

中国数字乡村发展统计测度与时空异质性研究

毛 军, 梁沐晗, 谢嘉豪

(海南师范大学 数学与统计学院, 海南 海口 571158)

摘要: 中国政府积极推动着数字技术与乡村发展融合发展的数字乡村建设。引入因果推断方法研究影响因素对数字乡村建设在不同时期中的平均处理效应 (ATE) 的重要程度, 并通过构建空间杜宾模型对其时空特征与溢出效应进行探析。研究表明: 第一, 科技创新水平对数字乡村建设的影响, 由前期的负向作用 ($ATE = -1.908$) 转为后期的正向作用 ($ATE = 1.936$), 后期中低科技创新水平地区的 $ATC > ATT$, 说明中低科技创新水平地区创新效果会优于科技创新水平高的地区; 第二, 数字乡村发展具有较强的时间积累性, 受到经济要素流动等影响, 呈现出明显的“棘轮效应”特征和空间依赖性, 本地科技创新水平、产业结构升级将对本地数字乡村发展产生正面影响, 邻近地区的科技创新水平、产业结构升级对本地数字乡村发展并未产生实际影响。由此可见, 加强数字化农业技术创新供给能力、稳步推进产业结构升级对中国数字乡村的区域协调发展具有重要意义。

关键词: 数字乡村发展; 统计测度; 时空异质性; 因果推断

中图分类号: F323

文献标识码: A

文章编号: 1005-0566(2024)05-0080-10

Statistical measurement and spatial-temporal heterogeneity of digital rural development in China

MAO Jun, LIANG Muhan, XIE Jiahao

(School of Mathematics and Statistics, Hainan Normal University, Haikou 571158, China)

Abstract: The Chinese government actively promotes the construction of digital villages that integrate digital technology and rural development. This paper introduces causal inference method to study the importance of influencing factors to the average treatment effect (ATE) of digital rural construction in different periods. This paper constructs a spatial Dubin model to analyze its spatial and temporal characteristics and spillover effects. The research shows that: First, the impact of scientific and technological innovation level on digital rural construction has changed from the negative effect of the previous stage ($ATE = -1.908$) to the positive effect of the later stage ($ATE = 1.936$); in the later stage, $ATC > ATT$ in the regions with low and medium technological innovation level, indicating that the innovation effect of the regions with low and medium technological innovation level will be better than that of the regions with high technological innovation level. Secondly, the development of digital villages has a strong time accumulation, showing obvious 'ratchet effect' characteristics; the development level of digital villages is affected by the flow of economic factors and thus forms obvious spatial dependence and spatial spillover effects; the level of local scientific and technological innovation and the upgrading of industrial structure will have a positive impact on the development of local digital villages. The level of scientific and technological innovation and the upgrading of industrial structure in neighboring areas have no actual

收稿日期: 2023-10-20 修回日期: 2024-04-10

基金项目: 国家哲学社会科学基金项目“数字财政赋能农业低碳绿色转型的理论机制与实现路径研究”(23BJY026)。

作者简介: 毛军(1986—), 男, 海南琼中人, 海南师范大学数学与统计学院副教授, 研究方向为社会经济统计。通信作者: 谢嘉豪。

impact on the development of local digital villages. It can be seen that strengthening the supply capacity of digital agricultural technology innovation and steadily promoting the upgrading of industrial structure are of great significance to the regional coordinated development of digital villages in China.

Key words: digital village development; statistical measurement; spatio-temporal heterogeneity; causal inference

一、引言与文献综述

数字经济高速发展所带来的“扩散效应”和“普惠效应”为中国数字乡村建设带来了机遇^[1],数字乡村以农民为主体、以互联网为载体,日益改变着农村媒介生活和发展方式^[2]。然而,2021年数字农业发展水平占数字化发展总量的占比仅为7.3%,这反映出中国目前较低的农业数字化水平,数字基础设施建设和人才资源不足等现实问题^[3],导致数字技术发展在乡村发展的渗透与扩散存在明显障碍^[4]。在此背景下,中央和各级地方政府高度重视数字乡村发展。2024年中央一号文件提出:“以提升乡村建设水平为重点,强化科技和改革双轮驱动,推进乡村全面振兴。”中共中央、国务院印发的《数字中国建设整体布局规划》指出:“要求推进数字社会治理精准化,深入实施数字乡村发展行动,以数字化赋能乡村产业发展、乡村建设和乡村治理。”由此可见,数字乡村建设的政策实施在时间维度呈现出螺旋式演进特征,为推进中国数字乡村建设提供政策依据。在实践层面,国外数字乡村建设重点聚焦在宽带基础设施、数字公共服务和数字创新技能。美国数字乡村建设内容包括乡村通讯设施、乡村电子医疗网络和远程教育网络设施等。英国农村数字化建设主要实施宽带接入、数字培训、社区数字中心和电子健康等创新措施。德国联邦资助建设乡村高速宽带基础设施,并试验推广智能农业领域5G应用。日本大力发展以农业机器人为核心的无人农场,提供远程看护、远程课程培训和互联网医疗咨询等服务应用,来补偿乡村人口外迁和老龄化问题。国内诸多企业推出创新性的建设模式,阿里巴巴的“乡村钉”数字乡村治理模式、腾讯的“移动互联网+乡村”的赋能模式,以数据要素赋能推动农业数字化生产力发展,以平台化经济组织形式推动农村生产关系变革。然而,随着中国数字乡村建设不断推进,区域间数字乡村发展不平衡、不充分凸显出来。因此,借助数字乡村建设实施的政策红利,加快推进农业农村信息化进程,运

用数字科技创新发展的扩散效应,有效弥合区域间数字鸿沟,为乡村振兴的全面“布局”和重点“破局”提供重要保障。

城乡发展不平衡和农村发展不充分等问题,导致当前农村资源资产存在着盘活与流动障碍和普遍利用率不高,制约着乡村振兴战略的实施,阻碍农村经济高质量发展。而数字经济以提供的要素和数据深刻变革实体经济和社会生活,重塑社会生产、交换、分配、消费新格局,数字技术赋能乡村振兴计划由此实现产村融合,推进乡村振兴战略实施。2023年中央经济工作会议提出“全面推进乡村振兴。坚持农业农村优先发展,坚持城乡融合发展,畅通城乡要素流动。扎实推动乡村产业、人才、文化、生态、组织振兴”。因而,如何全面推进乡村振兴、有效加快数字乡村建设,建立现代区域乡村社会治理体系,开创城乡融合发展和现代化建设新局面,成为数字中国建设的重要内容。

