

doi:10.13582/j.cnki.1672-7835.2021.06.015

要素禀赋与中西部地区数字经济可持续发展

——基于机器学习方法

曾祥炎,李姣,曾小明

(湖南科技大学 商学院,湖南 湘潭 411201)

摘要:在地区间数字经济投资“竞赛”日益激烈,并且增速上形成中西部对东部地区赶超之势的背景下,中西部地区的要素禀赋是否可以支撑其数字经济可持续发展是值得深入研讨的问题。运用机器学习方法中的随机森林算法,基于地级市层面数据对24种投入要素进行分析,发现职工平均工资、人均社会消费品零售额、人均科学技术支出、城镇单位从业人员中信息传输计算机服务和软件业从业人员占比4种要素是支撑数字经济发展最为重要的投入要素。进一步运用4种最重要要素的偏效应曲线图进行分析,发现在要素禀赋上,中西部地区数字经济的发展已经同东部地区站到了“同一起跑线上”,且大多数城市的重要投入要素已经具备了支撑数字经济可持续发展的基本条件,但城市之间的差距仍然明显。因此,要实现中西部地区数字经济的可持续发展,既要根据不同城市的要素禀赋构成情况实施差异化的补“短板”政策,又要通过推进数字资源的平衡布局来改变城市之间的数字经济不平衡发展状况,以构筑中西部地区数字经济发展的良性循环。

关键词:要素禀赋;数字经济;可持续发展;机器学习

中图分类号:F49 **文献标志码:**A **文章编号:**1672-7835(2021)06-0080-10

一 问题的提出

近年来我国数字经济发展迅速,2020年规模达到39.2万亿元,占GDP比重达38.6%,增速达GDP增速3倍以上,成为拉动经济增长的关键动力。在增速比较上,贵州、重庆、福建数字经济增速位列全国前3名,增速前10名中有8个是中西部地区省份^①。可以说,数字经济在增速上已经形成了中西部对东部的赶超之势,表明中西部地区有望通过大力发展数字经济来缩小与东部地区的经济差距,从而为加快实现我国区域协调发展创造积极条件。只是,要实现这一点,必须有一个前提:中西部地区数字经济增速赶超东部地区是可持续性的,换言之,中西部地区的要素禀赋与制度设计可以持续支撑这些地区通过发展数字经济

来缩小与东部地区的差距。

一般而言,一个地区的产业要可持续发展,与其拥有的要素禀赋存在密切关联。根据新经济地理学的解释,地理位置和自然要素禀赋是产业集聚的起始条件,规模报酬递增和正反馈效应则可以进一步导致产业集聚的自我强化,使得优势地区保持领先。因此,地区产业可持续发展既与其自然要素禀赋相关联,也与其“后天”的要素禀赋,如人口与经济规模、科技、制度等相关联。相比较而言,数字经济的发展似乎更依赖后者而不是前者,因为与传统制造业的空间聚集规律不同,数字企业有明显摆脱港口、道路交通或原材料产地等传统经济地理因素束缚的趋势^②。

收稿日期:2021-07-26

基金项目:国家社科基金一般项目(20BJL096)

作者简介:曾祥炎(1975—),男,湖南新宁人,教授,博士生导师,主要从事宏观经济理论与政策、区域发展规划与新型城镇化建设、数字经济与数字信息资源配置研究。

^①根据中国信通院《2021年中国数字经济发展白皮书》,2020年数字经济增速前10的省份分别为:贵州、重庆、福建、湖南、四川、江西、浙江、广西、安徽、河北。

^②参见乔尔·科特金:《新地理:数字经济如何重塑美国地貌》,王玉平、王阳译,社会科学文献出版社2001年版,第1页;钟业喜,毛炜圣:《长江经济带数字经济空间格局及影响因素》,《重庆大学学报(社会科学版)》2020年第1期。

影响数字经济发展的原因远比影响传统制造业的原因要复杂得多,对其产生关键影响的要素禀赋也更难明确,这就需要在分析要素禀赋与地区数字经济发展之间的关系时,尽可能选择足够的对数字经济产出可能有影响的投入要素,以便从中获取更多有价值的信息。只是,囿于目前大多数研究使用的是传统的回归分析方法,现有文献在投入要素数量的选择上尚难有显著突破,一般只纳入5—7个投入要素进行分析。为突破现有文献在投入要素数量选择上的局限,尽可能全面地考虑各地区的要素禀赋对数字经济发展的影响,避免一些重要的投入要素被遗漏。本文拟采用机器学习方法,尽可能尝试加入更多的对数字经济产出有影响的投入要素,从而系统分析要素禀赋与地区数字经济发展之间的关系,更精准地找出影响数字经济发展的最重要资源要素,然后进一步探讨地区应如何依据自身要素禀赋找到数字经济发展的最优路径,进而实现地区数字经济可持续发展。

二 理论分析框架

为厘清要素禀赋与数字经济发展水平之间的关系,本文参照Feenstra两种要素两种产品的理论分析框架^①。假设全国仅存在A、B两个地区,地区的数字经济发展水平用地区的数字经济产量代替,表示为字母Y,若A地的数字经济发展水平比B地高,则 $Y_A > Y_B$ 。再假设决定数字经济产量的要素仅有两种,且要素价格在A、B两地相同,则对于两地区和两种要素的经济体来说,生产一单位数字经济产量的成本分别为:

$$C_A(P_1, P_2) = A_1P_1 + A_2P_2 \quad (1)$$

$$C_B(P_1, P_2) = B_1P_1 + B_2P_2 \quad (2)$$

其中, C_A 、 C_B 是A、B两地各生产一单位数字经济产量的成本, A_1 、 A_2 是A地使用两种要素的数量, B_1 、 B_2 是B地使用两种要素的数量, P_1 、 P_2 是两种要素的价格,经济体的总成本为:

$$C(P_1, P_2) = A_1P_1 + A_2P_2 + B_1P_1 + B_2P_2 \quad (3)$$

基于微观经济学的成本最小化原则,应使总成本达到最低,再根据数学线性规划问题中的多元函数求极值的原理,可构造拉格朗日函数L:

$$L = A_1P_1 + A_2P_2 + B_1P_1 + B_2P_2 +$$

$$\lambda_1 F_A(A_1, A_2) + \lambda_2 F_B(B_1, B_2) \quad (4)$$

其中, $A_1P_1 + A_2P_2 + B_1P_1 + B_2P_2$ 是目标函数, $F_A(A_1, A_2)$ 、 $F_B(B_1, B_2)$ 是A、B两地关于第一、二种要素的生产函数,也是约束条件,上式分别对 A_1 、 A_2 、 B_1 、 B_2 、 λ_1 、 λ_2 求偏导,可以得到:

$$\begin{cases} \frac{\partial L}{\partial A_1} = P_1 + \lambda_1 \frac{\partial F_A}{\partial A_1} = 0 \\ \frac{\partial L}{\partial A_2} = P_2 + \lambda_1 \frac{\partial F_A}{\partial A_2} = 0 \\ \frac{\partial L}{\partial B_1} = P_1 + \lambda_2 \frac{\partial F_B}{\partial B_1} = 0 \\ \frac{\partial L}{\partial B_2} = P_2 + \lambda_2 \frac{\partial F_B}{\partial B_2} = 0 \\ \frac{\partial L}{\partial \lambda_1} = F_A(A_1, A_2) = 0 \\ \frac{\partial L}{\partial \lambda_2} = F_B(B_1, B_2) = 0 \end{cases} \quad (5)$$

以上6个方程可解出最关键的4个未知数 a_1 、 a_2 、 b_1 、 b_2 ,其中 a_1 与 a_2 是A地生产一单位数字经济产量所使用的两种要素的最优组合数量, b_1 与 b_2 是B地生产一单位数字经济产量所使用的两种要素的最优组合数量。若整个经济体的要素资源充分利用,则:

$$\begin{cases} a_1Y_A + b_1Y_B = E_1 \\ a_2Y_A + b_2Y_B = E_2 \end{cases} \quad (6)$$

其中, E_1 、 E_2 分别是经济体中第一种要素和第二种要素的投入总量, Y_A 、 Y_B 分别为A、B两地的数字经济产量。再对上述方程组两边全微分,可以得到:

$$\begin{cases} a_1dY_A + b_1dY_B = dE_1 \\ a_2dY_A + b_2dY_B = dE_2 \end{cases} \quad (7)$$

进一步地,上式可以转换为:

$$\begin{cases} \frac{a_1Y_A}{E_1} \cdot \frac{dY_A}{Y_A} + \frac{b_1Y_B}{E_1} \cdot \frac{dY_B}{Y_B} = \frac{dE_1}{E_1} \\ \frac{a_2Y_A}{E_2} \cdot \frac{dY_A}{Y_A} + \frac{b_2Y_B}{E_2} \cdot \frac{dY_B}{Y_B} = \frac{dE_2}{E_2} \end{cases} \quad (8)$$

其中, $\frac{a_1Y_A}{E_1}$ 、 $\frac{a_2Y_A}{E_2}$ 分别表示A地所使用的第一种要素和第二种要素的量占整个经济体中所使用相应

^①Feenstra R.C. “Advanced International Trade: Theory and Evidence Second Edition”, *Economics Books*, 2015, 66(2): 541-544.

要素总量的比例,如果 $\frac{a_1 Y_A}{E_1} > \frac{a_2 Y_A}{E_2}$, 则说明 A 地第一种要素相对于第二种要素对数字经济发展的影响更为关键;相反,如果 $\frac{a_1 Y_A}{E_1} < \frac{a_2 Y_A}{E_2}$, 则第二种要素影响更为关键。同时, B 地亦同理。另外, $\frac{dY_A}{Y_A}$ 、 $\frac{dY_B}{Y_B}$ 分别表示 A、B 两地数字经济产量的百分比变化, $\frac{dE_1}{E_1}$ 、 $\frac{dE_2}{E_2}$ 分别表示第一、二种要素的百分比变化。

以上两要素两地区的分析框架可以进一步推广到多要素多地区,因为多要素多地区的要素禀赋与数字经济产量之间的关系,同理存在:

$$\begin{cases} \frac{a_{11} Y_1}{E_1} \cdot \frac{dY_1}{Y_1} + \frac{a_{21} Y_2}{E_1} \cdot \frac{dY_2}{Y_2} + \dots + \frac{a_{i1} Y_i}{E_1} \cdot \frac{dY_i}{Y_i} = \frac{dE_1}{E_1} \\ \frac{a_{12} Y_1}{E_2} \cdot \frac{dY_1}{Y_1} + \frac{a_{22} Y_2}{E_2} \cdot \frac{dY_2}{Y_2} + \dots + \frac{a_{i2} Y_i}{E_2} \cdot \frac{dY_i}{Y_i} = \frac{dE_2}{E_2} \\ \dots \\ \frac{a_{1j} Y_1}{E_j} \cdot \frac{dY_1}{Y_1} + \frac{a_{2j} Y_2}{E_j} \cdot \frac{dY_2}{Y_2} + \dots + \frac{a_{ij} Y_i}{E_j} \cdot \frac{dY_i}{Y_i} = \frac{dE_j}{E_j} \end{cases} \quad (9)$$

其中, a_{ij} 代表第 i 个地区生产一单位数字经济产量所使用的第 j 种要素的最优数量, Y_i 代表第 i 个地区的数字经济产量, E_j 代表整个经济体中第 j 种要素的投入总量,那么 $\frac{a_{ij} Y_i}{E_j}$ 则代表第 i 个地区第 j 种要素的重要程度,记为 β_{ij} 。根据这一比例可以得知各个地区中各个要素对数字经济发展的贡献大小。

