

【区域经济与产业发展】

# 数据要素流动视角下省域数字产业集聚的 差异化路径研究<sup>\*</sup>

黄晓婷 孙久文 韩雅清

**摘要:**在“数字中国”战略与数据要素市场化进程加速的背景下,数字产业集聚已成为推动区域经济高质量发展的重要路径。基于数据要素流动的全过程视角,整合基础支撑、价值转化与环境适配三大维度的理论框架,运用必要性分析(NCA)与动态定性比较分析(QCA)方法,对2014—2023年中国30个省份(不含西藏及港澳台地区)的面板数据进行组态路径识别与演化分析。研究发现,数字产业集聚不存在单一必要条件,而是多重要素协同联动的结果,不同条件在不同发展阶段具有差异化约束强度;存在三条驱动高数字产业集聚的差异化路径,体现了“多重并发”与“殊途同归”的组态特征;路径演化呈现动态性,市场—技术驱动型稳健性最强,生态引领型存在边际效用递减趋势,发达省份呈现出从要素驱动向生态驱动的转换趋势。

**关键词:**数据要素流动;数字产业集聚;差异化路径;组态分析;省域差异

中图分类号:F127;F49 文献标识码:A 文章编号:2095-5766(2025)06-0132-12 收稿日期:2025-10-14

**\*基金项目:**教育部人文社科青年基金项目“数字化转型中农业产业链韧性测度及其提升路径研究”(23YJC790040);福州市社会科学规划项目“福州都市圈数字经济与实体经济融合水平测度及协同路径优化研究”(2025FZY243);福建江夏学院校级科研人才培育项目“数字经济对城市经济韧性的影响效应和作用机制研究”(JXS2023006)。

**作者简介:**黄晓婷,女,福建江夏学院金融学院讲师(福州 350108)。

孙久文,男,中国人民大学应用经济学院教授、博士生导师,通信作者(北京 100872)。

韩雅清,女,福建江夏学院金融学院副教授(福州 350108)。

数据作为关键生产要素,正深刻重塑着区域经济发展格局与产业空间布局。在“数字中国”战略和“数据二十条”政策的推动下,数据要素市场化进程不断加快,为实体经济创新注入持续动能。作为数字经济的高级空间组织形态,数字产业集聚凭借其高创新性、强渗透性和广覆盖性特征,正成为培育新质生产力、构建现代化产业体系的战略支点。当前,我国数字产业集聚呈现多元化发展态势。截至2023年,全国已形成200余个数字产业集群,构建起以京津冀、长三角、粤港澳三大核心区为引领,成渝、贵州、内蒙古等中西部

集聚区协同发展的空间格局(马文娟,2024)。值得关注的是,不同省份在资源禀赋、发展阶段和制度环境存在显著差异的背景下,都能够实现高水平的数字产业集聚。这引发了一个深层次的理论思考:在数据要素驱动的背景下,各省份究竟是通过何种差异化路径实现数字产业集聚的?这些成功路径背后是否存在某些共性的驱动逻辑,还是完全依赖各异的要素组合?现有研究尚未系统揭示这种差异化发展的内在规律,而这对于理解数据要素如何重塑区域产业格局、推动数字经济高质量发展具有重要理论价值。

## 一、文献综述

作为数字经济的核心载体与重要表现形式,数字产业的蓬勃发展已引发学术界广泛关注。通过对现有文献的系统梳理,可将与本文相关的研究归纳为以下三个主要方面。

关于数字产业集聚的理论内涵,学界已形成基本共识。数字产业的概念主要由数字经济演化而来,通常在中国学界和业界使用,其内涵经历了从狭义的信息通信技术产业向广义的数据要素、数字技术驱动的经济活动的扩展(卢福财等,2024)。现有研究指出,数字产业集聚是在数字技术发展背景下,以数据要素流动与整合为纽带,推动数字企业在地理空间和数字网络平台高度集中的现象(赵放等,2025)。从本质上讲,它属于技术集群的范畴,是高端生产要素在空间上的集中,推动产业链上下游实现协同发展(Kerr et al., 2020; 辛璐璐, 2023)。相较于传统产业集聚,数字产业集聚呈现出虚实双重集聚、知识溢出范围扩展及协作方式网络化三个典型特征。

关于数字产业集聚的测度研究,现有文献主要采用两类方法。区位熵作为被广泛应用的测度方法,通常选取能够体现数字产业特征的关键指标进行计算,如数字产业就业人数、产业增加值规模(辛璐璐, 2023; 李栋等, 2023)。基于此方法的研究表明,中国主要城市群数字产业集聚水平呈现不均衡的发展态势,整体上从多极分化逐步演变为“高值降低、低值追赶”的收敛趋势,各城市群间的空间差距正在缓慢缩小(赵放等, 2024)。在微观企业层面,学者们尝试利用更细化的数据测度数字产业集聚。

关于影响因素的研究,学者们主要从外部环境与内部动能两个维度展开探讨。在外部环境层面,区域发展基础与制度政策构成关键支撑。研究表明,良好的经济基础(叶堂林等, 2023)、高效的数据要素市场(赵放等, 2024)、合理的产业结构与充足的科技人力资本(赵放等, 2024)为数字产业集聚提供重要推动力。政府通过实施工业4.0政策(Rudy et al., 2023)、培育信息产业新业态(林挺, 2023)以及推进“宽带中国”等战略(余壮雄等, 2024),能够有效引导数字生产要素的区域集聚,促进数字产业

集聚高地的形成。刘淑春(2019)进一步指出,政府可通过靶向建设高质量数字产业聚集区,系统提升数字经济集聚能级。在内部动能层面,企业行为与创新活动发挥着核心作用。企业的数字创新(夏杰长等, 2025)、平台共建(Cenamor et al., 2019)和资源共享(胡乾韬等, 2024)等战略行为,通过促进知识溢出、优化资源配置和强化网络协同,从微观层面推动产业集群的形成与演化。

既有研究为本文奠定了良好基础,但仍存明显局限:一方面,多数文献侧重探讨单一因素的“净效应”,忽视了数据要素多维度、协同联动的本质特征;另一方面,传统线性模型难以揭示多重并发因果关系及背后的组态路径,尤其缺乏对各省份驱动模式动态演化的审视。这些不足限制了学界对省域数字产业发展绩效差异的深层理解。本文从数据要素流动全过程出发,构建一个整合基础支撑、价值转化与环境适配三大维度的理论框架,并创新性地采用动态定性比较分析(QCA)方法,系统识别导致高数字产业集聚的多元驱动组态及其历时演变规律。本研究旨在从组态比较的视角揭示省域数字产业集聚的复杂驱动机制,为理解数据要素重塑区域产业格局提供新学理依据,也为各省因地制宜推进数字经济发展提供实证支撑。

