}{1 + e^{-c(x-d)}} \right) * M & \text{其他} \\ 0 & \text{如果 } DN \leq 0 \end{cases}$

其中,  $DN$  为拟合的 DMSP 数据,  $x$  为输入的对数化 VIIRS 数据,  $a, b, c, d$  为函数参数,  $M = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$  是一个二维高斯函数, \* 表示空间卷积,相应的窗口半径根据惯例设定为3。由于 DMSP 和 VIIRS 数据在2012年和2013年具有重叠,本文首先参考吴等(2022)为参数选取初始设定范围,并参考吴和王(2019)的做法,借助网格搜索法与交叉验证法,选取拟合 DMSP 与真实 DMSP 的均方误差(RMSE)最小的参数。在获取相关参数后,本文对 VIIRS 进行拟合得到最终的 DMSP 数据。

⑨经济显著性的计算方式:自变量的回归系数/因变量的平均值。

⑩由于不同的公共数据开放平台对开放类型的定义名称不同,本文所界定的开放门槛较低的数据集包括无条件开放、完全开放、注册后开放等,不包括申请后开放、审批后开放、受限开放等。

⑪其中实线为每一期的估计系数,虚线为90%显著性水平上的置信区间。

⑫JF、JFE、RFS 分别指 Journal of Finance、Journal of Financial Economics、Review of Financial Studies。

⑬回归中加入城市冲击前年份固定效应将损失较多的自由度,因此本文的基准回归中不加入这项固定效应。

⑭根据2009年统计用行政区划代码和城乡划分代码,本文将111个主城区视为发展程度较高的主城区,其余视为发展程度较差的非主城区,具体包括112个城乡结合区、121个镇中心区、122个镇乡结合区、123个特殊区域、210个乡中心区、220个村庄。

⑤ 中外文人名(机构名)对照:纳加拉吉(Nagaraj);戈德法布(Goldfarb);塔克(Tucker);罗杰斯(Rogers);约特(Hjort);保尔森(Poulsen);查克费(Chaqfeh);福尔曼(Forman);加塞马加海(Ghasemaghaei);卡利克(Calic);布林约尔松(Brynjolfsson);麦克埃尔赫兰(McElheran);弗霍尔斯斯特(Verhulst);安德鲁(Andrew);弗尔曼(Furman);铃村(Suzumura);布埃拉(Buera);杰内奥利(Gennaioli);莱斯曼(Lessmann);阿莱西纳(Alesina);南比桑(Nambisan);吉布森(Gibson);吴(Wu);陈(Chen);诺德豪斯(Nordhaus);贝拉哈(Beraja);阿格拉沃尔(Agrawal);伊文斯(Ewens);姜(Jiang);曹(Cao);格拉泽尔(Glaeser);王(Wang)。

