

DOI: 10.16538/j.cnki.fem.20240517.401

## 唤醒“沉睡”的数据: 公共数据开放与新企业进入

郑志强<sup>1</sup>, 何佳俐<sup>2</sup>

(1. 安徽财经大学 安徽经济社会发展研究院, 安徽 蚌埠 233030;  
2. 安徽农业大学 经济管理学院, 安徽 合肥 230036)

**摘要:** 推动公共数据开放是释放数据要素红利的重要举措。本文利用2009—2019年中国所有新建企业的注册信息, 构建城市—行业层面的新企业进入数据, 并以公共数据开放平台上线作为准自然实验, 使用双重差分法实证检验公共数据开放对新企业进入的影响。研究发现, 公共数据开放显著促进了新企业进入, 该结论在经过了一系列稳健性检验后仍然成立。机制检验表明, 降低创业者的信息搜寻成本和提高政府透明度是公共数据开放促进新企业进入的主要渠道。进一步分析发现, 公共数据开放对新企业进入的促进作用主要体现在非国有企业、服务业(尤其是生活性服务业)以及互联网普及率较高的城市; 同时, 公共数据开放质量越高, 对新企业进入的促进效果越强。本文不仅丰富了关于数据生产要素经济价值的理论分析与经验研究, 对于畅通数据资源大循环和推动数字经济高质量发展也具有重要的政策启示。

**关键词:** 公共数据开放; 数据要素; 新企业进入

**中图分类号:** F270 **文献标识码:** A **文章编号:** 1001-4950(2025)04-0082-16

### 一、引言

数字经济时代, 数据不仅仅是信息的载体, 更是一种新型生产要素, 以及推动经济和社会发展的关键性资源(Jones和Tonetti, 2020)。特别地, 政府在履行公职和提供公共服务过程中所创造、采集、获取的数据, 不仅种类丰富、量级庞大, 更是领域内重要的权威数据, 也被视为国家发展的战略性资源(彭远怀, 2023)。公共数据具有明显的公共品属性, 是社会公共利益的重要组成部分, 对于不涉及国家安全和隐私保护的公共数据理应开放共享(郑磊, 2015)。然而, 在过去相当长一段时间内, 公共数据由于政府部门不愿公开、不会公开或不便公开而一直“沉睡”。为唤醒“沉睡”的公共数据, 2015年8月, 国务院发布《促进大数据发展行动纲要》, 提出要稳步推动公共数据资源开放, 由此公共数据开放上升为国家层面的战略举措。2023年2月, 中共中央、

收稿日期: 2023-12-21

基金项目: 国家社会科学基金重大项目(21&ZD086); 安徽省社会科学创新发展研究课题(2023CX031); 安徽农业大学人才科研资助项目(RC402403)

作者简介: 郑志强(1995—), 男, 安徽财经大学安徽经济社会发展研究院助理研究员;

何佳俐(1996—), 女, 安徽农业大学经济管理学院讲师(通信作者, [hejl1103@163.com](mailto:hejl1103@163.com))。

国务院颁布《数字中国建设整体布局规划》，规划进一步指出推动公共数据汇聚利用是畅通数据资源大循环的关键环节。当前，随着公共数据资源的不断增长和积累，评估公共数据开放的经济价值已成为亟需研究的重要课题。这不仅对于优化数据要素的资源配置和促进数据资源大循环具有重要意义，还为推动数字经济的高质量发展提供了重要启示。

政府所拥有的公共数据资源涉及经济建设、公共安全、资源环境、道路交通、民生服务、信用服务、教育科技、卫生健康以及文化休闲等诸多领域，开放公共数据资源不仅是提高政府透明度的重要举措，也是推动市场主体挖掘潜在商业需求进而创造经济价值的重要驱动力（方锦程等，2023）。近年来，我国各地各部门高度重视公共数据资源的开发应用，不断推进公共数据开放工作。2012年，上海率先在中国探索开放公共数据资源，并上线“上海市政府数据服务网”。根据上海公共数据开放平台显示，截至目前，平台已开放19.9亿条数据，平台访问量和数据下载次数分别达6.18百万次和1.89百万次。创业实践者可从开放的公共数据资源中获取丰富的知识信息、市场信息以及政策信息等，以挖掘潜在的市场机会，进而开展创新创业。例如，美国哈佛大学毕业生Alex Laskey和Daniel Yates利用美国政府公布的能源消费数据创建了为家庭提供能源消费计划服务的Opower公司，以帮助家庭降低能源消费浪费（Jetzek等，2014）。旨在解决城市供水管网漏损与水质安全问题的“水网卫士”项目在获得2018 SODA (shanghai open data apps) 开放数据创新应用大赛后，获得了创投机构的投资，并顺利进入产业孵化阶段<sup>①</sup>。

作为一种非稀缺性资源，公共数据与劳动力、资本以及土地等传统资源有所不同，具有增值性、非竞争性、高度融合性等特征（黄先海等，2023）。增值性是指相对于其他传统生产要素，公共数据在使用中非但不会被消耗，反而能在积累和循环过程中通过协同其他生产要素实现价值的几何增长（Carriere-Swallow和Haksar，2019）。非竞争性体现在公共数据要素具有公共品属性，可同时供多个社会主体重复使用，并且复制成本低。高度融合性是指公共数据要素能够与公共部门的公共服务供给过程以及私人部门的商业模式变革充分融合，缓解信息不对称问题（Zuiderwijk等，2014）。这些特征使得公共数据的开放过程不仅能够为用户的经济活动赋能，也能够倒逼提供者转变职能。一方面，增值性和非竞争性使得公共数据开放降低了数据资源的使用门槛，在一定程度上破除了创业者面临的信息壁垒，从而降低其信息搜寻成本，促使其进入市场。另一方面，高度融合性使得公共数据开放提高了政府的开放性和透明度，这有利于营造良好的创业生态环境，进而激励新企业进入。

与本文直接相关的文献主要考察了数字经济对创业的影响。例如，已有研究发现，数字金融发展、数字乡村建设，以及数字经济发展等对个人创业、家庭创业或新企业进入具有显著的推动作用（谢绚丽等，2018；张勋等，2019；赵涛等，2020；Wang和Shao，2023；田鸽等，2023）。鉴于数字技术是数字经济发展的底层基础，部分学者还考察了数字技术应用对创业的影响。例如，互联网使用能够通过增加个体的信息可得性进而提高其创业概率（Fairlie，2006；Tan和Li，2022）。李磊和何艳辉（2024）发现机器人使用对创业活动具有显著的促进作用，并且这主要体现在机会型创业方面。莫怡青和李力行（2022）以外卖平台为例，探究了数字经济背景下零工经济对创业的影响，发现外卖平台进入总体上会减少大约4.7%的企业注册数量。何雨可等（2024）基于2014年实施的“信息惠民国家试点”政策，考察了数字政府建设对创业的影响，发现数字政府建设通过降低制度性交易成本和不确定性感知提升城市创业活跃度。综上，现有文献主要考察数字经济的整体发展状况、数字经济的典型样态或数字技术应用对创业的影响，而相对忽视了作为数字经济发展的基础要素——数据——对创业的直接作用。在数据生产要素化

<sup>①</sup>“水网卫士”团队应用了长三角三省一市政府开放的温度、交通、用水等方面数据，通过水务建模和大数据分析，自动定位漏损点、水质异常点，从而精准解决管网漏损问题。

的背景下,探究释放数据要素对创业的影响有利于深入理解数字经济高质量发展的内涵。

基于此,本文利用2009—2019年中国所有新建企业的注册信息,构建城市—行业层面的新企业进入数据,并以公共数据开放平台上线作为准自然实验,使用双重差分法实证检验公共数据开放对新企业进入的影响。研究发现,公共数据开放显著促进了新企业进入,并且该结论通过了一系列稳健性检验。影响机制检验表明,降低创业者的信息搜寻成本和提高政府透明度是公共数据开放促进新企业进入的两个主要渠道。进一步分析发现,公共数据开放对新企业进入的促进作用主要体现在非国有企业、服务业(尤其是生活性服务业)以及互联网普及率较高的城市;同时,公共数据开放质量越高,对新企业进入的促进效果越强。