推进数字乡村建设是顺应信息全球化和数字全球化的趋势,伴随着数字化技术在数字发展过程中的融合渗透,目前关于数字乡村问题的研究主要包括以下4个方面。①关于如何界定数字乡村发展内涵的研究。数字乡村能够整合数字化资源^[5-6],是数字化技术在农业农村的融合渗透,利用信息技术创新驱动,提高农村现代化发展的内生动力^[7-8]。②数字经济赋能农业农村数字化建设研究。数字乡村建设能够通过人才、技术、资本、需求等方面的提高,弥补乡村经济信息化转型的技术短板^[9-10],充分发挥对乡村治理、农业生产、产业转型、公共服务的积极赋能作用与普惠效应^[11-12],消弭城乡“数字鸿沟”,农村公共服务体系实现重构性健全,能够培育新型乡村数字产业,推动农业农村多元化发展,形成驱动农村产业兴旺、城乡融合发展以及乡村全面振兴的新动能^[13-14]。③数字乡村发展评价体系构建研究。夏显力等^[15]、汪亚楠等^[16]围绕产业、生产、经营三大体系构建数字乡村发展的指标。Li等^[17]构建基础设施、资金投入、环境

开发、投资效益的数字乡村发展指标体系。许敬辉等^[18]则从数字乡村建设参与度出发,构建包括应用环境、基础设施、治理力度、政策支持的数字乡村发展指标体系,并应用熵权法进行测度与分析。④关于数字乡村发展现实困境与提升途径。中国数字乡村建设主要劣势体现在信息生产、接入、支付和使用,突出症结是区域间存在数字鸿沟问题^[19-20]。因此,数字乡村发展应保持乡村性价值,分层次实施农业农村现代化建设^[21-22],培育农民主体意识、释放农业发展能级、颠覆重构农村社会关系,实现乡村善治与共治^[23-24]。

综上所述,现有文献围绕中国数字乡村发展的战略意义、发展现状与模式、优化策略进行了探讨。然而,对于如何构建综合全面的数字乡村发展水平指标体系,以及数字乡村发展的区域异质性及时空分布特征鲜有涉及。因此,本文尝试以乡村数字基础设施指数、乡村数字产业化指数、乡村数值化参与指数、乡村数字化经济指数、乡村产业数字化指数等 5 个维度,构建数字乡村发展水平评价体系,并运用熵权法对中国数字乡村发展水平进行定量测度和评估;运用 Dagum 基尼系数法,深入剖析中国数字乡村建设的发展现状和区域特征。并且本文引入因果推断方法研究影响因素对数字乡村建设在不同时期中的重要程度,运用空间杜宾模型实证剖析空间因素对数字乡村发展水平的作用机理,希冀为中国数字乡村建设的区域协调发展战略的优势扩散提供有益借鉴。

二、中国数字乡村发展的统计测度和地区差异分解

(一) 中国数字乡村发展的统计测度

本文选取 2011—2022 年中国 31 个省份(未包含港澳台地区)作为研究样本,数据来源于《中国统计年鉴》《中国经济普查年鉴》《中国城市统计年鉴》《中国农村统计年鉴》《中国社会统计年鉴》以及各省份统计年鉴,并运用插值法对个别缺失数据进行补充处理。本文运用熵权法对数字乡村指数进行测算,充分考虑我国数字乡村建设基本现状以及数据可得性,主要从乡村数字基础设施水平、乡村数字产业化水平、乡村数值化参与水平、乡村数字化经济水平、乡村产业数字化水平等 5 个

维度构建数字乡村发展指标体系(见表 1)。其中,乡村数字基础设施是持续推动农业农村现代化改革的本质要求,主要包括乡村的网络信息通信等方面。乡村数字产业化水平是建设数字乡村的有力保证,乡村数字产业化水平越高,则其数字产业的发展水平越高。乡村数值化参与度用于衡量数字乡村建设中居民的参与情况,主要包括为居民从事数字化产业的情况。乡村数字化经济水平是数字乡村建设的重要组成部分,可以有效改善和提升乡村地区数字建设各方面需求。乡村产业数字化水平是指数字乡村将传统产业升级为数字化产业的能力,也是数字乡村发展必由之路^[25-26]。

表 1 中国数字乡村发展指标体系

	一级指标	二级指标	权重
数字乡村发展水平	乡村数字基础设施指数	农村宽带用户接入数量(万户)	0.351 250
		移动电话普及率(部/百人)	0.088 822
		快递网点数量(处)	0.224 743
		光缆线路长度(公里)	0.225 368
		互联网普及率(%)	0.109 823
	乡村数字化经济指数	农村居民消费水平(元)	0.121 635
		快递业务收入(万元)	0.495 800
		地方财政科学技术支出(亿元)	0.266 843
		非农业收入比例	0.020 973
		从事数字化产业人员平均工资(元)	0.094 750
	乡村数字产业化指数	信息技术产业收入(万元)	0.130 689
		电信业务总量(亿元)	0.243 862
		网站数(万个)	0.286 857
		软件业务收入万元)	0.131 681
		软件产品收入(万元)	0.206 911
	乡村产业数字化指数	电子商务销售额(亿元)	0.239 830
		电子商务的采购额(亿元)	0.259 068
快递件数(万件)		0.392 003	
专利授权数量(件)		0.109 099	
乡村数字化参与度指数	农村信息化产业人员数量(万人)	0.395 916	
	金融业单位农村就业人员数量(万人)	0.287 483	
	农村信息化从业人员比重	0.316 601	

