

国民账户体系视角下数据资产统计测度研究*

朱发仓 石秋乐 苏为华 许宪春

内容提要:数据资产统计与核算是当前经济统计领域亟待突破的理论与实践难题。为刻画数据作为生产要素的作用与经济贡献,本文基于其提供财富储备和服务经济运行的双重功能,构建了涵盖数据资本存量与服务流量的全流程测度框架,并将数据资本服务量纳入增长核算框架以探讨其经济贡献。研究表明:数据资本存量净额与总额均呈现初期迅速扩张、中期高速发展以及后期高质量发展的三阶段演进特征,数据资本服务量则由高速增长转向中高速增长;考察期内数据资本财富性指标与服务性指标的年均增速均远超同期GDP的增速,凸显其相较于传统物质资本更大的价值累积潜力;数据资本的经济贡献潜力逐步显现并趋于主导地位,2020—2022年数据资本、R&D资本、计算机及通信设备三者的平均贡献率之和达54.57%,有效缓冲了建筑物及房地产的负向贡献。中国亟须以数据为核心优化资本结构,促进新旧增长动能的接续转换。本文的研究方法有助于推动数据资产统计核算的理论研究与实务工作进展,研究结论为优化中国经济增长动能提供了实证依据。

关键词:国民账户体系 数据资产 统计测度

作者简介:朱发仓,浙江工商大学统计与数学学院教授、博士生导师,310018;

石秋乐(通讯作者),浙江工商大学统计与数学学院博士研究生,310018;

苏为华,浙江工商大学统计与数学学院教授、博士生导师,310018;

许宪春,中国金融四十人论坛学术顾问、统计学者,100080。

中图分类号:F222.3 **文献标识码:**A **文章编号:**1002-8102(2025)08-0128-17

一、引言

20世纪90年代后期,互联网、云计算等数字技术加速发展,腾讯、阿里巴巴等互联网企业应运而生,推特、微信等社交工具将人类生活网络化。全球互联网用户数由2013年的27亿人增长至

* 基金项目:国家社会科学基金重点项目“人工智能驱动新质生产力发展的统计测度及实现路径研究”(24ATJ002);浙江工商大学“数字+”学科建设项目“数据资产:经济理论、价值核算、市场交易与政策创新”(SZJ2022A001);浙江省登峰学科(浙江工商大学统计学)和“统计数据工程技术与应用协同创新中心”资助项目。感谢匿名审稿专家的宝贵意见,文责自负。石秋乐电子邮箱:qiuleshi007@163.com。

2022年的53亿人,约占全球人口总数的2/3。^①数据量也呈指数级增长,Statista预计全球数据总量将于2025年达181ZB。^②微观层面,企业间数据竞争态势愈发激烈,如2014年谷歌斥资32亿美元收购智能家居公司Nest Labs,将其海量用户数据内部化。宏观层面,各国将发展数字经济提升至重要战略高度。2022年美国数字经济增加值为2.60万亿美元(占美国GDP的10%),2017—2022年平均增长率达7.10%;^③2020年中国数字经济核心产业增加值达79637.90亿元,较上年增长8.45%(鲜祖德、王天琪,2022)。在全球经济增长乏力情形下,数字经济之所以呈逆势增长态势,得益于数据这一新型生产要素加快了人流、资金流、信息流、技术流等的流动速度(余东华、李云汉,2021)。理论层面,蔡继明等(2022)论证数据要素自身、物化在数据收集处理中的前期劳动以及本期用于数据收集处理的活劳动同样创造价值,揭示数据要素创造价值的微观机制;徐翔和赵墨非(2020)、杨艳等(2023)认为数据既能直接作为生产要素推动经济增长,也可通过促进技术进步间接驱动增长。

尽管数据的经济贡献得到广泛认可,数据一词持续频繁地出现在各国政府工作报告中,但《2008年国民账户体系》(以下简称SNA2008)和我国现行的《中国国民经济核算体系(2016)》(以下简称CSNA2016)均未单独识别数据生产活动,《统计用产品分类目录》也没有单独的数据类别,数据的重要性与其在国民账户体系中“不可见”的地位形成鲜明反差。因此,联合国国民经济核算工作组(Inter-secretariat Working Group on National Accounts, ISWGNA)于2018年开始探讨如何将数据纳入国民账户体系(Ahmad和Ven,2018)。直至2025年3月,联合国统计委员会第56次会议审议通过了SNA2025,明确将数据纳入知识产权产品范畴并提供了核算建议(UN等,2025),同时发布的数据测算指南(ISWGNA,2025)为各国开展数据核算工作提供可操作的方法论。现有实证研究中,Statistics Canada(2019b)拓宽SNA2008的资产范围,测算了数据、数据库和数据科学的价值,但在国际通用性和可比性问题上存在争议;Goodridge等(2022)扩大SNA2008关于计算机软件和数据库的边界,主观设定各职业从事数据生产活动的时间使用系数,利用欧洲国家劳动力调查资料测算数据投资额;Calderón和Rassier(2022)利用机器学习模型估计不同职业从事数据生产活动的时间比例,但未能完全剔除各职业用于软件开发的时间;刘涛雄等(2023)利用成本法、增值法和存储规模法估算中国数据资本的价值量和物量,并进一步探讨数据资本的产出弹性。

现有文献虽为数据资产核算研究提供了重要理论基础和方法论参考,但大多聚焦于数据投资额与资本存量的测算,忽视数据资本服务量的测算,致使数据作为生产要素的经济贡献难以准确衡量。为全面刻画数据在国民经济中的作用,本文构建数据资产全流程测度框架,利用机器学习模型和在线招聘广告估计数据生产活动从业人员的员工比例和时间使用系数,在SNA框架下测算了数据资本存量与服务流量并量化其经济贡献。与既有研究相比,本文边际贡献体现在以下三方面:一是理论上,探讨将数据作为生产资产纳入国民账户体系的途径,基于提供财富储备和服务经济运行双重功能视角构建涵盖产出、资本形成、资本存量净额与总额、资本服务流量的测度指标体系;二是资料来源上,将《中华人民共和国职业分类大典(2022年版)》(以下简称《职业大典2022》)、海量在线招聘广告等文本资料运用于国民经济统计与核算,拓展传统以统计报表为核心的资料来源类型,为数字经济时代统计实务提供新思路;三是方法上,基于SNA框架提出数据生产活动产

① 资料来源:国际电信联盟(ITU)网站, <https://www.itu.int/itu-d/reports/statistics/facts-figures-2022/>。

② 资料来源:Statista网站, <https://www.statista.com/statistics/871513/worldwide-data-created/>。

③ 资料来源:美国经济分析局(BEA)网站, <https://www.bea.gov/data/special-topics/digital-economy>。

出、数据资本存量和服务流量的测算技术方法,量化数据资本经济贡献,为全面研究数据核算问题提供新参考。

二、理论框架

(一)基于SNA视角的数据与数据生产活动

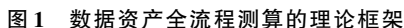
1.数据的核算概念

数据并非新生事物,但其在实践中常与信息、知识等概念混淆。为厘清概念间的差异,Ackoff(1989)提出从数据(Data)经由信息(Information)、知识(Knowledge)最终形成智慧(Wisdom)的转换过程。Rowley(2007)在此基础上构建DIKW层次结构,强调数据不代表知识,未经系统化组织与存储的数据无法直接使用,其价值仅能在转化为信息或知识后得以体现。Varian(2018)提出涵盖数据、信息、知识及行为的数据金字塔,明晰数字化存储的数据、文档记录的信息与人脑存储的知识之间的关系。随着数字技术的发展,数据得以即时且高效地采集与存储,并有效转化为具有经济价值的信息与知识。基于此,Rassier等(2019)、Statistics Canada(2019a)及许宪春等(2022)均从经济活动视角界定数据的概念,并强调数据应具有应用场景。SNA2025将数据定义为“通过访问和观测可观察现象,并以数字化形式记录、组织与存储,进而在生产活动中创造经济价值的信息内容”(UN等,2025)。本文遵循此界定,以区分自然存在的可观察现象与生产活动生成的数字化数据,且仅将后者视为劳动力与资本投入所生产的无形产品。

