

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

刘涛雄 戎珂 张亚迪

摘要：随着数字经济的蓬勃发展，数据资本对中国经济增长的重要性日益显现。数据经过采集、清洗、存储、加工等各环节，不断积累价值形成数据资本。基于数据价值链的视角，使用成本法和增值法估算价值层面的数据资本形成额与数据资本存量，用存储规模法估算物理层面的数据规模，基于数据资本估算结果利用省级面板模型估计其对经济增长的贡献，可以发现：2003—2020年间，我国数据资本形成额、数据资本存量在总量和人均上均实现大幅度增长，增速明显快于GDP增速；2011年后，数据资本对经济增长的产出弹性和对经济增长率的贡献均明显超过之前阶段，已成为中国经济增长重要动能之一。

关键词：数据资本 数据价值链 增值法 经济增长

作者刘涛雄，清华大学社会科学学院教授（北京 100084）；戎珂，清华大学社会科学学院教授（北京 100084）；张亚迪，北京交通大学经济管理学院助理教授（北京 100044）。

引言

随着数字经济的蓬勃发展，数据更加深入地参与到各类生产活动之中。数据可以在研发部门参与知识的生产，也可以用于生产部门提升生产效率，^① 被称为“数字经济时代的石油”。数据是指利用一定的信息技术手段所形成的对信息的数字化记载。^② 物理意义的数据经过采集、清洗与存储、加工等步骤，逐步形成具有生产价值的数据。从经济学视角来讲，具有生产价值的数据为“数据资本”，其积累的存量为“数据资本存量”，每年新形成的数据资本则为“数据资本形成额”。要进一步证

① W. C. Lin, D. Xie and L. Zhang, “Knowledge Accumulation, Privacy, and Growth in a Data Economy,” *Management Science*, vol.67, no.10, 2021, pp.80-92; C. I. Jones and C. Tonetti, “Nonrivalry and the Economics of Data,” *American Economic Review*, vol.110, no.9, 2020, pp.19-58.

② M. Farboodi and L. Veldkamp, “A Growth Model of the Data Economy,” National Bureau of Economic Research Working Papers, no.28427, 2021.

实和度量数据在经济活动中的作用，一个基础性工作就是理清一个经济体中到底有多少数据资本。因此，理解数据资本形成的过程，科学度量我国数据资本的存量、分布及走势，进而客观评价其对经济增长的贡献，引导其更好地发挥经济价值，在理论或实践层面都已成为一个十分重要和迫切的课题。

当前，国内数据资本的估算尚处于探索阶段，鲜有公开发表的估算成果。现有研究借鉴关于物质资本、人力资本、基础设施资本等的已有方法，^①认为数据资本估算可采用市场法、收益法和成本法等三类方法。^②市场法基于与数据相关的同类物品的近期市场交易价格或者市场价格等价物来估算数据资本形成额。但当前数据要素市场并不成熟，可供参考的市场交易价格甚至市场价格等价物难以获取。收益法基于数据在未来能够获取到的预期收益，通过折现来估算数据资本。而当前数据缺乏成熟数据要素市场的支撑，实际操作过程中对数据预期收入的估算存在诸多不确定性，无法进行准确预测。成本法基于数据价值获取过程中的各类成本来估算数据资本形成额。从数据的可获取性以及既有实践来看，成本法是现阶段对数据资本形成额进行估算最为可行的方法。目前加拿大统计局（Canada Statistics）以及美国商务部经济分析局（BEA）的相关估算都是基于成本法，中国资产评估协会以及中国信息通信研究院也同样建议采纳成本法作为数据资本估算的主要方法。

国际上关于数据资本的估算也处于探索阶段，少量可参考的工作都是基于成本法。加拿大统计局在2019年^③和美国BEA在2022年^④的估算逻辑类似，都是通过对数据价值链的分析，分别估算从事数据价值链中相关工作的就业人员的劳动力成本以及生产数据资本的非劳动力成本，进而用每一年总和的成本去度量新增的数据资本形成额。但目前各国都没有关于数据生产所消耗的劳动力成本和资本成本的调查统计，因而必须借助其他方法。其中，美国BEA在劳动力成本估算方面相对精细一些，采用机器学习对不同职业与数据的相关程度进行了估算，从而确定这些职业的劳动力成本中有多大比例可以算入数据相关的劳动力成本之中。

① 参见张军、吴桂英、张吉鹏：《中国省际物质资本存量估算：1952—2000》，《经济研究》2004年第10期；金戈：《中国基础设施与非基础设施资本存量及其产出弹性估算》，《经济研究》2016年第5期；李海崢等：《中国人力资本测度与指数构建》，《经济研究》2010年第8期。

② 参见许宪春、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；M. Reinsdorf and J. Ribarsky, “Measuring the Digital Economy in Macroeconomic Statistics: The Role of Data,” International Monetary Fund, 2019.