本研究的边际贡献主要体现在以下两个方面:第一,本文丰富了关于数据生产要素经济价值的理论分析与经验研究。数字技术的快速发展和广泛应用衍生出数字经济,而数据是数字化、网络化、智能化的基础以及数字经济发展的关键生产要素。虽然已有文献充分探究了数字经济发展的创业效应,但关于数据生产要素本身如何影响新企业进入,目前现有研究尚未提供直接证据和系统性分析。本文从数据要素视角探究释放数据要素的新企业进入效应,这不仅丰富了有关创业影响因素的研究文献,也有效地补充了数据生产要素经济价值的理论分析和经验研究。第二,本文是较早从实证视角对公共数据开放经济后果进行评估的文献。现有关于公共数据开放的文献多从案例研究和理论分析的视角讨论公共数据开放面临的挑战、制度基础、影响因素,以及经济后果(Attard等,2015;Hughes-Cromwick和Coronado,2019;朱崢,2020;Ansari等,2022;马龙和陈奕博,2023),少数实证方面的研究也仅探讨了其对区域协调发展、企业生产效率以及城投债定价等的影响(方锦程等,2023;彭远怀,2023;欧阳伊玲等,2024)。而本文使用工商企业注册数据,实证检验了公共数据开放对新企业进入的影响,从而对评估公共数据开放经济效应的相关研究形成了有益补充。

## 二、制度背景与理论分析

### (一)制度背景

推动公共数据开放是释放数据要素红利的重要举措。而公共数据开放平台是开放和获取公共数据的载体,也是连接数据供给侧和需求者的桥梁,用于公开国民经济和社会发展等各个方面的公共数据。公众可在公共数据开放平台上公开检索、浏览、获取和使用公共数据,并与政府部门进行良性互动。2012年,上海率先探索推进公共数据开放工作,上线“上海市政府数据服务网”,该网站是中国第一个省级公共数据开放平台。随后,各省市开始大力推进公共数据开放平台建设,截至2022年10月,中国已上线208个省级、副省级以及地级市的公共数据开放平台:省级平台25个(含省、自治区以及直辖市,不包括港澳台),副省级平台12个,地级市平台171个。其中,省级地区中云南、西藏、内蒙古、黑龙江、吉林、山西尚未开通公共数据开放平台,副省级城市中西安、长春、大连尚未上线公共数据开放平台,而地级市中尚未开通公共数据开放平台的城市大多数位于中西部和东北部地区。

### (二)理论分析

公共数据开放是指政府或公共机构将其掌握的数据资源统一整合到开放平台上,并免费开放给社会主体下载和使用的过程,从而保障了社会主体对公共数据的知情权、查询权以及使用权(方锦程等,2023)。一方面,公共数据开放提供了一个以较低成本获取行业和市场数据的途径。通过分析和利用这些数据,潜在的创业者可以更好地理解市场需求,优化商业决策,从而提高其创业项目的成功率。另一方面,公共数据开放提高了政府的开放性和透明度,由此有利

于营造良好的创业生态环境,进而促进新企业进入。基于上述两方面的影响,本文分别从降低创业者面临的信息搜寻成本和提高政府透明度两个方面展开理论分析。

第一,公共数据开放能够通过降低创业者的信息搜寻成本进而促进新企业进入。创业者在选择进入市场前,需要深入了解目标市场的需求偏好、竞争格局以及政策环境,即进行充分的市场调研和信息搜寻。然而,每项信息收集活动都涉及大量的信息搜寻成本。政府开放的各个领域的结构化和非结构化公共数据资源可能蕴含着创业者需要的知识信息、市场信息以及政策信息等,这为创业者提供了一个较低成本的信息源,创业者可在公共数据开放平台上查询、下载、使用以及分析这些数据(Hughes-Cromwick和Coronado,2019),从而帮助其识别出潜在的市场机会,制定更具针对性的市场进入策略。例如,上海市公共数据开放平台分主题列示了可供下载的15项领域的的数据,包括经济建设、城市建设、教育科技、民生服务、资源环境、信用服务以及文化休闲等,数据种类较为齐全且信息量丰富。此外,公共数据开放平台上的公共数据均由各级相关公共部门提供,与线下搜寻或线上其他渠道搜寻相比,其来源具有高度的权威性和可信性,极大地降低了创业主体在信息搜寻过程中的辨别真伪的核查成本(张莉和林安然,2023)。由此可见,开放具有权威性的公共数据极大地提高了公共数据资源的普惠化水平,降低了数据资源的使用门槛,进而一定程度上破除了创业者面临的信息壁垒(方锦程等,2023),降低了其信息搜寻成本。创业者面临的信息壁垒的降低,以及信息搜寻成本的下降有利于其准确识别商业机会,并将更多的资源分配至生产、研发活动以及新企业的建设与运营等方面(Shane,2000),从而有利于提高创业者的进入意愿以及新企业成功进入的可能性。基于上述分析,本文提出如下研究假说:

假说H1:公共数据开放能够通过降低创业者的信息搜寻成本进而促进新企业进入。

第二,公共数据开放能够通过提高政府透明度进而促进新企业进入。数字经济时代,公共数据不仅是推动数字经济发展的关键生产要素,也是推进国家治理体系和治理能力现代化的治理要素(陈潭,2023)。通常来说,政府部门掌握的公共数据不仅包括其在提供公共服务过程中所采集和获取的企业、行业以及市场数据,还涵盖政策制定、政策实施和执行过程以及政策实施效果等公共信息资源,这部分数据是增加私人部门对公共部门了解的重要信息来源。然而,在过去相当长一段时间内,由于公共部门的不愿、不会或不便公开,私人部门对公共部门存在着较高的信息不对称,从而阻碍了政府透明度的提升(彭鐔,2023)。而公共数据开放有利于缓解这一情况。一方面,公共数据开放后,创业主体可直接在公共数据开放平台上下载、获取公共数据和公共信息资源,这显著增加了创业主体对公共部门的信息可得性,使得其更加了解政府的行为和决策过程,从而提高了政府透明度(Lourenço,2015)。另一方面,政府公开的原始数据中详细地披露了数据的提供方和联系方式,并且大部分公共数据开放平台提供了关于披露数据的评论和咨询功能,这有利于创业者和公共部门之间形成良性互动。此外,通过分析公共数据,创业者可结合自身发展需要评估相关政策的有效性,进而对政府政策制定和公共资源管理提供有效反馈,这有利于提高创业者在政府决策过程中的参与度,并进一步提高政府运作的透明度(Park和Gil-Garcia,2022)。政府透明度的提升有利于营造良好的创业生态环境,一方面可增加创业者对政府部门的廉洁感知,增强创业者信心并减少创业者进入市场所面临的制度性成本(Fabro和Aixalá,2009),另一方面也有利于降低创业者由于政策不确定性感知而带来的创业风险(于文超等,2020),进而促进新企业进入市场。基于上述分析,本文提出如下研究假说:

假说H2:公共数据开放能够通过提高政府透明度进而促进新企业进入。

### 三、研究设计

#### (一) 样本选择与数据来源

本文利用中国全量工商企业注册数据检验公共数据开放对新企业进入的影响。新企业进入数据来源于中国国家企业信用信息公示系统,该数据库包含了1949年以来在中国注册过的近两亿家企业。为排除2008年金融危机和2020年新冠疫情事件对企业注册的影响,本文基于该数据库整理了2009年到2019年城市—行业层面的新建企业数量,用于分析的企业总量超过4000万。为提升数据质量,本文剔除了企业注册时间、所属行业、所有制类型缺失或存在明显错误的样本。最终,城市—年份—行业层面的样本观测数为61 920个。城市层面的特征数据来源于《中国城市统计年鉴》。