## 二、理论分析框架

本研究认为,数据要素从原始状态到最终驱动产业集聚,是一个依次经历基础支撑、价值转化和环境适配三个阶段的连续动态过程。这一过程体现了数据要素价值实现的完整生命周期,其中每个阶段都具有独特的特征和功能,同时又与其他阶段保持密切关系。这三个维度相互关联、协同作用,共同构成数字产业空间格局形成的完整因果链。各省份因其在三个维度上的资源禀赋与制度条件不同,形成异质性的比较优势,进而演化出多条差异化的数字产业集聚路径。这种路径分异不仅反映了各地区资源条件的差异,更体现了数字经济发展中的多元均衡特征。

在基础支撑维度,数据基础设施与数字经营成本共同构成了影响数字产业空间布局的基础性条件。数据基础设施作为数据要素流通的物理载体,其建设水平直接决定了数据流动的效率与质量,为

数字企业的地理集中提供必要的物质保障(钞小静等,2024)。高速网络、算力中心等设施的完善不仅能够降低数据传输成本,还能吸引依赖高质量数据服务的企业集聚,从而形成数字产业发展的区位优势。数字经营成本则从经济维度制约数字产业的空间分布。其中,高素质数字人才的薪酬水平、商业用电价格等关键成本因素直接影响企业的区位决策与运营效率。在人才竞争日益激烈的背景下,合理的成本结构能够增强区域对数字企业的吸引力,而过高的经营成本则可能削弱区域的竞争优势。该维度是数据价值链运作的先决条件,缺乏良好的基础支撑,数据流动与价值转化将难以实现。

在价值转化维度,数字技术人才、数字市场需求和数字技术创新构成了将原始数据转化为经济价值和产业活动的核心能动系统。相较于传统产业集聚对普通劳动力规模的依赖,数字产业集聚表现出对高素质人力资本的强烈偏好(丛屹等,2022)。数字技术人才作为价值创造的能动主体,不仅是高端人力资本的核心载体,更是将数据要素转化为技术解决方案与商业价值的智力资本,其规模与质量直接决定数据要素的价值转化效率。数字市场需求反映区域市场对数字化产品和服务的吸收能力,形成拉动企业集聚的重要市场力量,为数据要素价值实现提供市场空间和应用场景。数字技术创新是确保价值转化持续性与塑造产业空间格局的核心。它不仅体现为将数据转化为新技术、新业态的长期机制,通过专利产出与知识积累构建技术壁垒与持续竞争优势,并引导产业空间布局(李彦臻等,2020);在更深层次上,作为数据爆炸与技术进步的直接产物,数字产业集聚的核心正依赖这种持续迭代的技术支撑(李研,2021)。此维度是驱动产业集聚的内在引擎,直接决定数据要素的价值实现效率,三个因素相互促进,共同构成价值创造的完整闭环。

在环境适配维度,数字政策支持与数字生态共同为数据要素流动提供了不可或缺的制度保障与市场环境,二者相辅相成,构成数字产业可持续发展的两大支柱。数字政策支持侧重政府“有形之手”的引导作用。省级政府通过数字战略顶层设计、数据确权与交易规则构建以及产业发展规划等机制,有效降低数据要素市场制度的不确定性,有效引导高质量资本与技术要素向数字产业领域集

中,从而激励数字企业自发形成产业集群、优化区域数字产业布局,推动数字产业在空间上的协同发展与能级提升(毛丰付等,2022)。数字生态指支撑数字产业发展的金融市场环境与生态系统成熟度,侧重市场“无形之手”的滋养作用,其核心体现为数字金融的发展水平。一个覆盖广、深度高、数字化程度强的金融生态能够为数据要素的价值转化提供灵活的资本支持、便捷的支付结算和丰富的风险管理工具。风险投资、产业基金等多层次资本市场体系为数字企业,特别是中小企业和高科技企业的孵化与成长注入持续动力,是数字产业生态健康与活跃的晴雨表。

本框架强调,数字产业集聚是三个维度多重条件彼此联动匹配与协同互动的结果(见图1)。不同省份可依据其独特的初始禀赋,在上述价值链的不同环节形成比较优势,进而通过多样化的条件组态产生多重并发等效路径。

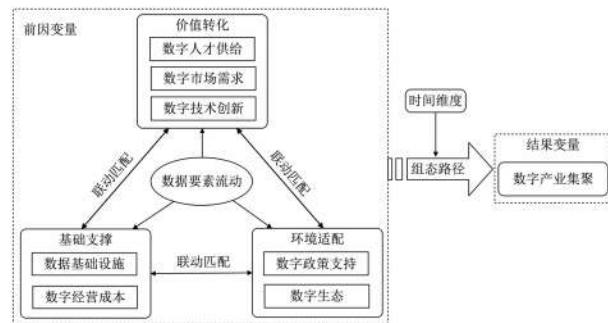


图1 理论分析框架

资料来源:作者绘制。

### 三、研究设计

QCA是一种基于集合论与布尔代数的案例导向型研究方法,它通过识别不同前因条件的组合,即“组态”,来阐释结果的产生(Ragin, 2008)。与传统回归分析关注变量的独立“净效应”不同,QCA的核心优势在于处理因果复杂性,包括多重并发因果关系、因果非对称性以及组态等效性(杜运周等,2017)。这些特征与本研究的核心目标高度契合。然而,传统QCA对时间维度的考察不足,难以揭示驱动路径的动态演变规律(杜运周等,2021)。为此,本文引入动态QCA方法,通过分析跨时期面板数据,考察条件组态的稳定性与演化性,从而弥补静态截面分析的缺陷,更深入地揭示省域数字产业发展战略的调整与转换过程。另外,必要性条件分

析(NCA)是近年来兴起的一种互补性方法,它专门用于识别和检验产生特定结果的必要条件(Dul et al., 2020)。相较于QCA内部的必要性检验,NCA不仅能够定性判断一个条件是否必要,还能通过上限回归和上限包络分析技术,定量衡量该条件的必要程度。更重要的是,NCA独特的瓶颈水平分析能够揭示为实现不同水平的结果,各个前因条件所需达到的最低水平,这为理解不同发展阶段的关键约束提供了动态和精细化的视角(杜运周等,2022)。因此,本文在方法论上融合了NCA和动态QCA,以兼顾对必要条件与充分组态的识别。具体而言,先运用NCA判别构成高数字产业集聚的单个必要条件及其瓶颈水平;继而,采用基于时间序列的汇总型动态QCA,对2014—2023年中国30个省份的面板数据进行分析,以揭示驱动高数字产业集聚的多元等效路径及其核心条件组态。

### (一)变量设定

#### 1.结果变量

本文的结果变量为数字产业集聚(*agglom*),其测度综合依据《数字经济及其核心产业统计分类(2021)》与《国民经济行业分类》,将数字产业界定为信息传输、软件与信息技术服务业以及计算机、通信和其他电子设备制造业两大核心门类。在此基础上,借鉴范剑勇等(2021)与袁歌骋等(2023)的研究方法,采用区位熵来量化省域层面的产业集聚水平。具体计算公式如下:

$$agglom_{it} = \frac{emp\_dig_{it}/emp\_total_{it}}{\sum emp\_dig_{it}/\sum emp\_total_{it}} \quad (1)$$