(二) 中国数字乡村发展水平地区差异性分解

为研究数字乡村发展水平的区域性差异,本文使用 Dagum 基尼系数法测算了 2011—2022 年中国数字乡村发展水平的基尼系数。Dagum 基尼系数的测算公式为:

$$G = \frac{\sum_{j=1}^k \sum_{h=1}^k \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}{2\mu n^2} \quad (1)$$

$$\overline{Y_1} \leq \overline{Y_h} \leq \dots \leq \overline{Y_j} \leq \dots \leq \overline{Y_k} \quad (2)$$

式(2)中, μ 表示为所有地区数字乡村发展水平的平均值; n 表示地区数量; k 表示划分的区域

数量; $y_{ji}(y)$ 表示 $j(h)$ 地区内的数字乡村发展水平; $n_j(n_h)$ 表示 $j(h)$ 地区内所包含的省份数量。基尼系数测算前提, 需要依据公式(2)对各地区的数字乡村发展水平的平均值进行排序, 进而划分各地区。

Dagum 基尼系数 $G = G_w + G_{nb} + G_t$ 。其中, G_w 表示地区内差异、 G_{nb} 表示地区间差异; G_t 表示超变密度贡献。 G_{ji} 和 G_{jh} 分别代表 j 地区内基尼系数与 j 、 h 地区间基尼系数。

$$G_w = \sum_{j=1}^k G_{ji} P_j S_j \quad (3)$$

$$G_{ji} = \frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_j} |y_{ji} - y_{jr}|}{2 \bar{Y}_j n_j^2} \quad (4)$$

$$G_{nb} = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (P_j S_h + P_h S_j) D_{jh} \quad (5)$$

$$G_{jh} = \frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{jr}|}{n_j n_h (\bar{Y}_j + \bar{Y}_h)} \quad (6)$$

$$G_t = \sum_{j=1}^k \sum_{h=1}^{j-1} G_{jh} (P_j S_h + P_h S_j) (1 - D_{jh}) \quad (7)$$

$$D_{jh} = \frac{d_{jh} - P_{jh}}{d_{jh} + P_{jh}} \quad (8)$$

$$d_{jh} = \int_0^\infty dF_j(y) \int_0^y (y-x) dF_h(x) \quad (9)$$

$$P_{jh} = \int_0^\infty dF_h(y) \int_0^y (y-x) dF_j(x) \quad (10)$$

式(3)—式(10)中, $P_j = \frac{n_j}{n}$, $S_j = \frac{n_j \bar{Y}_j}{n \bar{Y}}$, $j = 1, 2, 3, \dots, k$; D_{jh} 表示 j 与 h 两地区之间数字乡村发展水平的相对影响; $F_j(F_h)$ 表示地区 $j(h)$ 的累积分布函数; d_{jh} 表示 j 地区与 h 地区之间数字乡村发展水平贡献度的差值, 即 j 与 h 区域内所有 $y_{ji} - y_{hr} > 0$ 的样本值加总的加权平均; P_{jh} 表示超变一阶矩, 即 $y_{hr} - y_{ji} < 0$ 的样本值加总的数学期望。

为进一步刻画中国数字乡村发展水平分布的区域差异特征及其来源, 本文在我国省域数字乡村发展水平的基尼系数测算基础上, 将 31 个省份划分为东中西部子群进行分析(见表 2), 为实现区域协同发展提供政策制定的事实依据。①从总体

差异上看, 中国数字乡村发展的总体差异呈现出下降态势, Dagum 基尼系数由 2011 年的 0.067 3 下降至 2022 年的 0.060 1。由于国家财政倾斜政策、产业结构转型升级提升着数字乡村欠发展区域的整体动能, 从而带动了全域数字乡村的均衡发展, 中国区域数字乡村水平发展非均衡现状得到改善, 说明数字乡村建设正趋于高质量协同发展。②研究期内贡献率方面, 区域间差异数值高于区域内差异数值, 说明区域间差异是中国数字乡村发展总体差异的主要原因。因此, 数字乡村试点的开展和提升数字乡村建设整体动能, 急需重视区域间的协调发展问题。

此外, 研究期内数字乡村发展的东西部差距、东中部差距、中西部差距均呈现出下降趋势。随着陆海新通道建设以及乡村振兴水平的提升, 中部地区和西部地区政府试图通过信息与设备投资, 来弥合区域间的数字鸿沟, 致使各区域间数字乡村发展水平差异稳步下降。研究期内东西部差距、东中部差距整体均高于中西部差距。究其原因, 中国乡村幅员辽阔, 地理分布的差异性大, 区域间数字基础设施和技术使用方面存在不均衡发展态势, 基础设施建设、信息化人才、科技技术是造成区域间“数字鸿沟”的主要原因, 已成为制约中西部地区数字乡村建设的主要瓶颈, 数字鸿沟在地区间形成了获取和使用信息资源的机会与能力的的不平等状态。

研究期内东部地区内部差距、中部地区内部差距、西部地区内部差距均呈现出下降趋势。具体来看, 东部地区从 2011 年的 0.049 5 逐年下降到 2022 年的 0.041 1; 中部地区从 2011 年的 0.119 3 上涨到 2014 年的 0.123 8(峰值), 而在 2014 年之后从峰值下降到 2022 年的 0.098 7; 西部地区整体变化不大。此外, 中部地区内部差距最大, 远远高于东部地区和西部地区的内部差距。究其原因, 各省份数字乡村发展水平处于不同阶段, 中部地区处于东部地区与西部地区的衔接地带, 各省份数字乡村发展不平衡, 与长三角地区邻近的地区由于极化效应发展较快, 而与西部省份邻近的地区则多受“塌陷”影响发展较慢。