2.数据生产活动的界定

现有研究提出一系列涵盖数据收集、清洗、存储、处理及挖掘等过程的数据价值链(Rassier等,2019;刘涛雄等,2023),以描述从价值很小的非结构化数据到可用于商业用途的结构化数据的生产过程,但未能直观体现数据的价值演变。而Statistics Canada(2019a)的信息价值链又侧重反映从观察、数据、数据库到数据科学的形态演变,OECD(2022)提出由原始记录、数据库到数据智能的数据价值增值过程也未体现生产环节。

本文基于形态演化与价值增值两个维度,将数据生产活动解构为数据生产链与价值增值链,内容详见图1。数据生产链是指数据生产者运用机器设备、数字技术等手段对数据进行处理的一系列加工处理;而价值增值链展现“可观察现象—原始数据—数据资源—数据产品”的价值提升过程。具体而言,可观察现象本身不涉及任何生产活动,其本质是人类、物体及环境相互作用的客观表征,涵盖自然事件(如下雨)及人类行为(如走路)等一切现象。传统数据记录方式依赖纸质载体,随着键盘、智能传感器等数据基础设施的普及,观察结果的可获取性与可处理性得以增强,促使可观察现象以数字化形式传输、存储和操作,形成原始数据。根据特定目标和现实需求,原始数据经过标准化清洗与结构化存储,形成可高效访问和复用的数据集合即数据资源。进一步地,通过算法、可视化工具等技术手段对数据资源进行深度管理与开发,数据资源实现价值跃迁,转化为具有实际应用价值的“新数据”即数据开发产品,如交互式图表、咨询报告等。本文认为数据生产活动应包含数据收集、数据整理和数据加工三个阶段,其产出包含数字化转化后的各类数据衍生品,而可观察现象作为生产起点不纳入最终产出。尽管数字技术和数据存储能力的飞速发展使得数据收集和存储能够瞬时完成,但这一过程仍需投入人力和物力,应视为生产活动的结果(许宪春等,2022;刘涛雄等,2023)。从事数据记录、清洗入库、管理开发等数据生产全过程的专业人员,即为数据生产活动从业人员。



资料来源:改编自 OECD(2009)。

(二)数据资产的记录方式与核算范围

数据既像货物一样可存储,又像服务一样无形,Goodridge等(2022)、彭刚等(2022)、许宪春等(2022)结合数据的特征,建议将其纳入与货物和服务并列的知识产权产品范畴。最终,SNA2025明确将数据视作生产资产并归入知识产权产品的一个子类,将SNA2008“知识产权产品”类别下的“计算机软件与数据库”,拆分为“数据和数据库”与“计算机软件”两个类别,数据作为“数据和数据库”的小类(UN等,2025)。

本文基于国民经济核算原则,从机构部门视角界定数据资产的核算范围:(1)企业部门作为数据生产的核心主体,其自给型生产数据(用于内部决策优化)与交易型生产数据(通过市场交易变现)均可获得潜在或显性的经济利益,应全部纳入核算范畴;(2)政府部门通过行政记录形成的专业数据库(如气象、交通等数据)兼具社会效益和市场价值,属于SNA生产边界;(3)住户部门的数据呈零散化、非市场化特征,对宏观经济影响甚微,参照SNA2008未将住户部门自给型知识产权产品纳入生产范围的处理原则,暂不将其纳入核算范畴。鉴于当前我国乃至全球范围内数据交易市场面临制度规范不健全、市场生态欠完善、数据流通困难等问题,大多数数据仍为自给型(许宪春等,2022),本文重点考察企业部门自给型数据生产活动。

(三)数据积累与使用核算框架

1.数据资本形成

基于SNA基本原则,固定资产确认需满足在生产过程中被反复或连续使用一年以上的条件,此为区分产出究竟作为中间消耗还是固定资本形成的关键。实践中,统计人员和数据生产者难以准确判断数据是否使用超过一年。ISWGNA(2023)提出两项解决方案:一是按固定比例将总产出部分资本化,二是将全部产出资本化。从实际情况考虑,方案一虽能剔除部分时效敏感型数据,但数据通常具备存储持久性、用途可扩展性等特征,且数据生产者普遍期望数据能够作为长期生产要素投入。鉴于此,SNA2025建议采纳方案二,对全部自给型数据生产活动产出进行资本化处理,同时采用相对较短的使用寿命测算数据资产组的存量价值,以体现在一年内快速消耗或过时的部分数据。故本文测算数据资本形成时遵循全部产出资本化的原则。

2.数据资产核算指标体系

依据《资本测度手册》(OECD, 2009), 数据资本兼具提供财富储备与服务经济运行的双重功

能。在财富维度,生产者将筹集资金投入数据生产活动,其收益通过资本报酬与消费者储蓄的再投资形成闭环,数据资本成为可产生经济利益的国民财富储备。数据资本存量分为总额和净额,数据资本存量总额指由过去数据投资形成的且按当前购买者价格重新估价的资本存量;数据资本存量净额是在总额基础上扣除数据资本消耗,表示经济所有者特定时点的实际财富拥有量,故又称财富性存量。从资本运动过程看,期初数据资本存量净额经历期间数据资本形成和数据资本消耗后,形成期末数据资本存量净额。在服务维度,数据资本通过资本市场向生产者部门提供服务流量(即数据资本服务量),其规模由生产性数据资本存量(反映不同役龄数据资产的服务供给能力)与数据资本租赁价格(衡量数据资本投入生产的收益水平)共同决定。在数据市场均衡状态下,生产者部门以租金形式向资本市场支付数据资本服务报酬,形成“服务供给—价值实现”闭环。基于此,本文将数据资本财富性指标与服务性指标连接至国民账户体系。其中,数据生产活动投入源自收入账户,数据生产活动产出衔接生产账户,数据资本形成衔接积累账户,而数据资本存量净额与总额等计入资产负债表,内容详见图1。

三、核算技术方法

(一)数据生产活动产出的核算方法

企业生产数据大多为自身最终使用,SNA2025认为应采用成本总和法核算此类产出(UN等,2025)。借鉴Calderón和Rassier(2022)的做法,将数据生产活动成本细化为:从事数据生产活动的劳动者报酬(以下简称数据劳动者报酬)、用于数据生产活动的中间投入(C_I)、与数据生产活动相关的资本服务成本(C_S)、生产税净额(C_T)。令 ω 表示某一职业, Ω 表示所有职业集合(即 $\omega \in \Omega$),数据生产活动产出(I_{output})则表示为:

$$I_{output} = \sum_{\omega \in \Omega} \rho_{\omega} \tau_{\omega} H_{\omega} M_{\omega} + C_I + C_S + C_T \quad (1)$$

其中, ρ_{ω} 、 τ_{ω} 分别表示职业 ω 从事数据生产活动的员工比例和时间使用系数, H_{ω} 、 M_{ω} 分别表示职业 ω 的就业人数和平均工资。

1.数据劳动者报酬的估计方法

现代劳动者普遍呈现多任务特征,其工作难以明确归属于某一特定生产活动。不同职业在职责范围、任务复杂度等方面具有显著差异,对数据生产活动的时间分配情况也有所不同。早期研究如Statistics Canada(2019b)、Goodridge等(2022)依赖专家主观设定各职业数据生产参与度。典型案例如日本内阁通过互联网专项调查获取数据相关职业的时间投入占比(Japanese Cabinet Office,2024)。在官方资料尚不完善的情况下,Calderón和Rassier(2022)、Schmidt等(2023)采用机器学习模型间接推算职业参与度。基于国际经验,ISWGNA(2025)建议各国优先采用企业调查或机器学习模型两种方式测算职业参与度。由于我国官方现有统计资料不足以估计各职业的数据劳动者报酬,本文利用《职业大典2022》、在线招聘广告及机器学习模型,间接推算各职业从事数据生产活动的员工比例及其分配给数据生产活动的时间比例。