#### (二) 实证模型构建

本文以公共数据开放平台上线作为准自然实验,使用双重差分法(DID)实证检验公共数据开放对新企业进入的影响。公共数据开放平台上线由当地政府决定,对于创业主体来说是一个外生事件。具体地,以上线公共数据开放平台的城市作为实验组,以未上线公共数据开放平台的城市作为控制组,并构建如下模型检验公共数据开放对新企业进入的影响:

$$Entry_{i,j,t} = \alpha + \beta Dataopen_{i,t} + \gamma Control_{i,t} + City_i + Year_t + City_i \times Ind_j + Ind_j \times Year_t + \varepsilon_{i,j,t} \quad (1)$$

其中,被解释变量 $Entry_{i,j,t}$ 表示*i*城市*j*行业在*t*年的新企业进入,借鉴Kaplan等(2011)和李硕等(2022)的研究,以新注册企业数目加1的自然对数来衡量。核心解释变量 $Dataopen_{i,t}$ 为*i*城市在*t*年的政策哑变量,各城市上线公共数据开放平台的当年及以后, $Dataopen$ 赋值为1,其他为0。 $Control$ 为控制变量,包括国内生产总值( $GDP$ )、第二产业增加值( $Add$ )、财政收入( $Revenue$ )、财政支出( $Expense$ )、总人口( $People$ )、对外开放程度( $Trade$ )、贷款余额( $Credit$ )、高等教育水平( $Edu$ )。 $City_i$ 为城市固定效应,以控制不随时间变化的城市层面的差异, $Year_t$ 为年份固定效应,以控制不随个体变化的时间层面的差异, $City_i \times Ind_j$ 为城市和行业的联合固定效应,以控制不随时间变化的城市—行业层面的差异, $Ind_j \times Year_t$ 为行业和年份的联合固定效应,以控制行业层面随时间变化的影响因素。 $\varepsilon_{i,j,t}$ 表示残差项。标准误聚类到城市层面。变量定义及其描述性统计结果如表1所示。

表1 变量定义及其描述性统计

| 变量名             | 变量定义                              | 观测数    | 均值     | 标准差   |
|-----------------|-----------------------------------|--------|--------|-------|
| <i>Entry</i>    | $\text{Log}(\text{新企业进入数}+1)$     | 61 920 | 3.164  | 2.625 |
| <i>Dataopen</i> | 平台上线的当年及以后为1,其他为0                 | 61 920 | 0.058  | 0.234 |
| <i>GDP</i>      | $\text{Log}(\text{国内生产总值})$       | 61 920 | 16.471 | 0.934 |
| <i>Add</i>      | $\text{Log}(\text{第二产业增加值})$      | 61 920 | 15.695 | 0.988 |
| <i>Revenue</i>  | $\text{Log}(\text{财政收入})$         | 61 920 | 13.828 | 1.075 |
| <i>Expense</i>  | $\text{Log}(\text{财政支出})$         | 61 920 | 14.736 | 0.767 |
| <i>People</i>   | $\text{Log}(\text{年末总人口})$        | 61 920 | 5.897  | 0.663 |
| <i>Trade</i>    | $\text{Log}(\text{进出口贸易总额})$      | 61 920 | 16.172 | 1.975 |
| <i>Credit</i>   | $\text{Log}(\text{年末金融机构各项贷款余额})$ | 61 920 | 16.286 | 1.172 |
| <i>Edu</i>      | $\text{Log}(\text{普通高等学校数}+1)$    | 61 920 | 1.702  | 0.909 |

### 四、实证结果与分析

#### (一) 基准回归结果

表2报告了基准模型的估计结果。其中,列(1)为公共数据开放对新企业进入影响的估计效应,且控制城市固定效应、年份固定效应、城市和行业的联合固定效应以及行业和年份的联合

固定效应, *Dataopen* 的估计系数在1%的水平上显著为正。然后,依次在列(1)的基础上加入多个控制变量,列(2)至(4)的估计结果显示, *Dataopen* 的估计系数始终在1%的水平上显著为正,表明公共数据开放与新企业进入具有正相关关系<sup>①</sup>。根据理论分析,公共数据开放可通过降低创业者的信息搜寻成本和提高政府透明度,从而对新企业进入产生促进效应。后文会展开更为详细的机制检验。

表2 基准回归结果

| 变量                       | (1)<br><i>Entry</i> | (2)<br><i>Entry</i> | (3)<br><i>Entry</i> | (4)<br><i>Entry</i> |
|--------------------------|---------------------|---------------------|---------------------|---------------------|
| <i>Dataopen</i>          | 0.304*** (0.086)    | 0.315*** (0.084)    | 0.316*** (0.086)    | 0.299*** (0.085)    |
| <i>GDP</i>               |                     | -0.238 (0.355)      | -0.274 (0.339)      | -0.256 (0.334)      |
| <i>Add</i>               |                     | -0.004 (0.207)      | -0.029 (0.199)      | -0.045 (0.196)      |
| <i>Revenue</i>           |                     |                     | 0.156 (0.101)       | 0.171* (0.101)      |
| <i>Expense</i>           |                     |                     | -0.153 (0.144)      | -0.147 (0.140)      |
| <i>People</i>            |                     |                     | 0.068 (0.583)       | 0.020 (0.574)       |
| <i>Trade</i>             |                     |                     |                     | -0.013* (0.008)     |
| <i>Credit</i>            |                     |                     |                     | -0.034 (0.095)      |
| <i>Edu</i>               |                     |                     |                     | 0.102 (0.099)       |
| <i>Cons</i>              | 3.147*** (0.005)    | 7.127** (3.288)     | 7.814* (4.466)      | 8.349* (4.539)      |
| <i>City</i>              | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>Year</i>              | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>City</i> × <i>Ind</i> | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>Ind</i> × <i>Year</i> | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>N</i>                 | 61 920              | 61 920              | 61 920              | 61 920              |
| <i>R</i> <sup>2</sup>    | 0.946               | 0.946               | 0.946               | 0.946               |

注: \*、\*\*、\*\*\*分别表示1%、5%和10%的显著性水平,括号内为聚类到城市层面的标准误,下表同此。

## (二)平行趋势检验

在本文的研究情境下, DID方法运用的前提是实验组和控制组满足平行趋势假设,即未上线公共数据开放平台的城市在不上线情况下的新企业进入的变化,是上线公共数据开放平台的城市假如不上线公共数据开放平台时的新企业进入变化的反事实。为此,本文使用事件研究法对平行趋势假定进行检验,即估计模型(2)。此外,为了缓解处理样本分布不均衡的影响,参考白俊红等(2022)以及张子尧和黄炜(2023)的研究,在实际操作中将早于-5的观测值的相对时期重新设定为-5期,并以-1期作为基期。

$$Entry_{i,j,t} = \alpha + \sum_{K=-5}^{K=4} \beta_K Treat_{i,t}^K + \gamma Control_{i,t} + City_i + Year_t + City_i \times Ind_j + Ind_j \times Year_t + \varepsilon_{i,j,t} \quad (2)$$

其中,  $Treat^K$  表示为上线公共数据开放平台这一“事件”的虚拟变量,  $K$  为相对于事件的年份,具体而言,当  $K=-1$ , 表示上线公共数据开放平台的前一年,  $Treat^{-1}=1$ , 其他为0; 当  $K=1$ , 表示上线公共数据开放平台后的第一年,  $Treat^1=1$ , 其他为0。以此类推。回归结果中  $Treat^K$  的一系列估计系数反映了公共数据开放对新企业进入影响的动态效应。图1描述了95%置信区间下  $Treat^K$  的估计系数, 可以看到, 在公共数据开放之前,  $Treat^K$  的估计系数均不显著, 这意味着新企业进入随时间的变动趋势在实验组和控制组中无明显差异; 而在公共数据开放后,  $Treat^K$  的系数显著为正, 这说明在政策干预后实验组和控制组的新企业进入变化趋势具有显著差异, 由此平行趋势假设成立。此外, 动态效应检验结果还表明在政策实施后,  $Treat^K$  的系数值逐渐增大, 即公共

