

【专题:高质量发展研究】

# 中国经济增长:数据要素的“双维驱动”

杨艳 王理 李雨佳 廖祖君

**【摘要】**数据要素已成为5大基础性生产要素之一,但其对经济增长的作用需进一步明晰。本文从人工智能技术入手,建立数据要素影响经济增长的内生增长模型,构建数据要素的产生路径和价值路径,并据此估算1999-2018年我国省级层面的新增数据要素价值。在理论分析和实证检验的基础上,明确了数据要素对我国经济增长具有“双维驱动”作用,即数据要素既能通过自身的经济增长效应直接驱动经济增长,也能通过促进技术进步间接驱动经济增长。进一步,本文进行三项扩展性讨论,结果表明,地方政府大数据交易平台能够提升数据要素的规模价值进而促进经济增长;在省级层面,数据要素可以通过降低经济政策的不确定性以促进经济增长;此外,1999-2018年数据要素的产出弹性呈“倒U型”演化趋势。

**【关键词】**数据要素;经济增长;人工智能;经济增长效应

**【作者简介】**杨艳,四川大学经济学院教授,系主任,博士生导师,研究方向为国民经济学;王理(通讯作者),四川大学经济学院博士研究生,四川省社会科学院大数据科学与创新发展研究院执行院长助理,研究方向为宏观经济与区域经济学,E-mail:447161772@qq.com;李雨佳,四川大学经济学院助理研究员,研究方向为宏观经济学;廖祖君,四川省社会科学院区域经济研究所研究员,所长,四川农业大学经济学院博士生导师,四川省社会科学院大数据科学与创新发展研究院执行院长,研究方向为农业经济与区域经济学。

**【原文出处】**《统计研究》(京),2023.4.3~18

**【基金项目】**2022年度国家社会科学基金项目“‘市场+政府+社群’协同治理框架下数据要素高质量供给的系统规制研究”(22BJL033);四川省社会科学研究“十四五”规划2021年度课题重大项目“资源要素市场化配置下城乡融合发展的‘成渝方案’研究”(SC21ZDCY007);四川省社会科学研究“十四五”规划2021年度课题“要素市场化配置视角下四川省政府数据与企业数据的开放、融合及高效应用研究”(SC21TJ007)。

## 一、引言

生产要素的核心作用在于促进经济增长。2020年4月《关于构建更加完善的要素市场化配置体制机制的意见》将数据与土地、劳动力、资本、技术并列列为5大基础性生产要素。现阶段,土地、劳动力、实物资本和技术进步对经济增长的正面作用已被大量文献证实,然而学界关于数据要素对经济增长的影响,数据要素是否具有经济增长效应等问题尚未有明确结论。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》对建立健全数据要素市场规则和国家公共数据资源体系做出部署。《“十四五”数字经济发展规划》明确提出数据要素是数字经济深化发展的核心

引擎。此外,得益于新一代信息技术的迅猛发展,近年来数据被采集、转换为数字格式、存储和分析以提取新知识(新数据)的效率迅速提升,数据体量及价值均呈现爆炸式增长。目前,我国数据市场交易量位居世界前五位<sup>①</sup>。鉴于以上背景,加快认清数据要素对经济增长的影响,有利于我国更好地通过数据红利推动经济增长和数字经济高质量发展。

国外经济学家率先对数据如何影响经济增长展开理论研究,一般认为数据主要通过促进知识和想法的产生与积累来间接助推经济增长。Nielsen(2011)指出大数据及其相关技术正在改变科学知识发现的机制,强调数据驱动智能(Data-driven Intelligence)可以提升技术从而促进重要知识和数据

的获取。Agrawal 等(2018)认为大数据背景下的经济增长依赖于计算机发现和处理数据的“元技术”(Meta Technologies)提升,该类技术的关键作用是产生新知识。Jones 和 Tonetti(2020)建立了一个数据经济学的理论框架,认为想法和数据都是信息,想法是生产商品的指令,而数据则是企业生产或创造想法中的投入要素,其作用是提升想法质量且对经济增长只有水平效应。Farboodi 和 Veldkamp(2021)的进一步补充研究表明,数据与有效信息一样具有减少不确定性的作用,但不确定性下限为零将注定数据难以单独维持长期经济增长,仍需通过促进想法积累来驱动经济增长。Cong 等(2021)认为数据是产生新知识的关键投入且具有动态非竞争性,其构建的内生增长模型表明,中间厂商使用数据进行创新可助力最终商品生产。

国内关于数据与经济增长关系的研究虽起步较晚,但我国首次提出数据要素的概念并将其纳入到基础性生产要素中,助推了国内学者关于数据要素影响经济增长的研究。徐翔和赵墨非(2020)构建数据资本的经济增长模型<sup>②</sup>,通过数值模拟发现数据资本不仅可以通过提升企业生产要素配置效率间接促进技术进步,还能直接作为生产要素推动经济增长。蔡跃洲和马文君(2021)指出数据要素作为重要的生产要素,其低成本、大规模可得、非竞争性等特征可以提升经济的微观运行效率,进而促进经济增长。王谦和付晓东(2021)认为数据要素可以凭借其经济技术特征,通过要素驱动、协同提升、促进结构优化等推动生产、组织、交易效率提升,赋能经济实现增长。杨艳等(2021a)立足数据要素的流动环境视角,发现数据要素不仅可以提升劳动力、资本等要素的生产率,对实际人均产出也具有显著正向影响。区别于国外研究,上述学者均认为数据要素作为基础性生产要素,可直接推动经济增长。

从现有研究可以看出,数据要素可能会通过“双维驱动”来影响经济增长:一维是数据要素直接驱动经济增长,另一维则是通过促进新知识和想法的产生与积累,即促进技术进步来间接驱动经济增长。若要实证检验数据要素双维驱动的存在,一是需要获取数据要素的“数据”,二是要找到与数据要

素间接驱动紧密相关的“技术”。现阶段,人工智能技术对经济增长的促进最能体现数据要素的间接驱动,理由如下:第一,人工智能需要不断依托数据实现自我积累和更新迭代(Lu,2021),数据要素是人工智能发展的绝对动力(Jones 和 Tonetti,2020);第二,人工智能技术的进步可以帮助社会更有效地获取和使用数据要素(Veldkamp 和 Chung,2019; Farboodi 和 Veldkamp,2021);第三,诸如物质资本、普通劳动力等传统生产要素可能不再对经济增长有显著性推动,而人工智能却是决定未来经济增长的重要因素之一(Aghion 等,2019;Lu,2021)。

鉴于此,本文从数据要素的直接驱动和促进人工智能技术进步的间接驱动入手,综合考虑数据要素对我国经济增长的双维驱动影响:首先,设定基本的理论分析框架,构建一个包含数据要素和人工智能积累的内生增长模型,理论上描述并推导数据要素对经济增长的直接驱动和间接驱动影响;其次,借助 Statistics Canada(2019a,2019b)、徐翔和赵墨非(2020)的相关研究,构建数据要素的产生路径和价值路径,据此并结合我国实际初步估算 1999-2018 年省级新增数据要素价值来表征新增数据要素的规模,以充当本文后续实证分析中数据要素的“数据”;最后,结合理论推导结果和数据要素的“数据”,实证检验了数据要素能够通过“双维驱动”促进经济增长,由此证明数据要素同样具有经济增长效应,进一步说明数据要素同土地、劳动力、资本和技术一样,均为基础性生产要素。

本文的主要贡献体现在以下三方面。一是理论层面,从数据来源于居民消费的基本假设入手,将数据要素同时纳入到人工智能技术积累方程和企业生产函数,推导了数据要素对经济增长的双维驱动影响及其机制,从理论层面证明了数据要素兼具“水平效应”和“增长效应”。二是在数据要素价值测算方面,本文参考现有文献并结合我国社会经济现实环境,构建出数据要素的产生路径和价值路径,据此估算了省级层面的年度新增数据要素价值,以期为本文后续及其他学者开展相关研究提供实践参考。三是在现实意义上,证实了数据要素对经济增长具有明显的促进作用,这为我国将数据与土地、劳动力、资本和技术并列为基础性生产要素