③ Statistics Canada, “The Value of Data in Canada: Experimental Estimates,” <https://www150.statcan.gc.ca/nl/pub/13-605-x/2019001/article/00009-eng.htm>, July 10, 2019.

④ J. B. S. Calderón and D. G. Grassier, “Valuing the US Data Economy Using Machine Learning and Online Job Postings,” BEA Working Paper, 2022.

现有成本法也存在明显不足。首先,从数据价值形成的规律看,在数据生产的不同阶段,同样成本所积累形成的价值量可能并不相同,比如从事数据采集和从事数据分析各自耗费的一个单位劳动对数据价值量的影响可能十分不同,现在对中国数据资本估算的讨论都忽略了这一点。其次,更重要的是,数据资本估算的基本逻辑是价值量估算,采用成本法进行估算的前提是数据生产的成本大体和数据的价值量相当,但这一点难以实现。一个突出问题是,数据的价值和数据分析的技术水平密切相关,在数字经济发展早期,算力和算法水平相对较低,用成本法估算的数据资本,其价值明显会被低估。再次,实际估算工作中也面临数据来源不足的问题。比如,相比发达国家,我国关于数据生产的劳动力成本的统计数据基本没有直接的统计资料,现有文献没有针对我国的情况提出可行的解决方案。最后,除了在价值层面估算数据资本,在物理层面对数据存储规模进行估算同样有必要,现有研究忽视了这一点。对物理层面的数据资本进行估算,可以最终获取单位存储的数据资本价值,从而为数据要素市场的构建提供更好支撑。由以上可见,基于现有框架对中国数据资本进行估算,难以得到关于数据资本存量的科学度量,进而难以得到关于数据资本对经济增长贡献的客观评价。特别是,很可能低估数据资本存量,进而低估其对经济增长的贡献。

本文将数据价值链划分为数据采集、数据清洗与存储、数据加工三个环节,基于三个环节构建数据生产的经济学模型,论证成本法的经济学原理及其存在的问题。进而基于价值链的视角,考虑数据分析技术发展带来的数据价值增值,在成本法基础上提出了增值法。增值法可以更直接地反映数据要素的本质特征,即数据分析技术的进步会提升数据的价值,解决成本法会低估数据资本存量这一问题。本文利用成本法、增值法及存储规模法等三种方法,从价值层面和物理层面对我国数据资本存量做出系统估算,刻画其随时间的动态演进特征。其中,本文采用机器学习的方法,基于我国职业大典中各类职业的详细描述,估算了数据生产三个环节中各个职业分配给数据相关工作的时间比例,对三个环节中的劳动力成本进行了详细估算。使用分省数据资本存量的估算结果,本文对中国经济增长过程中数据资本的产出弹性及其对中国经济增长率的贡献率进行实证检验,发现在数字经济快速发展的2011—2019年,数据要素无论在经济增长的产出弹性还是对经济增长率的平均贡献率上,均已成为中国经济增长重要的动能。本文研究发现,过往估算所采用的简单成本法存在低估数据资本贡献的可能。

一、基于价值链的数据资本估算方法

(一) 数据资本和数据价值链

在经济活动中,数据经过采集、清洗、存储、加工等各个环节,其规模和价值

不断积累,形成数据资本。数据价值链反映数据资本形成过程中所经历的环节。一些学者、国家及国际组织均对数据价值链进行定义,比如经济合作与发展组织(OECD)基于个人数据将数据价值链分解为数据的“采集/授权”“储存与聚合”“分析与销售”和“利用”等四个环节;^①美国 BEA 认为数据要素参与价值创造的价值链可以被分为数据的“采集”“存储”“加工”“销售”和“利用”等五个环节;^②加拿大统计局同样基于数据要素参与社会生产活动的过程提出“信息价值链”,可以被划分为“观测”“数据”“数据库”和“数据科学”四个环节。^③类似地,国内最新研究在探讨数据要素价值的估算逻辑时,将数据价值链划分为了“数据收集”“数据存储”和“数据分析”三个环节。^④尽管这些研究对数据价值链的划分各有差异,但是基本逻辑相通,均考虑了从原始数据的采集和收集到数据的分析利用这一完整的流程。

基于数据资本积累过程中的重要步骤,借鉴加拿大统计局、美国 BEA 等研究,本文将数据价值链划分为了数据采集、数据清洗与存储、数据加工三个阶段,见图 1。其中,数据采集将原始信息转换为数字形式的观察结果,可以储存、传输或处理。数据清洗与存储是对采集到的数据以一定的标准进行初步处理,并有组织地进行数据存储,以方便后续的检索和操作。数据加工指的是基于存储的数据,有系统地进行创造性活动,将初步处理的数据加工成表征一定模式的,可以更好地在各种数字经济场景中发挥作用的数据。