<sup>①</sup>值得注意的是,从经济意义看,表2列(4)中 *Dataopen* 的系数值为0.299,这一数值相对较大,可能的原因在于:较多的新企业进入可能会倒逼政府开放公共数据(在新企业进入较多的地区,当地政府的公共服务质量和效率可能较高,政府拥有的公共数据量也可能更大,这会促使政府加快推进公共数据开放,以释放更多的经济价值),从而导致上述结果具有内生性,因此应谨慎解释0.299这一回归系数的经济意义,后文将通过工具变量检验来加强因果识别。

数据开放对新企业进入的促进作用在政策后的一段时间内逐步增强,即促进效果是逐渐释放的。

### (三)稳健性检验

1.遗漏变量分析。城市层面的经济发展变量与新企业进入相关,并可能对公共数据开放平台政策具有选择性,如果遗漏了这些变量,那么*Dataopen*的估计系数反映的就是包含了遗漏变量在内的影响因素对新企业进入的复合效应。为此,在表3列(1)中剔除一线城市和新一线城市的样本<sup>①</sup>,

可以看到,*Dataopen*的估计系数仍显著为正。其次,在表3列(2)中控制“是否是一线或新一线城市”与“是否上线‘公共数据开放平台’”这两个虚拟变量的交互项(*City\_first*×*Dataopen*),可以看出,即便控制了一线或新一线城市随着公共数据开放政策的实施而可能存在的影响,反映公共数据开放政策效果的*Dataopen*的系数仍然显著为正。再者,在表3列(3)中剔除省会城市样本,可以看到,*Dataopen*的估计系数仍显著为正。最后,在表3列(4)中进一步控制省份和年度的联合固定效应以排除省份层面随时间变化的因素对估计结果造成的干扰,可以看到,*Dataopen*的估计系数仍显著为正,进一步验证了公共数据开放对新企业进入的促进作用。

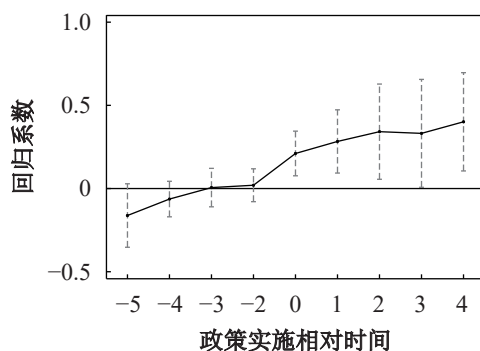


图1 平行趋势检验

表3 遗漏变量分析

| 变量                                  | (1)<br><i>Entry</i> | (2)<br><i>Entry</i> | (3)<br><i>Entry</i> | (4)<br><i>Entry</i> |
|-------------------------------------|---------------------|---------------------|---------------------|---------------------|
| <i>Dataopen</i>                     | 0.308*** (0.094)    | 0.287*** (0.084)    | 0.338*** (0.090)    | 0.264*** (0.073)    |
| <i>City_first</i> × <i>Dataopen</i> |                     | 0.168 (0.132)       |                     |                     |
| <i>Cons</i>                         | 10.144** (4.808)    | 8.664* (4.504)      | 9.206** (4.528)     | 12.854*** (3.902)   |
| <i>Control</i>                      | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>City</i>                         | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Year</i>                         | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>City</i> × <i>Ind</i>            | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Ind</i> × <i>Year</i>            | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Province</i> × <i>Year</i>       |                     |                     |                     | Yes                 |
| <i>N</i>                            | 57 960              | 61 920              | 56 960              | 61 920              |
| <i>R</i> <sup>2</sup>               | 0.945               | 0.946               | 0.945               | 0.958               |

2.异质性处理效应。同一政策对不同个体的差异化效果可能体现在不同时间受到处理的组别和受到处理的时长两个方面。在这种多维异质性下,使用传统双向固定效应(TWFE)模型进行估计会产生“坏控制组”问题,即先受到政策干预的样本会成为后受到政策干预样本的对照组,此时若政策效果在时间或个体维度上具有异质性,则会造成“负权重”问题出现,从而导致估计结果有偏(De Chaisemartin和D’Haultfoeuille, 2020;刘冲等, 2022)。为此,参考已有文献的做法,通过计算“组别—时期平均处理效应”以得到稳健估计(Callaway和Sant’Anna, 2021),图2列展示了采用逆概率加权法对“组别—时期平均处理效应”进行估计的动态趋势图,可以看到,公共数据开放平台上线前,各估计系数均不显著,这与图1的动态趋势一致,从而说明本文中传统双向固定效应(TWFE)模型的估计偏误并不严重,基准模型的估计结果基本可靠。

<sup>①</sup>一线城市和新一线城市名单来自《第一财经新闻》旗下的新一线城市研究所。

3.工具变量检验。在新企业进入较多的地区,当地政府的公共服务质量和效率可能较高,政府拥有的公共数据量也可能更大,这会促使政府加快推进公共数据开放,以释放更多的经济价值。因此,公共数据开放和新企业进入之间可能存在着双向因果关系。对此,本文使用工具变量法来缓解上述可能存在的内生性问题。借鉴欧阳伊玲等(2024)和Du等(2024)的研究,本文选取1984年各城市邮电业务量作为工具变量。

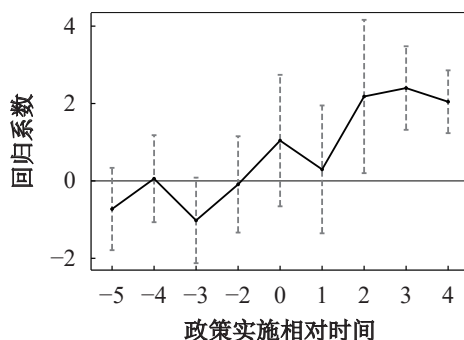


图2 异质性处理效应动态趋势

城市过去使用的通信方式将从基础设施、技术方式和社会偏好等方面影响样本期内各地区的数字基础设施建设和数字经济发展,因此历史邮电业务发展水平越高,当地的数字基础设施建设和数字经济发展可能越好,政府就越可能搭建公共数据开放平台,从而满足工具变量选取的相关性要求。同时,历史邮电数据并不直接影响样本期的新企业进入,因此也满足工具变量选取的外生性要求。此外,考虑到历史邮电业务量仅为截面数据,难以直接作为面板数据的工具变量,我们使用1984年各城市人均邮电业务量和滞后一期的全国互联网上网人数的交互项作为公共数据开放的工具变量(IV)。表4列(1)工具变量第一阶段的回归结果显示,IV的估计系数在1%的水平上显著为正,说明工具变量与解释变量具有较好的相关性。表4列(2)第二阶段的回归结果显示,Dataopen的估计系数显著为正,LM统计量在1%的水平显著为正,F统计量的值为2006.855,即同时拒绝了不可识别检验和弱工具变量检验的原假设,说明在使用历史邮电数据作为工具变量缓解内生性问题后,本文基准回归结果依然成立。