的合理性提供了有力支撑,也有助于各级政府进一步加强对数据要素的重视和运用。

## 二、理论分析

### (一) 数据要素产生与使用的环境设定

遵循“数据是经济活动的副产品”(Jones 和 Tonetti, 2020; Farboodi 和 Veldkamp, 2021; Cong 等, 2021)以及我国对“数据”赋予的新型生产要素内涵,本文将数据要素产生、数据要素促进创新与人工智能技术积累(进步)等问题转化为宏观经济模型来表征经济增长。为简化分析,假设经济体由大量的代表性家庭和代表性企业组成,企业归家庭所有,经济环境是确定的且家庭存续时间连续无限。

#### 1. 代表性家庭。

本文将数据要素设定为消费的副产品<sup>③</sup>:  $D = Nc$ 。现有国外文献对数据要素和数据的概念并未加以区分。然而,在社会经济活动产生的大量数据中,仅部分有价值的“有效数据”能够助力社会经济运行并符合数据要素的中国内涵。因此,本文与 Jones 和 Tonetti(2020)、Cong 等(2021)关于数据产生设定的区别是,1 单位的消费  $c$  能产生  $N \in [0, \infty)$  单位的数据要素  $D$ ,而不是产生 1 单位数据。家庭直接拥有数据要素的所有权,并且数据要素在每个时期使用后完全折旧<sup>④</sup>。

假定同质的代表性家庭拥有 1 单位时间,其中  $u$  用于学习, $v$  用于研发人工智能技术(AI),则  $1-u-v$  用于生产商品。家庭将  $s$  部分的物质资本投入到 AI 研发, $1-s$  用于生产商品。值得注意的是,由于数据的非竞争性,家庭可以将全部数据要素同时用于 AI 研发和商品生产。与 Lucas(1988)类似,家庭的人力资本通过学习进行积累。结合 Lu(2021)的观点,人工智能可以依靠深度学习实现自我强化和积累。此外, AI 研发不仅需要计算机(物质资本)和程序员(人力资本),更需要借助海量数据进行算法训练。因而,人力资本和 AI 的积累过程如下:

$$h = Buh \quad (1)$$

$$A = M(vh)^\phi (sk)^\theta D^\varphi A^{1-\phi-\theta-\varphi} \quad (2)$$

其中, $B, M > 0$  分别衡量了人力资本和 AI 的积累效率, $\phi, \theta, \varphi, 1-\phi-\theta-\varphi$  取值范围为  $(0, 1)$ 。 $k, h, A$  分别表示物质资本、人力资本和 AI,其初始存量分别

为  $k(0), h(0), A(0)$ 。代表性家庭的预算约束为:

$$\dot{k} = w(1-u-v) + r(1-s)k + r_A + r_D + \pi - c \quad (3)$$

其中, $\pi$  是企业利润, $w, r$  是工资率和租金率, $r_A, r_D$  分别是 AI 和数据要素的回报。代表性家庭的效用为:

$$U = \int_0^\infty e^{-\rho t} \frac{c(t)^{1-\sigma} - 1}{1-\sigma} dt \quad (4)$$

其中, $c(t)$  是家庭消费, $\rho > 0$  是家庭贴现因子, $\sigma$  的倒数是消费跨期替代弹性。

#### 2. 代表性企业。

与传统机器取代人类进行简单的体力劳动不同,依托数据要素不断学习和强化的 AI,可以逐渐代替人类面对越来越广泛的复杂现实场景,例如,自动驾驶算法、医学图像处理等。同时, AI 的非竞争性可以使其在每一时期同时用于研发和生产。在竞争性环境中,代表性企业通过雇佣劳动力并租用物质资本、AI 和数据要素来生产商品,其生产函数设定为:

$$y = F[(1-s)k]^{1-\alpha-\delta} \{(1-\alpha)[(1-u-v)h]^\beta + \alpha A^\beta\}^{\alpha/\beta} D^\delta \quad (5)$$

其中, $F > 0$  是企业生产效率, $\alpha, \delta, 1-\alpha-\delta \in (0, 1)$  分别是劳动力、数据要素和物资资本的份额。 $\{(1-\alpha)[(1-u-v)h]^\beta + \alpha A^\beta\}^{\alpha/\beta}$  表明 AI 和人类劳动可以相互替代, $\beta < 1$  是人力资本和 AI 的替代参数, $\alpha \in (0, 1)$  衡量 AI 相对于人力资本的重要性。在单一时间段  $t$  内,代表性企业的利润  $\pi$  为:

$$\pi = y - w(1-u-v)h - r(1-s)k - r_A A - r_D D \quad (6)$$

#### (二) 平衡增长路径上的“双维驱动”

平衡增长路径由数量路径  $\{c(t), v(t), u(t), y(t), k(t), h(t), A(t), D(t)\}_{t=0}^\infty$  和价格路径  $\{r(t), r_k(t), r_A(t), r_D(t)\}_{t=0}^\infty$  组成,使得家庭终生效用和企业利润最大化。同时,商品市场、资本市场、劳动力市场和数据市场均满足市场出清条件。由此可以得到经济增长率  $g_y$ <sup>⑤</sup>:

$$g_y = Bu = \psi \left(\frac{k}{h}\right) \times \left(\frac{D}{h}\right)^\delta + \chi \left(\frac{k}{h}, \frac{D}{h}\right) \times \left(\frac{A}{h}\right)^\beta \quad (7)$$

其中, $\psi \left(\frac{k}{h}\right) = \frac{1-\alpha}{\alpha} Fu\sigma (1-s)^{-\alpha-\delta} (1-\alpha) (1-u-v)^\beta$

$$\left(\frac{k}{h}\right)^{-\alpha-\delta} \chi \left(\frac{k}{h}, \frac{D}{h}\right) = \frac{1-\alpha}{a} Fu\sigma (1-s)^{-\alpha-\delta} \alpha \left(\frac{k}{h}\right)^{-\alpha-\delta}$$

$$\left(\frac{D}{h}\right)^\delta, \frac{k}{h} = \left\{ \frac{B^\varphi}{M} w^{-\phi} S^{-\theta} N^{-\varphi} \left(\frac{1-s}{1-\alpha-\delta}\right)^{-\varphi} \left[\frac{(1-\alpha)v}{\alpha\phi}\right] \right.$$

$$\left(\frac{1}{u} - 1 + \phi + \theta + \varphi\right) (1-u-v)^{\beta-1} \Big|_{\frac{\phi+\theta+\varphi}{\beta}}^{\frac{1}{\theta+\varphi}}$$

数据要素对经济增长具有“双维驱动”的影响在式(7)中得以体现:一方面,  $(D/h)^\delta$  的指数为正, 这意味着经济体中的数据要素越多, 经济增长速率会越快, 这反映了数据要素的直接驱动影响, 同时表明数据要素具有经济增长效应; 另一方面,  $(A/h)^\beta$  的指数为正且  $A$  会受到式(2)中  $D$  的正向影响, 这表明当经济体中的数据要素不断增加时, 人工智能技术进步越快, 经济增长速率也会越快, 这体现了数据要素的间接驱动影响。

### 三、数据要素的价值估算

#### (一) 数据要素的产生路径与价值路径

我国现行的《中国国民经济核算体系(2016)》并未建立关于数据要素价值的核算体系。数据可以被复制并且能够轻松地从一个地点或经济领土转移到另一个地点或经济领土, 给数据要素价值核算带来了较大困难。我国国民账户体系主要记录的是生产, 而其中与数据价值紧密相关的经济生产性概念, 例如“相关研究与开发”“矿藏勘探与评估”“计算机软件与数据库”等虽早已列入国民账户核算体系, 但却被归并到知识产权产品一类纳入核算。目前, 我国数据要素市场建设刚刚起步, 难以实现数据要素的市场定价。顶层核算框架以及数据要素市场化配置体制机制尚未完善, 使得数据要素价值的利益主体与合理分配无法被明确, 进而使得准确测算社会经济活动中的数据要素价值仍面临诸多现实挑战。