#### 参考文献

- (1) 蔡庆丰、陈熠辉、林焜:《信贷资源可得性与企业创新:激励还是抑制?——基于银行网点数据和金融地理结构的微观证据》,《经济研究》,2020年第10期。
- (2) 蔡庆丰、王瀚佑、李东旭:《互联网贷款、劳动生产率与企业转型——基于劳动力流动性的视角》,《中国工业经济》,2021年第12期。
- (3) 陈晓红、李杨扬、宋丽洁、汪阳洁:《数字经济理论体系与研究展望》,《管理世界》,2022年第2期。
- (4) 丁从明、吉振霖、雷雨、梁甄桥:《方言多样性与市场一体化:基于城市圈的视角》,《经济研究》,2018年第11期。
- (5) 郭家堂、骆品亮:《互联网对中国全要素生产率有促进作用吗?》,《管理世界》,2016年第10期。
- (6) 黄群慧、余泳泽、张松林:《互联网发展与制造业生产率提升:内在机制与中国经验》,《中国工业经济》,2019年第8期。
- (7) 贾俊雪、梁煊:《地方政府财政收支竞争策略与居民收入分配》,《中国工业经济》,2020年第11期。
- (8) 李成友、孙涛、王硕:《人口结构红利、财政支出偏向与中国城乡收入差距》,《经济学动态》,2021年第1期。
- (9) 李兰冰、张聪聪:《高速公路连通性对区域市场一体化的影响及异质性分析》,《世界经济》,2022年第6期。
- (10) 李彦龙、沈艳:《数字普惠金融与区域不平衡》,《经济学(季刊)》,2022年第5期。
- (11) 李志生、金凌:《银行竞争提高了企业投资水平和资源配置效率吗?——基于分支机构空间分布的研究》,《金融研究》,2021年第1期。
- (12) 刘倩、朱书尚、吴非:《城市群政策能否促进区域金融协调发展?——基于方言视角下的实证检验》,《金融研究》,2020年第3期。
- (13) 刘修岩、李松林、陈子扬:《多中心空间发展模式与地区收入差距》,《中国工业经济》,2017年第10期。
- (14) 马光荣、赵耀红:《行政区划壁垒、边界地区公共品提供与经济发展》,《金融研究》,2022年第8期。
- (15) 马红旗、黄桂田、王韧:《物质资本的积累对我国城乡收入差距的影响——基于资本—技能互补视角》,《管理世界》,2017年第4期。
- (16) 邱泽奇、张树沁、刘世定、许英康:《从数字鸿沟到红利差异——互联网资本的视角》,《中国社会科学》,2016年第10期。
- (17) 曲永义、王可:《中国政务服务信息化及其对企业创新的影响研究》,《数量经济技术经济研究》,2022年第4期。
- (18) 石大千、丁海、卫平、刘建江:《智慧城市建设能否降低环境污染》,《中国工业经济》,2018年第6期。
- (19) 石军伟、付海艳:《企业的异质性社会资本及其嵌入风险——基于中国经济转型情境的实证研究》,《中国工业经济》,2010年第11期。
- (20) 唐要家、王钰、唐春晖:《数字经济、市场结构与创新绩效》,《中国工业经济》,2022年第10期。
- (21) 万海远:《城市社区基础设施投资的创业带动作用》,《经济研究》,2021年第9期。
- (22) 王京滨、李博:《银行业务地理集中是否降低了金融风险?——基于中国城市商业银行微观数据的研究》,《管理世界》,2021年第5期。
- (23) 温家宝:《中国农业和农村的发展道路》,《求是》,2012年第2期。
- (24) 文雁兵、张梦婷、俞峰:《中国交通基础设施的资源再配置效应》,《经济研究》,2022年第1期。
- (25) 吴群峰、刘冲、刘青:《国内市场一体化与企业出口行为——基于市场可达性视角的研究》,《经济学(季刊)》,2021年第5期。
- (26) 徐业坤、马光源:《地方官员变更与企业产能过剩》,《经济研究》,2019年第5期。
- (27) 杨继东、罗路宝:《产业政策、地区竞争与资源空间配置扭曲》,《中国工业经济》,2018年第12期。
- (28) 余华义:《城市化、大城市化与中国地方政府规模的变动》,《经济研究》,2015年第10期。
- (29) 袁淳、肖土盛、耿春晓、盛誉:《数字化转型与企业分工:专业化还是纵向一体化》,《中国工业经济》,2021年第09期。
- (30) 张勋、万广华、吴海涛:《缩小数字鸿沟:中国特色数字金融发展》,《中国社会科学》,2021年第8期。
- (31) 张勋、万广华、张佳佳、何宗樾:《数字经济、普惠金融与包容性增长》,《经济研究》,2019年第8期。
- (32) 庄毓敏、储青青、马勇:《金融发展、企业创新与经济增长》,《金融研究》,2020年第4期。
- (33) Agrawal, A., Galasso, A. and Oettl, A., 2017, "Roads and Innovation", *Review of Economics and Statistics*, 99(3), pp.417~434.
- (34) Alesina, A., Michalopoulos, S. and Papaioannou, E., 2016, "Ethnic Inequality", *Journal of Political Economy*, 124(2), pp.428~488.
- (35) Beraja, M., Yang, D. Y. and Yuchtman, N., 2022, "Data-Intensive Innovation and the State: Evidence From AI Firms in China", *Review of Economic Studies*, 90(4), pp.1701~1723.
- (36) Brynjolfsson, E. and McElheran, K., 2016, "The Rapid Adoption of Data-Driven Decision-Making", *American Economic Review*, 106(5), pp.133~139.
- (37) Buera, F. J., Kaboski, J. P., Rogerson, R. and Vizcaino, J. I., 2022, "Skill-Biased Structural Change", *Review of Economic Studies*, 89(2), pp.592~625.
- (38) Cao, Y. and Chen, S., 2022, "Rebel on the Canal: Disrupted Trade Access and Social Conflict in China, 1650~1911", *American Economic Review*, 112(5), pp.1555~1590.
- (39) Chaqfeh, M., Asim, R., AlShebli, B., Zaffar, M. F., Rahwan, T. and Zaki, Y., 2023, "Towards a World Wide Web without Digital



Inequality", *Proceedings of the National Academy of Sciences*, 120(3), e2082318176.