表4 工具变量检验和倾向得分匹配法

| 变量             | (1)<br>Dataopen  | (2)<br>Entry    | (3)<br>Entry     |
|----------------|------------------|-----------------|------------------|
| IV             | 0.028*** (0.007) |                 |                  |
| Dataopen       |                  | 0.169** (0.077) | 0.322*** (0.120) |
| Cons           | -1.670 (1.503)   |                 | -20.140 (17.687) |
| Control        | Yes              | Yes             | Yes              |
| City           | Yes              | Yes             | Yes              |
| Year           | Yes              | Yes             | Yes              |
| City×Ind       | Yes              | Yes             | Yes              |
| Ind×Year       | Yes              | Yes             | Yes              |
| LM             |                  | 2121.029***     |                  |
| Wald F         |                  | 2006.855        |                  |
| N              | 47 040           | 47 040          | 4 820            |
| R <sup>2</sup> | 0.469            | 0.009           | 0.977            |

4.倾向得分匹配法。上线公共数据开放平台的城市和未上线公共数据开放平台的城市间存在诸多可观测和不可观测的特征差异,很可能是这些差异带来了新企业进入结果的不同,从而导致样本存在选择性偏差,影响基准回归结果的稳健性。为此,本文使用倾向得分匹配法(PSM)缓解上述可能存在的样本选择问题。首先,以基准模型中国内生产总值(GDP)、年末总人口(People)、高等学校数(Edu)、财政收入(Revenue)、贷款余额(Credit)为匹配变量对是否上线公共数据开放平台的分组变量进行logit回归,并通过1:2的近邻匹配法为实验组样本匹配具有共同支撑的控制组样本。匹配平衡性检验的结果显示,匹配后实验组与控制组的匹配变量



均值并无显著差异,即支持使用倾向得分匹配方法。然后,使用匹配后的样本重新进行回归,表4列(3)的估计结果显示,*Dataopen*的估计系数显著为正,进一步验证了公共数据开放对新企业进入的促进作用。

5.安慰剂检验。公共数据开放和新企业进入的相关性可能存在伪回归问题,即上线公共数据开放平台城市的新企业进入本身可能就比较,不受是否开放公共数据的影响。因此,本文进一步通过安慰剂检验方法对估计结果进行检验,以确保政策处理效应的有效性。具体地,首先为每个城市随机选取一个年份作为虚拟的公共数据开放平台上线时间点,再从总共286个城市

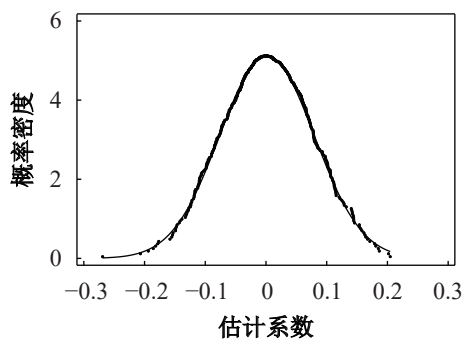


图3 安慰剂检验

中随机选取88个城市<sup>①</sup>。其次,将得到的城市—年份数据与原样本匹配。其中,随机选取的88个城市作为新的实验组,其他城市则作为控制组。最后,按照基准回归模型估计政策效果。重复上述置换检验500次,得到500个随机政策的估计系数,并绘制随机估计系数的概率密度分布图。如图3所示,随机政策的估计系数分布在0附近,且其数值均显著小于真实的估计系数(0.299),进一步验证了真实的公共数据开放政策的有效性。

6.排除竞争性解释。为进一步排除其他相关政策对前文估计结果可能造成的干扰,本文进一步在表5列(1)至(4)中分别控制国家电子商务示范城市政策(*Ecomm*)、“宽带中国”示范城市政策(*Internet*)、“智慧城市”试点政策(*Wisdom*)和大数据试验区政策(*Bigdata*),并在表5列(5)中同时控制上述所有竞争性政策。可以看到,上述回归中*Dataopen*的估计系数均显著为正,进一步验证了基准回归结果的稳健性。

表5 排除竞争性解释

| 变量                       | (1)<br><i>Entry</i> | (2)<br><i>Entry</i> | (3)<br><i>Entry</i> | (4)<br><i>Entry</i> | (5)<br><i>Entry</i> |
|--------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| <i>Dataopen</i>          | 0.290*** (0.085)    | 0.287*** (0.085)    | 0.296*** (0.084)    | 0.293*** (0.083)    | 0.270*** (0.081)    |
| <i>Cons</i>              | 8.961* (4.578)      | 8.575* (4.559)      | 8.220* (4.452)      | 7.766* (4.482)      | 8.393* (4.434)      |
| <i>Ecomm</i>             | Yes                 |                     |                     |                     | Yes                 |
| <i>Internet</i>          |                     | Yes                 |                     |                     | Yes                 |
| <i>Wisdom</i>            |                     |                     | Yes                 |                     | Yes                 |
| <i>Bigdata</i>           |                     |                     |                     | Yes                 | Yes                 |
| <i>Control</i>           | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>City</i>              | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Year</i>              | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>City</i> × <i>Ind</i> | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Ind</i> × <i>Year</i> | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>N</i>                 | 61 920              | 61 920              | 61 920              | 61 920              | 61 920              |
| <i>R</i> <sup>2</sup>    | 0.946               | 0.946               | 0.946               | 0.946               | 0.946               |

## 五、影响机制检验

前文实证结果表明,公共数据开放显著促进了新企业进入,并且该结论在经过一系列稳健性检验后仍然成立。接下来,结合理论分析,本文对降低创业者的信息搜寻成本和提高政府透

①本文样本期间总共有88个城市上线政府数据开放平台。

明度机制进行实证检验。

### (一)降低信息搜寻成本

由于信息搜寻成本难以直接度量,已有文献通常寻找相关替代指标,并且研究对象多聚焦于微观企业。例如, Broekel和Boschma(2012)、张敏等(2018)从搜寻主体和搜寻对象之间的地理距离刻画企业的搜寻成本。李青原等(2023)以客户年报的信息冗余度作为供应商的信息搜寻成本。本文关注的新企业进入为宏观城市层面的变量,因此微观层面的相关指标并不适用于本研究。就本文的研究主题而言,从信息可得性的角度度量创业者的信息搜寻成本更加合理。因此,参考欧阳伊玲等(2024)的做法,使用各地区单位人口的媒体报纸数和报纸总印数作为创业者信息搜寻成本的代理变量。较多的媒体报纸和报纸总印数代表着较高的媒体报道程度与信息传播程度,创业者可以从媒体报道中获取相关信息,从而有利于降低其信息搜寻成本。而公共数据开放平台的信息增益可以有效缓解媒体报道不足地区的信息劣势,缓解创业者所面临的信息不对称。因此,可以预期,如果公共数据开放能够通过降低创业者的信息搜寻成本进而促进新企业进入,那么公共数据开放对新企业进入的促进效应在媒体报道相对不足的地区更为明显。具体地,首先构建人均媒体报纸数和人均报纸总印数的分组变量 $News1$ 和 $News2$ ,当其数值小于样本期间的中位数时,将样本划分为媒体报道不足的组别, $News1$ 和 $News2$ 赋值为1,否则为0。然后,将 $News1$ 和交互项 $News1 \times Dataopen$ 以及 $News2$ 和交互项 $News2 \times Dataopen$ 分别加入基准回归模型中,回归结果如表6列(1)和列(2)所示。可以看到,交互项 $News1 \times Dataopen$ 和 $News2 \times Dataopen$ 的估计系数均显著为正,表明公共数据开放对新企业进入的促进效应在媒体报道相对不足的地区更为明显,从而在一定程度上验证了公共数据开放能够通过降低创业者面临的信息搜寻成本进而促进新企业进入这一结论。由此,研究假说H1成立。