再对各个地区中各个要素的贡献大小进行归一化处理:

$$\beta_{1j} = \frac{\beta_{1j}}{\sum_{k=1}^j \beta_{1k}}, \beta_{2j} = \frac{\beta_{2j}}{\sum_{k=1}^j \beta_{2k}}, \dots, \beta_{ij} = \frac{\beta_{ij}}{\sum_{k=1}^j \beta_{ik}} \quad (10)$$

最后,综合所有地区对 j 种要素的贡献大小进行均值处理:

$$\gamma_1 = \frac{\beta_{11} + \beta_{21} + \dots + \beta_{i1}}{i}, \gamma_2 = \frac{\beta_{12} + \beta_{22} + \dots + \beta_{i2}}{i}, \dots, \gamma_j = \frac{\beta_{1j} + \beta_{2j} + \dots + \beta_{ij}}{i} \quad (11)$$

其中, $\gamma_1, \gamma_2, \dots, \gamma_j$ 分别代表 j 种要素的贡献大小,对其进行相互对比就可以得到投入要素重要程度的排序。

三 数据与经验分析方法

(一) 变量选取

由于数字经济发展不仅与地区“先天”的自然要素禀赋,如地理区位、气候条件等有关,更与地区“后天”的要素禀赋,如技术、人力资本、经济发展、信息化、产业结构、人口规模、交通基础设施、制度条件、科教支出、金融发展等相关联。为纳入更多投入要素进行分析,本文综合参考了中国信息化百人会课题组^①、王彬燕等^②、甘星和刘成昆^③、欧阳志刚和陈普^④等学者的做法,最终选择了自然资源、人力资源、物质资本、科技创新、结构、制度和市场驱动等 7 个类别共 24 种要素,具体如表 1 所示。其中需要特别说明的是,关于人力资本的度量有基于成本的教育年限法和基于收入的劳动者平均工资法^⑤,而我国城市平均受教育年限的测度数据在全国人口普查以及地区抽查数据中有详尽的记录,又因全国人口普查每十年进行一次,最后无法获取完整的地区连续数据,故参考何秋琴等^⑥与徐康宁和王剑^⑦等学者的做法采用职工平均工资、中等职业技术学校在校学生数占总人口比重和高等学校在校学生数占总人口比重来表示人力资本水平。

①中国信息化百人会课题组:《信息经济崛起:区域发展模式、路径与动力》,电子工业出版社 2016 年版,第 118、120 页。

②王彬燕,田俊峰,程利莎,等:《中国数字经济空间分异及影响因素》,《地理科学》2018 年第 6 期。

③甘星,刘成昆:《区域金融发展、技术创新与产业结构优化——基于深圳市 2001—2016 年数据的实证研究》,《宏观经济研究》2018 年第 11 期。

④欧阳志刚,陈普:《要素禀赋、地方工业行业发展与行业选择》,《经济研究》2020 年第 1 期。

⑤黎金玲,诺克军,等:《基于劳动者平均工资的人力资本计量模型研究》,《中国地质大学学报》2005 年第 2 期。

⑥何秋琴,郭美晨,等:《品牌资本、R&D 资本和全要素生产率》,《科学学研究》2019 年第 3 期。

⑦徐康宁,王剑:《自然资源充裕程度与经济发展水平关系的研究》,《经济研究》2006 年第 1 期。

表 1 要素禀赋类别

要素禀赋类别	一级测度指标	二级测度指标
自然资源要素	气候资源	年平均最低气温、年平均最高气温
	教育投入	人均教育支出
人力资源要素	劳动力投入	城镇单位从业人员中信息传输计算机服务和软件业占比
	人力资本水平	职工平均工资、中等职业技术学校在校学生数占总人口比重、 高等学校在校学生数占总人口比重
	人口规模	人口密度
物质资本要素	信息化基础设施建设水平	电信业务收入占 GDP 比重
	交通基础设施建设水平	公路密度、地级市到最近港口的距离
科技创新要素	科技投入	人均科学技术支出
	科技产出	每十万人人口专利授权数
结构要素	经济结构	城镇化率
	产业结构	第二产业占 GDP 比重、第三产业占 GDP 比重
制度要素	对外开放度	进出口总额占 GDP 比重
	政府干预度	地方财政支出占 GDP 比重
	外资依存度	外商直接投资额占 GDP 比重
市场驱动要素	市场规模	人均国内生产总值
	金融发展水平	金融机构各项存贷款总额占 GDP 比重
	生活应用水平	人均社会消费品零售额、移动电话用户数占比、互联网宽带用户数占比

(二) 数据说明

本文采用腾讯研究院公布的 2015—2018 年各地级市数字经济发展指数表示数字经济发展水平。除自然资源相关测度指标的基础数据来源于中国气象局官网,其他的人力资源、物质资本、科技创新、结构、制度和市场驱动等 6 个类别相关测度指标的基础数据均来源于 2016—2019 年《中国城市统计年鉴》。另外,几个自行计算的指标分述如下:年平均最低、最高气温,通过日最低、最高气温求年平均值得到;公路密度,每平方公里所拥有的公路总里程数表示;人均科学技术支出,通过研究与试验发展(R&D)经费内部支出除以总人口得到;城镇化率,用城镇人口在总人口中的比率表示。需要说明的是,中国目前有 293 个地级市,由于某些地级市相关指标的数据缺失,所以本文剔除 17 个地级市,保留 276 个地级市,再加上 4 个直辖市,最终选择 280 个城市。

(三) 方法选择与模型构建

随机森林是一种集成算法,是通过随机方式建立、包含多个决策树的分类器。其估计思想是每次以随机抽样的方式抽取样本数据,进行数据分类或回归从而形成一棵决策树,该决策树代表了样本数据中的自变量对因变量的一种估计。当人们对数据进行 N 次重复抽样,就会形成 N 棵决策树,从而形成样本数据中自变量对因变量的 N