式(1)中,*emp\_dig<sub>it</sub>*为*i*省份*t*年份数字产业从业人员数,*emp\_total<sub>it</sub>*为*i*省份*t*年份就业总人数,以年末单位从业人员数表征。各省数字产业就业人数采用信息传输、软件和信息技术服务业就业人数以及计算机、通信和其他电子设备制造业就业人数之和构建。该指标有效消除了地区规模差异,其数值越高,代表数字产业在空间上的集聚态势越显著。

#### 2.前因变量

本文的前因条件选取严格遵循上述理论分析框架。在基础支撑维度,数据基础设施(*infra*)的测度致力于全面反映其多维度特征,构建了一个综合指数体系,该体系由传统网络设施与新型数字设施两个子维度构成(王欣亮等,2023)。传统网络设施通过省际互联网宽带接入端口数、光缆线路长度及

移动电话普及率等指标合成,以表征基础通信能力;新型数字设施则通过5G基站密度、数据中心机架规模等指标合成,以表征先进算力与连接能力。各底层指标经标准化处理后,采用熵值法分别确定两个子维度内各指标的权重并计算最终的综合指数。数字运营成本(*cost*)的测度聚焦数字产业最核心的投入要素,采用“信息传输、软件和信息技术服务业”的城镇单位就业人员平均工资作为代理变量,该指标能够有效反映数字企业所面临的关键人力成本约束。

在价值转化维度,数字人才供给(*talent*)采用“信息传输、软件和信息技术服务业”的城镇单位从业人员数进行衡量,该数据能够准确表征一个地区为数字产业发展提供核心智力资本与专业劳动力的规模。数字市场需求(*demand*)的测度选用人均软件业务收入这一指标,软件作为数字产品与服务的核心价值载体,其业务收入直接反映了市场端的实际支付能力,采用人均形式旨在消除人口规模差异,从而更纯粹地衡量市场的需求强度与活跃度。数字技术创新(*innov*)的量化则通过识别与数字技术相关的发明专利并加总计数来实现,具体参考陶锋等(2023)的研究所建立的IPC专利分类号与数字技术领域的映射关系,在“省份-年份”层面进行专利计数,并优先采用发明专利,以体现实质性的技术创新水平。

在环境适配维度,数字政策支持(*policy*)的测度通过内容分析法对省级政府工作报告进行量化分析,该方法依据《“十四五”数字经济发展规划》等权威文件构建包含“数字经济”“数据要素”“人工智能”“工业互联网”等在内的核心关键词库,利用Python的Jieba库进行分词与词频统计,最终以总词频数来量化地方政府对数字经济的战略重视与政策推力。数字生态(*eco*)的测度采用北京大学数字金融研究中心发布的省级数字普惠金融指数,该指数从覆盖广度、使用深度和数字化程度三个维度综合评估了地区数字金融的发展水平,因其权威性、连续性与系统性能够很好地表征支撑数字企业创新与成长的金融生态环境成熟度。

### (二)数据来源

基于数据可得性与案例可比性,本研究将观测样本确定为中国30个省份,研究时段设定为2014—2023年。2014年是中央网络安全和信息化

领导小组成立之年,标志着数据资源正式成为国家战略核心资源,为数据要素的治理与数字产业的发展奠定了制度基础;终点2023年则完整覆盖了后疫情时代数字经济的深化发展阶段,使研究能够系统考察从“大数据元年”后续效应到数据要素化探索深化的完整演进周期。核心变量的原始数据主要取自历年《中国统计年鉴》及各省份统计年鉴。数字技术创新的相关专利数据来源于中国研究数据服务平台,数字政策支持变量通过分析各省份政府工作报告获取,数字生态的测度采用北京大学数字金融研究中心发布的“数字普惠金融指数”。为确保数据质量与可比性,本研究对原始数据进行了统一处理:针对个别缺失数据,采用线性插值法予以补全;所有涉及价格的指标均以2014年为基期,利

用各省份的消费者价格指数(CPI)进行平减处理,以消除价格因素的影响。

#### 四、结果与分析

在QCA方法中,为样本案例赋予集合隶属的过程被称为变量校准。根据本次变量的数值特点,采用直接校准法,以95%、50%及5%分位数值作为完全隶属、交叉点以及完全不隶属的定性锚点(张放,2023)。确定校准锚点后,对前因变量和结果变量进行校准,将原始数据转化为可用于分析的模糊集隶属分数。此外,本文将0.5隶属度均加上0.001常数,从而避免其组态归属问题(田庆锋等,2025)。各变量具体校准锚点及描述性统计特征见表1。

表1 模糊集校准与变量描述性统计

变量类型	变量名称	模糊集校准			描述性分析			
		完全隶属	交叉点	完全不隶属	平均值	标准差	最小值	最大值
结果变量	<i>agglom</i>	1.9126	0.5688	0.2275	1.3279	0.4386	0.7275	2.0775
前因变量	<i>infra</i>	0.5048	0.1558	0.0341	0.2889	0.2485	0.0772	0.7239
	<i>cost</i>	20.7391	9.7291	6.1091	17.1394	5.8569	8.4189	30.7072
	<i>talent</i>	60.7995	7.5299	0.9439	33.7867	34.0444	3.8477	101.1646
	<i>demand</i>	1.9592	0.1296	0.0067	2.3174	2.4168	0.2320	10.5314
	<i>innov</i>	64.2010	5.4540	0.3120	27.9204	28.6425	3.0810	106.6570
	<i>policy</i>	0.0069	0.0036	0.0015	0.0043	0.0019	0.0015	0.0087
	<i>eco</i>	420.8594	305.7874	175.5965	342.0050	82.3712	184.7100	473.8300

数据来源:作者计算。

#### (一)单个条件的必要性分析

##### 1.NCA分析

通过R软件运用NCA法对单个前因条件是否为结果变量的必要条件展开分析。NCA方法的判断标准是效应量( $d$ )须不低于0.1的经验阈值,且具有统计显著性( $p<0.05$ )(杜运周等,2022;Dul et al.,2020)。鉴于本研究涉及的前因变量均为连续型变量,主要采用上限回归分析技术(CR)进行估计,并辅以上限包络分析(CE)作为稳健性检验。表2报告了NCA的必要条件分析结果。分析结果表明,所有前因条件均未同时满足效应量  $d \geq 0.1$  且  $p < 0.05$  的判定标准。具体而言,数据基础设施、数字人才供给、数字市场需求与数字技术创新的效应量虽大于0.1,但未通过显著性检验;数字政策支持的效应量虽显著,但未达到0.1的阈值;数字经营成本与数字生态则既未达到效应量阈值也不显著。上限包络分析的结果与上限回归分析结论基本一致,进一