表6 影响机制检验

| 变量                              | (1)<br><i>Entry</i> | (2)<br><i>Entry</i> | (3)<br><i>Entry</i> | (4)<br><i>Entry</i> |
|---------------------------------|---------------------|---------------------|---------------------|---------------------|
| <i>Dataopen</i>                 | -0.204(0.166)       | 0.080<br>(0.114)    | 0.073<br>(0.111)    | 0.111*(0.061)       |
| <i>News1</i> × <i>Dataopen</i>  | 0.642*** (0.181)    |                     |                     |                     |
| <i>News1</i>                    | -0.048* (0.024)     |                     |                     |                     |
| <i>News2</i> × <i>Dataopen</i>  |                     | 0.369*** (0.126)    |                     |                     |
| <i>News2</i>                    |                     | -0.162** (0.070)    |                     |                     |
| <i>Trans1</i> × <i>Dataopen</i> |                     |                     | 0.291** (0.116)     |                     |
| <i>Trans1</i>                   |                     |                     | -0.051* (0.026)     |                     |
| <i>Trans2</i> × <i>Dataopen</i> |                     |                     |                     | 0.072** (0.032)     |
| <i>Trans2</i>                   |                     |                     |                     | 0.011 (0.017)       |
| <i>Cons</i>                     | 9.195** (4.348)     | 6.640 (4.468)       | 10.119* (5.161)     | 3.941 (3.615)       |
| <i>Control</i>                  | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>City</i>                     | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>Year</i>                     | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>City</i> × <i>Ind</i>        | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>Ind</i> × <i>Year</i>        | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          | <i>Yes</i>          |
| <i>N</i>                        | 61 920              | 61 920              | 50 880              | 17 020              |
| <i>R</i> <sup>2</sup>           | 0.946               | 0.946               | 0.949               | 0.989               |

### (二)提高政府透明度

政府透明度是现代民主治理的核心要素之一,它涉及政府信息的公开程度、政府决策过程的透明性以及政府行为的可预测性(Da Cruz等,2016)。参考高明等(2021)的做法,使用中国社

会科学院发布的《中国政府透明度年度报告》披露的政府透明度指数作为政府透明度的第一个度量指标,其数值越大代表政府透明度越高<sup>①</sup>。同时,进一步使用中国人民大学国家发展战略研究院发布的《中国城市政商关系评价报告》披露的政府透明指数作为政府透明度的第二个度量指标<sup>②</sup>。可以预期,如果公共数据开放能够通过提高政府透明度进而促进新企业进入,那么公共数据开放对新企业进入的促进作用应在政府透明度相对较低的地区更为明显。具体地,根据上述两份报告分别构建政府透明度的分组变量 $Trans1$ 和 $Trans2$ ,当其数值小于样本中位数时,将样本划分为政府透明度较低的组别, $Trans1$ 和 $Trans2$ 赋值为1,否则为0。然后,将 $Trans1$ 和交互项 $Trans1 \times Dataopen$ 以及 $Trans2$ 和交互项 $Trans2 \times Dataopen$ 分别加入基准回归模型中,回归结果如表6列(3)和列(4)所示,可以看到,交互项 $Trans1 \times Dataopen$ 和 $Trans2 \times Dataopen$ 的估计系数均显著为正,表明公共数据开放对新企业进入的促进作用主要体现在政府透明度较低的地区,从而在一定程度上验证了公共数据开放能够通过提高政府透明度进而促进新企业进入这一结论。由此,研究假说H2成立。

## 六、进一步分析

### (一)产权的影响

不同类型产权的新企业进入不仅对制度环境和市场环境变动的敏感性不同,其在信息获取、政策支持等方面也存在着显著差异,因此公共数据开放对新企业进入的影响效应可能在不同类型产权方面存在异质性。借鉴李硕等(2022)的研究,根据企业产权类型,将企业划分为民营企业、国有企业、外资企业、港澳台企业和其他企业(包括集体企业和合作社等)。表7列(1)~(5)的回归结果显示, $Dataopen$ 的估计系数在企业产权类型为民营企业、外资企业和港澳台企业时显著为正,而在产权类型为国有企业和其他企业时并不显著,这说明公共数据开放主要促进了民营企业、外资企业和港澳台企业的进入,而对国有企业和其他企业的进入并没有显著的促进作用。可能的原因在于,相较于民营企业、外资企业和港澳台企业,国有企业和其他企业的成立通常由政府部门直接或间接决定,其新建投资决策受诸多非市场化因素的影响,从而导致国有企业和其他企业的进入主体对公共数据开放并不敏感,由此对其进入市场的行为并没有产生显著的影响。

表7 企业产权类型异质性

| 变量                | (1)<br>民营企业      | (2)<br>国有企业     | (3)<br>外资企业      | (4)<br>港澳台企业     | (5)<br>其他企业   |
|-------------------|------------------|-----------------|------------------|------------------|---------------|
| $Dataopen$        | 0.313*** (0.085) | 0.057 (0.043)   | 0.275*** (0.053) | 0.420*** (0.059) | 0.055 (0.035) |
| $Cons$            | 7.930* (4.441)   | 4.247** (2.063) | -2.718 (2.925)   | -6.155* (3.441)  | 3.227 (2.147) |
| $Control$         | Yes              | Yes             | Yes              | Yes              | Yes           |
| $City$            | Yes              | Yes             | Yes              | Yes              | Yes           |
| $Year$            | Yes              | Yes             | Yes              | Yes              | Yes           |
| $City \times Ind$ | Yes              | Yes             | Yes              | Yes              | Yes           |
| $Ind \times Year$ | Yes              | Yes             | Yes              | Yes              | Yes           |
| $N$               | 61 920           | 61 920          | 61 920           | 61 920           | 61 920        |
| $R^2$             | 0.946            | 0.765           | 0.700            | 0.633            | 0.856         |

### (二)行业的影响

进一步地,从行业层面考察公共数据开放对新企业进入影响的异质性。首先,根据国民经济行业分类,将新企业按行业分类划分为第一产业、第二产业和第三产业,其中第一产业为农

①《中国政府透明度年度报告》公开可得的起始年份为2011年,与原数据合并后会损失一小部分样本。

②由于《中国城市政商关系评价报告》于2017年开始公布,与原数据合并后会损失较多样本,但仍可作为政府透明度机制的稳健性检验。

林牧渔业,第二产业为工业,第三产业为服务业。表8列(1)至(3)的回归结果显示, *Dataopen* 的估计系数在第一产业样本中并不显著,在第二产业和第三产业样本中均显著为正,且系数值在第三产业样本中更大,显著性也更高。费舍尔组合检验结果显示第二产业和第三产业分组回归后的组间系数存在显著差异( $P=0.000$ ),表明相对于第二产业,公共数据开放更加促进了第三产业的新企业进入。此外,参考李平等(2017)的研究,本文进一步将第三产业划分为生产性服务业和生活性服务业。表8列(4)和列(5)报告了服务业异质性回归结果,可以看到, *Dataopen* 的估计系数在生活性服务业样本中显著更大,并且分组回归结果通过了组间系数差异检验,表明相对于生产性服务业企业,公共数据开放更加促进了生活性服务业的新企业进入。导致上述结果可能的原因在于:一方面,根据中国信息通信研究院发布的《中国数字经济发展研究报告(2023年)》显示,2022年,我国第一、二、三产业数字经济渗透率分别为10.5%、24.0%和44.7%。因此,相对于第二和第三产业,第一产业的数字化程度相对较低,政府所拥有的第一产业方面的公共数据资源相对较少,从而导致公共数据开放并未显著促进第一产业的新企业进入。另一方面,相对于第二产业,第三产业的数字化程度明显更高,同时生活性服务业的数字化程度也要明显高于生产性服务业。并且,公共数据是政府为公众提供公共服务过程中产生和搜集的数据,这部分数据更多地涉及生活性服务业,由此导致公共数据开放更加促进了第三产业(尤其是生活性服务业)的新企业进入。