种估计,再通过所有估计的平均值得到因变量的最后估计值。随机森林方法适用于多变量,可处理很高维度的数据以及不平衡的数据集,对整体模型的结果也具有很好的拟合效果、较高的精确度和泛化性能。本文中,选择的自变量达 24 个,变量较多且变量本身变化及变量之间的关系无规则,使用传统回归模型将会导致自由度下降、多重共线性等估计问题,所以选择随机森林中的回归树模型比较合适,而且随机森林在估计过程中会按照各要素对残差平方和减少的贡献大小自动对变量的重要性进行排序,这就有利于我们进一步分析不同投入要素对数字经济发展的影响程度。

回归树模型按照自变量要素的数量、种类和结构等特征进行数据分割,要素特征相似的地区被回归树归为同一组,然后这些地区数字经济指数的平均值将作为输出变量,即该组投入要素所对应的估计值。在这里,将第 i 个地区的观测数据记为 $(Y_i, E_{i1}, \dots, E_{ij})$, 其中, Y_i 代表输出变量, E_{ij} 代表第 i 个地区第 j 种要素。对于第 j 种要素 E_j , 若已知阈值 t , 数据将按照与阈值的大小关系进行分割,若 $E_j \leq t$, 则分割进入左节点,若 $E_j > t$, 则分割进入右节点,最后产生左右两个数据集 R_1 与 R_2 :

$$R_1(j, t) = \{Y_i, E_{i1}, \dots, E_{ij} \mid E_j \leq t\} \quad (12)$$

$$R_2(j, t) = \{Y_i, E_{i1}, \dots, E_{ij} \mid E_j > t\} \quad (13)$$

这样,每个数据集中分别包含 I_{R_1} 和 I_{R_2} 个样

本观测值,每个数据集都将获得输出变量均值:

$$\bar{Y}_{R_1} = \text{ave}(Y_i | Y_i \in R_1(j,t)) \quad (14)$$

$$\bar{Y}_{R_2} = \text{ave}(Y_i | Y_i \in R_2(j,t)) \quad (15)$$

若要素 j 与阈值 t 未知时,可按照残差平方和最小化的方式获得最佳的要素 j 与阈值 t :

$$\min_{\bar{Y}_{R_1}} \left[\min_{Y_i \in R_1(j,t)} (Y_i - \bar{Y}_{R_1})^2 + \min_{\bar{Y}_{R_2}} \sum_{Y_i \in R_2(j,t)} (Y_i - \bar{Y}_{R_2})^2 \right] \quad (16)$$

式(16)可实现对每个变量每个数据点逐一计算直到找到第一个最佳的要素和阈值分割点,数据被分裂成左集合 R_1 和右集合 R_2 。接着再从 R_1 中利用下式找寻最佳的第二个要素和阈值分割点:

$$\min_{\bar{Y}_{R_3}} \left[\min_{Y_i \in R_3(j,t)} (Y_i - \bar{Y}_{R_3})^2 + \min_{\bar{Y}_{R_4}} \sum_{Y_i \in R_4(j,t)} (Y_i - \bar{Y}_{R_4})^2 \right] \quad (17)$$

最佳的第二个要素和阈值分割点找到之后,数据再次被分割成 R_3 、 R_4 。以此类推,继续类似式(17)的最佳优化,对 R_2 、 R_3 、 R_4 进行数据分割,直到达到某种预设终止条件。例如:当要求树的深度达到8时,数据分割到8层后便结束运行。在节点分割的过程中,每次选取一个要素属性进行分割,即单个要素的阈值分割点类似树的分叉点,分割完成后的数据结构就形成一棵倒立的“决策树”。

以上随机森林方法研究了要素禀赋对数字经济的影响,为了进一步揭示随着投入要素增加而带来数字经济发展的动态变化,采用Friedman提出的方法来估计自变量对因变量的偏效应^①。假设共有 S 个自变量,要估计其中 m 个自变量 X_m 的偏效应,有以下定义:

$$f_m(X_m) = E_{x_k} f(X_m, X_k) \quad (18)$$

其中, X_m 为待估的自变量, R_k 为剩下的自变量, $m \cup k = \{1, 2, \dots, s\}$, $f(\cdot)$ 为随机森林估计的要素与数字经济均值的函数关系,具体到样本数据,上式可按如下方式获得估计:

$$f_m(X_m) = \frac{1}{n} \sum_{j=1}^n f(X_m, X_{jk}) \quad (19)$$

其中, n 为样本总数。通过上式可得到自变量与因变量的函数关系,并绘制出偏效应曲线图,可以更直观地看到随着投入要素的变化引起数字经济变化的情况。

四 经验分析结果

(一) 数字经济投入要素的重要性

运用随机森林估计方法,可以得到“要素禀赋决策树”,越接近树形最上面的要素对残差平方和减少贡献度越大,也就越重要。根据“要素禀赋决策树”,能够得到所有投入要素对数字经济发展重要程度的排列次序(见图1)。