(1)数据员工比例的估计方法。数据员工比例衡量某一职业就业人员总数中实际从事数据生产活动的人员数占比。本文利用词频-逆向文件频率算法从《职业大典2022》97个数字职业的描述文本中提取关键词,筛选数据收集、数据存储等作为数据生产技能,并与Calderón和Rassier(2022)

进行对照和补充,构建数据生产技能集。^①令集合 X 表示数据生产技能集的任一子集,构建示性函数($I_X(x_\omega^l)$)。其中, x_ω^l 表示职业 ω 第 l 条招聘广告的职业描述。当 $x_\omega^l \in X$ 时,表明职业 ω 第 l 条招聘广告职业描述中至少包含一项数据生产技能,则 $I_X(x_\omega^l) = 1$;反之,示性函数值 $I_X(x_\omega^l) = 0$ 。那么,职业 ω 所有招聘广告的示性函数值之和占该职业招聘广告总条数(L_ω)的比重即为该职业的数据员工比例($\hat{\rho}_\omega$):

$$\hat{\rho}_\omega = \frac{\sum_{l=1}^{L_\omega} I_X(x_\omega^l)}{L_\omega} \quad (2)$$

由2022年测算结果^②可知,不同职业从事数据生产活动的员工比例存在显著差异。数据分析处理工程技术人员、大数据工程技术人员及统计调查员等职业数据员工比例均超过90%,此类职业已形成“数据即生产”工作范式。数字化管理师、数字孪生应用技术员、全媒体运营师等职业数据员工比例在60%~80%,这类职业借助对数据的深度分析重构传统工作流程。环境监测员的数据员工比例为46.31%,其日常工作涉及统计监测污染源数据并编制实验报告;病案信息技师的数据员工比例为44.77%,主要工作包括收集、整理并加工病案信息。这类以数据为辅助支撑职业的数据员工比例通常在15%~60%,且占有数据相关职业的比重约55%。相比之下,营销员和照明设计师等职业数据员工比例仅约10%,这类职业大多集中于生产制造、基础服务以及传统手工艺等领域,主要从事设备参数记录等基础数据生产活动。

(2)时间使用系数的估计方法。时间使用系数衡量各职业总工作时间内专门用于数据生产流程的时间比例。既有研究如Goodridge和Haskel(2015)、Statistics Canada(2019b)大多主观设定各职业在某项工作中的时间投入比例,Calderón和Rassier(2022)利用招聘广告估计美国各职业从事数据生产活动的时间使用系数,其结果更贴近实际情况。本文借鉴该方法,将数据员工比例大于80%的20个细类职业界定为全时数据职业,并运用Doc2Vec模型计算各职业与全时数据职业广告文本的余弦相似度。余弦相似度反映不同职业工作任务特征与全时数据职业的相似程度,其数值越高,表明该职业工作任务特征越接近全时数据职业,投入数据生产活动的时间比例越大。那么,职业 ω 从事数据生产活动的时间使用系数($\hat{\tau}_\omega$)为:

$$\hat{\tau}_\omega = \max_{w \in M} \left\{ \frac{\hat{A}_\omega \times \hat{A}_w}{\|\hat{A}_\omega\| \times \|\hat{A}_w\|} \right\} \quad (3)$$

其中, w 表示全时数据职业, M 表示全时数据职业集合(即 $w \in M$), $\hat{A}_\omega$ 、 $\hat{A}_w$ 分别为职业 ω 和全时数据职业 w 的广告文本向量。由2022年测算结果^③可知,全时数据职业与自身招聘广告文本余弦相似度为1,故时间使用系数为1。信息管理工程技术人员及全媒体运营师等职业的时间使用系数均超0.8,其工作任务高度聚焦于数据采集、处理与分析等环节,数据生产活动在其职业活动中占据核心地位。地震监测预测工程技术人员、知识产权信息分析专业人员及信息安全测试员等职业时间使用系数在0.65~0.75区间内,由数据员工比例可知此类职业仅有50%左右的从业人员实际

① 《职业大典2022》依据职业完成主要工作所用的工具、工作环境、工作内容、工作过程等6个要素标识了97个数字职业。数字职业的数字特征显著,将其职业描述中的关键词作为数据生产技能具有代表性。限于篇幅,正文未报告数据生产技能集,详见线上附录。

② 限于篇幅,正文未报告各职业从事数据生产活动的员工比例,详见线上附录。

③ 限于篇幅,正文未报告各职业从事数据生产活动的时间使用系数,详见线上附录。

参与数据生产活动,其职业活动是专业领域技术与数据生产技术的综合应用。营销员、照明设计师等职业时间使用系数约为0.2,这类职业的数据员工比例仅为10%左右,最终用于剥离该类职业劳动报酬的比例仅约为2%(即 $0.2 \times 10\%$),表明数据生产并非其核心职业活动,与《职业大典2022》的职业界定相吻合。

2. 其他成本的估计方法

数据生产活动其他成本包括中间消耗、资本使用成本以及生产税净额。现有研究普遍采用简化方式,将劳动者报酬乘以主观设定的加成系数来间接估计总成本(Goodridge等,2022),但此方式缺乏客观依据且国际可比性较差。Corrado等(2022)利用计算机程序设计、咨询及相关活动和信息服务活动的成本结构推算加成系数。ISWGNA(2025)则建议采用《所有经济活动的国际标准行业分类(第五版)》的“计算机程序设计、咨询及相关活动(类62)”与“计算基础设施、数据处理、托管和其他信息服务活动(类63)”两个数据密集型行业总产出与劳动者报酬之比计算加成系数。当前我国统计调查制度尚未涉及数据专项调查,无法直接获取数据生产活动其他成本的统计资料。鉴于数据生产与软件开发各环节具有相似性,本文选取浙江省3061家限额以上软件和信息技术服务业企业作为研究样本,^①开展成本费用调查,利用软件开发活动的成本结构估计数据生产活动的成本结构。依据调查结果,测算得到中间消耗与劳动者报酬之比为30.25%,资本使用成本和生产税净额与劳动者报酬之比为9.37%。

(二)数据资本存量与流量的核算方法

1. 数据资本形成的核算方法

借鉴OECD(2010)的供需平衡法,时期内数据总供给由国内数据生产活动产出(I_{output}^t)和进口(I_{import}^t)组成;时期内数据总需求包括最终消费(I_{con}^t)、中间消耗(I_{incon}^t)、出口(I_{export}^t)、数据资本形成(I_{GFCF}^t)和存货变化(I_{stoc}^t)。数据资本形成可计算为:

$$I_{GFCF}^t = I_{output}^t - I_{con}^t - I_{incon}^t - I_{stoc}^t - I_{export}^t + I_{import}^t \quad (4)$$

由于国内外数据交易市场的交易量较低,从可行性角度考虑,将数据交易额和进出口额均视为0。数据的更新、维护视为数据资本形成,不记录中间消耗和存货变化。^②那么,数据资本形成可用数据生产活动产出近似估计: $I_{GFCF}^t = I_{output}^t$,此做法与ISWGNA(2025)建议一致。

2. 数据资产组的平均役龄-效率函数和平均役龄-价格函数

数据资产生产效率在其服役初期因技术垄断和差异化获取而呈缓慢下降趋势,服役后期受技术迭代等因素影响而加速下降,且大部分十年前采集的数据会被归为“冷数据”(刘涛雄等,2023),此变化规律比较接近双曲型效率衰减模式(OECD,2009)。鉴于一般磁、电存储寿命在5~10年,^③本文将数据资产最长使用寿命设为10年。基于双曲式衰减模式和正态退役模式^④,拟合得到数据

① 样本企业均属ISWGNA(2025)建议的数据密集型行业范畴。以浙江省作为典型样本主要基于三方面考量:浙江省作为我国首批国家数字经济创新发展试验区,率先开展数据要素市场化配置改革试点,能够作为观察数字经济政策效应的典型窗口;浙江省大数据行业生态链完整,样本企业覆盖杭州未来科技城、宁波软件园等6个国家级软件产业基地,与全国“东数西算”工程布局形成区域呼应;浙江省软件企业以杭州为核心,宁波、温州、嘉兴等地各具特色的发展格局,能够映射全国东部主导、中西部特色发展的特征。