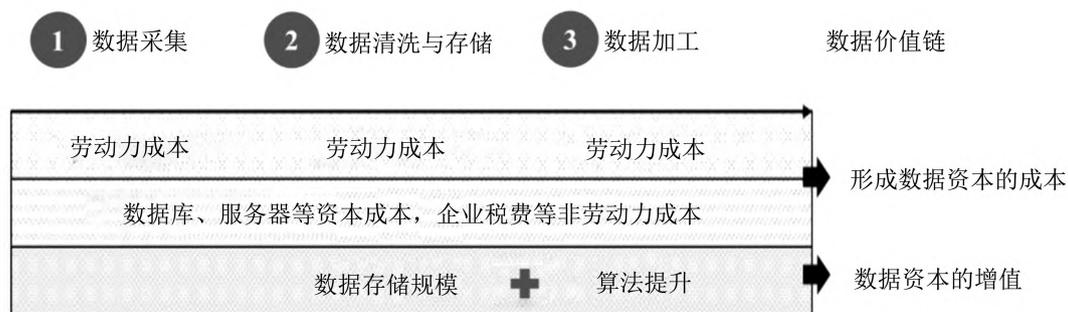


图 1 数据价值链与数据资本

- ① OECD, “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value,” Paris: OECD Publishing, 2013.
- ② R. Kornfeld, “Measuring Data in the National Accounts,” BEA Advisory Committee Meeting, 2019.
- ③ Statistics Canada, “The Value of Data in Canada: Experimental Estimates,” <https://www150.statcan.gc.ca/nl/pub/13-605-x/2019001/article/00009-eng.htm>, July 10, 2019.
- ④ 参见许宪春、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。

形成数据资本的主体是从事数据相关业务的企业。根据传统企业成本的核算范围，数据资本形成的成本可以拆解为：在数据采集、数据清洗与存储、数据加工三个环节中发生的一系列劳动力成本，以数据库、服务器、存储设备等为代表的资本成本，和企业税负及其他费用成本。其中，资本成本、企业税负及其他费用成本这两块成本可合称为非劳动力成本。从数据本身而言，在数据采集环节，主要是形成数据规模；在数据清洗与存储环节，主要是提升数据质量；而在数据加工环节，主要是融入了数据分析技术。^① 数据资本价值的形成，不仅仅来自各环节直接消耗的成本，另外一个重要来源是数据分析技术的进步会直接提升数据的价值。尤其在第三个环节数据技术提升的过程中，不同创新性企业针对数据加工的算法可以扩散到整个数据相关行业，让已存储的数据更有效地应用于更多经济场景之中，从而发挥出更大的价值。

基于图 1，本文的三种估算方法分别对应如下范畴。第一，成本法仅针对形成数据资本的成本；第二，增值法对应形成数据资本的成本和数据资本的增值；第三，存储规模法对应于数据存储规模。下一小节通过引入一个简单的经济学模型，介绍数据资本估算方法的理论逻辑。

（二）数据资本估算方法

1. 成本法

首先，假定市场中一共存在从事数据资本积累相关经济活动的同质化企业 N 家。为了便于分析，进一步假定每一家企业的经济活动均涉及数据价值链的三个环节。^② 企业 i 在数据采集、数据清洗与存储、数据加工三个环节的相关劳动力投入分别为 l_{i1} ， l_{i2} ， l_{i3} ，投入的资本为 k_i 。则可以假定企业 i 每期的数据资本形成额 V_i 有以下柯布一道格拉斯函数形式：

$$V_i = A_i k_i^\alpha l_{i1}^\beta l_{i2}^\gamma l_{i3}^{1-\alpha-\beta-\gamma} \quad (1)$$

就式 (1) 而言， l_{i1}^β 代表了在数据规模形成过程中的劳动力贡献， l_{i2}^γ 代表了在数据质量提升过程中的劳动力贡献， $l_{i3}^{1-\alpha-\beta-\gamma}$ 代表了数据技术提升过程中的劳动力贡献；同时， k_i^α 则代表了整个数据资本积累过程中资本所作出的贡献。

基于式 (1)，本文假定，企业 i 在数据采集、数据清洗与存储、数据加工三个环节的相关劳动力投入，对应的市场劳动力单位成本（即工资）分别为 s_1 ， s_2 ， s_3 。同时，投入的资本所对应的资本利率为 r ，那么可以进一步得到企业 i 在数据资本形