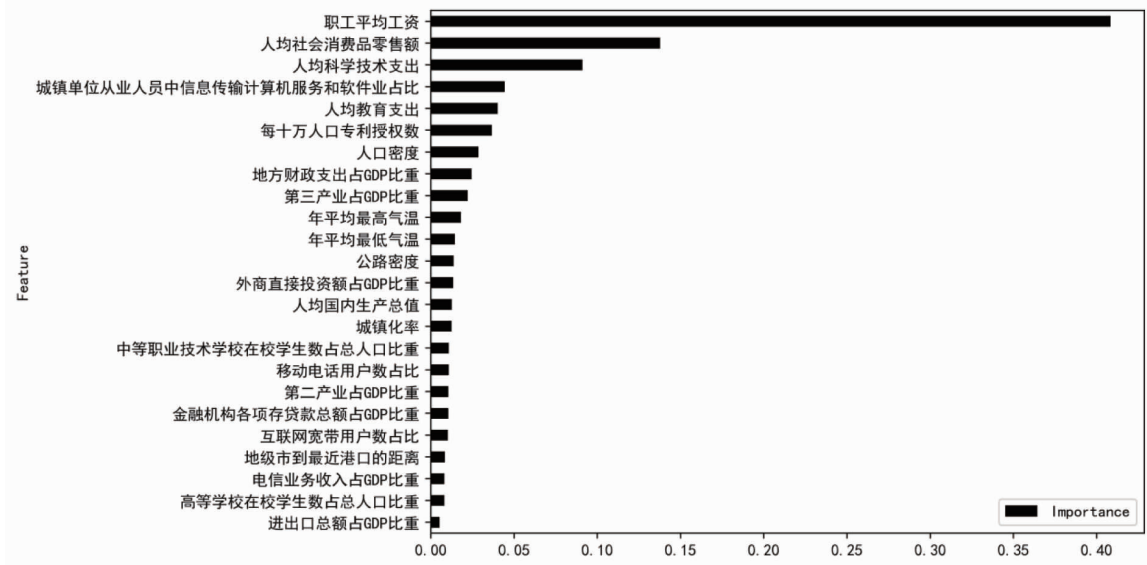


图1 数字经济投入要素的重要性

^①Friedman J.H. “Greedy Function Approximation: A Gradient Boosting Machine”, *Annals of Statistics*, 2001, 29(5):1189-1232.

由图 1 可知,正如我们前文指出,数字经济与传统制造业经济不同,它的发展更依赖“后天”的要素禀赋而非“先天”的要素禀赋。在重要性排序上,职工平均工资、人均科学技术支出、人口密度和地方财政支出占 GDP 比重等代表“后天”要素禀赋的投入要素排列次序比较靠前,而对传统经济有着决定意义的“先天”要素禀赋如到港口的距离、公路密度、平均气温等的重要性并不突出,甚至排列次序比较靠后。

影响数字经济发展的要素重要性虽然各不相同,但对数字经济发展影响最重要的 4 种投入要素分别为:职工平均工资、人均社会消费品零售额、人均科学技术支出、城镇单位从业人员中信息传输计算机服务和软件业从业人员占比。其中,职工平均工资代表的人力资本水平重要性超过了 0.4,成为影响地区数字经济发展的最关键要素,其重要程度位列首位,起主导作用;人均社会消费品零售额和人均科学技术支出的重要性处于 0.09—0.15 之间,对数字经济发展的影响力也较强,分别位列第二和第三;城镇单位从业人员中信息传输计算机服务和软件业从业人员占比的重要性接近 0.05,相比后面的 20 种投入要素而言也更重要一些。由于这 4 种要素的重要性合计占到了 0.68,也就是在影响数字经济发展的程度上达到了 68%,因此,接下来我们重点分析这 4 种投入要素对地区数字经济发展的影响。

(二) 重要要素对数字经济发展的偏效应

1. 职工平均工资的偏效应

图 2 是由职工平均工资所代表的人力资本水平对数字经济发展的偏效应曲线图,可以发现两者之间呈现非线性的正相关关系。总体上,随着职工平均工资的增加,数字经济边际产出也随之上升。数字经济的发展离不开高水平的人力资本驱动,其原因可以从供需两个方面加以理解:一方面,优质劳动力更容易支撑数字产业中硬件与软件的设计、研发与运行,促进数字化相关产业发展,提升数字化发展水平;另一方面,人力资本水平越高的地区,人们的需求层次也相对较高,会对更高的数字化技术应用和数字化服务水平形成更大的市场需求。所以,人力资本水平越高越利于数字经济的发展。

由图 2 可知,职工平均工资的偏效应曲线分别在 5.5 万元以及 7.75 万元左右形成拐点,可分别命名为“拐点 1”和“拐点 2”,当地区职工平均

工资跨过第一个拐点即 5.5 万元时,继续增长会带来较大的数字经济边际产出,说明此时该地区职工平均工资可以支撑数字经济较快增长;而当地区职工平均工资跨过第二个拐点即 7.75 万元时,继续增长会带来更大的数字经济边际产出,说明此时该地区职工平均工资可以支撑数字经济高速增长。同时,在 7.75 万元—8.25 万元区间内有一个斜率最大点,说明在 7.75 万元—8.25 万元作为职工平均工资支撑数字经济发展的“最优组合区间”,存在职工平均工资支撑数字经济发展的最佳组合点。另外,根据图 2,我国 280 个城市中,有 180 个城市已经越过了“拐点 1”,有 71 个城市已经越过了“拐点 2”,且有 47 个城市已经越过要素“最优组合区间”,说明我国大多数城市职工平均工资已经可以支撑数字经济加快发展。

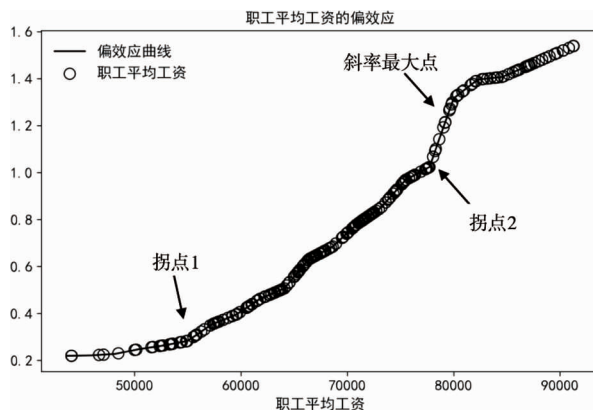


图 2 职工平均工资对数字经济发展的偏效应

2. 人均社会消费品零售额的偏效应

图 3 是由人均社会消费品零售额所代表的社会消费水平对数字经济发展的偏效应曲线图,可以发现两者之间呈现非线性的正相关关系。总体而言,随着人均社会消费品零售额的增加,数字经济边际产出起初缓慢上升,之后快速上升,最后呈“指数型增长”,前者对后者起着积极的推动作用。其原因可能在于数字经济的发展在一定程度上需要“消费红利”推动,可以从以下两个方面加以理解:一方面,消费水平提升刺激了“新经济”的增长,带动了地区新型基础设施建设投资,而数字基础设施又是新基建的关键内容和重要领域,所以消费水平的提升为更好地实现数字基建的规模化投资提供了良好支撑;另一方面,数字经济催生了新型消费方式,改变了以往消费者的消费习惯、偏好和行为,推动了消费升级,而众多消费主体所产生的消费数据又反过来促进数字经济往更