步验证了上述发现的稳健性。因此,本研究认为,在单一条件视角下,未发现任何变量是产生高水平省域数字产业集聚的必要条件。

表3呈现了基于NCA方法的瓶颈水平分析结果。瓶颈水平指为达到特定程度的结果,必要条件所需达到的最低水平。分析结果表明,各前因条件对数字产业集聚的约束作用呈现出明显的分层特征与阶段性规律。具体而言,数据基础设施(*infra*)与数字人才供给(*talent*)作为基础性门槛条件,在产业集聚水平达到30%时即开始产生约束效应,其初始瓶颈阈值分别为2.1%和0.7%。随着目标结果水平的提升,数字市场需求(*demand*)和数字技术创新(*innov*)在中等发展阶段(50%水平)逐步显现为关键驱动条件,阈值要求分别升至10.7%和13.0%。而数字政策支持(*policy*)与数字生态(*eco*)则属于高阶保障性条件,直至产业集聚水平达到60%以上才产生显著约束作用。尤为值得注意的是,数字经营

表2 NCA方法必要条件分析结果

变量	方法	效应量d	p值	精确度/%
infra	CR	0.309	0.067	100
	CE	0.176	0.183	100
cost	CR	0.006	0.912	96.70
	CE	0.004	0.849	100
talent	CR	0.286	0.131	88.30
	CE	0.184	0.194	100
demand	CR	0.225	0.126	86.70
	CE	0.198	0.095	100
innov	CR	0.233	0.091	92.30
	CE	0.194	0.105	100
policy	CR	0.084	0.007	93.70
	CE	0.017	0.008	100
eco	CR	0.046	0.414	96.70
	CE	0.030	0.008	100

数据来源:作者计算。

注:①采用校准后的模糊集隶属度数据。② $0.0 \leq d < 0.1$  表示效应量为“低水平”, $0.1 \leq d < 0.3$  为“中等水平”。③显著性检验采用置换检验,重复抽样 10000 次。

表3 上限回归法的瓶颈水平分析结果%

agglom	infra	cost	talent	demand	innov	policy	eco
0	NN	NN	NN	NN	NN	NN	NN
10	NN	NN	NN	NN	NN	NN	NN
20	NN	NN	NN	NN	NN	NN	NN
30	2.1	NN	0.7	NN	NN	NN	NN
40	14.1	NN	12.2	NN	NN	NN	0.4
50	26.2	NN	23.6	10.7	13	NN	2.8
60	38.2	NN	35.1	24.1	25.9	6.1	5.2
70	50.2	NN	46.6	37.4	38.8	13.2	7.7
80	62.2	NN	58.1	50.8	51.7	20.4	10.1
90	74.2	NN	69.6	64.1	64.6	27.6	12.5
100	86.2	74.8	81.1	77.5	77.5	34.7	14.9

数据来源:作者计算。

注:NN表示“不必要”。

成本(cost)表现为一种典型的极限约束,仅在对顶尖绩效(100%水平)的追求中才显现其必要性,阈值高达 74.8%。

## 2.QCA 分析

本文进一步通过 QCA 方法进行必要性检验。在基于截面数据的定性比较分析中,一致性系数大于 0.9 且覆盖度大于 0.5 是判断必要条件存在的通用标准 (Schneider et al., 2012; Hossain et al., 2022),这一标准在基于面板数据的定性比较分析中同样适用。不同的是,动态 QCA 在检验必要条件

时引入调整距离这一概念,当组间一致性调整距离大于 0.2,变量间关系可能随时间推移发生变化,需要进一步探究该条件的必要性(方芳等,2024)。

从表4可以看出,所有前因变量的汇总一致性均低于 0.9,初步判断不存在结果变量的必要条件。但数字经营成本、数字政策支持及数字生态存在组间一致性调整距离大于 0.2 的情况,需要结合面板数据逐年进行判断,相关结果呈现在表5中。通过对组间一致性调整距离大于 0.2 的因果关系组合进一步分析后发现,情况 2、情况 3 及情况 5 各年份均不满足组间一致性大于 0.9 且组间覆盖度大于 0.5,故不存在必要性关系。情况 1 与情况 4 虽然分别在 2021—2023 年、2023 年出现一致性大于 0.9 且覆盖度大于 0.5,但通过 X—Y 散点图(见图 2)进行可视化判别,发现前者超过 1/3 的案例点均分布在对角线以上,而后者大多数样本点分布于图形右侧纵轴周围,说明上述条件变量均未通过必要性检验(谭海波等,2019;荆玲玲等,2024)。

表4 QCA方法必要条件分析结果

条件变量	agglom			
	汇总一致性	汇总覆盖度	组间一致性调整距离	组内一致性调整距离
infra	0.859	0.780	0.036	0.270
~infra	0.565	0.474	0.069	0.472
cost	0.598	0.486	0.196	0.426
~cost	0.778	0.732	0.229	0.276
talent	0.860	0.844	0.040	0.236
~talent	0.624	0.489	0.073	0.403
demand	0.787	0.845	0.080	0.374
~demand	0.578	0.424	0.087	0.443
innov	0.852	0.897	0.116	0.242
~innov	0.596	0.443	0.131	0.408
policy	0.698	0.622	0.327	0.236
~policy	0.660	0.563	0.342	0.276
eco	0.733	0.637	0.501	0.167
~eco	0.587	0.514	0.567	0.328

数据来源:作者计算。

注:~ 表示集合运算非。

## (二)条件组态的充分性分析

参考既有研究并结合样本实际情况,本文将一致性阈值设为 0.8,将 PRI 阈值设为 0.7,频数阈值设为 2(武永超等,2025),最终涵盖 281 个案例。此

外,由于我国幅员辽阔,省域资源禀赋差异较大,前因条件对结果的作用难以统一判断,所以在反事实操作步骤中不进行方向预设,将7个前因条件状态全部选定为“存在或缺失”。本文主要汇报中间解,并基于中间解与简约解的嵌套关系区分核心条件和边缘条件(杜运周等,2022)。研究重点对高数字产业集聚水平的各组态展开理论内涵的讨论与对应案例的分析。

### 1. 汇总结果分析

通过表6可知,存在三种产生高数字产业集聚水平的组态,整体解的汇总一致性为0.957,高于可接受的最低标准0.75(Park et al., 2020; 杜运周等,2021),且单个组态的组内与组间一致性调整距离均低于0.2,表明汇总一致性具有较好解释力度,这三个组态可以视为省域高数字产业集聚水平产生的充分条件。整体解的汇总覆盖度为0.687,表明

表5 组间一致性调整距离大于0.2的因果组合情况

因果组合情况	情况1		情况2		情况3		情况4		情况5	
	$\sim cost/agglom$		$policy/agglom$		$\sim policy/agglom$		$eco/agglom$		$\sim eco/agglom$	
	一致性	覆盖度	一致性	覆盖度	一致性	覆盖度	一致性	覆盖度	一致性	覆盖度
2014年	0.500	0.931	0.318	0.905	0.923	0.449	0.151	0.993	0.975	0.433
2015年	0.572	0.821	0.530	0.724	0.866	0.522	0.305	0.916	0.932	0.453
2016年	0.649	0.829	0.601	0.625	0.818	0.582	0.361	0.917	0.922	0.467
2017年	0.707	0.803	0.554	0.627	0.853	0.588	0.545	0.834	0.819	0.525
2018年	0.793	0.794	0.594	0.630	0.770	0.575	0.542	0.776	0.680	0.569
2019年	0.809	0.762	0.718	0.580	0.646	0.642	0.755	0.710	0.522	0.581
2020年	0.850	0.724	0.883	0.573	0.435	0.613	0.785	0.637	0.399	0.566
2021年	0.931	0.671	0.961	0.471	0.345	0.590	0.797	0.534	0.254	0.634
2022年	0.950	0.656	0.910	0.438	0.469	0.594	0.828	0.533	0.246	0.721
2023年	0.971	0.621	0.865	0.632	0.522	0.622	0.929	0.510	0.201	0.808