表8 行业异质性

| 变量                    | (1)<br>第一产业       | (2)<br>第二产业     | (3)<br>第三产业      | (4)<br>生产性服务业    | (5)<br>生活性服务业    |
|-----------------------|-------------------|-----------------|------------------|------------------|------------------|
| <i>Dataopen</i>       | 0.203(0.140)      | 0.166*(0.093)   | 0.350*** (0.083) | 0.343*** (0.098) | 0.448*** (0.099) |
| <i>Cons</i>           | 20.371*** (7.176) | 9.797** (4.801) | 7.509* (4.379)   | 9.383* (5.088)   | 8.234 (5.075)    |
| <i>Control</i>        | Yes               | Yes             | Yes              | Yes              | Yes              |
| <i>City</i>           | Yes               | Yes             | Yes              | Yes              | Yes              |
| <i>Year</i>           | Yes               | Yes             | Yes              | Yes              | Yes              |
| <i>City×Ind</i>       | Yes               | Yes             | Yes              | Yes              | Yes              |
| <i>Ind×Year</i>       | Yes               | Yes             | Yes              | Yes              | Yes              |
| 组间差异检验                |                   |                 | 0.184***         |                  | 0.106***         |
| <i>N</i>              | 3 096             | 12 384          | 43 344           | 18 576           | 18 576           |
| <i>R</i> <sup>2</sup> | 0.882             | 0.936           | 0.950            | 0.941            | 0.930            |

### (三)互联网普及率的影响

政府数据开放网络平台是下载和获取公共数据的载体,是开放数据供给侧和需求侧沟通互动的节点和桥梁。因此,社会主体对公共数据的下载、获取以及使用情况一定程度上依赖于其所在城市的互联网普及率。当互联网普及率相对较低时,即使政府开放大量的公共数据,公众也可能无法进行下载和使用,即公共数据开放对新企业进入的影响效应受到互联网普及率的影响。本文使用各城市人均互联网宽带接入用户数作为互联网普及率的度量指标,具体地,根据各城市互联网普及率是否大于样本中位数将样本划分为互联网普及率较高的组别和较低的组别。表9列(1)和列(2)的分组回归结果显示, *Dataopen* 的估计系数仅在互联网普及率较高的样本中显著为正,表明相对于互联网普及率较低的城市,公共数据开放主要促进了互联网普及率较高的城市的新企业进入。

### (四)公共数据开放质量的影响

如前所述,公共数据开放的实践形式是政府将公共数据资源统一整合到开放平台上以供

社会公众下载和使用。目前,虽然各地区的公共数据开放平台都已提供了各种类型的公共数据资源,但是地区间存在着较大的差异,并且部分已上线的平台也存在有平台无数据、平台无法登录或数据杂乱无章的情况,这都会在一定程度上限制公共数据开放对新企业进入的促进效应。基于此,本文进一步使用复旦大学数字与移动治理实验室构建的中国开放数林指数作为公共数据开放质量的代理变量,对公共数据开放质量的新企业进入效应进行检验<sup>①</sup>。开放林指数对各地公共数据开放平台进行了较为全面的客观评价,除综合指数外还包括数据层指数、平台层指数、准备度指数等细分指数。其中,数据层指数是对数据开放范围、数据规范、数据数量等的评价。由此,本文使用综合指数(*Index1*)和数据层指数(*Index2*)进行检验。需要说明的是,《中国地方政府数据开放报告(城市)》中开放林指数自2017年开始公布,并且报告仅公开了指数排名前50城市的数据,因此匹配数据后样本量会大幅减少,但这可在一定程度验证本文想要考察的问题。表9列(3)和列(4)的回归结果显示,*Index1*和*Index2*的估计系数均显著为正,表明公共数据开放质量越高,新企业进入就越多。

表9 互联网普及率和公共数据开放质量的影响

| 变量                       | (1)<br>高互联网普及率  | (2)<br>低互联网普及率 | (3)<br><i>Entry</i> | (4)<br><i>Entry</i> |
|--------------------------|-----------------|----------------|---------------------|---------------------|
| <i>Dataopen</i>          | 0.237** (0.100) | 0.229 (0.144)  |                     |                     |
| <i>Index1</i>            |                 |                | 0.007** (0.003)     |                     |
| <i>Index2</i>            |                 |                |                     | 0.011* (0.006)      |
| <i>Cons</i>              | -0.913 (6.223)  | 5.187 (7.311)  | 9.982 (8.436)       | 10.228 (8.957)      |
| <i>Control</i>           | <i>Yes</i>      | <i>Yes</i>     | <i>Yes</i>          | <i>Yes</i>          |
| <i>City</i>              | <i>Yes</i>      | <i>Yes</i>     | <i>Yes</i>          | <i>Yes</i>          |
| <i>Year</i>              | <i>Yes</i>      | <i>Yes</i>     | <i>Yes</i>          | <i>Yes</i>          |
| <i>City</i> × <i>Ind</i> | <i>Yes</i>      | <i>Yes</i>     | <i>Yes</i>          | <i>Yes</i>          |
| <i>Ind</i> × <i>Year</i> | <i>Yes</i>      | <i>Yes</i>     | <i>Yes</i>          | <i>Yes</i>          |
| <i>N</i>                 | 31 300          | 29 000         | 2 140               | 2 140               |
| <i>R</i> <sup>2</sup>    | 0.949           | 0.947          | 0.994               | 0.994               |

## 七、结论与启示

本文利用2009—2019年中国所有新建企业的工商注册信息,构建城市—行业层面的新企业进入数据,并以公共数据开放平台上线作为准自然实验,使用双重差分法实证检验公共数据开放对新企业进入的影响。主要得到以下研究结论:(1)公共数据开放显著促进了新企业进入,并且该结论通过了遗漏变量分析、异质性处理效应、工具变量检验、倾向得分匹配、安慰剂检验、排除竞争性解释等一系列稳健性检验。(2)降低创业者的信息搜寻成本和提高政府透明度是公共数据开放促进新企业进入的两个主要渠道。(3)公共数据开放对新企业进入的促进作用主要体现在非国有企业、服务业(尤其是生活性服务业)以及互联网普及率较高的城市。(4)公共数据开放质量越高,对新企业进入的促进效果越强。

新企业进入是经济增长和就业创造的重要来源,而公共数据开放对新企业进入具有促进作用,这充分验证了公共数据要素的资源配置效应,对于畅通数据资源大循环具有重要意义。基于上述研究结论,本文提出如下政策启示:第一,各地方政府应主动依法依规地开放公共数据资源,在保障数据安全和隐私保护的情况下做到全面开放。特别是,还未上线公共数据开放平台的省份或城市应积极争取早日上线。第二,进一步强化农业和工业部门的网络化、信息化和智能化建设,以整合更多农业和工业方面的公共数据资源。同时,持续完善网络基础设施建