表 2 中国数字乡村发展的基尼系数及其分解结果

年份	总体	地区内差异			地区间差异			贡献度/%		
		东部	中部	西部	东部—中部	东部—西部	中部—西部	地区间	超变密度	地区内
2011	0.067 3	0.049 5	0.119 3	0.045 1	0.063 6	0.085 5	0.048 8	51.71	14.95	33.33
2012	0.067 2	0.050 1	0.119 3	0.044 1	0.064 4	0.085 1	0.047 8	52.77	13.90	33.33
2013	0.068 7	0.049 6	0.122 3	0.046 8	0.063 8	0.087 6	0.050 6	53.11	13.56	33.33
2014	0.069 4	0.049 4	0.123 8	0.047 8	0.063 4	0.089 1	0.051 5	52.59	14.08	33.33
2015	0.068 1	0.048 1	0.121 7	0.047 3	0.061 6	0.087 8	0.050 9	53.17	13.50	33.33
2016	0.066 8	0.046 5	0.119 0	0.047 6	0.059 4	0.086 1	0.051 3	52.90	13.76	33.33
2017	0.066 2	0.045 3	0.118 0	0.048 2	0.057 8	0.085 6	0.051 8	52.95	13.72	33.33
2018	0.063 5	0.044 3	0.112 7	0.045 7	0.056 5	0.081 5	0.049 3	52.68	13.99	33.33
2019	0.062 9	0.043 4	0.111 2	0.046 0	0.055 4	0.080 4	0.049 6	52.49	14.18	33.33
2020	0.061 3	0.042 5	0.108 2	0.044 9	0.054 0	0.078 5	0.048 5	51.47	15.20	33.33
2021	0.060 8	0.041 7	0.104 4	0.043 1	0.053 4	0.072 4	0.047 7	50.55	16.12	33.33
2022	0.060 1	0.041 1	0.098 7	0.042 6	0.052 9	0.071 8	0.046 1	50.32	16.35	33.33

三、中国数字乡村发展的异质性政策效应研究

(一) 数据说明

数字化科技创新举措赋能乡村建设,是数字乡村发展过程中的核心内容。抓准数字化科技创新在不同时期对数字乡村建设的影响效应,是我国数字乡村建设道路中的关键环节。本文主要是在政策效应的背景下,研究各协变量在不同时间段对数字乡村建设的影响情况,以及各协变量在数字乡村发展中的重要程度。本文主要从科技创新指标、对外开放指标、产业结构指标、城镇化指标、人力资源指标等方面进行探究。自 2017 年国务院提出创新驱动发展的战略方针以来,我国数字化乡村建设飞速发展。因此,本文将 2011—2016 年设定为前期,2017—2022 年设定为后期,由于国家政策不同,从而导致这两个时期之间数字乡村的发展水平存在着较大差异。科技创新水平 (innovation level) 是用科研经费支出来表示,科研经费支出能够直接反映该地区科研创新能力。产业结构水平 (industrial structure) 是用第三产业增加值与国内生产总值的比值来衡量。人力资源水平 (population quality) 用高校本科招生人数与高校总招生人数的比值来表示。城镇化水平 (urbanization level) 是用城镇人口数量与总人口数量的比值来表示。对外开放水平 (open level) 用外商投资总额来表示。

(二) 因果森林估计处理效果

传统的研究方法针对影响因素对数字乡村发展水平的效果评估,会面临内生性的难题以及多重共线性的影响,而因果森林算法则可以很好的

缓解这些问题,并且在传统的研究方法上,因果森林还能够计算出科技创新水平的平均效应 (ATE) 以及其效应的置信区间。本文将科技创新水平作为主要影响因素,将数字乡村水平 (digital village) 设置为目标变量,研究地区 i 科技创新水平的影响效应。处理变量 (treatment variables) 是根据科技创新水平的强弱来判断,假如科技创新水平的值高于所有观测值的平均值,则将 Z_i 的值赋为 1,否则 $Z_i = 0$ 。假设用 $Y_i(1)$ 与 $Y_i(0)$ 分别表示为地区 i 高科技创新水平与低科技创新水平情况下,数字乡村发展水平的情况,则处理效应 (treatment effect) 为两者的差值。然而,对于反事实的是不易观测得到,但通过观察 treatment 组与 control 组的样本可以得到以下结果:

$$Y_i = Y_i(0) + Z_i[Y_i(1) - Y_i(0)] \quad (11)$$

式(11)中, $Y_i(1)$ 表示 treatment 组样本 $Z_i = 1$ 时的结果; $Y_i(0)$ 表示 control 组样本 $Z_i = 0$ 时的结果。估计式子如下:

$$E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0) = \{E(Y_i(1) | Z_i = 1) - E(Y_i(0) | Z_i = 1)\} + \{E(Y_i(0) | Z_i = 1) - E(Y_i(0) | Z_i = 0)\} \quad (12)$$

式(12)为 treatment 组的样本平均处理效果加上选择偏差 (selection bias)。当科技创新水平越高时,其被处理的倾向值的得分也会随之变高,也就说明偏差不为 0,即 $E(Y_i(0) | Z_i = 1) - E(Y_i(0) | Z_i = 0) \neq 0$ 。

因此,针对估计量 $E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)$ 而言,并不能测算出其真实的处理效应 (TE)。使用因果森林的方法来消除这种选择偏差以及因果推断的难题,其通过对比 treatment 组与 control

组的平均差异,得到条件平均处理效应(CATE)的估计:

$$\tau(x) = E[Y_i(1) - Y_i(0) | X_i = x] \quad (13)$$

对于 CATE 的估计效果而言,只有在重叠(overlap)与无混淆(unconfoundedness)的情况下,其估计效果才能更好。CATE 常常通过建立分类与回归树(CARTs)的办法进行估计,但在估计的过程中容易产生偏差估计以及出现过拟合的情况。本文基于地区特征最终分区 L,得到 CATE 的估计值为:

$$\hat{\tau}_0(x) = \frac{1}{\{i: Z_i = 1, x_i \in L\}} \sum_{\{i: Z_i = 1, x_i \in L\}} y_i - \frac{1}{\{i: Z_i = 0, x_i \in L\}} \sum_{\{i: Z_i = 0, x_i \in L\}} y_i \quad (14)$$

式(14)中的处理效应 $\hat{\tau}_0(x)$ 是树叶内处理(treated)和未处理(untreated)观测的平均结构的差值。这里的因果树有 3 个特征:一是与普通的回归树不同,不以预测结果的 MSE 为导向,而是关注处理效应的 MSE;二是因果树能确保置信区间的涉及范围合适化;三是因果树相比回归树而言考虑了参数的渐进性。