数据来源:作者计算。

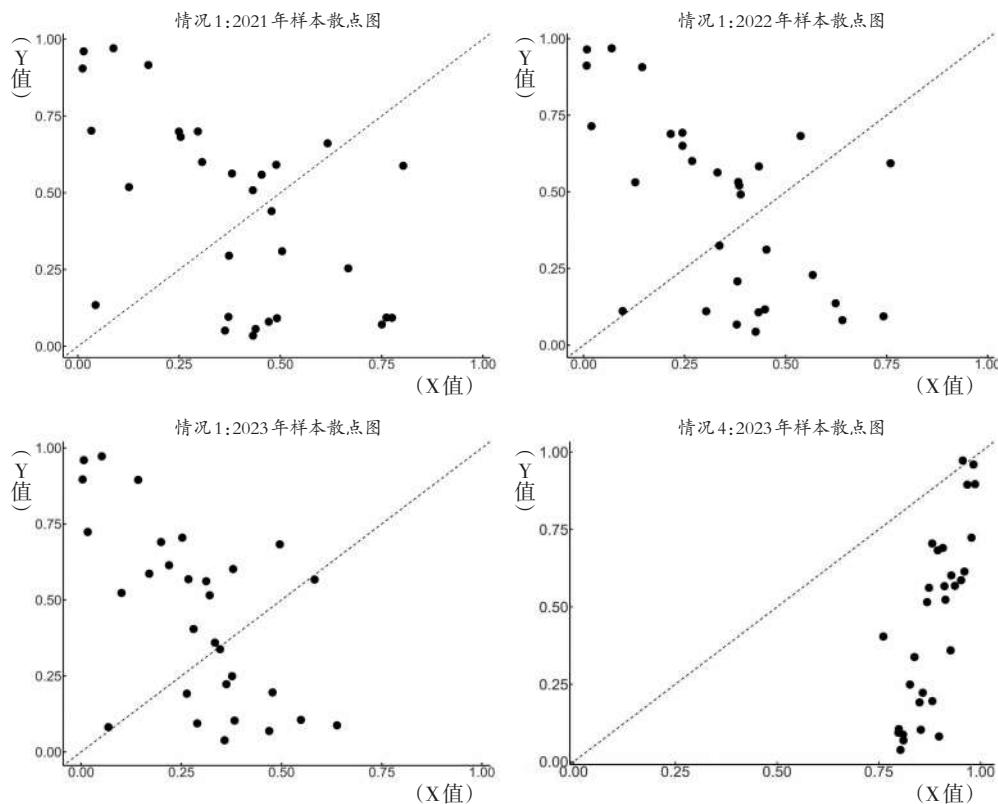


图2 必要性条件检验散点图组图

资料来源:作者绘制。

表6 实现高、非高数字产业集聚水平的组态

前因条件	高数字产业集聚水平 <i>agglom</i>			非高数字产业集聚水平~ <i>agglom</i>			
	H1	H2	H3	NH1	NH2	NH3	NH4
<i>infra</i>	●	●		⊗	⊗	⊗	
<i>cost</i>			●		⊗	⊗	●
<i>talent</i>	●	●	●	⊗	⊗		⊗
<i>demand</i>	●	●	●	⊗		⊗	⊗
<i>innov</i>	●	●	●	⊗	⊗	⊗	⊗
<i>policy</i>	⊗				●	●	●
<i>eco</i>		⊗	●		●	⊗	●
一致性	0.974	0.963	0.966	0.962	0.978	0.982	0.981
PRI	0.927	0.870	0.918	0.932	0.933	0.950	0.937
覆盖度	0.470	0.386	0.568	0.733	0.355	0.368	0.350
唯一覆盖度	0.011	0.011	0.182	0.273	0.008	0.003	0.021
组间一致性 调整距离	0.007	0.022	0.025	0.015	0.018	0.022	0.044
组内一致性 调整距离	0.104	0.115	0.121	0.144	0.121	0.115	0.197
汇总一致性	0.957			0.962			
汇总PRI	0.905			0.930			
汇总覆盖度	0.687			0.766			

案例频数为2,原始一致性阈值为0.8,PRI阈值为0.7

数据来源:作者计算。

注:●代表核心条件存在;⊗代表核心条件缺失;●代表边缘条件存在;⊗代表边缘条件缺失;空白代表该条件对于结果的发生可有可无;下表同。

三个组态可解释 68.7% 的高水平案例,满足动态 QCA 的解释力要求。本文遵循组态理论化的流程,对上述发现的省域数字产业集聚组态命名:市场—技术双轮驱动型路径、基础支撑引领型路径以及生态引领克服成本约束型路径,各组态路径的具体情况分析如下。

首先,市场—技术双轮驱动型路径(组态 H1)展现了内源性增长的特点。该路径以完善的数据基础设施、活跃的数字市场需求和强劲的数字技术创新为核心驱动力,辅以良好的数字人才供给基础,而数字政策支持在此路径中相对弱化。这一模式在 2014—2019 年尤为显著,北京、上海、江苏、山东和广东等经济发达省份(直辖市)通过市场机制与技术创新双轮驱动,形成了自下而上的产业发展路径。这些地区通常具备完善的数字经济生态体系,民营企业活跃,自主研发能力较强,市场需求旺盛,使得数字产业能够在较少政策干预的情况下实现内生性增长。

其次,基础支撑引领型路径(组态 H2)凸显了以资源与能力为本位的发展逻辑。该路径的充分

性由数据基础设施、数字市场需求与数字技术创新三大核心条件的存在所驱动。值得注意的是,数字生态在该组态中呈现“边缘缺失”,表明其并非此路径成功的核心驱动因素。这意味着浙江、福建、湖北、四川等典型案例省份在 2014—2018 年实现数字产业集聚,主要依托的是其雄厚的工业基础、显著的区位优势以及由大规模基础设施投入和市场主体构成的压倒性比较优势。在此模式下,强劲的硬件资源与市场规模有效地主导了发展进程,使得数字生态即便具有一定水平,也未成为关键性的制约或推动力量。然而,这种对基础资源高度依赖的模式,在面对外部冲击或产业升级要求时,其韧性与可持续性可能面临挑战。