① 值得注意的是，数据价值链的三个环节并不是严格独立的。比如，在数据采集环节也可能存在数据质量的提升或者数据技术的提升，本文对这三个环节的划分主要还是为了便于后文构建经济学模型，以更好地从经济学理论层面分析数据资本的形成过程。

② 现实中可能存在很多企业仅涉及数据价值链的某一个环节，本文的这一假设便于后续分析中使用一个统一的生产函数。

成过程中所花费的成本 c_i 如下:

$$c_i = s_1 l_{i1} + s_2 l_{i2} + s_3 l_{i3} + r k_i \quad (2)$$

若企业在数据生产过程中其所追求的目标为利润最大化, 将数据资本的单位价值标准化为 1 后, 可以得到企业 i 的最优化问题:

$$\max_{l_{i1}, l_{i2}, l_{i3}, k_i} A_i k_i^\alpha l_{i1}^\beta l_{i2}^\gamma l_{i3}^{1-\alpha-\beta-\gamma} - (s_1 l_{i1} + s_2 l_{i2} + s_3 l_{i3} + r k_i) \quad (3)$$

如果进一步假定数据相关的市场是一个企业可以自由进入的完全竞争市场, 那么当同质化的企业在市场中有正利润时, 便会有其他企业不断地进入, 直到每个企业最终的利润为 0。假设在最终均衡时, 市场中的企业数量为 N , 那么利润为 0 就意味着每家企业都有:

$$V_i = c_i = s_1 l_{i1} + s_2 l_{i2} + s_3 l_{i3} + r k_i \quad (4)$$

若 l_1, l_2, l_3 分别为所有企业用于数据采集、数据清洗与存储、数据分析三个环节的相关劳动力总投入, k 为各企业总的资本投入。由于假定了企业是同质的, 必然有 $A=A_i, l_1=N l_{i1}, l_2=N l_{i2}, l_3=N l_{i3}, k=N k_i$ 。通过对式 (3) 的求解, 可以得到:

$$\begin{aligned} \alpha &= \frac{r k_i}{s_1 l_{i1} + s_2 l_{i2} + s_3 l_{i3} + r k_i} = \frac{r k}{s_1 l_1 + s_2 l_2 + s_3 l_3 + r k} \\ \beta &= \frac{s_1 l_{i1}}{s_1 l_{i1} + s_2 l_{i2} + s_3 l_{i3} + r k_i} = \frac{s_1 l_1}{s_1 l_1 + s_2 l_2 + s_3 l_3 + r k} \\ \gamma &= \frac{s_2 l_{i2}}{s_1 l_{i1} + s_2 l_{i2} + s_3 l_{i3} + r k_i} = \frac{s_2 l_2}{s_1 l_1 + s_2 l_2 + s_3 l_3 + r k} \end{aligned} \quad (5)$$

于是, 数据资本形成的总额 V 便可用总成本衡量为:

$$V = \int_0^N V_i di = A k^\alpha l_1^\beta l_2^\gamma l_3^{1-\alpha-\beta-\gamma} = s_1 l_1 + s_2 l_2 + s_3 l_3 + r k \quad (6)$$

引入年份 t 后, 通过对每一年 l_{1t}, l_{2t}, l_{3t} 以及 k_t 的估算以及相应的 s_{1t}, s_{2t}, s_{3t} 和 r_t 的估算, 即可得到测量区域内每一年的数据资本形成额 V_t 。

之后可利用永续盘存法估算数据资本存量 D_t 。首先, 使用价格指数将所得到的历年数据资本形成额 V_t 换算成基年不变价。因没有统计对口的“数据资本形成额价格指数”, 考虑到数据资本形成过程中, 涉及了劳动力成本、资本成本等, GDP 平减指数在一定程度上反映了这些成本的变化, 因此本文选取 GDP 平减指数作为价格指数。之后, 假定借在经济稳态增长时, 数据资本存量增长率等于数据资本形成额的增长率,^① 那么只需用折旧率与数据资本形成额的增长率之和去除基年($t=0$) 的数据资本形成额, 就能得到基年 ($t=0$) 年初的数据资本存量。

若数据资本的使用寿命为 τ , 且残值率为 0, 本文直接采用平均年限法估计数据资本的折旧率为 $\delta_v = 1/\tau$ 。同时, 求得基年数据资本形成额 V_0 到第 10 年数据资本形

① 参见单豪杰:《中国资本存量 K 的再估算: 1952~2006 年》,《数量经济技术经济研究》2008 年第 10 期; 金戈:《中国基础设施与非基础设施资本存量及其产出弹性估算》,《经济研究》2016 年第 5 期。