积累的数据是一种有价值的资产 (Farboodi 和 Veldkamp, 2021)。理论上, 通常用于评估任何资产价值的三种方法 (收入法、市场法和成本法) 也应适用于数据要素。具体而言, 当数据要素的净现金流收益能够被科学化量时, 收入法可以用来测度经济活动中数据要素的价值; 如果在数据要素市场进行数据交易的过程中, 其价值可以被清晰地记录, 则市场法是衡量数据要素价值的最佳方法; 当无法得知数据要素的所有者可以从数据中获得多少经济收益时, 也可使用成本法间接估算数据要素价值, 即将用于产生、收集、整理与使用数据要素相关的投入成本总和, 作为数据要素的价值表征。在市场中出售的数据, 其价值应是交易价格 (Statistics Can-

ada, 2019a), 但就我国当前现实而言, 除通过各类数据交易平台完成交易的少部分数据要素外, 社会中还有大量的数据要素未入市出售<sup>⑥</sup>, 这就使得收入法和市场法难以适用于数据要素的价值测算。数据是一种无形资产, 无形资产价值的测算一般采用成本法 (Veldkamp 和 Chung, 2019)。因此, 通过测算投入到产生、收集、整理与使用数据要素的总成本来间接估算, 是现阶段评估数据要素价值的可行方法。

估算数据要素的价值, 需厘清数据要素的产生路径和价值路径。在数据要素的产生路径中, 所有数据要素的逻辑起点是物理世界中的现实信息, 现实信息不由社会生产得到。当人类用笔、纸、电子设备等记录下被观察到的现实信息后, 现实信息就成了数据, 数据要素产生路径的起点便在于此。尽管一些现实信息本身具有价值 (比如苹果成熟掉落这一现实信息引发牛顿对万有引力的思考), 但其没有人类的投入成本, 所以数据要素价值路径的起点是数据化后的现实信息。数据要素是数据中有助于社会经济运行的部分, 但数据化后的现实信息含有大量对经济社会无用的或非结构化的数据<sup>⑦</sup>, 需经过数据整理才会成为数据要素。此外, 数据要素可以被用于进一步研究分析或交易, 该过程又会产生新的数据或知识, 这便是数据要素产生路径和价值路径的共同终点。随后, 新数据又会通过数据整理转化为新的数据要素, 循环往复。本文将数据要素的产生路径和价值路径总结为附图 1<sup>⑧</sup>。接下来, 本文将依托数据要素价值路径并借鉴 Statistics Canada (2019b) 关于数据价值的测算思想, 同时参考徐翔和赵墨非 (2020) 利用微观数据库估算数据资产存量价值的研究, 尝试对我国省级年度新增数据要素的价值进行初步估算<sup>⑨</sup>。

#### (二) 新增数据要素的价值估算及合理性检验

在数据要素的价值路径中, 若运用成本法估算数据要素的价值, 则应该包含投入到数据产生、收集、数据要素整理与使用等活动中的工资成本以及相关的非直接人工与其他成本<sup>⑩</sup>, 但从事这些活动的劳动者并不会将 100% 的工作时间用于产生数据价值。为此, 本文先以中国家庭追踪调查数据库中的职业行业代码为基础, 将与数据收集、数据要素

整理、数据要素使用等活动类型相关的职业进行筛选归类<sup>①</sup>；再对上述职业劳动者相对集中的国内 20 家机构进行线下或线上访谈调研<sup>②</sup>，以大致获取这些职业在现实中参与数据收集、数据要素整理、数据要素使用等活动所直接产生数据要素价值的日均有效工作时间（按一天 8 小时工作制计算，暂不考虑加班）以及相关的平均非直接工资和其他成本。尽管该做法存在一定误差，且访谈的样本不能代表所有年份的普遍状况，但受限于客观条件，本文采用调研访谈获取成本数据的做法，已然是对 Statistics Canada(2019b)中“非直接工资和其他成本占工资成本的 50%”“生产活动数据份额为 10% ~ 100%”等主观设定的改进，更加符合我国现实环境。此外，对最终估算价值的一系列合理性检验，也在一定程度上证实了本文做法的理论合理性。附表 1 展示了从事数据收集、数据要素整理、数据要素使用等活动相关职业的平均有效工作时间、非直接工资和其他成本占工资的平均比重，年度新增数据要素的价值估算公式可设定为：

$$Value\_Data_t = \sum_j \sum_i wage_{ij} \left( \frac{work\_time_j}{8} + other_j \right) \quad (8)$$

其中，下标  $i$  表示职业， $j$  表示数据活动类型， $t$  表示年份。 $Value\_Data_t$  表示  $t$  年新增数据要素的价值。 $wage_{ij}$  表示活动  $j$  中职业  $i$  在  $t$  年的工资。 $work\_time_j$  表示活动  $j$  中所有职业的平均每天有效工作时间， $other_j$  表示活动  $j$  中与所有职业相关的非直接工资和其他成本占对应职业工资比重的平均值<sup>③</sup>。

表 1 新增数据要素价值的合理性检验指标

指标类别	相关指标	时间跨度(年)	观测个数	最小值	最大值
个体类指标	移动电话基站(万个)	2013-2018	180	1.18	65.06
	信息技术服务收入(亿元)	2014-2018	150	0.36	6280.05
	有电子商务交易活动企业数(个)	2013-2018	180	41.00	12240.00
	规模以上工业企业研究与试验发展(R&D)经费(千万元)	2012-2018	210	65.03	21072.03
	移动互联网接入流量(百万 GB)	2014-2018	150	106.97	84584.12
	地理信息数据(幅)	2001-2011	330	49.00	264247.00
综合类指标	数字普惠金融指数	2011-2018	240	18.33	377.74
	数字经济综合发展指数	2011-2018	240	-1.03	5.80

根据年度新增数据要素的价值估算公式(8)，利用各地方政府统计年鉴、《中国劳动统计年鉴》、中国家庭追踪调查(CFPS)、中国劳动力动态调查(CLDS)、中国综合社会调查(CGSS)等相关数据，依据附表 1 指标对 1999-2018 年我国省级新增数据要素价值进行初步估算<sup>④</sup>。可以看出，在进入大数据元年(2013 年)<sup>⑤</sup>后，各个地区新增数据要素的价值均出现了显著攀升，这与我国大数据发展进程相吻合。

为进一步验证估算结果是否合理，本文采用两大类指标(表 1<sup>⑥</sup>)对估算的新增数据要素价值进行分类分时段合理性检验。第一类指标是个体类指标，其中前 4 个指标代表数据要素的规模体量，是数据要素规模价值高低的间接体现；后 2 个指标本身即为数据，与经济中数据要素的真实价值直接相关。第二类指标属于综合类指标，较第一类指标而言，可以在更大程度、更广范围上体现出数据要素的流动、分析应用与价值增值。其中，“数字普惠金融指数”来自北京大学数字金融研究中心发布的《北京大学数字普惠金融指数(2011-2020)》，“数字经济综合发展指数”则是参考赵涛等(2020)的做法，利用主成分分析法和标准化降维处理得到。所有变量指标在检验中均做标准化处理。从合理性检验的结果可以看到，新增数据要素估算价值与第一、二类指标均具有强正相关性<sup>⑦</sup>，表明通过构建数据要素价值路径估算出的 1999-2018 年我国省级新增数据要素价值，在理论与现实层面上具有一定的合理性，后续实证分析中本文会将年度新增数据要素价值作为数据要素的衡量指标。

## 四、实证策略、变量与数据

### (一) 研究思路与模型设定

本文研究目的是探究数据要素对经济增长的作用,数据要素既具有直接经济增长效应,也可以通过促进人工智能技术进步来间接促进经济增长。因此,采用“三步法”来直观展现数据要素的“双维驱动”作用与机制。