深更广的领域发展,因此消费升级对数字经济释放出了强大的“指数倍增效应”。

由图3可知,人均社会消费品零售额的偏效应曲线的“拐点1”和“拐点2”分别出现在3.5万元和6万元左右,说明当地区人均社会消费品零售额在超过3.5万元和6万元时,可分别支撑数字经济较快速增长与高速增长。同时,在6万元以后有一个斜率最大点,说明在6万元以后存在人均社会消费品零售额支撑数字经济发展的最佳组合点。另外,根据图3,我国280个城市中,只有58个城市已经越过了“拐点1”,另有23个城市已经越过了“拐点2”,而有222个城市则位于“拐点1”的左边,说明我国大多数城市人均社会消费品零售额还不足以支撑数字经济快速发展。

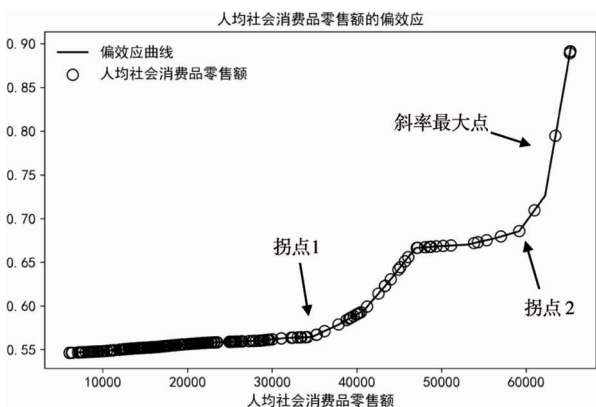


图3 人均社会消费品零售额对数字经济发展的偏效应

3.人均科学技术支出的偏效应

图4是人均科学技术支出对数字经济发展的偏效应曲线图,可以发现两者之间呈现非线性的正相关关系。总体而言,随着人均科技支出的增加,数字经济边际产出随之上升。其原因可能在于,数字经济作为创新引领和驱动转型的先导力量,它的发展离不开科学技术的支撑。一个地区的研究经费投入越多,越有利于激发科技创新活力和提高创新活动效率,为加快实现地区数字产业化和产业数字化转型发展创造条件,并进一步吸引更多的数字经济企业入驻,形成数字经济产业的良性循环,产生“科技创新—数字经济发展”的“螺旋式上升效应”。

由图4可知,人均科学技术支出的偏效应曲线的“拐点1”和“拐点2”分别出现在2000元和3400元左右,说明人均科学技术支出在超过2000元和3400元时,可以分别支撑数字经济较快速增长与高速增长。同时,在3400元—3700元

区间内有一个斜率最大点,说明在3400元—3700元“最优组合区间”,存在人均科学技术支出支撑数字经济发展的最佳组合点。另外,根据图4,我国280个城市中,只有46个城市已经越过了“拐点1”,另有23个城市已经越过了“拐点2”,而有234个城市位于“拐点1”的左边,说明我国大多数城市人均科学技术支出还不足以支撑数字经济快速发展。

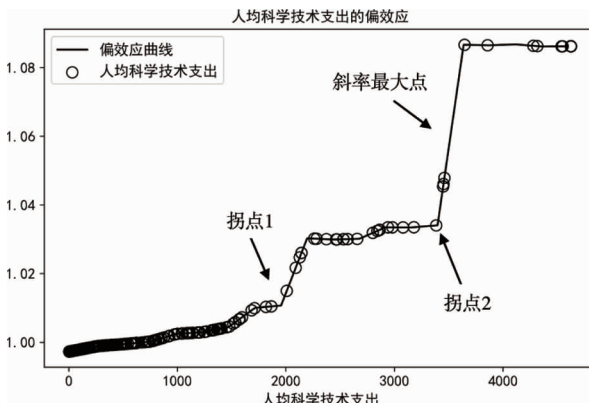


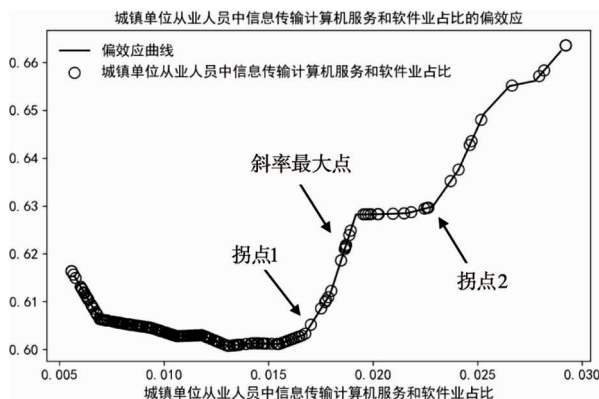
图4 人均科学技术支出对数字经济发展的偏效应

4.城镇单位从业人员中信息传输计算机服务和软件业占比的偏效应

图5是城镇单位从业人员中信息传输计算机服务和软件业占比对数字经济发展的偏效应曲线图,可以发现两者之间呈现左低右高的非对称“U型”关系。总体而言,随着信息传输计算机服务和软件业从业人员占比的增加,数字经济边际产出起初下降,但是,当占比达到一定大小时,数字经济边际产出随之上升。这种影响可能是因为该要素的滞后效应,从信息传输计算机服务和软件业从业人员占比增加到数字经济加速增长存在一个时间差,导致人员占比未能提升当期数字经济的边际产出,但随着该要素的不断增加并超过一定水平后,其对数字经济边际产出将产生正效应。