② 此处借鉴朱发仓和杨诗淳(2020)的做法:研发有新知识产生,资本化核算时不记录存货变化。

③ 资料来源:国际电子商情网站, <https://www.esmchina.com/news/7736.html>。

④ 正态退役模式假设数据资产役龄与其对应退役概率服从正态分布。限于篇幅,正文未报告效率衰减与退役模式的基础函数公式,可参见OECD(2009)。

资产组的平均役龄-效率函数(h_n):

$$h_n = \sum_{t=n}^T g_n F_n \quad (5)$$

其中, g_n 和 F_n 分别代表单一数据资产的效率衰减函数和退役分布函数; T 表示数据资产最长服役年限;假设数据投资发生在期中, $n = 0.5, \dots, T - 0.5$ 表示数据资产役龄,即 $n=1.5$ 意味着数据资产服役2年。令 tB 、 tE 分别表示 t 年年初和年末, P^{tB} 、 P^{tE} 分别表示年初和年末数据资产价格,市场均衡时平均役龄-价格函数(ψ_n)与平均役龄-效率函数具有如下关系:

$$\psi_n = \frac{P_n^{tB}}{P_0^{tB}} = \frac{h_n + h_{n+1}(1+i)(1+r)^{-1} + \dots + h_{n+T}(1+i)^{T-1}(1+r)^{1-T}}{1 + h_1(1+i)(1+r)^{-1} + \dots + h_T(1+i)^{T-1}(1+r)^{1-T}} \quad (6)$$

其中, i 和 r 分别表示数据资产平均价格上涨率和收益率,借鉴王春云和王亚菲(2019)的研究结果,平均役龄-折旧率函数(δ_n)可表示为: $\delta_n = (P_n^{tB} - P_{n+1}^{tB})/P_n^{tB} = 1 - \psi_{n+1}/\psi_n$ 。

3. 数据资本存量的核算方法

基于平均役龄-价格函数(ψ_n)和不变价数据资本形成序列(I_{GFCF}^t),分别测算年初数据资本存量净额(W^{tB})和年末数据资本存量净额(W^{tE}),即:

$$W^{tB} = \psi_{0.5} I_{GFCF}^{t-1} + \psi_{1.5} I_{GFCF}^{t-2} + \psi_{2.5} I_{GFCF}^{t-3} + \dots + \psi_{T-0.5} I_{GFCF}^{t-T} \quad (7)$$

$$W^{tE} = \psi_{0.5} I_{GFCF}^t + \psi_{1.5} I_{GFCF}^{t-1} + \psi_{2.5} I_{GFCF}^{t-2} + \dots + \psi_{T-0.5} I_{GFCF}^{t-T+1} \quad (8)$$

数据资本存量净额年均余额($\bar{W}^t$)为年初和年末数据资本存量净额的简单平均,即满足 $\bar{W}^t = (W^{tB} + W^{tE})/2$ 。由于投资发生在年中,则以 t 年平均价格计价的数据资本消耗(E^t)为: $E^t = 0.5P_0^t \delta_0 I_{GFCF}^t + P_{0.5}^t \delta_{0.5} I_{GFCF}^{t-1} + P_{1.5}^t \delta_{1.5} I_{GFCF}^{t-2} + \dots + P_{T-0.5}^t \delta_{T-0.5} I_{GFCF}^{t-T}$ 。可得数据资本存量总额:

$$S^{tE} = W^{tE} + E^t = W^{tB} + I_{GFCF}^t = S^{t+1,B} \quad (9)$$

其中, S^{tE} 、 $S^{t+1,B}$ 分别表示第 t 年年末和第 $t+1$ 年年初的数据资本存量总额。同理,数据资本存量总额年均余额($\bar{S}^t$)也为年初和年末的简单平均: $\bar{S}^t = (S^{tB} + S^{tE})/2$ 。

4. 数据资本服务量的核算方法

基于平均役龄-效率函数(h_n)和不变价数据资本形成序列(I_{GFCF}^t),可测算得到生产性数据资本存量(K^t),即 $K^t = 0.5I_{GFCF}^t + h_{0.5} I_{GFCF}^{t-1} + h_{1.5} I_{GFCF}^{t-2} + \dots + h_{T-0.5} I_{GFCF}^{t-T}$,则 t 年数据资本服务量(D^t)可由数据资本使用成本对生产性数据资本存量加权得到:

$$D^t = f_0^t K^t = f_0^t [0.5I_{GFCF}^t + h_{0.5} I_{GFCF}^{t-1} + \dots + h_{T-0.5} I_{GFCF}^{t-T}] \quad (10)$$

其中, $f_0^t = P_0^{tB}[r + d(1+i) - i]$ 为新资产的单位使用成本, d 为数据资产平均折旧率。

四、测算结果分析

(一)数据来源及价格指数

1. 数据来源及预处理

本文基础数据源自《中国统计年鉴》《中国劳动统计年鉴》《中国人口普查年鉴》。由于《中国劳动统计年鉴》公布分行业而非分职业的平均工资,本文基于《职业大典2022》职业描述和《国

民经济行业分类》行业描述,利用文本描述之间的联系获得各职业从业人员的平均工资。在线招聘广告源自智联招聘、前程无忧以及BOSS直聘等国内主流招聘网站,共计1.6亿余条招聘信息。

在线招聘广告文本预处理过程如下。(1)去除冗余招聘广告。鉴于招聘信息通常存放在招聘网站一定周期且同一招聘广告可能被多次发布,通过“统一社会信用代码”“城市分类名称”“工作名称”三个字段进行去重,仅保留符合字段组合唯一性约束的首条招聘广告。(2)两步匹配细类职业。一是运用模糊字符串匹配技术将“工作名称”字段与《职业大典2022》细类职业名称进行名称匹配,并设定80%阈值;二是对名称匹配未达阈值的招聘广告,借助BERT-Base-Chinese模型实现招聘广告“职责描述”字段与《职业大典2022》细类职业描述的深层语义匹配。经名称和语义匹配,实现所有招聘广告与《职业大典2022》1601个细类职业^①精准映射。(3)数据相关职业筛选。依据《职业大典2022》职业界定,剔除化妆师、消防员、保洁员等427个与数据生产活动关联甚微的细类职业。(4)语义无关性清洗。删除剩余1174个细类职业招聘广告中诸如企业宣传用语、薪酬福利条款等可能干扰实证分析的非职能描述。

2. 数据资产价格指数

由于数据交易市场活跃度不高,难以获得数据产品的交易价格。鉴于数据生产活动与软件开发活动的相似性,本文选取通信设备、计算机及其他电子设备制造业工业生产者出厂价格指数作为数据资产的价格指数。

(二) 数据资本形成的测算结果及分析

以当年价计价的数据资本形成测算结果及其占GDP的比重如图2所示。从整体趋势看,数据资本形成呈多阶段演进特征。初始积累阶段(2000—2009年),数据资本形成由2000年的383.66亿元增至2009年的1949.06亿元,年均增速达19.79%,占GDP比重由0.38%升至0.56%。此时期我国信息化战略处于初步实施阶段,互联网基础设施逐步普及,创造了数据产生、传输和存储的必要条件,故数据资本形成呈高速扩张态势。快速发展阶段(2010—2019年),数据资本形成年均增速达15.47%,虽较上一阶段下降4.32个百分点,但规模显著扩大。究其原因,一是我国大力实施“宽