第一步,不考虑人工智能技术,直接考察数据要素对经济增长的影响,系数  $\alpha_1$  显著为正是验证数据要素具有直接驱动作用的前提。计量模型设定如下:

$$Gpgdp_{it} = \alpha_0 + \alpha_1 \ln Data_{it} + \alpha_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (9)$$

其中,下标  $i$  表示省份, $t$  表示年份; $\mu_i$  是省份固定效应,表示省份在样本期内不随时间变化的特征; $\lambda_t$  是年份固定效应,用来控制样本省份随年份变化的共同遗漏变量; $\varepsilon_{it}$  是随机扰动项。 $Gpgdp_{it}$  是以 1998 年为不变价所衡量的实际人均地区生产总值( $pgdp_{it}$ )增长率,用来度量本文的经济增长,计算公式为  $Gpgdp_{it} = \ln pgdp_{it} - \ln pgdp_{i,t-1}$ 。 $Data_{it}$  是新增的数据要素,用前文中估算的新增数据要素价值表示,是本文的核心解释变量之一。 $X_{it}$  是控制变量集,包括人力资本存量、人口规模、研发水平、市场化程度、城乡结构、资本要素、政府规模、对外贸易。

第二步,分析数据要素对人工智能技术进步的作用,借此验证本文理论分析中式(2)设定的合理性。计量模型设定如下:

$$\ln AI_{it} = \beta_0 + \beta_1 \ln Data_{it} + \beta_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (10)$$

其中, $AI_{it}$  表示省份  $i$  在  $t$  年产生的人工智能技术进步,是本文的另一个核心解释变量,系数  $\beta_1$  显著为正是验证数据要素具有间接驱动作用的前提。

第三步,在式(9)基础上引入人工智能技术进步,即:

$$Gpgdp_{it} = \gamma_0 + \gamma_1 \ln Data_{it} + \gamma_2 \ln AI_{it} + \gamma_3 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (11)$$

此时,由式(10)和式(11)可以识别出数据要素对经济增长的双维驱动影响。具体而言, $\gamma_1$  是数据要素对经济增长的直接驱动影响, $\beta_1 \times \gamma_2$  则在一定程度上可以体现数据要素对经济增长的间接驱动影响。

### (二) 人工智能技术进步的度量

人工智能技术的发展离不开其专利的保护,人工智能专利的开发更离不开企业的投资、运用和推广。鉴于此,从数据可得性和现有研究出发,本文使用年度人工智能专利申请量与人工智能企业注册量的交互项,来度量人工智能技术进步。首先,以人工智能、机器学习、机器(计算机)视觉、计算机识别、虚拟现实、神经网络、自然语言处理、机器人、生物识别等为关键词,在《中国专利全文数据库(知网版)》<sup>⑧</sup>中搜索 1999-2018 年的相关专利申请量,然后将其加总并匹配到各样本省份,由此得到省份  $i$  在  $t$  年的新申请人工智能技术专利数  $patent\_ai_{it}$ ;其次,为得到与专利类型一致的人工智能企业,本文仍基于上述关键词,通过“企查查”<sup>⑨</sup>搜索 1999-2018 年省份  $i$  在  $t$  年的新注册人工智能企业数  $firm\_ai_{it}$ ;最终,省份  $i$  在  $t$  年发生的人工智能技术进步  $AI_{it}$  表示为  $AI_{it} = patent\_ai_{it} \times firm\_ai_{it}$ 。

### (三) 控制变量与数据说明

模型的主要解释变量除  $Data_{it}$  和  $AI_{it}$  两个核心解释变量外,为缓解遗漏变量偏误,本文进一步加入可以影响经济增长的控制变量集:人力资本存量( $h$ ),参照彭国华(2005)计算得到省级层面的人力资本存量;人口规模( $peo$ ),用各省份年末人口总数表示;研发水平( $rd$ ),用当年高技术产业的从业人员年平均人数表示;市场化程度( $market$ ),用市场化指数表示<sup>⑩</sup>;城乡结构( $urb$ ),用当年的常住人口城镇化率表示;资本要素( $k$ ),用当年资本形成总额占地区生产总值(GDP)的比重来表示;政府规模( $gov$ ),用当年地方政府财政支出占 GDP 比重来测度;对外贸易( $trade$ ),用当年地方进出口总额占 GDP 比重来表示。控制变量的数据,除 1999-2016 年市场化指数外,所有数据均来自《中国统计年鉴》和各地政府统计年鉴。本文采用 1999-2018 年我国 30 个省份的面板数据作为研究样本<sup>⑪</sup>。附表 2 报告了各变量的描述性统计结果,除比例型变量外的连续型变量均取自然对数。

## 五、实证分析

### (一) 初步分析

首先采用年份固定效应和省份固定效应对模型(9)~(11)进行回归分析,结果见表 2。列(1)~

表 2 数据要素的“双维驱动”影响:初步估计

	<i>Gpgdp</i>		<i>ln AI</i>		<i>Gpgdp</i>	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>ln Data</i>	0.1630*** (0.0259)	0.0715*** (0.0193)	0.5910*** (0.1216)	0.5110*** (0.1245)	0.1180*** (0.0231)	0.0515* (0.0200)
<i>ln AI</i>					0.0759*** (0.0076)	0.0392*** (0.0076)
<i>ln h</i>		0.0624** (0.0201)		0.3690** (0.1249)		0.0479* (0.0198)
<i>ln peo</i>		-0.8240*** (0.0572)		-1.3850*** (0.3336)		-0.7690*** (0.0564)
<i>ln rd</i>		1.7990 (1.0919)		45.8300*** (5.1484)		0.0020 (1.1213)
<i>market</i>		0.0242*** (0.0062)		0.2220*** (0.0323)		0.0155** (0.0060)
<i>urb</i>		0.0043*** (0.0009)		0.0320*** (0.0056)		0.0031*** (0.0008)
<i>k</i>		0.1720*** (0.0344)		0.0243 (0.1843)		0.1710*** (0.0342)
<i>gov</i>		-0.6520*** (0.1541)		2.5830*** (0.6745)		-0.7530*** (0.1608)
<i>trade</i>		0.0222 (0.0241)		0.2580 (0.1330)		0.0121 (0.0234)
直接驱动影响	0.0515					
间接驱动影响	0.0200					
省份固定效应	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
观测值	600	600	600	600	600	600
Within R <sup>2</sup>	0.1110	0.5503	0.0597	0.3929	0.2438	0.5732

注:括号内是稳健标准误。\*、\*\*、\*\*\* 分别表示在 10%、5%、1% 统计水平上显著。下同。

(2) 是基于模型(9)的回归,估计结果显示,无论是否加入控制变量集,数据要素对我国经济增长都具有非常明显的促进作用,但未知该作用是来自人工智能技术进步的间接溢出还是数据要素作为基础性生产要素的直接推动。列(3)~(4)是基于模型(10)的估计结果,从其估计系数的显著性来看,数据要素的确可以促进人工智能技术进步,本文关于

人工智能技术创新边界(式(2))的设定合理。列(5)~(6)对应模型(11)的回归,从估计结果来看,数据要素对我国经济增长的双维驱动影响显著为正,具体而言,数据要素规模每增加 1%,可以直接驱动我国经济增长率提升 0.0515%;数据要素规模每增加 1%,可以促进人工智能技术进步 0.511%,进而间接驱动我国经济增长率提升 0.02%。

## (二) 内生性处理与“双维驱动”的重新识别

上一部分采用“三步法”的目的在于直观了解数据要素对经济增长的双维驱动机制,但数据要素与人工智能技术进步所采用的代理指标,基本都内生于社会经济运行,加之模型(9)~(11)的内生性、遗漏变量、乘积误差等问题,使得初步回归中的结论并不严谨。为此,本文借鉴 Beck 等(2010)的机制识别思想,构建如下实证模型并基于两阶段最小二乘法(2SLS)框架重新估计数据要素对经济增长的双维驱动影响:

$$Gpgdp_{it} = \xi_0 + \xi_1 \ln Data_{it} + \xi_2 \ln Data_{it} \times \ln AI_{it} + \xi_3 X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (12)$$