由图5可知,城镇单位从业人员中信息传输计算机服务和软件业占比的偏效应曲线的“拐点1”和“拐点2”分别出现在0.016和0.023左右,说明城镇单位从业人员中信息传输计算机服务和软件业占比在超过0.016和0.023时,可以分别支撑数字经济较快速增长与高速增长。同时,在0.016—0.019区间内有一个斜率最大点,说明在0.016—0.019“最优组合区间”,存在城镇单位从业人员中信息传输计算机服务和软件业占比支撑数字经济发展的最佳组合点。

根据图 5,我国 280 个城市中,只有 57 个城市已经越过了“拐点 1”,另有 26 个城市已经越过了“拐点 2”,而有 223 个城市位于“拐点 1”的左边,说明我国大多数城市城镇单位从业人员中信息传输计算机服务和软件业占比还不足以支撑数字经济快速发展。



(三) 关于中西部地区实证结果的进一步分析

为探讨中西部地区目前的要素禀赋是否能够支撑数字经济的可持续发展,可以进一步实证分析各城市最重要投入要素处于偏效应曲线的位置。根据前文,最终统计样本共有 280 个城市,其中 191 个为中西部城市。在 191 个中西部城市中,可考虑两种情况:一是位列中国 70 个大中城市名单的共有 39 个城市,其余 152 个城市为中小城市;二是省会城市 20 个,非省会城市 171 个。根据这些城市在 4 个偏效应曲线图的位置,可以得出不同城市要素禀赋对数字经济发展的支撑情况。

第一,与东部地区相比,中西部地区站到了与东部地区数字经济发展的“同一起跑线上”。一方面,依据最为重要的投入要素职工平均工资的偏效应曲线图(图 2),280 个城市中,有 25 个城市位于 5.5 万元“拐点 1”的左边,其中 18 个中西部城市,7 个东部城市;有 71 个城市已经越过了 7.75 万元的“拐点 2”,其中 37 个中西部城市,34 个东部城市。另一方面,综合观察 4 个偏效应曲线图,280 个城市中,4 种最重要投入要素均低于其对应要素“拐点 1”的城市共有 15 个,其中 11 个中西部城市,东部仍然有 4 个城市(汕头市、鞍山市、营口市、铁岭市);4 种最重要投入要素均超

过对应要素“拐点 1”的城市共有 21 个,其中 14 个东部城市,7 个中西部城市。以上充分说明,从数字经济发展最重要的 4 种要素而言,虽然中西部地区与东部地区获得的支撑仍有一定差距,但相对来说差距不大,基本位于“同一起跑线上”。

第二,从总体上看,大多数中西部城市的数字经济可持续发展已经具备了投入要素支撑的基本条件。一方面,依据最为重要的投入要素职工平均工资的偏效应曲线图,191 个中西部城市中,只有 18 个城市低于 5.5 万元的“拐点 1”,有 173 个城市(超过 90%)职工平均工资已经超过了 5.5 万元,说明这些城市的职工平均工资已经能够较好地支撑数字经济较快发展。尤其是,有 40 个城市职工平均工资已经超过 7.75 万元的“拐点 2”,说明这些城市工资水平已经可以支撑数字经济高速增长。另一方面,综合观察 4 个偏效应曲线图,191 个中西部城市中,4 种最重要投入要素均低于其对应要素“拐点 1”的城市只有 11 个;4 种最重要投入要素均超过对应要素“拐点 1”的城市达到 7 个。说明对多数中西部城市来说,虽然大部分城市在投入要素上还存在着一些“短板”,但还是具备了支撑数字经济可持续发展的基本条件。

第三,分情况来看,中西部城市之间的差距还较为明显,地区差距有可能会成为中西部地区通过大力发展数字经济来缩小与东部地区经济差距的“障碍”。一方面,依据最为重要的投入要素职工平均工资的偏效应曲线图,191 个中西部城市中,18 个未能越过“拐点 1”的城市均为非省会城市同时也是中小城市;在 133 个职工平均工资处于 5.5 万元—7.75 万元区间的城市中,只有 20 个大中城市,省会城市只有石家庄、哈尔滨和呼和浩特 3 市,中小城市多达 126 个。说明大中城市和中小城市、省会城市和非省会城市之间,职工平均工资存在差距,导致这一最为重要的投入要素对数字经济发展的支撑作用,大中城市明显优于中小城市、省会城市明显优于非省会城市。另一方面,综合观察 4 个偏效应曲线图,191 个中西部城市中,4 种最重要投入要素均低于其对应要素“拐点 1”的 11 个城市均为中小城市和非省会城市;4 种最重要投入要素均超过对应要素“拐点 1”的 7 个城市均为大城市和省会城市,没有中小城市和非省会城市。说明大中城市和中小城市、省会城市和非省会城市之间数字经济发展的要素条件还存在明显差距,相比较而言,非省会城市尤

其是中小城市的要素“短板”更为突出。

五 结论与政策建议

(一) 研究结论

本文运用机器学习方法中的随机森林算法,基于地级市层面数据,在24种投入要素中找出对数字经济发展起至关重要作用的4种投入要素,运用它们对数字经济发展的偏效应曲线图,揭示了重要投入要素对数字经济尤其是中西部地区数字经济可持续发展的影响。

第一,影响数字经济发展的投入要素与传统制造业存在明显差异。尽管影响数字经济发展的投入要素多样且错综复杂,但与传统制造业更依赖地理位置和自然要素等“先天”的要素禀赋不同,数字经济发展更依赖人口、经济规模、科技等“后天”的要素禀赋。表现在,在投入要素的影响中,如城市到最近港口的距离、年平均气温等影响传统制造业的自然要素禀赋,在数字经济发展中已经居于后列,而职工平均工资、信息计算机相关就业人员、科技、政府干预度等,已经成为影响地区数字经济发展的更重要因素。利用机器学习中的随机森林方法得到所有要素对数字经济发展贡献的排列次序,最重要的前4种要素为:职工平均工资、人均社会消费品零售额、人均科学技术支出、城镇单位从业人员中信息传输计算机服务和软件业从业人员占比。