本文在研究科技创新水平与数字乡村水平之间的联系时,就理论而言,因果树算法可以防止过度拟合的情况以及出现偏差的情况,但在现实中所用的样本量并没有那么大,因此我们无法避免估计量方差的扩大。因果森林(causal forest)却可以缓解这一情况,因果森林是由众多因果树组合形成的,假如因果森林由 B 棵因果树构成, $\hat{\tau}_b$ 则为第 b 棵树所对应的处理效应的估计值,则因果森林处理后的 CATE 的估计值为:

$$\hat{\tau}_1(x) = B^{-1} \sum_{b=1}^B \hat{\tau}_b(x) \quad (15)$$

式(15)的加总估计量就是 leaf-level CATE,能够有效缓解引入“honesty”而增大方差的现象。

(三)异质性因果效应

选择偏差(selection bias)与溢出效应(spillover effect)在个体间并不能完全排除,因此对于同质性处理效应的估计结果,从个体平均效应的角度出发研究科技创新水平对数字乡村发展的影响更加合理,这过程也叫异质性处理效应(HTE)估计。HTE 指的是城市层面处理效应(city-level treatment effects)的均值,是建立在叶层级的条件平均处理效

应(leaf-level CATE)的基础上。其详细的思路如下。

假设 CATE 不变,则在无混淆(unconfoundedness)条件下,处理效应 τ 的半参数估计结果为:

$$\hat{\tau} = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}^{(-i)}(X_i))(Z_i - \hat{e}^{(-i)}(X_i))}{\frac{1}{n} \sum_{i=1}^n (Z_i - \hat{e}^{(-i)}(X_i))^2} \quad (16)$$

式(16)中, $e(x) = P[Z_i | X_i = x]$ 为倾向值得分, $m(x) = E[Y_i | X_i = x]$ 为科技创新水平期望值。-i 为“袋外(out-of-bag)”预测,即 Y_i 并不用于测算 $\hat{m}^{(-i)}(X_i)$ 。

式(16)被用于产生“R-learner”函数,以便用于 HTE 估计:

$$\hat{\tau}(\cdot) = \arg \min_{\tau} \left\{ \sum_{i=1}^n ((Y_i - \hat{m}^{(-i)}(X_i)) - \tau(X_i)(Z_i - \hat{e}^{(-i)}(X_i)))^2 + \Lambda_n(\tau(\cdot)) \right\} \quad (17)$$

$\Lambda_n(\tau(\cdot))$ 为控制 $\hat{\tau}(\cdot)$ 函数的正则化因子(regularizer)。将式(17)中“R-learner”函数整合到下式因果森林中:

$$\hat{\mu}(x) = \sum_{i=1}^n \alpha_i(x) Y_i, \alpha_i(x) = \frac{1}{B} \sum_{b=1}^B \frac{1(\{X_i \in L_b(x), i \in S_b\})}{|\{i: X_i \in L_b(x), i \in S_b\}|} \quad (18)$$

式(18)中, $\hat{\mu}(x)$ 为 $\mu(x) = E[Y_i | X_i = x]$ 的估计值; $L_b(x)$ 代表因果森林中第 b 棵树包含 x 的因果树; S_b 为与树相关的子样本; B 是因果数目; $\alpha_i(x)$ 为权重,是指由数据驱动形成的核(kernel)。本文对城市特征变量进行了协变量的空间处理,通过因果森林算法计算得出权重 $\alpha_i(x)$,因此科技创新水平的异质性处理效果(HTE)估计可以变为:

$$\hat{\tau} = \frac{\sum_{i=1}^n (Y_i - \hat{m}^{(-i)}(X_i))(Z_i - \hat{e}^{(-i)}(X_i))}{\sum_{i=1}^n \alpha_i(x) (Z_i - \hat{e}^{(-i)}(X_i))^2} \quad (19)$$

本文运用因果森林算法计算得到倾向得分与 HTE 结果。平均处理效应(ATE)是对 HTE 进行平均处理得到,本文也测算了基于子样本的平均处理效应,ATT 代表的是处理组样本的平均处理效应,ATC 为控制组样本的平均处理效应,ATO 为重叠校正的平均处理效应。本文中使用的因果森林

算法是用于分析城市的处理效应,在实证分析中,为减小过拟合带来的偏差情况,将城市比例(honesty fraction)与样本比例(sample fraction)均设为 0.5。另外,本文为使 HTE 的变化都源于数据,而非随树随机生长过程,将随机森林中的树设置为 2 000 棵,本文得到的估计结果如表 3 所示。

由表 3 的 ATE、ATT、ATC、ATO 估计结果可以看到,存在以下现象:一是在前后两个阶段内科技创新水平对数字乡村发展的影响效应存在明显差别,ATE 的估计结果由前期 -1.908 转变为后阶段的 1.936;二是在后阶段(2017—2022 年)中,ATC 的值大于 ATT 的值,这说明在后阶段当中低科技创新水平的地区如果加大创新力度,其数字乡村的发展效果会优于科技创新水平高的地区;三是对于 ATO 的分析,重叠组的影响效应在前期为 -0.311,后期为 0.684,这与 ATO 的估计方向相同,说明科技创新水平对于数字乡村的影响在前期中产生负效应,后期中产生正效应,原因可能是由于科技创新水平在前期中需要投入大量的资金,在某种程度上是不利于数字乡村水平的建设,而在后期中,科技创新带来的产出效果对于数字乡村的建设有着很强的正向促进。

表 3 不同样本的处理效应

Treatment Effects	前阶段(2011—2016 年)	后阶段(2017—2022 年)
ATE	-1.908 +/- 3.622	1.936 +/- 2.741
ATT	-1.144 +/- 4.337	1.53 +/- 2.946
ATC	0.895 +/- 0.534	1.766 +/- 1.09
ATO	-0.311 +/- 1.806	0.684 +/- 0.689

注:表中给出了 4 个组的处理效果,“±”前的数据代表估计值,“±”之后的数据代表在 0.05 显著性水平下的误差。ATE、ATT、ATC、ATO 分别对应“all”“treated”“control”“overlap”4 个样本组的处理效果。