其中, $\xi_1$ 代表数据要素对经济增长的直接驱动影响, $\xi_2 \times \ln AI_{it}$ 代表数据要素通过促进人工智能技术进步对经济增长产生的间接驱动影响, $\overline{\ln AI_{it}}$ 是  $\ln AI_{it}$  的样本均值。尽管固定效应面板模型可适度缓解遗漏变量偏误,但逆向因果问题依然会对“双维驱动”识别带来干扰,仍需借助与内生变量( $\ln Data_{it}$ 和  $\ln Data_{it} \times \ln AI_{it}$ )相关,且又不直接影响被解释变量( $Gpgdp_{it}$ )的工具变量。为此,本文先找到数据要素( $\ln Data_{it}$ )和人工智能技术进步( $\ln AI_{it}$ )的工具变量,即国企单位人均工资( $wage$ )与其他省份人力资本总和( $exh$ ),然后参考 Goldsmith - Pinkham 等(2020)的方法,用交互项  $wage \times exh$  进一步作为  $\ln Data_{it} \times \ln AI_{it}$  的工具变量。

1. 数据要素的工具变量:国企单位人均工资( $wage$ )。

本文使用 1959-1978 年国企单位人均工资作为数据要素工具变量<sup>②</sup>的原因,第一,改革开放前,我国实行的是重工业优先发展战略,国有企业是这一战略实施的重要组织形式,得益于高度集中的计划经济体制与国家大力扶持,改革开放前 20 年的国企在整体上不会受到太大冲击,较为稳定的内外部环境可以保障国企长期、完整地记录自身工资数据,而在前文估算新增数据要素价值时,用到了国企单位的劳动力工资数据,因此国企单位工资与本文的数据要素价值紧密相关。第二,工资收入能够通过消费作用于经济增长,1959-1978 年间国企单位人均工资虽然可以对当年及以后一段时间的经济运行产生影响,但对 1999-2018 年间我国经济增

长影响甚微,满足工具变量的排他性要求。

2. 人工智能技术进步的工具有变量:其他省份人力资本总和( $exh$ )。

本文借鉴于文超和梁平汉(2019)的思想,用当年其他省份人力资本总和作为本省份人工智能技术进步的工具有变量。一方面,人力资本是推动人工智能技术产生和积累的重要因素,由于技术同知识一样具有非竞争性,本省份人工智能技术的创新除了受当地人力资本助推外,还会受到其他省份人力资本研发出的人工智能技术溢出影响。另一方面,能成功研发人工智能技术的人力资本一般属于高层次人才,在短期内实现跨省份完全自主流动存在一定难度,所以当年其他省份人力资本难以对本省份经济增长产生显著的直接驱动影响。因此,其他省份人力资本总和满足工具有变量的相关性和排他性条件。

表 3 报告了基于模型(12)的 2SLS 回归结果。第一阶段的 F 检验值远大于 10,估计系数的显著性表明,工具有变量  $wage$  和  $wage \times exh$  能够有效剥离出数据要素与人工智能技术进步中的外生信息,进而证明了工具有变量的有效性。第二阶段的回归结果显示,无论是否加入控制变量集,数据要素对我国经济增长双维驱动的影响在方向上和显著性上均与表 2 初步回归结果一致,与采用“三步法”的初步回归结果相比,直接驱动影响和间接驱动影响分别提升至 0.6805 与 0.0294,进一步验证了数据要素不仅能够通过促进技术进步来推动我国经济增长,而且数据要素作为基础性生产要素之一,具有经济增长效应。

## (三) 异质性与稳健性检验

### 1. 双维驱动影响的时空差异。

本部分在分区域与全样本两个层面上,考察数据要素双维驱动影响的时空异质性。分区域时,本文采用东、中、西部和东北地区的划分方法<sup>③</sup>,并将中、西部与东北地区合并(以下称中西部及东北地区);全样本的时间划分上,本文以大数据元年作为时间节点,分为 1999-2012 年和 2013-2018 年两个样本时间段进行检验。

表 4 列(1)~(4)报告了分区域的时空异质性检验结果。由列(1)~(2)可知,东部地区数据要素

表 3 数据要素的“双维驱动”影响:基于 2SLS 模型的工具变量估计

第一阶段回归	(1)		(2)	
	ln Data	ln Data×ln AI	ln Data	ln Data×ln AI
wage	0.0015*** (0.0002)		0.0005** (0.0002)	
wage×exh		-0.0767*** (0.0038)		-0.0629*** (0.0042)
第一阶段 F 值	549.7300		530.4800	
第二阶段回归	Gpgdp		Gpgdp	
ln Data	0.5372*** (0.0885)		0.6805** (0.3269)	
ln Data×ln AI	0.0044*** (0.0015)		0.0047* (0.0026)	
直接驱动影响	0.5372		0.6805	
间接驱动影响	0.0275		0.0294	
控制变量	否		是	
省份固定效应	是		是	
年份固定效应	是		是	
Within R <sup>2</sup>	0.9686		0.9756	

表 4 数据要素的“双维驱动”影响:时空差异

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	分区域的时空异质性				全样本的时间异质性			
	东部地区		中西部地区(含东北)		1999-2012 年		2013-2018 年	
	1999-2018 年		2003-2018 年					
	ln AI	Gpgdp	ln AI	Gpgdp	ln AI	Gpgdp	ln AI	Gpgdp
ln Data	0.6582** (0.1795)	0.1815*** (0.0269)	1.6117*** (0.2142)	0.0722** (0.0357)	0.2803** (0.1170)	0.0985*** (0.0225)	1.1781* (0.6535)	0.1424*** (0.0473)
ln AI		0.0713*** (0.0111)		0.0336*** (0.0091)		0.0554*** (0.0089)		0.0140* (0.0077)
直接驱动影响	0.1815		0.0722		0.0985		0.1424	
间接驱动影响	0.0469		0.0542		0.0155		0.0165	
控制变量	是	是	是	是	是	是	是	是
省份固定效应	是	是	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是	是	是
观测值	200	200	340	340	420	420	180	180
Within R <sup>2</sup>	0.4920	0.3645	0.0597	0.3929	0.1638	0.3843	0.0699	0.4621

规模每增加 1%,将直接和间接驱动该地区经济增长提升 0.1815% 和 0.0469%。由列(3)~(4)可知,中西部及东北地区的数据要素双维驱动影响在

2003 年前并不显著,在 2003 年之后,当中西部及东北地区的数据要素规模每增加 1% 时,将直接和间接驱动该地区的经济增长率提升 0.0722% 和

0.0542%。尽管在间接驱动影响上,中西部及东北地区比东部地区高0.0073个百分点,但在“双维驱动”的整体影响上,东部地区要明显高出中西部及东北地区0.102个百分点。列(5)~(8)报告了全样本的时间异质性检验,结果显示,数据要素对我国经济增长的双维驱动影响在大数据元年前后均显著存在,在进入大数据元年后(2013-2018年),数据要素的直接驱动影响和间接驱动影响都明显增加。其中,直接驱动影响较大数据元年前(1999-2012年)显著提升了44.6%。

### 2. 经济增长的新指标:夜间卫星灯光数据。

夜间灯光亮度可以反映社会经济的活跃程度。相较于传统GDP数据,夜间卫星灯光数据基本不受主观统计误差、市场价格波动等因素干扰,常常被用作经济增长的代理指标(Henderson等,2012)。为确保研究结论更加严谨,本文交替使用校正后的两类夜间卫星灯光数据(DMSP/OLS和NPP/VIIRS)<sup>29</sup>,通过分时间段对表2中列(2)、(4)、(6)的初步估计结果进行稳健性检验,灯光数据校正采用廖祖君和王理(2019)的做法。附表3<sup>30</sup>列(1)~(3)与列(4)~(6)分别汇报了基于DMSP/OLS与NPP/VIIRS两类夜间灯光数据的稳健性检验结果。其中,*Glight*表示土地面积上夜间人均灯光亮度的增长率,是实际人均GDP增长率的替代指标。从估计结果来看,在采用夜间平均灯光亮度增长率后,数据要素依然能够通过“双维驱动”来推动经济增长。