第二,数字经济的发展给中西部地区带来了很好的发展机遇。一方面,由于数字经济发展不再像传统制造业那样强调地理区位优势,使得中西部与东部站在“同一起跑线上”成为可能,无论是基于数字经济发展中最重要的4种投入要素的分析,还是近年来中西部地区数字经济发展增速对比的现实状况,都表明中西部完全可以通过大力发展数字经济重塑比较优势,进而缩小与东部的差距。另一方面,针对西部地区数字经济发展中最重要的4种投入要素的分析表明,大多数中西部城市的数字经济可持续发展已经具备了投入要素支撑的基本条件,说明中西部地区完全可以“撸起袖子加油干”,加快推进数字经济的发展。

第三,中西部地区数字经济可持续发展仍然面临“瓶颈”挑战。一方面,数字经济的可持续发展离不开与工农业的融合发展,虽然在投入要素支撑上,中西部与东部站在了“同一起跑线上”,

但制造业发展的差距仍然可能对中西部数字经济的可持续发展带来负面影响,如何在发展过程中构筑自身优势,形成和东部地区的“错位竞合”格局,是中西部推进数字经济发展需要重点考虑的。另一方面,在投入要素支撑上,中西部地区大中城市和中小城市、省会城市和非省会城市之间的差距还较为明显,可能会成为其数字经济可持续发展的“障碍”,如何缩小内部差距,实现区域内数字经济平衡发展,也是中西部推进数字经济发展需要重点考虑的。

(二) 政策建议

数字经济对中西部地区来说,既是发展机遇也存在诸多挑战,根据中西部要素禀赋特点,可以找到其数字经济可持续发展的着力点与突破口。

第一,聚力数字产业化发展,形成与东部地区协同发展格局。中西部发展数字经济的投入要素支撑条件已经与东部地区较为接近,但制造业发展的差距仍然可能制约地区间数字经济差距的缩小,因此中西部数字经济可持续发展需要打造自身比较优势。一是在政策上,要抢先通过增加数字科研专项支出、大力培养数字经济相关人才、营造数字经济发展的良好创新环境等抢占“政策先机”。二是在技术研发上,要重点突破大数据、信息传输、云计算、人工智能等核心关键技术,为大力培育平台型企业推进数字产业化打下坚实基础。三是在产业数字化方面,在重视制造业数字化的同时,重点挖掘农业数字化与服务数字化空间,形成与东部地区“错位”发展格局。

第二,聚力投入要素补“短板”,对不同城市实施差异化政策。在4种重要投入要素支撑上,191个中西部城市只有7个均超过对应要素“拐点1”,说明其余184个城市存在要素支撑的“短板”,但不同的城市“短板”各不相同:有的城市的“短板”可能只有1种要素,而有的城市的“短板”可能有3种甚至4种要素;有的城市的“短板”可能离对应要素“拐点1”较近,很容易就能补上,而有的城市的“短板”可能离对应要素“拐点1”较远,要花大力气才能补上。针对不同的情况,需要出台不同的政策举措,因地制宜补“短板”。例如:云南省的昆明市,虽然还有人均科学技术支出未超过“拐点1”,但是距离“拐点1”很近,因此在政策上可以适宜实施以增加科技投入为主、其他要素投入同步推进的政策;四川省的绵阳市,虽然职工平均工资、人均科学技术支出、城镇单位从业

人员中信息传输计算机服务和软件业从业人员占比都已经越过“拐点 1”,但人均社会消费品零售额却未能越过“拐点 1”且离“拐点 1”还比较远,因此在政策上更应该多关注如何提高人们的消费水平,在补好这一“短板”后再考虑各要素的同步推进问题;安徽省的淮北市,除了职工平均工资越过“拐点 1”,其他三种要素都未超过“拐点”,因此有必要在提高其社会消费水平、增加科研经费投入、培养信息化人才等方面全面发力。

第三,聚力顶层设计,打造区域数字经济“一

盘棋”。中西部地区投入城市之间要素支撑上的差距,必然会导致数字经济发展的差距,在这种情况下,需要在更高层面进行统筹布局。建议细化国家相关举措,形成中西部地区数字经济统筹发展的顶层设计,重点突出几个方面:一是合理谋划中西部数字经济发展的技术创新路径与产业发展路径;二是推进优化中西部数字资源空间配置;三是出台扶持中西部数字经济欠发达地区的政策措施,尤其是如何帮这些地区补齐要素“短板”,争得发展数字经济的主动权。

Factor Endowment and the Sustainable Development of Digital Economy in the Middle and West of China: Based on Machine Learning Methods

ZENG Xiang-yan, LI Jiao & ZENG Xiao-ming

(School of Business, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: Under the background of increasingly fierce digital economy investment “competition” between regions, and the middle and west of China catching up with the east of China in growth rate, whether the factor endowment of the middle and west of China can support the sustainable development of digital economy is worth further study. The random forest algorithm of machine learning methods are used to analyze the 24 kinds of input factors based on the data of prefecture-level cities. This paper finds that the four kinds of the most important input factors supporting the development of digital economy are the average salary of employees, retail sales of consumer goods per capita, expenditure of science and technology per capita, the proportion of information transmission, computer service and software industry employees in urban units. By further analyzing the skew effect curves of the four kinds of the most important input factors, we find that in terms of factor endowment, the development of digital economy in the middle and west of China has reached the “same starting line” with that in the east of China, and most of the important input factors of cities have the basic conditions to support the sustainable development of digital economy, but the gap between cities is still obvious. Therefore, to realize the sustainable development of digital economy in the middle and west of China, China should implement differentiated policy of “shoring up weak links” according to the composition of factor endowment of different cities, and change the unbalanced development of digital economy among cities by promoting the balanced distribution of digital resources., so as to form a benign cycle of the development of digital economy in the middle and west of China.

Key words: factor endowment ; digital economy ; sustainable development ; machine learning

(责任校对 朱春花)