(40) Chen, X. and Nordhaus, W. D., 2011, "Using Luminosity Data as a Proxy for Economic Statistics", *Proceedings of the National Academy of Sciences*, 108(21), pp.8589~8594.

(41) Ewens, M., Nanda, R. and Rhodes-Kropf, M., 2018, "Cost of Experimentation and the Evolution of Venture Capital", *Journal of Financial Economics*, 128(3), pp.422~442.

(42) Forman, C., Goldfarb, A. and Greenstein, S., 2012, "The Internet and Local Wages: A Puzzle", *American Economic Review*, 102(1), pp.556~575.

(43) Furman, J. L., Nagler, M. and Watzinger, M., 2021, "Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program", *American Economic Journal: Economic Policy*, 13(4), pp.239~270.

(44) Gennaioli, N., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A., 2013, "Human Capital and Regional Development", *Quarterly Journal of Economics*, 128(1), pp.105~164.

(45) Ghasemaghaei, M. and Calic, G., 2019, "Does Big Data Enhance Firm Innovation Competency? The Mediating Role of Data-Driven Insights", *Journal of Business Research*, 104, pp.69~84.

(46) Gibson, J., Olivia, S., Boe-Gibson, G. and Li, C., 2021, "Which Night Lights Data Should We Use in Economics, and Where?", *Journal of Development Economics*, 149, pp.292~309.

(47) Glaeser, E. L., Rosenthal, S. S. and Strange, W. C., 2010, "Urban Economics and Entrepreneurship", *Journal of Urban Economics*, 67(1), pp.1~14.

(48) Goldfarb, A. and Tucker, C., 2019, "Digital Economics", *Journal of Economic Literature*, 57(1), pp.3~43.

(49) Hjort, J. and Poulsen, J., 2019, "The Arrival of Fast Internet and Employment in Africa", *American Economic Review*, 109(3), pp.1032~1079.

(50) Jiang, W., 2017, "Have Instrumental Variables Brought Us Closer to the Truth", *Review of Corporate Finance Studies*, 6(2), pp.127~140.

(51) Lessmann, C., 2014, "Spatial Inequality and Development — Is There an Inverted-U Relationship?", *Journal of Development Economics*, 106, pp.35~51.

(52) Nagaraj, A., 2022, "The Private Impact of Public Data: Landsat Satellite Maps Increased Gold Discoveries and Encouraged Entry", *Management Science*, 68(1), pp.564~582.

(53) Nagaraj, A., Shears, E. and de Vaan, M., 2020, "Improving Data Access Democratizes and Diversifies Science", *Proceedings of the National Academy of Sciences*, 117(38), pp.23490~23498.

(54) Nambisan, S., Lyytinen, K., Majchrzak, A. and Song, M., 2017, "Digital Innovation Management: Reinventing Innovation Management Research in a Digital World", *MIS Quarterly*, 41(1), pp.223~238.

(55) Rogers, E. M., Singhal, A. and Quinlan, M. M., 2014, *Diffusion of Innovations*, Routledge.

(56) Suzumura, K., 1992, "Cooperative and Noncooperative R&D in an Oligopoly with Spillovers", *American Economic Review*, 82(5), pp.1307~1320.

(57) Verhulst, S. and Andrew, Y., 2016, *The Global Impact of Open Data*, O'Reilly Media, Inc.

(58) Wu, K. and Wang, X., 2019, "Aligning Pixel Values of Dmsp and Viirs Nighttime Light Images to Evaluate Urban Dynamics", *Remote Sensing*, 11(12), pp.1463.

(59) Wu, Y., Shi, K., Chen, Z., Liu, S. and Chang, Z., 2022, "Developing Improved Time-Series DMSP-OLS-Like Data (1992–2019) in China by Integrating DMSP-OLS and SNPP-VIIRS", *IEEE Transactions On Geoscience and Remote Sensing*, 60, pp.1~14.