### 3. 人工智能的新指标:工业机器人安装密度。

本文参考闫雪凌等(2020)的做法,通过构建省级工业机器人安装率(*Robort*)作为人工智能技术的替代指标,以进一步验证数据要素的双维驱动影响是否稳健。省级层面的工业机器人安装率测算方式如下:

$$Robort_{it} = \sum_{j=1}^N \frac{Rob\_ind_{jt}}{L_{jt}} \times \frac{L_{ijt}}{L_{it}} \quad (13)$$

其中,下标*i*表示省份,*j*表示制造业行业,*t*表示年份;*Rob\_ind<sub>jt</sub>*是全国*j*行业的工业机器人存量,*L<sub>jt</sub>*是*j*行业的全国就业总人数;*L<sub>ijt</sub>*是*i*省*j*行业的就业人数,*L<sub>it</sub>*是*i*省的就业总人数。我国行业机器人存量的原始数据来源于国际机器人联盟(International Federation of Robotics, IFR),分地区分行业的就业

人员数据来源于《中国劳动统计年鉴》。由于国际机器人联盟自2006年起才开始记录我国分行业工业机器人的存量数据,因此本文仅测算2006-2018年我国省级工业机器人安装率。附表4<sup>31</sup>报告使用工业机器人安装率作为衡量人工智能技术新指标后的估计结果。从关键变量的估计系数显著性来看,数据要素对经济增长具有双维驱动影响的结论依然稳健。

## 六、扩展性讨论

本部分将进行三项扩展性讨论,即在地方政府大数据交易平台的建设、数据要素的信息本质、数据要素的产出弹性这三个研究环境中,聚焦数据要素对经济增长的整体影响。本部分的结果是对前文内容的补充与扩展,也是对后续研究的探索与延伸。

### (一) 政府大数据交易平台的影响

随着北京国际大数据交易所和上海数据交易所于2021年相继成立,建设大数据交易平台已成为近年来地方政府加快探索和推进数据要素市场化配置的常见举措之一。依托具有政府背景的地方大数据交易平台既可以保障数据要素供需双方的切身利益,也能较好满足数据经纪人利用平台开发数据产品,以更好挖掘数据要素价值的诉求。结合数据交易平台的设计理念,地方政府建设的大数据交易平台应该具有促进数据要素融合、汇聚、流动、价值增值及实现的作用。故本文提出假设H1:地方政府大数据交易平台能够提升数据要素的规模价值进而促进经济增长。

当前,我国地方政府大数据交易平台主要有数据交易中心和数据交易所。在杨艳等(2021b)的基础上,本文将1999-2018年间是否建设政府大数据交易平台作为一项拟自然实验,已建成政府大数据交易平台的省份作为实验组,其余未建平台的省份作为控制组,然后运用基于倾向得分匹配的双重差分法(PSM-DID)实证检验上述假设的正确性。构建如检验模型:

$$Gpgdp_{it} = \alpha_0 + \alpha_1 \ln platform_{it} + \alpha_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

$$\ln AI_{it} = \beta_0 + \beta_1 \ln platform_{it} + \beta_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

$$Gpgdp_{it} = \gamma_0 + \gamma_1 platform_{it} + \gamma_2 \ln Data_{it} + \gamma_3 X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

其中,下标  $i$  表示省份,  $t$  表示年份。  $platform_{it}$  是平台虚拟变量,如果省份  $i$  在  $t$  年建成了政府大数据交易平台,则  $platform_{it}=1$ , 否则为 0。其余控制变量设定与前文一致。

表 5 报告了关于假设 1 的检验结果。列(1)平台虚拟变量的估计系数显著为正,表明已建成政府大数据交易平台的省份经济增长率比未建政府大数据交易平台的省份平均高出约 0.0365%。列(2)平台虚拟变量的估计系数在 10% 水平下显著为正,表明相较于未建政府大数据交易平台的省份,已建政府大数据交易平台省份的平均数据要素规模价值会额外增加 0.0842%。列(3)平台虚拟变量和数据要素的估计系数均显著为正,表明地方政府数据交易平台能够提升数据要素的规模价值进而促进经济增长。至此,假设 1 得到验证。

(二)数据要素的本质:有效信息

在数据要素的产生路径中,数据要素源自数字化的现实信息,但数据要素并不等同于数字化信息。例如,一张人像数码照片与一首网络数字音乐都是数字化信息,照片可以用来训练人脸识别算法继而应用于银行、公安、教育等行业。数字音乐更多的是音乐平台提供的最终产品,用来供大众消遣娱乐。此时,训练人脸识别算法的人像数码照片是数据要素,是数字化信息中有助于社会经济发展的有效信息,但数字音乐属于不直接服务社会经济发展的一般数字化信息。可以看出,数据要素的本质是有效信息,而有效信息最关键的作用是减少不确定性。在宏观地区层面,地方政府在掌握更多的有效信息后,可以降低经济政策制定的不确定性进而

促进区域经济增长(饶品贵等,2017)。基于此,本文提出假设 H2:在省级层面,数据要素可以降低经济政策不确定性进而促进经济增长。构建如下检验模型:

$$Uncertainty_{it} = \alpha_0 + \alpha_1 \ln Data_{it} + \alpha_2 X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

$$Gpgdp_{it} = \beta_0 + \beta_1 \ln Data_{it} + \beta_2 Uncertainty_{it} + \beta_3 X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

其中,下标  $i$  表示省份,  $t$  表示年份。  $Uncertainty_{it}$  表示省级层面的经济政策不确定性,用《财经文本大数据基础数据库》中经济政策不确定性指数(EPU)的对数衡量<sup>⑤</sup>,时间跨度为 2000-2017 年。其余控制变量设定与前文一致。

表 6 报告了关于假设 2 的检验结果。列(1)显示,数据要素规模每增加 1%,省级层面的经济政策不确定性将降低 0.6298%。列(2)经济政策不确定性的估计系数在 5% 水平下显著为负表明,经济政策不确定性升高会对经济增长产生抑制作用。综上,数据要素可以降低省级层面的经济政策不确定性从而促进经济增长,假设 2 得到验证。

(三)数据要素产出弹性的初步考察

数据成为基础性生产要素后,考察总量生产函数中数据要素的产出弹性,是研究高质量发展阶段我国经济增长绩效与可持续性的重要一环。鉴于此,本文基于前文的新增数据要素价值,对数据要素的产出弹性进行推导和估算<sup>⑥</sup>。

表 7 报告了 1999-2018 年我国数据要素产出弹性的初步估计结果。可以发现,数据要素的产出弹性随时间呈现出明显的“倒 U 型”演化趋势,其均值约为 0.3046。在维持了近 13 年(1999-2012 年)

表 5 数据要素、政府大数据交易平台与经济增长

	$Gpgdp$ (1)	$Gpgdp$ (2)	$Gpgdp$ (3)
$platform$	0.0365*** (0.0125)	0.0842* (0.0420)	0.0307** (0.0132)
$\ln Data$			0.0690* (0.0404)
控制变量	是	是	是
省份固定效应	是	是	是
年份固定效应	是	是	是
观测值	600	600	600
Within R <sup>2</sup>	0.9835	0.9633	0.9842

表 6 数据要素、经济政策不确定性与经济增长

	<i>Uncertainty</i> (1)	<i>Uncertainty</i> (2)
<i>ln Data</i>	-0.6298*** (0.2162)	1.6558*** (0.5735)
<i>Uncertainty</i>		-0.3285** (0.1586)
控制变量	是	是
省份固定效应	是	是
年份固定效应	是	是
观测值	540	540
Within R <sup>2</sup>	0.1119	0.2217