因果森林得到的 CATE 放宽了无混淆性(unconfoundedness)假定,但并不能完全放弃对该假定的考虑,进一步研究预测处理状态(treatment status)的协变量与决定处理效果(treatment effect)的协变量之间的重叠是有意义的。表 4 为各协变量在不同阶段的重要性情况。为了消除各协变量存在的异方差和量纲问题,本文对所有协变量取自然对数,以进一步增加数据的平稳性。由表 4 的两阶段协变量重要性排序可以看出,处理状态(treatment status)重要性的变化主要是集中在中间部分,城镇化(urbanization level)的重要性由前期

的第四位变化为第二位,人力资源(population quality)与对外开放水平(open level)在原有排名上均下降一位。同时,在表 4 中也给出了各阶段处理效果(treatment effect)的重要性排序结果。由结果可以看出,人力资源(population quality)与城镇化(urbanization level)在处理效果的作用在明显下降,这说明人力资源水平与城镇化水平在后期当中对于数字乡村发展水平的作用力在不断降低。对外开放水平(open level)、产业结构(industrial structure)和科技创新(innovation level)的处理效果在明显上升,这表明对外开放水平、产业结构水平以及科技创新水平对数字乡村发展水平的作用力在不断扩大,产生这一变化的原因可能是由于人力资源水平以及城镇化的发展在后期当中对于数字乡村的影响作用的直接力较小,导致其的处理效果降低,而对于对外开放水平、产业结构与科技创新等变量而言,数字乡村的发展在后期更加依赖于数字化技术的内生动力,以至于这些变量的处理效果在不断增加。

表 4 协变量重要程度排序

排序	2011—2016 年(前期)				2017—2022 年(后期)			
	处理状态	处理效应	处理状态	处理效应	处理状态	处理效应	处理状态	处理效应
1	科技创新	0.690 5	人力资源	0.613 8	科技创新	0.687 6	对外开放	0.408 6
2	对外开放	0.087 4	对外开放	0.255 9	城镇化	0.088 9	产业结构	0.330 6
3	人力资源	0.083 9	城镇化	0.088 6	对外开放	0.082 7	人力资源	0.124 4
4	城镇化	0.077 3	产业结构	0.037 6	人力资源	0.077 9	科技创新	0.069 4
5	产业结构	0.060 8	科技创新	0.004 1	产业结构	0.062 7	城镇化	0.066 7

四、扩展性分析

中国数字乡村发展水平呈现区域差异化的发展趋势,区域间数字乡村发展水平呈现出向下转移的“马太效应”和向上转移“辐射效应”特征,需要通过提高政策实施效果的正外部性和削弱自身负向空间外溢性。因此,考察影响因素对地区数字乡村发展水平的作用机制时,需要考虑区域内影响因素的空间溢出效应。本文构建静态空间杜宾模型(SDM)和动态空间杜宾模型(DSDM)进行实证分析:

$$digital_village_{it} = \rho Wdigital_village_{it} + \beta X_{it} + \theta WX_{it} + \mu_i + \nu_t + \varepsilon_{it} \quad (20)$$

$$digital_village_{it} = \alpha digital_village_{it-1} + \rho Wdigital_village_{it} + \beta X_{it} + \theta WX_{it} + \mu_i + \nu_t + \varepsilon_{it} \quad (21)$$

其中, i 表示地区, t 表示时间, μ_i 与 ν_t 表示地区与时间的扰动项, ε_{it} 表示随机扰动项。 ρ 表示空间滞后系数, 主要是反映邻近地区数字乡村发展水平对本地区数字乡村发展水平的影响程度。 W 表示空间权重矩阵, 本文采用近邻权重矩阵和地理权重矩阵进行分析, 其中近邻权重矩阵 W_{0-1} 地区相邻为 1, 否则为 0; 地理权重矩阵 $W_d = 1/d_{ab}^2$, $a \neq b$, 否则为 0。 X_{it} 表示控制变量, 包括科技创新水平、产业结构升级、人力资源水平、城镇化水平、对外开放水平。

本文采用 ADF 检验和 PP 检验对各变量进行平稳性检验, 由结果可知, 各变量 P 值均小于 0.05, 说明原序列中不存在单位根, 可以进行回归分析。从调整后 R^2 来看, 各空间权重矩阵下中国数字乡村发展水平模型的拟合优度较好, 这证实了中国数字乡村发展在各省域间确实存在空间溢出效应。Hausman 检验的 P 值小于 0.1, 文章选择固定效应模型的估计结果进行实证分析; 通过模型的误差项建立 LMerror 和 LMlag 统计量的显著性检验 P 值均小于 0.1, 故本文构建 SDM 模型进行研究分析。数字乡村发展水平的空间滞后系数 (ρ) 显著为正, 表明中国数字乡村发展水平和相邻省域间数字乡村发展水平确实存在显著的空间正向依赖性与空间外溢效应, 说明数字乡村发展水平较高地区会明显带动相邻省域数字乡村发展。数字乡村发展水平滞后一期 ($l.digital_village$) 的估计系数均在 1% 的显著性水平下通过检验, 且估计系数大于零, 说明各省域数字乡村发展水平具有较强的时间积累特征。“十四五”期间数字乡村建设作为实现“乡村振兴”的必然要求, 带动数字乡村建设的各项因素叠加推进, 更有利于激发当期数字乡村发展水平的活力, 使其呈现出明显的“棘轮效应”特征。科技创新水平 ($innovation_level$) 的估计系数显著为正。依托“宽带中国”战略和区块链技术在乡村的创新应用, 推动着乡村 5G 基站建设、智慧物流等数字化发展, 为农业技术进步、农村经济增长与创新驱动发展提供强劲推力, 促进传统农村产业与数字信息产业融合渗透, 强化数字乡村建设的内生动力, 这些“拉力”共同作用下促进着中国数字公共资源均衡配置和数字乡村发展水平的提升。邻近地