<sup>①</sup>部分研究也使用这一指数作为公共数据开放质量的代理变量,例如彭远怀(2024)和欧阳伊玲等(2024)。

设,逐步提高互联网普及率,以提高公众对公共数据的可得性。第三,注重公共数据开放平台体系建设,扩大数据开放范围和数据更新频率,以不断提高公共数据开放质量。

### 主要参考文献

- [1]陈潭. 国家治理的数据赋能及其秩序生产[J]. *社会科学研究*, 2023, (6): 12-30.
- [2]方锦程, 刘颖, 高昊宇, 等. 公共数据开放能否促进区域协调发展?——来自政府数据平台上线的准自然实验[J]. *管理世界*, 2023, 39(9): 124-142.
- [3]高明, 艾美彤, 贾若. 家庭金融参与中的信任重建——来自农村社会养老保险的证据[J]. *经济研究*, 2021, 56(8): 174-191.
- [4]何雨可, 牛耕, 逮建, 等. 数字治理与城市创业活力——来自“信息惠民国家试点”政策的证据[J]. *数量经济技术经济研究*, 2024, 41(1): 47-66.
- [5]黄先海, 虞柳明, 戴岭. 政府数据开放与创新驱动: 内涵、机制及实践路径[J]. *东南学术*, 2023, (2): 102-113.
- [6]李磊, 何艳辉. 机器人使用与创业选择[J]. *南方经济*, 2024, (1): 124-141.
- [7]李青原, 李昱, 章尹赛楠, 等. 企业数字化转型的信息溢出效应——基于供应链视角的经验证据[J]. *中国工业经济*, 2023, (7): 142-159.
- [8]李硕, 王敏, 张丹丹. 中央环保督察和企业进入: 来自企业注册数据的证据[J]. *世界经济*, 2022, 45(1): 110-132.
- [9]马龙, 陈奕博. 基于技术的治理: 隐私计算技术赋能政府数据开放的价值与路径研究[J]. *中国行政管理*, 2023, 39(9): 105-113.
- [10]莫怡青, 李力行. 零工经济对创业的影响——以外卖平台的兴起为例[J]. *管理世界*, 2022, 38(2): 31-45.
- [11]欧阳伊玲, 王愉靖, 李平, 等. 数据要素与城投债定价: 基于公共数据开放的准自然实验[J]. *世界经济*, 2024, 47(2): 174-203.
- [12]彭远怀. 政府数据开放的价值创造作用: 企业全要素生产率视角[J]. *数量经济技术经济研究*, 2023, 40(9): 50-70.
- [13]彭箴. 论政务数据共享的推进与边界[J]. *交大法学*, 2023, (6): 63-77.
- [14]田鸽, 黄海, 张勋. 数字金融与创业高质量发展: 来自中国的证据[J]. *金融研究*, 2023, (3): 74-92.
- [15]谢绚丽, 沈艳, 张皓星, 等. 数字金融能促进创业吗?——来自中国的证据[J]. *经济学(季刊)*, 2018, 17(4): 1557-1580.
- [16]于文超, 梁平汉, 高楠. 公开能带来效率吗?——政府信息公开影响企业投资效率的经验研究[J]. *经济学(季刊)*, 2020, 19(3): 1041-1058.
- [17]张莉, 林安然. 中国地方政府数据开放与经济发展质量[J]. *公共行政评论*, 2023, 16(2): 120-137.
- [18]张敏, 刘耀淞, 王欣, 等. 企业与税务局为邻: 便利避税还是便利征税?[J]. *管理世界*, 2018, 34(5): 150-164.
- [19]张勋, 万广华, 张佳佳, 等. 数字经济、普惠金融与包容性增长[J]. *经济研究*, 2019, 54(8): 71-86.
- [20]赵涛, 张智, 梁上坤. 数字经济、创业活跃度与高质量发展——来自中国城市的经验证据[J]. *管理世界*, 2020, 36(10): 65-75.
- [21]郑磊. 开放政府数据研究: 概念辨析、关键因素及其互动关系[J]. *中国行政管理*, 2015, (11): 13-18.
- [22]朱峥. 政府数据开放的权利基础及其制度构建[J]. *电子政务*, 2020, (10): 117-128.
- [23]Ansari B, Barati M, Martin E G. Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research[J]. *Government Information Quarterly*, 2022, 39(1): 101657.
- [24]Attard J, Orlandi F, Scerri S, et al. A systematic review of open government data initiatives[J]. *Government Information Quarterly*, 2015, 32(4): 399-418.
- [25]Broekel T, Boschma R. Knowledge networks in the Dutch aviation industry: The proximity paradox[J]. *Journal of Economic Geography*, 2012, 12(2): 409-433.
- [26]Carriere-Swallow M Y, Haksar M V. The economics and implications of data: An integrated perspective[M]. Washington: International Monetary Fund, 2019.
- [27]Da Cruz N F, Tavares A F, Marques R C, et al. Measuring local government transparency[J]. *Public Management Review*, 2016, 18(6): 866-893.
- [28]Du J Y, Gao H Y, Wen H Y, et al. Public data access and stock price synchronicity: Evidence from China[J]. *Economic Modelling*, 2024, 130: 106591.

- [29]Fabro G, Aixalá J. Economic growth and institutional quality: Global and income-level analyses[J]. *Journal of Economic Issues*, 2009, 43(4): 997-1023.
- [30]Fairlie R W. The personal computer and entrepreneurship[J]. *Management Science*, 2006, 52(2): 187-203.
- [31]Hughes-Cromwick E, Coronado J. The value of US government data to US business decisions[J]. *Journal of Economic Perspectives*, 2019, 33(1): 131-146.
- [32]Jetzek T, Avital M, Bjorn-Andersen N. Data-driven innovation through open government data[J]. *Journal of Theoretical and Applied Electronic Commerce Research*, 2014, 9(2): 100-120.
- [33]Jones C I, Tonetti C. Nonrivalry and the economics of data[J]. *American Economic Review*, 2020, 110(9): 2819-2858.
- [34]Kaplan D S, Piedra E, Seira E. Entry regulation and business start-ups: Evidence from Mexico[J]. *Journal of Public Economics*, 2011, 95(11-12): 1501-1515.
- [35]Lourenço R P. An analysis of open government portals: A perspective of transparency for accountability[J]. *Government Information Quarterly*, 2015, 32(3): 323-332.
- [36]Park S, Gil-Garcia J R. Open data innovation: Visualizations and process redesign as a way to bridge the transparency-accountability gap[J]. *Government Information Quarterly*, 2022, 39(1): 101456.
- [37]Shane S. Prior knowledge and the discovery of entrepreneurial opportunities[J]. *Organization Science*, 2000, 11(4): 448-469.
- [38]Tan Y, Li X Y. The impact of internet on entrepreneurship[J]. *International Review of Economics & Finance*, 2022, 77: 135-142.
- [39]Wang L H, Shao J. Digital economy, entrepreneurship and energy efficiency[J]. *Energy*, 2023, 269: 126801.
- [40]Zuiderwijk A, Janssen M, Davis C. Innovation with open data: Essential elements of open data ecosystems[J]. *Information Polity*, 2014, 19(1-2): 17-33.

## Awakening “Sleepy” Data: Public Data Openness and New Firm Entry

Zheng Zhiqiang<sup>1</sup>, He Jiali<sup>2</sup>

(1. *Institute of Anhui's Economic and Social Development Research, Anhui University of Finance & Economics, Bengbu 233030, China*; 2. *School of Economics and Management, Anhui Agricultural University, Hefei 230036, China*)

**Summary:** With the continuous growth and accumulation of public data resources, evaluating the economic value of public data openness has become an important issue. It is of great significance for optimizing data resource allocation and facilitating data circulation. Based on this, this paper adopts the registration information of all newly established firms in China from 2009 to 2019 to construct new firm entry data at the city-industry level, and takes the launch of public data open platforms as a quasi-natural experiment to empirically test the impact of public data openness on new firm entry using the DID method. The study finds that public data openness significantly promotes new firm entry. Mechanism testing indicates that reducing entrepreneurs' information search costs and improving government transparency are the two main channels through which public data openness promotes new firm entry. Further analysis reveals that the promotion effect of public data openness on new firm entry is mainly observed in non-state-owned enterprises, service industries (especially living services), and cities with higher Internet penetration rates. Additionally, the higher the quality of public data openness, the stronger the promotion effect on new firm entry. New firm entry is an important source of economic

growth and job creation, and the promotion effect of public data openness on new firm entry fully validates the resource allocation effect of public data. The rapid development and widespread application of digital technology have led to the emergence of the digital economy. Data is the foundation of digitization, networking, and intelligence, and a key production factor in the development of the digital economy. Although existing literature has thoroughly explored the entrepreneurial effect of digital economic development, there is currently no direct evidence or systematic analysis on how data as a production factor itself affects new firm entry. This paper investigates the effect of public data openness on new firm entry from the perspective of data as a production factor, effectively supplementing the theoretical analysis and empirical research on the economic value of data as a production factor.

**Key words:** public data openness; data factor; new firm entry

(责任编辑:王雅丽)