成额 V_0 的几何年均增长率 \bar{g}_v 。本文确定基年年末的数据资本存量 D_0 如下：

$$D_0 = \frac{V_0}{\delta_v + \bar{g}_v} + V_0 \quad (7)$$

相应地，后续年份的数据资本存量则为：

$$D_t = D_{t-1}(1 - \delta_v) + V_t, \quad t \in N^+ \quad (8)$$

2. 增值法

数据资本形成过程中一个十分重要的特征便是，数据的价值是十分依赖算法等数据分析技术水平的，算法的进步显然会带来数据价值的增加。以电商应用场景为例，数据技术的不断提升可以让企业更好地利用所采集的用户个人数据和行为数据，更好地对用户的购物偏好进行预测，从而更好地服务于电商平台上买卖双方的匹配。这一点是十分重要的，实际上数据早已有之，但数据的巨大价值则是近年来随着数据分析技术的巨大进步才被认识到的。而且，算法等数据分析技术作为知识具有很强的正外部性，一个企业的算法提升可以扩散至整个行业。比如，一个电商平台数据分析技术的提升可以被模仿和外溢到各类不同的电商平台，让整个电商行业的数据价值得到进一步提升。尽管式 (1) 代表了数据分析技术的劳动力投入 l_{i3} 会创造价值，但该式没有考虑一个企业数据分析技术的进步对整个社会的数据价值通过知识外溢而产生的增值作用。因此，用基本的成本法进行估算在一定程度上低估了数据资本形成额。

因此，本文在前述成本法的基础上，提出一种简洁的、将算法等数据技术进步所带来的数据价值提升包含进来的数据资本估算方法，并称之为“增值法”。考虑在式 (1) 中， $l_{i3}^{1-\alpha-\beta-\gamma}$ 代表了企业 i 数据技术提升过程中的企业自身劳动力的贡献。如果假设个别企业的数据技术的提升可以一定程度上扩散至全行业，则企业 i 实际的数据技术提升应该和全行业的数据分析劳动力投入相关，即在式 (1) 中将 $l_{i3}^{1-\alpha-\beta-\gamma}$ 换成 $(\xi l_3)^{1-\alpha-\beta-\gamma}$ 的形式。其中， l_3 是整个行业在数据加工环节的劳动力投入， ξ 则反映了可以扩散至全行业的数据技术的比例。考虑数据要素的这一特征后，企业 i 形成数据资本 V'_i 的过程需由式 (1) 调整为：

$$V'_i = A'_i k_i^\alpha l_{i1}^\beta l_{i2}^\gamma l_3^{1-\alpha-\beta-\gamma} \quad (9)$$

其中， $A'_i = \xi^{1-\alpha-\beta-\gamma} A_i$ ，仍为一个常数。

单个企业在决策时仍为求解第 (3) 式。对于 α 、 β 、 γ 仍有表达式 (5)。在成本法的估算过程中，可以得到 l_1 ， l_2 ， l_3 ， k 以及对应的 s_1 ， s_2 ， s_3 ， r 。从而可以通过式 (5) 计算得到 α ， β 和 γ 的值。之后，便可以基于式 (9)，进一步求得测量区域内数据资本形成额 V' ：

$$V' = \int_0^N V'_i di = A' N^{1-\alpha-\beta-\gamma} k^\alpha l_1^\beta l_2^\gamma l_3^{1-\alpha-\beta-\gamma} \quad (10)$$

与式 (6) 相比，可以发现在同样劳动力和资本投入下，考虑算法进步的数据资本形成额 V' 要大于没考虑算法进步的资本形成额 V ，且市场中企业的数量越多，产

生的技术外溢越大, 考虑算法进步的资本形成额 V' 也越大。引入年份 t , 那么第 t 年的价值量则为 V'_t , 对比式(6)和式(10), 可以得到考虑算法进步的数据资本形成额的增长率 g_t 如下:

$$g_t = \frac{dV'_t}{V'_t} = (1 - \alpha_t - \beta_t - \gamma_t) \frac{dN_t}{N_t} + \frac{dV_t}{V_t} \quad (11)$$

如果基期年的数据资本形成额 V'_0 采用由成本法求得的 V_0 , 那么可以得到第 t 年基于算法进步的数据资本形成额为:

$$V'_t = \prod_0^t (1 + g_t) V_0, \quad t \in N^+ \quad (12)$$

类似地, 同样可以根据成本法中介绍的永续盘存法, 求得考虑算法进步后的历年数据资本存量 D'_t 。

3. 存储规模法

以上两种方法主要是从数据资本形成额的视角开展估算, 而对数据在物理意义上的存储规模进行估算同样很有必要。本质上, 如果一种投入的要素虽然其价值很大, 但是其数量很少, 那么这种要素很难成为可以在经济活动中大规模应用的生产要素。在农业经济时代, 尽管也有一些生产工具被应用于农业生产过程之中, 但是劳动力始终是最主要的生产要素; 到了工业经济时代, 机器和厂房被大规模地生产和建设, 物质资本成为经济发展过程中至关重要的生产要素。同样, 在数字经济时代, 数据要素能起到重要作用的一个基础便是数据能被大规模地存储和应用。