区科技创新水平 ($w \times innovation_level$) 的影响效应并不显著。这是由于我国地区间农业科技资源差异明显, 科技资金与技术未能发挥显著的“虹吸效应”和“邻居效应”, 从而造成各地区间未能呈现出数字乡村发展的集群特征。产业结构升级 ($industrial_structure$) 的估计系数显著为正。产业结构升级推动农村产业结构由单一化转向多元化, 释放农业农村现代化建设新动能, 能够有效增强农村数字产业对生产要素的吸纳能力, 降低农村数字产业发展的门槛, 进而持续推进数字乡村发展。邻近地区产业结构升级 ($w \times industrial_structure$) 的影响效应并不显著, 说明邻近产业结构升级对本地区数字乡村发展水平未能产生显著的影响。中国产业结构存在区域差异化特征, 行业壁垒致使形成自上而下的封闭性垂直传导路径, 横向部门间的数据壁垒未能有效发挥空间集聚效应与产业辐射效应(见表 5)。

表 5 空间杜宾模型估计结果

变量	模型			
	SDM 模型		DSDM 模型	
	模型 1	模型 2	模型 3	模型 4
$l.digital_village$	—	—	0.748 *** (0.048)	0.807 *** (0.049)
$innovation_level$	0.114 *** (0.015)	0.128 *** (0.017)	0.053 ** (0.014)	0.043 *** (0.017)
$industrial_structure$	0.030 ** (0.012)	0.025 ** (0.012)	0.006 * (0.003)	0.003 * (0.011)
$population_quality$	0.024 (0.022)	0.006 (0.023)	0.029 (0.019)	0.046 (0.020)
$urbanization_level$	-0.007 *** (0.002)	-0.003 * (0.002)	-0.004 ** (0.001)	-0.003 * (0.001)
$open_level$	0.004 *** (0.001)	0.002 ** (0.001)	0.003 *** (0.001)	0.002 * (0.001)
$W \times innovation_level$	0.039 (0.137)	0.007 (0.003)	0.001 (0.025)	0.051 (0.032)
$W \times industrial_structure$	0.019 (0.022)	0.042 (0.025)	0.003 (0.019)	0.019 (0.023)
$W \times population_quality$	-0.052 (0.041)	-0.002 (0.046)	-0.082 (0.035)	-0.006 (0.039)
$W \times urbanization_level$	0.009 ** (0.004)	0.002 * (0.005)	0.008 ** (0.004)	0.006 * (0.005)
$W \times open_level$	-0.005 ** (0.002)	-0.003 ** (0.003)	-0.004 ** (0.001)	-0.003 * (0.003)
ρ	0.348 *** (0.067)	0.398 *** (0.087)	0.207 *** (0.066)	0.179 ** (0.092)
R^2	0.965	0.964	0.973	0.972
Log_L	1 240.884	1 233.065	1 240.884	1 233.065

注: *、**、*** 分别表示在 $p < 0.10$ 、 $p < 0.05$ 、 $p < 0.01$ 时有统计学意义; 括号中的数值是回归系数的标准误。模型 1 与模型 3 采用邻接权重矩阵; 模型 2 和模型 4 采用地理权重矩阵。

五、结论与启示

随着产业数字化深入推进,以数据为关键要素的数字经济正在稳步发展,数据要素乘数效应也逐渐显现出来。数字技术与乡村发展融合发展的数字乡村建设,能够有效统筹数据资源整合共享和开发利用,共同激活数据要素价值,培育新质生产力,助力经济社会高质量发展。本文运用因果推断方法研究发现,科技创新水平对数字乡村建设的影响,由前期的负向作用($ATE = -1.908$)转为后期的正向作用($ATE = 1.936$);后期中低科技创新水平地区的 $ATC > ATT$,表明中低科技创新水平地区创新效果会优于科技创新水平高的地区。通过构建空间杜宾模型研究发现:数字乡村发展具有较强的时间积累性,呈现出明显的“棘轮效应”特征;数字乡村发展水平受到经济要素流动等影响因而形成明显的空间依赖性空间外溢效应;本地科技创新水平、产业结构升级将对本地数字乡村发展产生正面影响,邻近地区的科技创新水平、产业结构升级对本地数字乡村发展并未产生实际影响。基于上述研究结论,本文建议需要充分考虑中国数字乡村发展的内生动力与空间影响因素,提出以下有针对性的政策与建议。

第一,突破“中心—边缘”格局和地域空间限制,分层次、分区域统筹协调数字乡村发展。经济基础、地理位置、人力资源等条件决定着地区数字乡村发展模式,全力打通“三农”领域的数字壁垒,积极实施“雪亮工程”向乡村延伸,循序渐进推进农业农村现代化建设。而数字乡村建设避免实施“一刀切”政策。由于中国幅员辽阔和乡村类型多样性,资源禀赋和区位条件存在不均衡现象,呈现出多中心垂直扩散式发展形态,导致数字乡村发展水平存在区域差异。现阶段农村数据碎片化、数据壁垒等信息不对称问题制约数据开放共享水平,限制了数字乡村勾连互通、协同作出科学决策。针对数字乡村发展水平较高的地区应该加强其辐射作用,打破资源要素双向流动壁垒,积极引领数字乡村水平较低地区的发展,重点针对数字

乡村发展水平较为落后的地区进行政策帮扶,促进资源以及技术等生产要素流动性,进而产生即时性的同步共振效应,推动数字鸿沟向数字红利转变,形成全面、多领域的区域化数字乡村的集聚圈层,进而带动各地区数字乡村的发展,从而实现我国数字乡村发展水平的均衡化发展。

第二,发挥科技创新技术的吸虹效应,加强数字化农业技术创新供给能力。数字化技术是数字乡村发展的内生动力,技术创新是全面推进数字乡村发展的重要驱动力。加强地区之间资源的整合以及数字化技术推广,是提高乡村数字化水平、实现农业农村现代化建设的关键因素。强化乡村数字基建的技术支撑效用,统筹推进硬件基础设施与软件技术创新协同发展。硬件基础设施建设方面,东部沿海地区应借助自身经济发展优势着力改善传统硬件基础设施,中西部区域应持续加强农村物联网、大数据、千兆光纤以及 5G 基站等数字信息硬件基础设施建设,且积极响应“产学研用”合作政策,充分发挥科研机构在数字乡村建设中技术创新能力,积极推进地方高校做好农业创新型科技人才培养工作与乡村数字化教育平台建设,鼓励相关企业建立农业科技创新成果孵化与培育基地,提升乡村产业现代化、数字化、智能化水平。软件技术创新方面,以“数字技术”赋能农村产业结构优化升级,推动产业深度融合、加强数据流通与共享,探索新型农村“互联网+”业务模式,实现农村产业由“高速度”转向“高质量”数字化发展,加快乡村数字化建设进程。