最后,生态引领克服成本约束型路径(组态 H3)代表了环境赋能的新型发展模式。该路径以数字人才、市场需求和数字生态为核心优势,通过优越的创新环境和金融支持抵消经营成本压力。2018—2023 年,北京、上海、江苏、山东、湖北、广东、四川和陕西等省份(直辖市)相继转向这一发展模式。这一转变反映了数字经济发展阶段的演进:随着产业升级深入,单纯依靠市场和技术已不足以维持竞争优势,而构建良好的数字生态成为吸引高端要素、提升产业能级的关键。特别是陕西等省份的加入,表明通过积极构建数字金融体系和优化创新环境,相对欠发达地区也能够突破成本约束,实现数字产业追赶发展。

值得注意的是,北京、上海、江苏、广东等发达省份(直辖市)在不同时期呈现出路径转换的特征,从初期的市场技术驱动逐步转向生态引领驱动,这一现象深刻反映了数字经济发展阶段的内在规律:初期依靠要素投入和市场扩张,中期注重基础设施和能力建设,后期则强调生态构建和环境优化。这种路径演化不仅体现了各省份根据发展阶段调整战略重点的理性选择,也揭示了数字产业从规模扩张向质量提升演进的内在要求。

## 2.组间结果分析

为改善传统 QCA 的组态时间盲区问题,本研究进一步进行了组间一致性时序分析(见图 3)探讨时间效应。分析发现,三条路径的一致性在整个研究期内均高于 0.93,表明所有组态均具有较高的充分性,能够有效解释高数字产业集聚的形成。但其动态趋势揭示了迥异的演化特征与稳健性。组

态H1(市场—技术驱动型)的一致性始终保持高位平稳,波动极小。这表明依赖市场主体活力与技术创新内生动力的增长模式具有强大的稳健性与路径依赖性,受外部环境冲击的影响较小。组态H2(基础支撑引领型)的一致性波动相对明显,尤其在2021年出现显著下滑。这反映出该路径高度依赖固定资产投资与政策资源投入,其效能易受宏观经济周期与地方政府债务约束等外部因素的影响,抗风险能力较弱。组态H3(生态引领型)的演化趋势最具理论启示,其在2014—2017年作为先锋路径的一致性均高于0.99,彰显了其作为稀缺战略的极高效力;然而,自2020年起其一致性呈现持续下降趋势,由0.993降至2023年的0.939。这一看似矛盾的发现,即路径被广泛采纳和其解释效力逐渐递减并存,实则深刻揭示了数字经济发展中“战略趋同”与“边际效用递减”的规律。早期,少数省份通过构建数字生态形成的制度红利与创新租金获得超常规发展;但随着这一策略被竞相模仿并成为全国性的普适性政策,其先行者优势被稀释,各省份在数字金融、人才政策、园区建设等方面陷入同质化竞争,导致该路径的边际效用显著下降。

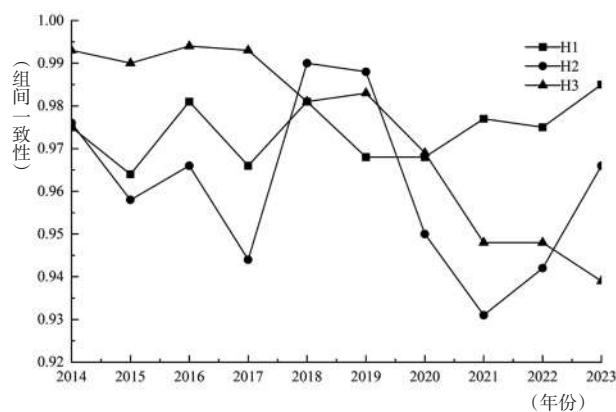


图3 2014—2023年组间一致性趋势图

资料来源:作者绘制。

### 3.组内结果分析

组内一致性计算结果显示,除贵州、新疆外,其余28个省份在三种组态上的一致性均高于0.75的可接受标准,多数省份达到0.95以上。同时,3个条件组态的组内一致性调整距离均低于0.2,表明无显著个体(集群)效应,即每个组态在各省域解释力度基本一致。在组态解释力无显著个体效应的情况下,需要进一步分析各组态的组内覆盖度,以考察组态所能解释案例的地区分布情况。鉴于QCA

还未开发专门针对该情况的测算指标,本文参考张放(2023)的研究方法,进一步采用单因素方差分析与Kruskal-Wallis秩和检验,研究三种组态覆盖度在我国四大经济区域(东部、中部、西部、东北)的差异情况,深入探究各组态解释力的空间分布特征。经检验,各组态的覆盖度数据均满足正态性与方差齐性前提,故采用单因素方差分析进行统计检验,结果如表7所示。单因素方差分析结果表明,三种组态的覆盖度在四大经济区域间均呈现出一定的差异性,但显著程度各不相同。组态H2的区域差异最为显著,且其平均覆盖度最高。这表明该路径的解释力存在强烈的空间不平衡性,对某些特定区域的案例,如资源禀赋雄厚或工业基础较好的地区,具有最强的解释力度。组态H1的覆盖度在区域间也存在统计学上的显著差异。这意味着该路径的有效性也显示出一定的地域依赖性。组态H3的覆盖度区域差异未达到常规的显著性水平,处于边缘显著状态。这表明该路径的解释力在不同经济区域间的分布相对更为均衡,普适性较强。

表7 单因素方差分析结果

分类	H1	H2	H3
总体均值	0.428	0.675	0.329
总体标准差	0.143	0.225	0.110
F	3.266	6.668	2.929
p(渐进显著性)	0.037*	0.002**	0.052
覆盖度均值			
东部	0.512	0.402	0.627
中部	0.313	0.267	0.378
西部	0.576	0.605	0.641
东北	0.732	0.72	0.662

数据来源:作者计算。

注: \*p<0.05, \*\*p<0.01。

对各组态在四大经济区域覆盖度的方差分析表明,其空间分布存在显著差异。深入分析覆盖度均值(见表7),可清晰识别各组态的“优势区域”。组态H1与H2在东北地区的覆盖度最高,显示出与传统工业禀赋的强关联性。组态H3在东、西、东北部均保持较高覆盖度,显示出作为新兴主导路径的广泛适应潜力,有效弥补了其在显著性上的边界水平。中部地区在所有路径上的覆盖度均为最低,提示其数字产业集聚可能受制于未被现有组态捕捉的独特因素,或正处于发展模式的转型阵痛期。

### 4.非高数字产业集聚水平组态分析

为检验数字产业集聚驱动机制中可能存在的

因果非对称性,本文对导致非高数字产业集聚水平的组态进行分析。如表6所示,共识别出四种充分条件组态,其整体解的一致性为0.962,覆盖度为0.766,表明这些路径能够稳健地解释大部分非高集聚案例的发生。通过对四条非高集聚路径的归纳分析,发现其呈现出两种典型的失灵模式:

其一为“要素—市场双重匮乏型”失灵模式,包含组态NH1和NH3。这一模式的核心特征在于数据基础设施与数字市场需求这两个关键条件的系统性缺失。这一发现表明,当地区同时面临基础硬件支撑不足和市场消化能力有限的双重制约时,数字产业将难以突破发展的初始瓶颈。特别是数字市场需求的持续疲软,使得数字企业缺乏足够的价值实现空间,导致产业生态系统无法形成良性循环。其二为“政策与生态双重失灵型”失灵模式,包含组态NH2和NH4。该模式揭示了政策与生态支持的局限性:在数据基础设施、数字人才供给等多维条件严重缺失的背景下,即使存在高数字政策支持和高数字生态等有利条件,仍难以有效推动产业集聚。这一发现具有重要的政策启示,表明单纯依靠政策引导和生态建设而忽视基础要素培育的发展策略,往往难以取得预期效果。

值得注意的是,数字技术创新的缺失在四条路径中普遍存在,凸显其在驱动集聚中的核心地位。而高经营成本仅在组态NH4中作为核心条件出现,说明其影响具有条件特定性。

### (三)稳健性检验

为评估研究结论的可靠性,本文对产生高数字产业集聚水平的组态进行稳健性检验。首先,考虑到北京、上海、天津、重庆四个直辖市在行政层级、资源集聚程度等方面与普通省级行政区存在系统性差异,其发展模式可能具有独特性。为确保前述组态路径源于更普适的驱动机制而非个别特殊案例的特征,本研究将4个直辖市从样本中剔除,并保持与主分析完全一致的校准锚点与分析参数,对剩余的26个省份重新进行模糊集定性比较分析(杜雯秦等,2021)。结果如表8所示,调整样本空间后组态数量减少,但所得组态与原组态存在清晰子集关系,整体解与各条路径的一致性水平均保持稳定,且覆盖度未出现大幅波动,表明上述研究结果具有充分的稳健性(张明等,2019;程恋军等,2025)。

另外,依据审慎性原则,将真值表构建过程中

表8 改变样本空间的稳健性检验

条件变量	组态分析-高	
	H1'	H2'
infra	●	●
cost		
talent	●	●
demand	●	●
innov	●	●
policy		
eco		⊗
一致性	0.954	0.964
PRI	0.885	0.872
覆盖度	0.566	0.509
唯一覆盖度	0.024	0.067
总体一致性		0.944
总体PRI		0.861
总体覆盖度		0.733

数据来源:作者计算。

的一致性阈值由0.80调高为0.85(张明等,2020),频数与PRI阈值不做变动,发现调整策略下组态数量、核心条件组合均与原始结果保持一致,这进一步证实了研究结果的稳健性。

## 五、结论与启示

本研究从数据要素流动的全过程视角出发,构建了一个整合基础支撑、价值转化与环境适配三大维度的理论框架,并运用NCA与动态QCA相结合的方法,基于2014—2023年中国30个省份的面板数据,系统分析了省域数字产业集聚的复杂驱动机制与路径演化规律。主要研究结论如下。

第一,数字产业集聚不存在单一的必要条件,而是多重要素协同联动的结果。必要性分析表明,在单一条件视角下,未有任何变量构成高水平数字产业集聚的必要条件。然而,瓶颈水平分析进一步揭示,不同条件在不同发展阶段具有差异化的约束强度:数据基础设施与数字人才供给构成基础性门槛,数字市场需求与数字技术创新在中高发展阶段作用凸显,而数字政策支持与数字生态则更多表现为高阶保障性条件。第二,存在三条驱动高数字产业集聚的差异化组态路径,体现了“多重并发”与“殊途同归”的特征。具体包括:市场—技术双轮驱动型路径,以完善的基础设施、活跃的市场需求与强劲的创新能力为核心,政策干预相对弱化;基础

支撑引领型路径,依托大规模的硬件投入与市场主体形成压倒性比较优势;生态引领克服成本约束型路径,以优越的数字生态与人才供给抵消运营成本压力,实现高质量发展。这三条路径分别对应不同的资源禀赋与发展逻辑,为不同条件的省份提供了多元化的战略选择。第三,数字产业集聚路径具有显著的动态演化特征。时序分析表明,市场—技术驱动型路径稳健性最强,基础支撑引领型路径易受外部冲击,而生态引领型路径则呈现“边际效用递减”趋势,反映出数字经济发展中“战略趋同”的内在规律。同时,北京等发达省份呈现出从“要素驱动”到“生态驱动”的路径转换,体现了数字产业从规模扩张向质量提升演进的内在要求。

基于上述研究结论,本文提出如下政策启示。

第一,因地制宜选择差异化发展路径。各省份应基于自身资源禀赋和发展阶段选择适合的集聚路径:东部发达地区可持续优化市场环境和创新生态,巩固生态引领型发展模式;中西部省份可立足基础设施和成本优势,通过靶向投入培育特色数字产业集群,避免盲目模仿和同质化竞争。第二,推动多维度条件协同发展。政策制定应注重基础支撑、价值转化和环境适配三个维度的系统协同,特别是在加强数据基础设施建设、培育数字人才、激发市场需求和促进技术创新的同时,要通过制度设计和金融支持优化数字生态系统,形成政策合力。第三,关注发展阶段的动态转换。地方政府应根据数字产业发展阶段及时调整战略重点:在发展初期可侧重于基础设施建设和市场培育,在中后期则应更加注重创新生态构建和营商环境优化,实现从要素驱动向创新驱动和生态驱动的适时转换。第四,避免“政策万能论”的误区。研究结果表明,在基础要素系统缺失的情况下,单纯依靠政策扶持难以有效推动数字产业集聚。因此,政策制定应更加注重培育内生发展能力,通过制度建设、环境营造和要素培育激发市场活力,实现可持续发展。