存储规模法主要从数据本身客观被存储在各种介质中的规模或者容量出发, 去估算数据资本的水平。比如一个16GB的U盘中储存了8GB的各类数据, 那么这个U盘的数据存储规模便为8GB。本文继续使用永续盘存法估算各年数据存储规模。假定历年数据存储形成规模为 d_t , 从基年至第10年的数据存储形成规模的几何平均增速为 \bar{g}_d , 数据存储的折旧率为 δ_d , 则基年年末的数据存储规模 S_0 为:

$$S_0 = \frac{d_0}{\bar{g}_d + \delta_d} + d_0 \quad (13)$$

后续第 t 年的数据存储规模 S_t 为:

$$S_t = S_{t-1} (1 - \delta_d) + d_t, \quad t \in N^+ \quad (14)$$

二、中国数据资本估算过程与结果

基于上述部分介绍的三种估算理论与方法, 本文对中国各省份2003—2020年的数据资本存量进行具体估算, 并展示基本的估算结果。

(一) 成本法估算

成本法需要分别估算数据资本形成过程中的劳动力成本和非劳动力成本。首先

对数据资本形成过程中的劳动力成本进行估算。本文先估算各职业分配给数据相关工作的时间，并通过各地区分行业就业人员和工资数据来计算劳动力成本。

由于不是所有的职业都和数据生产相关，与数据相关的职业也不一定全时从事数据相关工作，因此需要基于各职业分配给数据相关工作的时间来估算劳动力成本。若职业 ω 分配给数据相关工作的时间为 τ_ω ，该职业在第 t 年的就业人数为 $h_{\omega,t}$ ，单位时间的工资为 $w_{\omega,t}$ ，那么数据资本形成过程中的劳动力成本 cl_t 可以表示为：

$$cl_t = \sum \tau_\omega w_{\omega,t} h_{\omega,t} \quad (15)$$

本文参考美国 BEA 的做法，基于《中华人民共和国职业分类大典》（以下简称《职业大典》）中的职业和职业描述来估算相关职业分配给数据相关工作的时间。^①《职业大典》不仅包括了我国全部的职业分类，还描述了各个职业的职责范围和工作内容，对相关职业的定义均以最简练的语句、统一的表达方式进行描述，有利于做接下来的文本训练。基于前文提出的数据价值链，本文将数据生产活动相关人员定义为，从事任何涉及数据收集、数据清洗和存储、数据分析等工作的人员。结合《职业大典》中有关职业分类的说明，本文界定了 440 个数据相关的职业细类，分属于 159 个职业小类、19 个中类。^②

本文通过使用《职业大典》中的各个职业细类的职业信息描述来估算数据相关的工作时间 τ_ω 。首先，根据工作内容的描述确定从事数据相关活动比例最高的职业。通过给数据相关职业打分排序，参考美国 BEA 的方法，选出 20 个职业作为测试集，并标记为全时从事数据相关工作。本文选择使用 doc2vec 模型^③进行深度学习，通过计算相关职业描述与测试集中的职业描述的文本的相似度，来估算数据工作时间占比。经过对语料进行了基本的分词、分句以及降噪等预处理后，形成深度学习的基础数据。进一步，通过计算训练集与测试语料的余弦相似度来计算数据相关的工作时间 τ_ω ，设 X 、 Y 分别为测试集和训练集的职业描述特征向量，计算公式如下：

$$\tau_\omega = \cos(X, Y) = \frac{X * Y}{\|X\| \|Y\|} = \frac{\sum_{i=1}^n (X_\omega * Y_\omega)}{\sqrt{\sum_{i=1}^n (X_\omega)^2} * \sqrt{\sum_{i=1}^n (Y_\omega)^2}} \quad (16)$$

基于从事数据相关工作的从业人数、工资和工作时间占比可以得到数据资本的劳动力成本，其计算公式为：

$$cl_t = \sum \left[\left(\frac{\sum_{i=1}^n (X_\omega * Y_\omega)}{\sqrt{\sum_{i=1}^n (X_\omega)^2} * \sqrt{\sum_{i=1}^n (Y_\omega)^2}} \right) w_{\omega,t} h_{\omega,t} \right] \quad (17)$$

① 参见国家职业分类大典修订工作委员会编：《中华人民共和国职业分类大典》，北京：中国劳动社会保障出版社、中国人事出版社，2015 年。

② 由于篇幅所限，具体数据相关的职业读者如有兴趣，可向作者索取。

③ Q. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," *Proceedings of the 31st International Conference on Machine Learning, PMLR*, vol. 32, no. 2, 2014, pp. 1188-1196.