表 7 我国数据要素的产出弹性(1999–2018 年)

年份	产出弹性	年份	产出弹性	年份	产出弹性
1999	0.1852** (0.0753)	2006	0.3381*** (0.0774)	2013	0.3626*** (0.0774)
2000	0.2105*** (0.0709)	2007	0.3619*** (0.0765)	2014	0.3367*** (0.0820)
2001	0.2162*** (0.0719)	2008	0.3721*** (0.0755)	2015	0.3087*** (0.0819)
2002	0.2403*** (0.0683)	2009	0.3519*** (0.0745)	2016	0.28645*** (0.0872)
2003	0.2466*** (0.0800)	2010	0.3681*** (0.0775)	2017	0.2790*** (0.0918)
2004	0.2851*** (0.0776)	2011	0.3697*** (0.0776)	2018	0.2734*** (0.0960)
2005	0.3275*** (0.0726)	2012	0.3723*** (0.0764)	数据要素产 出弹性均值	0.3046

的持续增长后,自 2013 年开始,数据要素的产出弹性开始逐年降低。这可能是当时的大数据技术、数据基础设施以及不完善的数据要素市场体系,并不能有效支撑大数据元年后数据要素体量的爆炸式增长所致,Farboodi 和 Veldkamp(2021)的研究也提到了这一点。另外,得益于国家对大数据行业的大力支持以及对数据安全的逐渐重视<sup>⑨</sup>,由图 1 可以看到,2016 年起数据要素产出弹性的下降趋势开始放缓,这为当前政府应进一步加快建设数据要素市场体系、完善相应制度环境,以保障数据要素对经

济增长红利的有效释放提供了一定现实依据。

### 七、结论与政策启示

本文构建了数据要素的产生路径和价值路径,据此估算了 1999–2018 年间我国省级层面的新增数据要素价值,同时通过引入人工智能技术,从理论和实证两个层面探讨了数据要素对我国经济增长的“双维驱动”影响。本文的主要结论:通过数据要素价值路径估算所得省级新增数据要素价值具有一定的理论和现实合理性;数据要素能够直接驱动经济增长,即数据要素具有经济增长效应,是实质

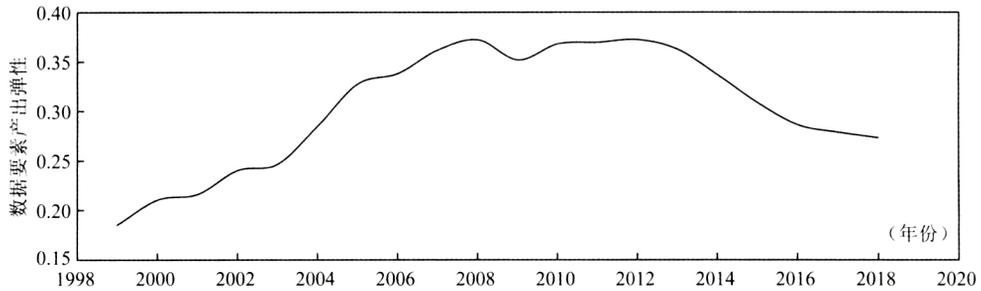


图1 我国数据要素产出弹性(1999-2018年)

的基础性生产要素;数据要素能够通过促进人工智能技术进步来间接驱动经济增长;地方政府大数据交易平台可以提升数据要素的规模价值;社会中的数据要素能够助力缓解经济政策不确定性;1999-2018年间,我国数据要素的产出弹性呈“倒U型”演化趋势。

上述研究结论具有重要的政策含义。首先,持续加强建设我国数据要素市场,充分发挥数据要素市场作用,促进社会中的数据要素自主充分有序流动,能够不断增强并更好发挥数据要素的“双维驱动”作用。其次,数据要素是数字经济的基础性资源,更是助力我国参与未来全球竞争的核心战略资源。目前,能够大力发展数字经济的发达地区应着力消除公共数据开放的制度性障碍,在做好数据隐私保护的同时充分提高社会经济中数据要素的规模和流动性;数字经济发展欠发达地区,可先行建设与数据要素相关的数字基础设施,重点推进数据要素的收集、存储和维护。最后,积极推动地方政府大数据交易平台建设,应注意的,通过平台进行数据要素交易不受空间限制,各地在政府大数据交易平台的建设中应做好统一规划,尽量避免重复建设造成资源浪费。

#### 注释:

①数据来源:中国信息通信研究院《数据价值化与数据要素市场发展报告(2021)》。

②徐翔和赵墨非(2020)认为数据资本是“以现代信息网络和各类数据库为重要载体,基于信息和通信技术的充分数字化、生产要素化的信息和数据”,这种定义更接近于数据要素内涵。根据中国信息通信研究院《数据价值化与数据要素市场发展报告(2021)》,原始数据需要依次通过资源化、资产化和资本化才能成为数据资本。数据资本主要包括数据信

贷融资与数据证券化。

③部分文献也将数据设定为资本的产出。例如 Farboodi 和 Veldkamp(2021)提出了数据点的概念,认为企业在每一时期末观察到的数据点的数量取决于资本产出。本文沿用了 Jones 和 Tonetti(2020)的做法,将数据要素作为消费的副产品,主要目的在于简化分析,无需在资本积累过程考虑数据要素。

④尽管数据可以积累并且不会像机器设备、建筑物、自然资源等一样自然衰减或消耗(Farboodi等,2019),但对于大多数经济活动,数据要素的效用会随着时间的推移而降低,其经济价值可能会贬值。因此,为了避免分析中出现过多的状态变量,本文做出了数据要素在每一期使用后会完全折旧的假设。

⑤因篇幅所限,详细的推导过程以附录 A 展示,见《统计研究》网站所列附件。下同。

⑥社会中有海量的公共数据因涉密或涉及个人隐私,例如医保、社保、公共医疗等数据,暂无法在市场中进行出售。

⑦非结构化数据是数据结构不规则或不完整,没有预定义的数据模型,不方便用数据库二维逻辑表来表现的数据,包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像和音频/视频信息等。源自百度百科:<https://baike.baidu.com/item/%E9%9D%9E%E7%BB%93%E6%9E%84%E5%8C%96%E6%95%B0%E6%8D%AE/309808?fi=aladdin>。

⑧因篇幅有限,数据要素的产生路径与价值路径以附图 1 展示。

⑨基于第二部分中的理论假设:“每一期数据要素在使用后完全折旧,并不会积累成为数据资产”。因此,本文只估算每一年新增数据要素的价值,不考虑数据资产的存量价值,并将估算出的年度新增数据要素的价值作为后续实证分析中数据要素的代理指标。

⑩非直接人工和其他成本大致包括相关的水电、建筑维护与租用、电信服务、社保、医保、办公设施以及人力资源管理费用。

⑪因篇幅有限,职业、平均有效工作时间、非直接工资和其他成本平均占比以附表 1 展示。

⑫20 家机构包括大学、科研事业单位、政府部门、省级大数据中心、大数据企业、电信运营商和部分金融机构。

⑬从访谈内容来看,尽管各职业的工资会随年份不断提高,但历年来的“每天平均有效工作时间(小时)”与“非直接工资和其他成本占工资的平均比重(%)”这两个平均值始终在一定范围内,不会有大幅度的波动。因此,本文假定其不随时间变化,没有下标  $t$ 。

⑭因篇幅所限,所有省份年度新增数据要素价值的估算结果以附录 B 展示。

⑮关于 2013 年是大数据元年这一提法的来源:熊建,黄碧梅,林琳,等. 2013 大数据元年 [N]. 人民日报,2013-12-25 (10)。

⑯个体类指标的数据均来自于 EPS (Easy Professional Superior) 数据平台中的宏观经济数据库,有统计数据的年份尽管不同但时间连续,且数据无缺失。后 2 个综合类指标数据的年份是 2011-2018 年,且数据无缺失。因此,用不同类别不同时间段的指标检验年度新增数据要素价值估算结果的合理性,会更有说服力。