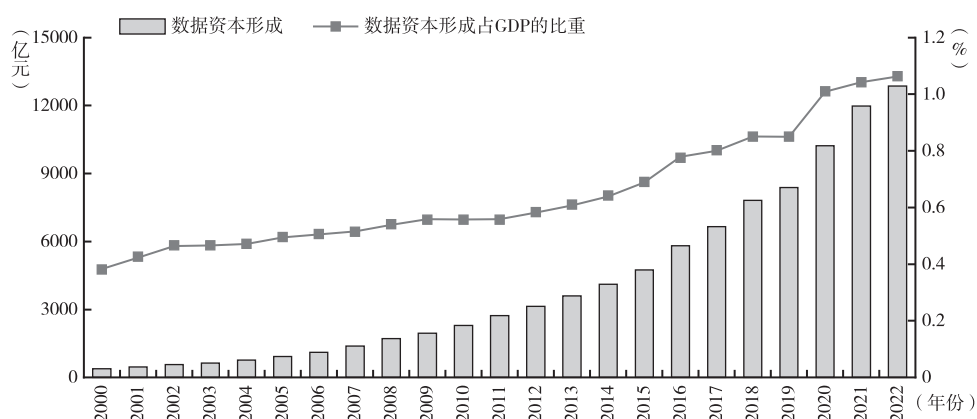


图2 2000—2022年数据资本形成及其占GDP的比重

注:图中数据均以当年价计价。

^① 《职业大典2022》实际包含1636个细类职业,其中35个细类职业涉及党政机关负责人、军队人员以及不便分类的其他人员,这类职业几乎不从事数据生产活动且招聘信息不公布于招聘网站,故将其排除在外。

带中国”战略、“互联网+”行动计划,移动互联网技术迅猛发展,为数据采集与处理提供了底层技术支撑;二是《促进大数据发展行动纲要》等政策相继出台,从顶层设计层面明确大数据发展任务和保障措施,激发市场主体参与数据开发和利用的积极性。高质量发展阶段(2020—2022年),数据资本形成由8384.08亿元增至12865.88亿元,年均增速为12.17%,增速虽较上一阶段进一步趋缓,但规模突破万亿元,数据资本形成由“量增”转向“质升”。尽管2019年数据资本形成占GDP比重较上年下降0.01个百分点,但2020年在“新基建”政策的推动以及疫情倒逼企业加速数字化转型的双重影响下,该比重突破1%的阈值,2022年达到1.06%,标志着我国正式进入以质量主导、技术驱动的数据资本发展新阶段。

美国经济分析局(Calderón和Rassier,2022)测算的企业部门数据资本形成由2002年的840亿美元(折合人民币约6953亿元)增至2021年的1860亿美元(折合人民币约11999.79亿元),^①早期规模显著领先我国而后期逐渐趋同,占美国GDP的比重维持在0.75%左右。这种先发优势源于美国作为全球数字经济策源地的技术积累,移动互联网与人工智能技术的持续迭代使得其数据资本形成规模持续扩增。加拿大统计局(Statistics Canada,2019b)公布2018年数据资本形成在294.55亿~400.25亿美元(折合人民币约1949.16亿~2648.61亿元),占加拿大GDP比重高达1.71%~2.32%。亚洲开发银行在2021年公布2019年印度数据资本形成成为236.38亿~284.85亿美元(折合人民币约1631亿~1965亿元),占印度GDP比重为0.83%~1.00%,测算方法同Statistics Canada(2019b)一致。对比发现,我国数据资本形成规模优势显著,但加拿大数据资本形成占GDP比重更为突出。究其原因,加拿大以服务业为主导,对数据依赖程度较高;而印度农业、制造业等传统产业占比较大,数字化转型相对滞后;我国虽处于新型工业化加速阶段,但传统制造业数字化转型仍需时日。此外,当前数据核算理论和方法尚处于探索阶段,各国核算体系差异也是导致测算结果差异的重要原因之一。^②

(三)数据资本存量和服务量的测算结果及分析

表1报告了数据资本存量与服务量测算结果。^③从财富视角,自2000年我国全面启动“企业上网工程”以来,数据资本存量净额和总额均呈跨越式积累态势。数据资本存量净额由2001年年初的281.33亿元增长至2022年年末的33298.68亿元,数据资本存量总额由2001年年初的383.66亿元跃升至2022年年末的50918.83亿元,年均增速分别为24.23%、24.88%,远超同期GDP年均增速,凸显数据资本相较于传统物质资本更为强劲的价值积累能力。2001—2022年数据资本形成和数据资本消耗也大幅增加,年均增速分别为19.84%、23.26%。从服务视角,数据资本服务量增长速度呈多阶段特征。2001—2008年数据资本服务量快速增长,年均增速高达39.86%,展现出蓬勃发展态势;受2008年国际金融危机等因素影响,2009—2015年数据资本服务量年均增速回调至13.36%,较上一阶段下降26.50个百分点;2016—2022年,以国务院印发《促进大数据发展行动纲要》和党的十九届四中全会将数据视作生产要素为重要契机,数据资本服务量年均增速回升至

① 本段使用的历年人民币汇率来源于中国国家外汇管理局,美国GDP数据来源于美国经济分析局(BEA)网站,加拿大和印度的GDP数据均来源于世界银行(World Bank)网站。

② 为验证测算结果的科学性,选取移动互联网接入流量、地理信息数据、互联网上网人数、网站数、互联网普及率、信息技术服务收入、数字经济核心产业增加值及信息与通信技术发展指数8项指标与数据资本形成序列进行相关性检验。结果显示,相关系数均在0.95以上,呈显著正相关性,验证了本文运用成本总和法和在线招聘广告的测算框架在理论与现实层面均具备科学合理。限于篇幅,相关性分析结果详见线上附录。

③ 考虑中国于2000年3月才正式启动“企业上网工程”,本文选取2001年作为数据资本存量和服务量核算的初始年,在此之前数据规模及经济价值忽略不计。限于篇幅,函数模拟值列于线上附录。

17.59%。从资本运动过程来看,2020年年初数据资本存量净额为21325.11亿元,在此期间数据资本形成与数据资本消耗分别为17023.95亿元、13196.62亿元,2020年年末数据资本存量净额为25152.44(即 $21352.11+17023.95-13196.62=25152.44$)亿元。

表1 2001—2022年数据资本存量和服务量 单位:亿元

年份	S^B	W^B	I_{GFCF}	E^U	W^E	S^E	D^I
2001	383.6620	281.3332	475.9093	218.0591	539.1834	757.2425	156.5285
2002	757.2425	539.1834	594.7965	348.8345	785.1453	1133.9799	252.8280
2003	1133.9799	785.1453	719.2828	487.0120	1017.4161	1504.4281	375.6553
2004	1504.4281	1017.4161	902.3592	638.4102	1281.3651	1919.7753	550.4297
2005	1919.7753	1281.3651	1148.6460	814.0655	1615.9455	2430.0110	681.9380
2006	2430.0110	1615.9455	1423.2364	1020.0453	2019.1366	3039.1819	891.5273
2007	3039.1819	2019.1366	1833.6276	1289.6806	2563.0836	3852.7642	1260.9016
2008	3852.7642	2563.0836	2304.9509	1628.5253	3239.5091	4868.0345	1638.7850
2009	4868.0345	3239.5091	2727.9837	2008.6985	3958.7943	5967.4928	1717.4672
2010	5967.4928	3958.7943	3272.0521	2448.2233	4782.6231	7230.8464	2251.9904
2011	7230.8464	4782.6231	3946.4488	2965.0873	5763.9846	8729.0719	2820.3699
2012	8729.0719	5763.9846	4645.3715	3547.7585	6861.5976	10409.3561	2895.5439
2013	10409.3561	6861.5976	5480.2414	4216.5392	8125.2998	12341.8390	3033.3016
2014	12341.8390	8125.2998	6366.3755	4966.0517	9525.6236	14491.6753	3303.7681
2015	14491.6753	9525.6236	7465.2598	5831.8053	11159.0781	16990.8833	3644.5931
2016	16990.8833	11159.0781	9278.4650	6978.0823	13459.4608	20437.5430	4247.6082
2017	20437.5430	13459.4608	10664.4838	8245.6227	15878.3219	24123.9446	5161.8713
2018	24123.9446	15878.3219	12704.9629	9772.6699	18810.6150	28583.2849	6184.3947
2019	28583.2849	18810.6150	13749.3909	11234.8960	21325.1099	32560.0059	7174.3896
2020	32560.0059	21325.1099	17023.9547	13196.6232	25152.4414	38349.0646	8300.8219
2021	38349.0646	25152.4414	19961.7538	15488.3392	29625.8560	45114.1952	9634.4297
2022	45114.1952	29625.8560	21292.9712	17620.1446	33298.69	50918.8272	11230.9778