由于本文计算的职业为《职业大典》中的细类，而省级层面的从业人数和工资数额数据为中类职业，为了计算各省份的劳动力成本，需要得到各职业中类中各细类的占比。本文利用2004年、2008年、2013年、2018年的《中国经济普查年鉴》中的就业人数得到各细类在中类中的占比。各地区分行就业人员和工资数据来源于2003—2020年《中国劳动统计年鉴》，经过计算得到省级层面数据资本的劳动力成本。采用同样的方法，可分别计算数据价值链中三个阶段的劳动力成本。

数据资本的价值链中除了劳动力成本，还包括非劳动力成本，本文分别采用两种方法对非劳动力成本进行估算。

第一，非劳动力成本按比例估算。假定非劳动力成本与劳动力成本的比重为一常数，从而得到数据资本形成额。加拿大统计局在调查数据的基础上采用的是非劳动力成本为劳动力成本的50%。本文在中国与数据相关的上市公司^①财务年报中，发现这些公司的固定资产相关折旧等成本与应付职工薪酬的比例平均在0.45左右。因此可以认为数据相关企业的资本成本约为劳动力成本的45%。进一步，根据中国财政科学研究院“企业成本”调查问卷结果，企业的税负及其他费用成本约为企业总体成本的14%。鉴于非劳动力成本由资本成本、税负及其他费用成本构成，经计算可知，在中国非劳动力成本约为劳动力成本的70%。按此可计算得到每年、每个省份的数据资本形成额。

第二，非劳动力成本分项估算。具体到不同行业、不同地区时，数据相关的非劳动力成本和劳动力成本的比例可能存在较大差异。因此，有必要考虑行业和地区的异质性，以得到更为精确的非劳动力成本估算。在上述劳动力成本估算的过程中，本文已经获得了历年每个省份每个行业中，数据相关劳动力成本 c_l 占总体劳动力成本 c_L 的比例。根据《中国固定资产投资统计年鉴》，可计算历年分省分行业的物质资本存量，进而能计算得到历年分省分行业的物质资本折旧额，该折旧即为总体资本成本，设为 c_K 。接下来，需要对总体资本成本中用于数据生产的资本成本 c_k 的比重进行估计。假定数据相关劳动力成本 c_l 占总体劳动力成本 c_L 的比例和数据相关资本成本 c_k 占总体资本成本 c_K 的比例存在如下关系：

$$\frac{c_k}{c_K} = \lambda \frac{c_l}{c_L} \quad (18)$$

在非劳动力成本按比例估算中，已经得到 $\frac{c_k}{c_K} = 0.45$ 。与此同时，每个行业总体劳动力成本 c_L 与总体资本成本 c_K 已知，通过求解便可得到 $\lambda = \frac{c_k/c_K}{c_l/c_L}$ 。将每个行业的总体劳动力成本 c_L 与总体资本成本 c_K 代入后，最终得到 λ 的行业平均取值为0.22。随后，可计算得到历年各省各行业中数据相关资本成本占总体资本成本的比

^① 这些上市公司包括南威软件、汇纳科技、初灵信息、华测导航等。

例，进而可得到各省各年用于数据生产的资本成本。最后，和非劳动力成本按比例估算中的做法一致，取除劳动力成本和资本成本以外的其余成本为劳动力成本的 25%。

估算发现两种方法估计出来的非劳动力成本几乎一致。其中在 2015 年之后，按比例估算的非劳动力成本略低于按照分项估算的非劳动力成本。可能的解释是，随着数据分析技术的提升，对算力的要求也不断提升，造成相关的资本成本比重实际上也在逐年提升。因此，按不变的比例估算出来的非劳动力成本可能会在前期的年份高估数据的资本成本，在后期的年份低估数据的资本成本。综合考虑后，本文后续的估算主要基于按分项估算的结果。