## 参考文献

- [1]马文娟.提升数字产业集群国际竞争力[EB/OL].(2024-12-28).[https://m.gmw.cn/toutiao/2024-12/28/content\\_37766012.htm](https://m.gmw.cn/toutiao/2024-12/28/content_37766012.htm).
- [2]卢福财,钟诗韵.数字产业集群发展水平评价及其时空演变[J].经济地理,2024,44(6).
- [3]赵放,张森.数字产业地理集聚与虚拟集聚融合推进中国式现代化[J].山西大学学报(哲学社会科学版),2025,48(1).
- [4]KERR W R, ROBERT-NICOUD F. Tech clusters [J]. Journal of economic perspectives, 2020, 34(3): 50-76.
- [5]辛璐璐.数字产业集聚、颠覆式技术创新与城市绿色经济效率[J].学习与实践,2023,(10).
- [6]李栋,张映芹,李开源.中国省际数字经济核心产业集聚度、非平衡性与动态演进[J].统计与决策,2023,39(18).
- [7]赵放,徐熠.我国数字产业集聚竞争态势比较与影响因素分析:以十九大城市群的实证数据为例[J].贵州社会科学,2024(6).
- [8]叶堂林,刘哲伟,张京亮.数字产业空间集聚影响因素探析[J].科技进步与对策,2023,40(15).
- [9]赵放,李文婷,马婉莹.数据要素市场化能否促进数字产业集聚:来自准自然实验的证据[J].浙江学刊,2024(3).
- [10]RUDY F, HÉCTOR C. The digital industrial cluster (DIC) in a post-pandemic era: exploring its theoretical deployment and potential benefits[J]. Procedia computer science, 2023, 221(8): 1131-1138.
- [11]林挺.培育信息产业新业态对数字经济空间格局的影响研究[J].当代财经,2023(5).
- [12]余壮雄,韩佳容,付锦华.“宽带中国”政策如何影响中国城市的数字产业[J].世界经济,2024,47(8).
- [13]刘淑春.中国数字经济高质量发展的靶向路径与政策供给[J].经济学家,2019(6).
- [14]夏杰长,陶鸿.数字产业集群创新与演化研究[J].财经问题研究,2025(1).
- [15]CENAMOR J, PARIDA V, WINCENT J. How entrepreneurial SMEs compete through digital platforms: the roles of digital platform capability, network capability and ambidexterity [J]. Journal of business research, 2019, 100: 196—206.
- [16]胡乾韬,王节祥,杨大鹏,等.集群中小企业协同新创平台的资源行动机制:构建数字产业集群的路径探索[J].管理案例研究与评论,2024,17(4).
- [17]钞小静,廉园梅,元茹静,等.数字基础设施建设与产业链韧性:基于产业链恢复能力数据的实证分析[J].数量经济技术经济研究,2024,41(11).
- [18]丛屹,闫苗苗.数字经济、人力资本投资与高质量就业[J].财经科学,2022(3).
- [19]李彦臻,任晓刚.科技驱动视角下数字经济创新的动力机制、运行路径与发展对策[J].贵州社会科学,2020(12).
- [20]李研.中国数字经济产出效率的地区差异及动态演变[J].数量经济技术经济研究,2021,38(2).
- [21]毛丰付,高雨晨,周灿.长江经济带数字产业空间格局演化及驱动因素[J].地理研究,2022,41(6).
- [22]RAGIN C C. Redesigning social inquiry: fuzzy sets and

- beyond[M]. Chicago: University of Chicago Press, 2008.
- [23]杜运周,贾良定.组态视角与定性比较分析(QCA):管理学研究的一条新道路[J].管理世界,2017(6).
- [24]杜运周,李佳馨,刘秋辰,等.复杂动态视角下的组态理论与QCA方法:研究进展与未来方向[J].管理世界,2021,37(3).
- [25]DUL J, VAN DER LAAN E, KUIK R. A statistical significance test for necessary condition analysis[J]. *Organizational research methods*, 2020, 23(2): 385—395.
- [26]杜运周,刘秋辰,陈凯薇,等.营商环境生态、全要素生产率与城市高质量发展的多元模式:基于复杂系统观的组态分析[J].管理世界,2022,38(9).
- [27]范剑勇,刘念,刘莹莹.地理距离、投入产出关系与产业集聚[J].经济研究,2021,56(10).
- [28]袁歌骋,潘敏,覃凤琴.数字产业集聚与制造业企业技术创新[J].中南财经政法大学学报,2023(1).
- [29]王欣亮,张家豪,刘飞.大数据是经济高质量发展的新引擎吗:基于数据基础设施与技术应用的双重效应解释[J].统计研究,2023,40(5).
- [30]陶峰,朱盼,邱楚芝,等.数字技术创新对企业市场价值的影响研究[J].数量经济技术经济研究,2023,40(5).
- [31]张放.影响地方政府信息公开的因素:基于省域面板数据的动态QCA分析[J].情报杂志,2023,42(1).
- [32]田庆锋,易磊.战略性新兴产业技术创新驱动新质生产力发展的组态研究[J].科研管理,2025,46(5).
- [33]SCHNEIDER C Q, WAGEMANN C. Set-theoretic methods for the social sciences: a guide to qualitative comparative analysis[M]. Cambridge: Cambridge University Press, 2012.
- [34]HOSSGIN M A, QUADDUS M, WARREN M, et al. Are you a cyberbully on social media? Exploring the personality traits using a fuzzy-set configurational approach[J]. *Internet Research*, 2022, 32(6): 2016—2038.
- [35]方芳,张立杰,赵军.制度组态视角下提升农业绿色全要素生产率的多元路径探析:基于动态QCA的面板数据分析[J].中国农村经济,2024(2).
- [36]谭海波,范梓腾,杜运周.技术管理能力、注意力分配与地方政府网站建设:一项基于TOE框架的组态分析[J].管理世界,2019,35(9).
- [37]荆玲玲,黄慧丽.时空双维下数字创新生态系统对区域创新能力的激发与影响研究:基于省域面板数据的动态QCA分析[J].科技进步与对策,2024,41(16).
- [38]武永超,王磊.制度逻辑视角下农业新质生产力培育路径研究:基于动态QCA的分析[J].管理学刊,2025,38(3).
- [39]PARK Y K, FISS P C, EL SAWY O A. Theorizing the multiplicity of digital phenomena: The ecology of configurations, causal recipes, and guidelines for applying QCA[J]. *MIS Quarterly*, 2020, 44(4): 1493—1520.
- [40]杜雯秦,郭淑娟.双碳目标下我国绿色能源效率提升路径研究[J].管理现代化,2021,41(6).
- [41]张明,杜运周.组织与管理研究中QCA方法的应用:定位、策略和方向[J].管理学报,2019,16(9).
- [42]程恋军,王琳茜.定性比较分析(QCA)中的稳健性研究:分析策略与未来方向[J].中国人力资源开发,2025,42(1).
- [43]张明,蓝海林,陈伟宏,等.殊途同归不同效:战略变革前因组态及其绩效研究[J].管理世界,2020,36(9).

## Research on Differentiated Pathways of Provincial Digital Industry Agglomeration from the Perspective of Data Factor Flow

Huang Xiaoting Sun Jiuwen Han Yaqing

**Abstract:** Against the backdrop of the “Digital China” strategy and the accelerated marketization process of data factors, digital industry agglomeration has become an important pathway for promoting high-quality regional economic development. Based on a holistic process perspective of data factor flow, this paper integrates a theoretical framework encompassing three major dimensions: foundational support, value transformation, and environmental adaptation. Employing Necessary Condition Analysis (NCA) and dynamic Qualitative Comparative Analysis (QCA), it conducts configurational path identification and evolutionary analysis on panel data from 30 Chinese provinces from 2014 to 2023. The findings are as follows: First, there is no single necessary condition for digital industry agglomeration; rather, it results from the synergistic interaction of multiple factors, with different conditions exhibiting varying constraining intensities at different development stages. Second, three differentiated pathways driving high-level digital industry agglomeration are identified, reflecting the configurational characteristics of “multiple concurrency” and “equifinality”. Third, pathway evolution demonstrates dynamism: the market-technology driver exhibits the strongest robustness, the ecology-led type shows a trend of diminishing marginal utility, and developed provinces demonstrate a transition trend from factor-driven to ecology-driven agglomeration.

**Key Words:** Data Factor Flow; Digital Industry Agglomeration; Differentiated Pathways; Configurational Analysis; Provincial Disparities

(责任编辑:江 夏)