第三,稳步推进农村产业结构升级,优化数字乡村发展资源要素。产业兴旺是持续推动数字乡村建设的基础,要着力推进数字经济与农村产业交叉融合、协同发展,牢固抓住农业农村现代化供给侧改革,依托区域优势强化自身特色产业链,构建数字一体化农村产业体系,规划好产业布局。同时,数字乡村建设关键在于资源要素整合与配置,增强人才、技术和资金等发展要素地自由流动,实现数字乡村资源合理配置,缓解各区域间数字乡村发展水平差距,为中国数字乡村建设营造

因地制宜的发展环境;优化农村产业人员结构,培养农民数字素养,极大地调动广大乡村居民参与乡村数字化建设的积极性,激发乡村数字化建设新动能,进而持续推进数字乡村发展,最终实现“乡村振兴”战略目标。

第四,针对不同阶段各变量对数字乡村发展的影响,需要及时调整不同时期数字乡村建设的侧重点。从随机因果推断中,科技创新水平对数字乡村建设在不同阶段的估计效应来看,本文认为在后期当中应该持续加大对低科技创新水平地区的科技创新生产要素的投入,提高地区科技创新的内生能力,为数字乡村的建设注入内在动力。针对各协变量在不同阶段对数字乡村建设的影响效应不同,应该在后期不断加强对外开放水平的深化建设以及产业结构的优化升级,对外开放水平的深入建设能有效推动数字化建设发展的多元化,有利于建设多层面协同、多平台支撑、多主体参与的数字化领域的交流体系。

参考文献:

- [1] 赵佳佳,魏娟,刘天军. 数字乡村发展对农民创业的影响及机制研究[J]. 中国农村经济,2023(5):61-80.
- [2] 薄文广,钱镱,屈建成,等. 新型城镇化与乡村振兴耦合协调及交互影响研究:基于156个地级市面板数据的实证分析[J]. 中国软科学,2023(9):106-116.
- [3] 杨柏,陈银忠,李海燕. 数字化转型下创新生态系统演进的驱动机制[J]. 科研管理,2023,44(5):62-69.
- [4] 焦豪,张睿,杨季枫. 数字经济情境下企业战略选择与数字平台生态系统构建[J]. 管理世界,2023,39(12):201-229.
- [5] EARL J, KIMPORT K. Digitally enabled social change: activism in the internet age[J]. Perspectives on politics, 2012, 10(2): 514-515.
- [6] 崔凯,冯献. 我国农业农村信息化的阶段性特征与趋势研判[J]. 改革,2020(6):125-135.
- [7] 王胜,余娜,付锐. 数字乡村建设:作用机理、现实挑战与实施策略[J]. 改革,2021(4):45-59.
- [8] 沈费伟,崔钰. 从“政府主导”到“村庄经营”:实现数字乡村高质量发展的策略选择[J]. 电子政务,2023(12):1-15.
- [9] 曾亿武,宋逸香,林夏珍,等. 中国数字乡村建设若干问题刍议[J]. 中国农村经济,2021(4):21-35.
- [10] 黄小勇,邹伟,李世成,等. 数字化产村融合激活农村闲置资源的演化博弈行为研究[J]. 中国软科学,2023(7):151-167.
- [11] 方堃,李帆,金铭. 基于整体性治理的数字乡村公共服务体系研究[J]. 电子政务,2019(11):72-81.
- [12] FU S, ZHUO H, SONG H, et al. Examination of a coupling coordination relationship between urbanization and the eco-environment: a case study in Qingdao, China[J]. Environmental science and pollution research, 2020, 27(19): 23981-23993.
- [13] 殷浩栋,霍鹏,汪三贵. 农业农村数字化转型:现实表征、影响机理与推进策略[J]. 改革,2020(12):48-56.
- [14] 韩清. 数字经济联动区域一体化发展的价值目标与规制路径[J]. 阅江学刊,2023(6):96-109.
- [15] 夏显力,陈哲,张慧利,等. 农业高质量发展:数字赋能与实现路径[J]. 中国农村经济,2019(12):2-15.
- [16] 汪亚楠,徐枫,叶欣. 数字乡村建设能推动农村消费升级吗?[J]. 管理评论,2021,33(11):135-144.
- [17] LI X J, et al. Analysis on regional differences and spatial convergence of digital village development level: theory and evidence from China[J]. Agriculture, 2022(12): 164-184.
- [18] 许敬辉,王乃琦,郭富林. 数字乡村发展水平评价指标体系构建与实证[J]. 统计与决策,2023,39(2):73-77.
- [19] 魏后凯,刘长全. 中国农村改革的基本脉络、经验与展望[J]. 中国农村经济,2019(2):2-18.
- [20] 李三希,王泰茗,刘小鲁. 数据投资、数据共享与数据产权分配[J]. 经济研究,2023,58(7):139-155.
- [21] 沈费伟,陈晓玲. 保持乡村性:实现数字乡村治理特色的理论阐述[J]. 电子政务,2021(3):39-48.
- [22] 丁建军,万航. 中国数字乡村发展的空间特征及其农户增收效应[J]. 自然资源学报,2023,38(8):2041-2058.
- [23] 石明明,张小军. 中国式现代化视域中的经济改革与制度建构[J]. 中国社会科学,2023(9):4-23,204.
- [24] 马香品,王星蕾. 数字乡村高质量发展的理论路径与影响因素[J]. 财经科学,2023(2):74-84.
- [25] 朱红根,陈晖. 中国数字乡村发展的水平测度、时空演变及推进路径[J]. 农业经济问题,2022(8):1-14.
- [26] 姬志恒. 中国农业农村高质量发展的空间差异及驱动机制[J]. 数量经济技术经济研究,2021,38(12):25-44.

(本文责编:辛 城)