## Does Public Data Access Promote Regional Harmonious Development?

### On a Quasi-natural Experiment of Government Data Platform Access

Fang Jincheng<sup>a,b</sup>, Liu Ying<sup>a,b</sup>, Gao Haoyu<sup>c</sup>, Dong Jichang<sup>a,b</sup> and Lv Benfu<sup>a,b</sup>

(a. School of Economics and Management, University of Chinese Academy of Sciences; b. MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Science;

c. The School of Finance, Renmin University of China)

**Abstract:** Public data serves as a crucial resource for public services. Evaluating its economic consequences on regional harmonious development is a prerequisite for scientific governance. Using the establishment of Public Data Access Platform by local governments as a quasi-natural experiment, this paper employs empirical analysis based on annual data from 295 cities in urban China from 2009 to 2021 to determine whether public data access promotes regional harmonious development. The main findings show that public data access can significantly reduce regional development inequality, which promotes regional harmonious development. The mechanism analysis verifies that public data access can break the information barrier and bridge the regional endowment gap, thus contributing regional harmonious development. Additionally, public data access can promote latecomer regions catch-up while contributing to the overall economy, and achieve the organic unity of development and common prosperity. This paper empirically evaluates the economic consequences of sharing data factor, providing significant implications for deepening the construction of the data resource system and unlocking the potential of data factor.

**Keywords:** public data access; data factor; regional harmonious development

# Does Public Data Access Promote Regional Harmonious Development? On a Quasi-natural Experiment of Government Data Platform Access

*Fang Jincheng<sup>a,b</sup>, Liu Ying<sup>a,b</sup>, Gao Haoyu<sup>c</sup>, Dong Jichang<sup>a,b</sup> and Lv Benfu<sup>a,b</sup>*

(a. School of Economics and Management, University of Chinese Academy of Sciences; b. MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Science; c. The School of Finance, Renmin University of China)

**Summary:** Public Data Access refers to the process of integrating data resources with public attributes from government and social entities and opening them for equal access, which not only provides the public with the right to be informed and to access data, but more importantly ensures their right to utilize data factors, thereby facilitating the release of data dividends. It has been confirmed that public data access helps integrate data factors into the production system and serves as a critical driver of economic growth. However, there is still a lack of micro-level empirical evidence on how public data access alters the spatial pattern of economic development and whether it promotes regional harmonious development.

This paper examines whether public data access can promote regional harmonious development and analyzes its underlying mechanism. Employing the launch of public data access platforms by local governments as a quasi-natural experiment, the empirical analysis is based on panel data from 295 cities over the period from 2009 to 2021. The research findings indicate that public data access reduces the regional development gap by an average of 4.23%, thus playing a significant role in promoting regional harmonious development. This effect is mainly attributed to the enhanced public resource efficiency brought about by public data access, which allows the opening of data factors to benefit all economic entities and alleviate the constraints imposed by unequal regional resource distribution on regional harmonious development. Furthermore, public data access breaks down regional information barriers, facilitating the rational allocation of resource factors, thus promoting regional harmonious development. Moreover, public data access strengthens the ability to explore and utilize scarce resources and reduces reliance on traditional resource-intensive economic development models, thus promoting regional harmonious development by bridging regional resource endowment gaps. Finally, public data access can promote latecomer regions catch-up while contributing to the overall economy, and achieve the organic unity of development and common prosperity.

The innovation of this paper is reflected in three aspects: Firstly, this paper provides micro-level evidence on the inclusion of data production factors in economic growth theories, directly evaluating the economic value of non-scarcity and non-exclusivity of data production factors. This paper also explores the role of data factors in optimizing allocation and enhancing the efficiency of traditional factors, revealing that public data access exerts its economic effects by breaking down regional information barriers and bridging regional resource endowment gaps. Secondly, this paper reveals the dual attributes of data factors in creating and sharing wealth, analyzes the impact of public data access from different perspectives of innovative technology and public resource, and verifies that public data access mainly exerts the public resource effect in promoting regional harmonious development. Finally, this paper employs more refined data at the city level to study China's regional economic inequality issues, thereby offering a clearer exploration of the transmission mechanism and economic impact of public data access in promoting regional harmonious development.

Based on our research results, we propose the following recommendations. Firstly, the government should actively promote the integration, opening and utilization of public data. Secondly, the government should build a data-centered market system and promote the empowerment of data factors in upgrading the real economy. Finally, the government should actively conduct public-interest digital skills training to ensure equal access to data dividends for the public.

**Keywords:** public data access; data factor; regional harmonious development

**JEL Classification:** O18, O33, R58