注:表中数据以2000年不变价计价。

五、进一步分析

(一)数据资本对经济增长的贡献分析

现有研究探讨经济增长时仍将资本存量视为资本投入的代理变量,混淆了资本存量与服务流量的概念。资本存量为特定时点的财富储备,不等同于时期内实际资本投入,且无法揭示不同类

型资本生产效率差异。正如劳动投入应是劳动者提供的实际劳动服务而非劳动者本身,资本投入应是资本在生产中的实际投入,即资本服务量。因此,Jorgenson和Griliches(1967)在SNA框架下测算资本服务量,并提出总量层面的增长核算框架。为探讨数据资本与其他投入的经济贡献差异,本文借鉴王维等(2017)和Zhu等(2020)的做法,将资本细分为农业资本品(K_1)、家具体育娱乐设施(K_2)、金属制品(K_3)、通用和专用机械设备(K_4)、交通设备(K_5)、电气设备(K_6)、计算机及通信设备(K_7)、仪器仪表(K_8)、建筑物及房地产(K_9)和其他无形资产(K_{10})十大类。此外,鉴于OECD(2010)建议不成功的研究和开发(Research and Development, R&D)和非市场生产者的R&D成果都应资本化,借鉴朱发仓和杨诗淳(2020)的做法测算R&D资本服务量(RD);^①数据资本投入则采用上文测算的数据资本服务量(D)。借鉴陈梦根和张鑫(2022)的方法,结合性别、年龄和学历等劳动特征以及年总工作时间测算得到实际劳动投入(L)。基于上述分类,借鉴蔡跃洲和张钧南(2015)的研究方法,将总量层面的增长核算框架构建为:

$$\frac{d\ln Y}{dt} = \frac{d\ln A}{dt} + \alpha_1 \frac{d\ln D}{dt} + \alpha_2 \frac{d\ln RD}{dt} + \alpha_3 \frac{d\ln L}{dt} + \sum_{j=1}^{10} \beta_j \frac{d\ln K_j}{dt} \quad (11)$$

其中, Y 表示总产出; A 表示全要素生产率; α_i ($i=1,2,3$)依次表示数据资本、R&D资本和劳动投入占总投入的份额, β_i ($i=1,2,\dots,10$)依次表示 K_1 至 K_{10} 各项投入占总投入的份额,且满足 $\sum_{i=1}^3 \alpha_i + \sum_{j=1}^{10} \beta_j = 1$ 。将式(11)两边同时除以产出增长率($d\ln Y/dt$),即可得到各项投入及全要素生产率对经济增长的贡献率。结合各项投入测算结果^②对我国经济增长来源进行分解,结果见图3。2002—2022年数据资本对经济增长的拉动作用呈现初期缓慢提升、中期波动调整及后期稳步上升的阶段性特征。在2002—2007年的初始阶段,数据资本拉动系数由0.13%上升至0.39%,反映互联网普及初期“数据红利”逐步释放。受2008年国际金融危机冲击,2009年数据资本拉动系数骤降至0.07%,随后2010—2015年剧烈震荡(峰值0.45%、谷值0.05%)。自2016年起,随着深度学习算法突破与5G商用,数据资本网络外部性和非竞争性得以充分释放,其推动经济增长的作用日益显著,2022年该系数已达0.74%。2002—2022年,全要素生产率始终是经济增长的核心驱动力,2007

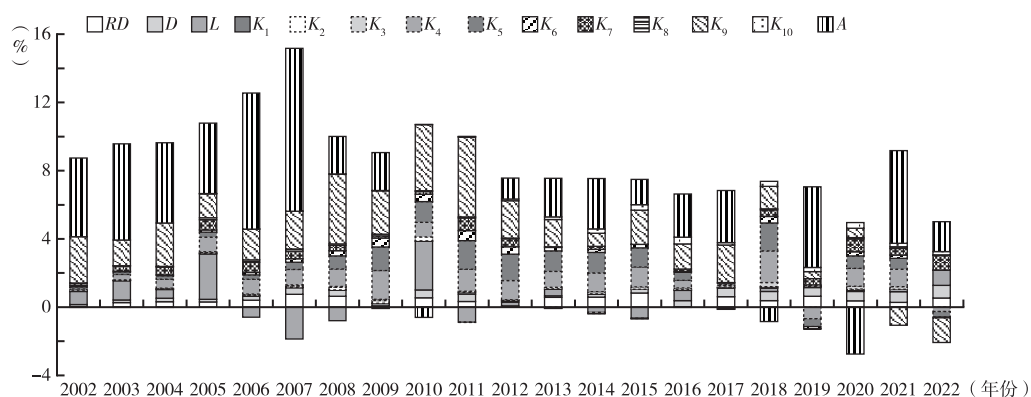


图3 2002—2022年中国经济增长的来源分解

① 王维等(2017)的十大类固定资本形成序列是基于CSNA2016对R&D资本化处理前的投入产出表,不包含R&D投资序列,故十大类资本品与R&D资本不存在重复计算问题。同时,现行R&D统计调查制度将R&D项目成果形式分为软件著作权、中间件或新算法、发明专利等14类,更偏向于制药、化学、芯片、电子设备等领域,并未包含与数据直接相关的成果。

② 限于篇幅,正文未报告除数据资本之外的其他各项投入,详见线上附录。

年拉动系数达9.55%。计算机及通信设备的经济作用日益凸显,2022年拉动系数增至0.85%,较2002年增加0.71个百分点。建筑物及房地产作为传统增长引擎展现逆周期特性,2008年其以4.09%的拉动系数有效缓冲外需萎缩冲击(经济增速仅为8.98%),2012年经济增速放缓至7.57%时仍拉动经济增长2.19个百分点。

受宏观政策、国际形势、疫情冲击等方面的影响,各项投入的经济贡献在个别年份异常波动。为消除此影响,进一步测算不同时段各项投入的平均贡献率。^①分时段来看,2002—2005年、2006—2010年及2011—2015年三个时段数据资本平均贡献率均低于3%,该阶段我国经济增长仍高度依赖通用和专用机械设备、交通设备、建筑物及房地产等传统资本投入,数据市场化配置与价值转化机制尚未成熟。2019年党的十九届四中全会将数据纳入生产要素范畴,并通过《关于构建更加完善的要素市场化配置体制机制的意见》明确其收入分配机制,政策红利加速释放。此背景下,2020—2022年数据资本平均贡献率高达19.36%,显著高于同期R&D资本、劳动投入的经济贡献。结合其他投入看,R&D资本经济贡献持续增强,平均贡献率由2002—2005年的2.27%逐步上升至2020—2022年的12.68%。劳动投入经济贡献则呈阶段性衰减,2002—2022年平均贡献率仅为3.91%,低于数据资本、R&D资本等的经济贡献,体现了传统人口数量红利消减与人力资本升级滞后的双重约束。农业资本品、家具体育娱乐设施、仪器仪表等的平均贡献率在各时段均低于1%。而通用和专用机械设备、交通设备、计算机及通信设备等的经济贡献较大,2020—2022年平均贡献率分别为17.02%、9.89%和22.53%,但各时段波动性显著,这是由于装备制造制造业受周期性产能调整、技术创新迭代等因素影响。2020—2022年数据资本、R&D资本与计算机及通信设备三者的平均贡献率之和达54.57%,形成对冲建筑物及房地产负向拖累(平均贡献率为-12.30%)的核心动力。此消彼长间,我国经济增长动力结构实现双重跃升:产业重心从房地产依赖型经济向数字技术融合型经济转变;发展路径由“要素堆积”模式向“数据与创新共同驱动”新范式变革。