⑰因篇幅所限,第一、二类指标的合理性检验结果以附录 C 展示。

⑱《中国专利全文数据库(知网版)》包含发明专利、实用新型专利、外观设计专利三个子库,准确地反映中国最新的专利发明,可以通过申请号、申请日、公开号、公开日、专利名称、摘要、分类号、申请人、发明人、优先权等检索项进行检索, <http://cn.oversea.cnki.net/kns55/brief/result.aspx?dbprefix=sepd>。

⑲“企查查”是官方认证的全国企业信用查询系统, <https://www.qcc.com/>。

⑳1999-2018 年本文各样本省份市场化指数详细来源:1999-2007 年数据源自《中国市场化指数——各地区市场化相对进程 2009 年报告》,2008-2016 年数据源自《中国分省份市场化指数报告(2018)》,2017-2018 年数据是基于《中国分省份市场化指数报告(2018)》和前文方法计算得到。

㉑本文研究样本不含西藏和港澳台地区。因篇幅所限,主要变量的描述性统计以附表 2 展示。

㉒1959-1978 年国企单位人均工资的数据来源于《新中国 60 年统计资料汇编》。其中,辽宁省的整体数据缺失,本文用吉林和黑龙江数据的均值作为替代。其余少量缺失数据通过插值法补齐。

㉓《东西中部和东北地区划分方法》,国家统计局,2011 年 06 月 13 日, <http://www.stats.gov.cn/zjtj/zthd/sjtjr/dejtkf/tjqp/201106/t2011061371947.htm>。

㉔两类卫星灯光影像数据来源于: <https://www.ngdc.noaa.gov/ngdc.html>。

㉕因篇幅所限,基于夜间人均灯光强度数据的检验结果以附表 3 展示。

㉖因篇幅所限,基于 IFR 工业机器人数据的检验结果以

附表 4 展示。

㉗《财经文本大数据基础数据库》地址: <http://cedcdata.cufe.edu.cn/cedc/metadata/list.html?>。

㉘因篇幅所限,推导过程和估算方法以附录 D 展示。

㉙2015-2017 年,国家开始采取一系列措施支持大数据行业发展;2015 年 4 月,全国首个大数据交易所(贵阳大数据交易所)正式挂牌运用;2015 年 8 月,国务院发布《促进大数据发展行动纲要》;2016 年 1 月,全国第一部大数据地方法规《贵州省大数据发展应用促进条例》出台;2016 年 10 月,国家同意在京津冀、珠三角、上海、重庆、河南等 7 个区域推进国家大数据综合试验区建设;2017 年 1 月,工业和信息化部印发《大数据产业“十三五”发展规划》。

#### 参考文献:

[1]蔡跃洲,马文君.数据要素对高质量发展影响与数据流动制约[J].数量经济技术经济研究,2021,38(3):64-83.

[2]廖祖君,王理.城市蔓延与区域经济高质量发展——基于 DMSP/OLS 夜间灯光数据的研究[J].财经科学,2019(6):106-119.

[3]彭国华.中国地区收入差距、全要素生产率及其收敛分析[J].经济研究,2005(9):19-29.

[4]饶品贵,岳衡,姜国华.经济政策不确定性与企业投资行为研究[J].世界经济,2017,40(2):27-51.

[5]王谦,付晓东.数据要素赋能经济增长机制探究[J].上海经济研究,2021(4):55-66.

[6]徐翔,赵墨非.数据资本与经济增长路径[J].经济研究,2020,55(10):38-54.

[7]闫雪凌,朱博楷,马超.工业机器人使用与制造业就业:来自中国的证据[J].统计研究,2020,37(1):74-87.

[8]杨艳,王理,廖祖君.数据要素:倍增效应与人均产出影响——基于数据要素流动环境的视角[J].经济问题探索,2021a(12):118-135.

[9]杨艳,王理,廖祖君.数据要素市场化配置与区域经济发展——基于数据交易平台的视角[J].社会科学研究,2021b(6):38-52.

[10]于文超,梁平汉.不确定性、营商环境与民营企业经营活力[J].中国工业经济,2019(11):136-154.

[11]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):65-76.

[12]Aghion P, Jones B F, Jones C I. Artificial Intelligence and Economic Growth [M]. University of Chicago Press, 2019.

[13]Agrawal A K, McHale J, Oettl A. Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth [J]. NBER Working Paper, 2018.

- [14] Beck T, Levine R, Levkov A. Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States[J]. *The Journal of Finance*, 2010, 65(5):1637–1667.
- [15] Cong L W, Xie D, Zhang L. Knowledge Accumulation, Privacy, and Growth in a Data Economy[J]. *Management Science*, 2021, 67(10):6480–6492.
- [16] Farboodi M, Veldkamp L. A Growth Model of the Data Economy[R]. National Bureau of Economic Research, 2021.
- [17] Farboodi M, Mihet R, Philippon T, et al. Big Data and Firm Dynamics[J]. *American Economic Association Papers and Proceedings*, 2019, 109:38–42.
- [18] Goldsmith-Pinkham P, Sorkin I, Swifi H. Bartik Instruments: What, When, Why, and How[J]. *American Economic Review*, 2020, 110(8):2586–2624.
- [19] Henderson J V, Storeygard A, Weil D N. Measuring Economic Growth from Outer Space[J]. *American Economic Review*, 2012, 102(2):994–1028.
- [20] Jones C I, Tonetti C. Nonrivalry and the Economics of Data[J]. *American Economic Review*, 2020, 110(9):2819–2858.
- [21] Lu C H. The Impact of Artificial Intelligence on Economic Growth and Welfare[J]. *Journal of Macroeconomics*, 2021, 69:103342.
- [22] Lucas Jr R E. On the Mechanics of Economic Development[J]. *Journal of Monetary Economics*, 1988, 22(1):3–42.
- [23] Nielsen M. Reinventing Discovery: The New Era of Networked Science[M]. Princeton, NJ: Princeton University Press, 2011.
- [24] Statistics Canada. Canada Data Framework Investment in Data, Databases and Data Science: Conceptual Framework [EB/OL]. <https://www150.statcan.gc.ca/nl/pub/13-605-x/2019001/article/00008-eng.htm>, 2019a.
- [25] Statistics Canada. The Value of Data in Canada: Experimental Estimates [EB/OL]. <https://www150.statcan.gc.ca/nl/pub/13-605-x/2019001/article/00009-eng.htm>, 2019b.
- [26] Varian H. Artificial Intelligence, Economics, and Industrial Organization[M]. University of Chicago Press, 2019.
- [27] Veldkamp L, Chung C. Data and the Aggregate Economy[J]. *Journal of Economic Literature*, 2019.

## China's Economic Growth: The "Two-dimension Driving Effect" of Data Factors

Yang Yan      Wang Li      Li Yujia      Liao Zujun

**Abstract:** Data factors have become one of the five most essential factors of production, but its role in economic growth has always been ambiguous. Starting from artificial intelligence technology, this paper establishes an endogenous growth model of data factors affecting economic growth, constructs the generation path and value path of data factors, and measures the value added of data factors at the provincial level in China from 1999–2018 accordingly. Based on theoretical analysis and a series of empirical tests, it is clear that the data factor has a "two-dimension driving effect" on China's economic growth, i. e., the data factor can drive growth both directly through its own economic growth effect and indirectly by promoting technological progress. Furthermore, this paper has three extended discussions. The results show that the local government big data trading platform can promote economic growth by increasing the scale value of the data factor; at the provincial level, the data factor can promote economic growth by reducing economic policy uncertainty. In addition, the output elasticity of factor for the years 1999–2018 showed an "inverted U-shaped" evolution trend.

**Key words:** data factor; economic growth; artificial intelligence; economic growth effect