(二)敏感性分析

借鉴现有研究(OECD,2009,2010;陈梦根、张鑫,2022;刘涛雄等,2023)相关参数设定,本文系统检验价格指数、服役年限及收益率对测算结果的敏感性。具体地,对比GDP缩减指数(指数Ⅰ)、通信服务类居民消费价格指数(指数Ⅱ)与通信设备、计算机及其他电子设备制造业工业生产者出厂价格指数(简称基准指数)的结果差异;对比服役年限设定为8年(年限Ⅲ)、12年(年限Ⅳ)与基准年限(10年)的测算结果;同时将历年收益率分别减少和增加1个百分点(分别为收益率Ⅴ和Ⅵ)进行比较。^②

数据资本存量与服务量测算结果对价格指数变动极为敏感。以2020年为例,由基准指数、指数Ⅰ和Ⅱ测得的数据资本存量净额年均余额分别为23238.78亿元、7535.29亿元和14784.51亿元,数据资本服务量分别为8300.82亿元、2732.76亿元和5289.35亿元。除初始若干年份外,指数Ⅰ和Ⅱ相对误差均超10%。究其原因,三项指数虽都能反映通货膨胀程度,但严格意义上,指数Ⅱ仅能反映消费领域的价格变动,而指数Ⅰ虽能体现国内所有最终货物和服务整体价格变动,但由CSNA2016可知当前货物和服务中尚未包含数据产品。基准指数能够反映信息通信类产品首次出售时的价格变动,本文用其近似反映数据产品的价格变动。

^① 限于篇幅,正文未报告不同时段各项投入对我国经济增长的平均贡献率,详见线上附录。

^② 限于篇幅,正文未报告稳健性测算结果,详见线上附录。为便于表述,仅分析相对误差的绝对值大小。

数据资本存量测算比数据资本服务量测算对服役年限的选取更敏感。从相对误差看,年限Ⅲ和Ⅳ的数据资本服务量相对误差在0.28%~7.99%,而数据资本存量净额年均余额和数据资本存量总额年均余额的相对误差大多在10%左右。从变动趋势看,考察期内年限Ⅲ资本存量净额年均余额和资本存量总额年均余额的年均增长率分别较基准结果减少0.71和0.61个百分点;相反,年限Ⅳ资本存量净额年均余额和资本存量总额年均余额的年均增速分别较基准结果增加0.59和0.51个百分点;不同服役年限测算的数据资本服务量变动趋势基本一致,但年限Ⅳ比年限Ⅲ测算的数据资本服务量更接近基准结果。

相较价格指数与服役年限,数据资本服务量测算对收益率的敏感性较低。从相对误差看,2001—2013年收益率Ⅴ和Ⅵ测得的数据资本服务量相对误差均小于5%,且2014—2022年相对误差在5.11%~5.54%,误差可忽略不计。从变动趋势看,收益率Ⅴ和Ⅵ的数据资本服务量变动趋势与基准结果一致,收益率下降1个百分点,数据资本服务量年均增速减少0.10个百分点;收益率上升1个百分点,数据资本服务量年均增速相应增加0.09个百分点。

进一步地,分别将价格指数(Ⅰ、Ⅱ)、服役年限(Ⅲ、Ⅳ)及收益率(Ⅴ、Ⅵ)测得的资本服务量代入公式(11),求解数据资本的平均贡献率。^①与数据资本服务量测算结果类似,数据资本的经济贡献对价格指数的选取较为敏感。相较于基准情形,基于指数Ⅰ、Ⅱ测算的数据资本服务量均低估了其对经济增长的贡献;数据资本的经济贡献对服役年限以及收益率的敏感性较低,2006—2010年数据资本服务量平均贡献率较基准情形的波动幅度均小于0.2个百分点。

六、结论与启示

数据资产统计核算问题是亟待突破的国际性、时代性难题。本文遵循SNA核算原则,构建兼具提供财富储备和服务经济运行双重功能的数据全流程测度框架。通过招聘广告和机器学习模型估计各职业数据员工比例和时间使用系数,测算2001—2022年我国数据资本形成、数据资本存量总额与净额及数据资本服务量;并将数据资本、R&D资本、劳动投入以及计算机及通信设备等传统物质资本共同纳入增长核算框架,分析我国经济增长动力来源。研究有以下发现。(1)各职业数据员工比例和时间使用系数存在显著差异。数据员工比例在15%~60%区间的职业占比超五成,此类职业从事数据生产活动多为专业领域工作的辅助环节,其时间使用系数均值约为0.55。(2)考察期内我国数据资本存量和流量增幅显著,年均增速显著超越同期GDP年均增速;数据资本服务量则呈“前期快速扩张、中期高速发展、后期高质量发展”三阶段特征。(3)随着国家大数据发展战略的纵深推进,数据资本对经济增长的贡献潜力逐步显现并趋于主导地位。2020—2022年数据资本对经济增长的平均贡献率高达19.36%,较同期R&D资本和劳动投入的经济贡献形成显著比较优势,标志着我国经济发展范式由“要素堆积”模式向“数据与创新共同驱动”转型。结合上述结论,主要有以下两方面启示。

第一,全面激发数据资本贡献潜力,助力经济高质量发展。研究发现,数据财富性存量和资本服务量均呈现前期迅猛累积、后续稳健增长的鲜明特征,数据资本的经济贡献也日益凸显。因此,在全球数字化转型浪潮席卷、数字经济方兴未艾的背景下,中国亟待全面解锁数据资本的深层价值。一方面,亟须强化跨部门、跨领域数据共享与整合,确立统一的数据标准与规范,并加强数据

① 限于篇幅,正文未报告不同参数测算所得的数据资本服务量对经济增长的平均贡献率,详见线上附录。

加密与安全审计,确保数据得以合理配置与高效运用;另一方面,积极推动数据技术的创新与应用,重点加强人工智能、大数据和云计算等前沿技术的研发投入,赋能企业和社会各界在数据采集、处理、分析与应用等环节的创新实践。

第二,深掘无形资本和劳动力的经济贡献潜力,塑造经济发展新动能、新优势。当前,中国经济已步入高质量发展轨道,传统以物质资本为核心驱动的经济增长模式面临严峻挑战。从增长核算分解结果来看,仪器仪表、金属制品等资本的经济贡献持续走低,通用和专用机械设备、计算机及通信设备等资本经济贡献呈现显著波动性,而数据资本与R&D资本的经济贡献潜力正逐步显现。鉴于此,中国亟须将战略重心转向现代服务业、高端制造业及数字经济新业态,优化资本结构,促进新旧动能接续转换;同时实施数据技能专项培训计划,促进劳动力向技术密集型行业有序流动。倡导“以数据流引领技术流、资金流、人才流、物资流”,打破传统资源要素对经济发展的束缚,为中国经济高质量发展注入不竭动力。

此外,合理公允的估值方法是数据得以充分开发与利用的前提,亟须加强数据资产核算领域的理论与实践探索。本文提出两点促进数据资产统计实务工作的可行性建议。(1)完善我国知识产权产品统计调查与核算制度。厘清数据与各知识产权产品之间的关系,构建包含数据在内的知识产权产品统计核算体系;充分汲取国际经验,积极开展数据等知识产权产品专项调查,并增加价格变动、资产使用寿命等调查内容。推动统计工作更好地服务宏观决策和满足社会需求。(2)统筹推进职业信息网络构建与劳动统计制度改革。整合招聘平台、行业协会及企业会计资料,构建动态更新的职业分类数据库,形成跨部门、跨行业的职业信息共享机制,保障官方统计资料能够准确反映新时代各领域的发展动态和趋势;参照国际经验构建符合我国经济发展实际情况的劳动统计制度。另外,增设数字经济活动类目,重点跟踪新兴岗位薪酬结构变化;细化时间利用调查的内容,增设工时分配情况的专项调查,为数字经济核算提供微观基础。

参考文献:

1. 蔡继明、刘媛、高宏、陈臣:《数据要素参与价值创造的途径——基于广义价值论的一般均衡分析》,《管理世界》2022年第7期。
2. 蔡跃洲、张钧南:《信息通信技术对中国经济增长的替代效应与渗透效应》,《经济研究》2015年第12期。
3. 陈梦根、张鑫:《数字经济要素投入核算框架及应用研究》,《统计研究》2022年第8期。
4. 联合国、欧盟委员会、经济合作与发展组织、国际货币基金组织、世界银行:《2008年国民账户体系》,中国国家统计局国民经济核算司、中国人民大学国民经济核算研究所译,中国统计出版社2012年版。
5. 刘涛雄、戎珂、张亚迪:《数据资本估算及对中国经济增长的贡献——基于数据价值链的视角》,《中国社会科学》2023年第10期。
6. 彭刚、李杰、朱莉:《SNA视角下数据资产及其核算问题研究》,《财贸经济》2022年第5期。
7. 王春云、王亚菲:《数字化资本回报率的测度方法及应用》,《数量经济技术经济研究》2019年第12期。
8. 王维、陈杰、毛盛勇:《基于十大分类的中国资本存量重估:1978~2016年》,《数量经济技术经济研究》2017年第10期。
9. 鲜祖德、王天琪:《中国数字经济核心产业规模测算与预测》,《统计研究》2022年第1期。
10. 徐翔、赵墨非:《数据资本与经济增长路径》,《经济研究》2020年第10期。
11. 许宪春、张钟文、胡亚茹:《数据资产统计与核算问题研究》,《管理世界》2022年第2期。
12. 杨艳、王理、李雨佳、廖祖君:《中国经济增长:数据要素的“双维驱动”》,《统计研究》2023年第4期。
13. 余东华、李云汉:《数字经济时代的产业组织创新——以数字技术驱动的产业链群生态体系为例》,《改革》2021年第7期。
14. 朱发仓、杨诗淳:《基于两种功能的中国R&D资本测度体系研究》,《统计研究》2020年第12期。
15. Ackoff, R. L., From Data to Wisdom. *Journal of Applied Systems Analysis*, Vol.16, No.1, 1989, pp.3-9.

16. Ahmad, N., & Ven, P., Recording and Measuring Data in the System of National Accounts. OECD Conference Centre, November 2, 2018.
17. Asian Development Bank, Capturing the Digital Economy: A Proposed Measurement Framework and Its Applications—A Special Supplement to Key Indicators for Asia and the Pacific. Asian Development Bank Working Paper, No.FLS210307-3, 2021.
18. Calderón, J. B. S., & Rassier, D. G., Valuing the US Data Economy Using Machine Learning and Online Job Postings. BEA Working Paper, 2022.
19. Corrado, C., Haskel, J., Iommi, M., & Jona-Lasinio, C., Measuring Data as an Asset: Framework, Methods and Preliminary Estimates. OECD Economics Department Working Papers, No.1731, 2022.
20. Goodridge, P., & Haskel, J., How Much is UK Business Investing in Big Data? . Imperial College Business School Working Papers, No.25159, 2015.
21. Goodridge, P., Haskel, J., & Edquist, H., We See Data Everywhere Except in the Productivity Statistics. *Review of Income and Wealth*, Vol.68, No.4, 2022, pp.862-894.
22. ISWGNA, Recording of Data in the National Accounts. Twenty-Second Meeting of the Advisory Expert Group on National Accounts, March 28, 2023.
23. ISWGNA, Handbook on Measuring Data. The 56th Session of the United Nations Statistical Commission, March 4, 2025.
24. Japanese Cabinet Office, Report of Research on the Measurement of the Digital Economy toward 2025SNA: Recording Methods for Data as Capital. Economic and Social Research Institute Discussion Paper Series No.88, 2024.
25. Jorgenson, D. W., & Griliches, Z., The Explanation of Productivity Change. *Review of Economic Studies*, Vol.34, No.3, 1967, pp.249-283.
26. OECD, *Measuring Capital—OECD Manual 2009 Second Edition*. Paris: OECD Publishing, 2009.
27. OECD, *Handbook on Deriving Capital Measures of Intellectual Property Products*. Paris: OECD Publishing, 2010.
28. OECD, Measuring Data as an Asset: Framework, Methods and Preliminary Estimates. OECD Economics Department Working Papers, No.1731, 2022.
29. Rassier, D. G., Kornfeld, R. J., & Strassner, E. H., Treatment of Data in National Accounts. Paper Prepared for the BEA Advisory Committee, 2019.
30. Rowley, J., The Wisdom Hierarchy: Representations of the DIKW Hierarchy. *Journal of Information and Communication Science*, Vol.33, No.2, 2007, pp.163-180.
31. Schmidt, J., Pilgrim, G., & Mourougane, A., What is the Role of Data in Jobs in the United Kingdom, Canada, and the United States? A Natural Language Processing Approach. Statistics and Data Directorate Working Paper No. 119, 2023.
32. Statistics Canada, Measuring Investment in Data, Databases and Data Science: Conceptual Framework. Statistics Canada Technical Report, 2019a.
33. Statistics Canada, The Value of Data in Canada: Experimental Estimates. Statistics Canada Technical Report, 2019b.
34. UN, IMF, WB, OECD, & EC, *System of National Accounts 2025*. New York: The United Nations, 2025.
35. Varian, H., Artificial Intelligence, Economics, and Industrial Organization. National Bureau of Economic Research Working Paper, No.24839, 2018.
36. Zhu, F. C., Hu, W. C., & Zhang, C. H., Revisiting China's Net Capital Stock and Fixed Capital Consumption for Ten Types of Assets: An Improved Methodology for 1978-2016. *Transformations in Business & Economics*, Vol.19, No.3C, 2020, pp.542-561.

Statistical Measurement of Data Assets: Based on the System of National Accounts

ZHU Facang, SHI Qiule & SU Weihua (Zhejiang Gongshang University, 310018)

XU Xianchun (China Finance 40 Forum, 100080)

Summary: Since the dawn of the 21st century, the Fourth Industrial Revolution, characterized by

digitalization and intelligent transformation, has rapidly swept across the globe. National strategic initiatives, such as the U.S. Industrial Internet, Germany's Industry 4.0, and Japan's Society 5.0, have all emphasized data as a pivotal driver for economic growth and social progress. However, despite its recognized importance, data remains largely invisible in the System of National Accounts. Addressing this gap through accurate and comprehensive statistical measurement of data is not only essential for effective policy regulation but also a crucial academic and practical challenge in understanding the trend of the times.

To comprehensively elucidate the pivotal role of data as a production factor and its economic impact, this study takes into account its dual functions of wealth accumulation and economic operations, constructs a whole-process measurement framework that covers data capital stock and service flows, and incorporates data capital services into the growth accounting framework to quantify its economic contribution. The research finds that both the net and gross capital stock of data exhibit a three-stage evolution: rapid expansion in the early stage, high-speed development in the middle stage and high-quality development in the later stage, while the data capital services transition from high-speed to moderate-high growth over time. During the study period, both the wealth indicators and service indicators of data capital grew faster than GDP during the same period, highlighting its stronger potential for value accumulation compared with traditional physical capital. The economic contribution potential of data capital has gradually emerged and tends to be dominant. From 2020 to 2022, data capital, R&D capital, and computer and communication equipment together contributed an average of 54.57% to economic growth, effectively offsetting the negative drag from the real estate and construction sectors.

Based on these findings, this paper puts forward the following policy suggestions. First, optimize the capital structure with data at the core to accelerate the transition from old to new growth drivers. Second, leverage intangible capital and labor force to unlock their economic potential and foster new drivers and new advantages for economic development. Third, develop sound and fair valuation methods for the full development and utilization of data assets and strengthen theoretical research in data asset accounting. It is imperative to improve the statistical investigation and accounting system for data assets, and advance the construction of the vocational information network and the reform of the labor statistics system in a coordinated manner.

Keywords: System of National Accounts, Data Assets, Statistical Measurement

JEL: E01, E22

责任编辑:原 宏