加总劳动力成本、资本成本和其余成本后，便可得到各年各省份的数据资本形成额，然后根据永续盘存法可计算得到相应数据资本存量。为了估算更加准确，本文对数据价值链不同阶段的劳动力成本、非劳动力成本分别假设了折旧率。沿用加拿大统计局的假定，数据资本的残值为 0，对各阶段劳动力所形成数据资本的使用寿命进行假定。其中第一阶段所形成数据资本的使用寿命主要指企业采集的数据被使用的平均时间长度，本文根据中国的情况假定为 10 年。^① 第二阶段和第三阶段采用加拿大统计局设定的 5 年和 6 年。采用平均年限法，可以得到三个阶段劳动力成本所形成数据资本的折旧率分别为 0.10、0.20 和 0.17。同时，鉴于每个阶段都会发生非劳动力成本，因此将非劳动力成本所形成数据资本的折旧率，定为这三个阶段劳动力成本所形成数据资本存量的折旧率的加权平均值。表 1 展示了基于成本法的数据资本估算结果，中国数据资本形成额从 2003 年的 3222.06 亿元增长至 2020 年的 30009.81 亿元，年均增长率为 14.03%；中国数据资本存量则从 2003 年的 14857.42 亿元增长至 2020 年的 132828.84 亿元，年均增长率为 13.75%。此外，数据资本存量与 GDP 的比例从 2003 年的 0.11 逐年波动增加至 2020 年的 0.21。而数据资本存量与数字经济增加值的比例则从 2007 年的 1.88 倍增加至 2.57 倍。

表 1 2003—2020 年基于成本法的数据资本估算结果（2003 年不变价）

| 年份 | 数据资本形成额 (亿元) | 数据资本存量 (亿元) | 数据资本存量/ GDP | 数据资本存量/ 数字经济增加值 |
|------|-----------------|----------------|----------------|--------------------|
| 2003 | 3222.06 | 14857.42 | 0.11 | |
| 2004 | 3524.73 | 16367.76 | 0.11 | |

① 加拿大统计局考虑到数据在一代人内保持其价值，假定其平均被使用的时间长度为 25 年。但是数据不仅有被使用的寿命，其使用价值也会随着时间的推移发生衰减。通过与腾讯、华为等企业的沟通，一般情况下，大部分十年前采集的数据会被归为“冷数据”进行存储，即不会再被大量地调用。考虑到中国数字经济的发展水平是高于加拿大的，对数据的利用周期也更快，因此本文将第一阶段的寿命假定为 10 年。

续表 1

| 年份 | 数据资本形成额 (亿元) | 数据资本存量 (亿元) | 数据资本存量/ GDP | 数据资本存量/ 数字经济增加值 |
|------|-----------------|----------------|----------------|--------------------|
| 2005 | 4021.44 | 18177.69 | 0.10 | |
| 2006 | 4762.58 | 20489.48 | 0.10 | |
| 2007 | 5568.88 | 23298.34 | 0.10 | 1.88 |
| 2008 | 6353.20 | 26514.67 | 0.10 | 1.89 |
| 2009 | 7472.22 | 30416.86 | 0.10 | 2.08 |
| 2010 | 8406.86 | 34727.23 | 0.10 | 2.18 |
| 2011 | 9594.44 | 39645.76 | 0.10 | 2.33 |
| 2012 | 11135.68 | 45444.24 | 0.11 | 2.51 |
| 2013 | 14106.57 | 53425.10 | 0.12 | 2.51 |
| 2014 | 16052.21 | 62258.48 | 0.13 | 2.62 |
| 2015 | 18346.94 | 72178.68 | 0.15 | 2.71 |
| 2016 | 20281.25 | 82676.31 | 0.16 | 2.79 |
| 2017 | 22456.26 | 93913.77 | 0.16 | 2.57 |
| 2018 | 24733.12 | 105888.31 | 0.17 | |
| 2019 | 27557.82 | 119038.90 | 0.18 | |
| 2020 | 30009.81 | 132828.84 | 0.21 | |

资料来源：GDP 数据来自 2003—2019 年《中国统计年鉴》（北京：中国统计出版社）。数字经济增加值数据来自许宪春、张美慧《中国数字经济规模测算研究——基于国际比较的视角》（《中国工业经济》2020 年第 5 期）。其余数据为本文估算结果。

（二）增值法估算

根据式（10），增值法数据资本形成额增长率需要在成本法的基础上，额外考虑数据相关企业的增长率。通过 IT 桔子数据库，^① 本文获取了 2003—2020 年各省与数据相关的新增企业和累计企业数，进而求得相应的企业数增长率 $\frac{d N_t}{N_t}$ 。然后，基于式（5）求得 2003—2020 年每个省份的 α_t 、 β_t 、 γ_t 。将 α_t 、 β_t 、 γ_t 、 $\frac{d N_t}{N_t}$ 、 $\frac{d V_t}{V_t}$ 代入式（11）后，最终得到基于增值法的数据资本形成额增长率 g_t 。其中根据式（10），因算法提升而增值的数据资本形成额由第三阶段数据加工的劳动所引致，因此将其折旧率定为成本法中第三阶段劳动力所形成数据资本的折旧率。根据式（11）、式（12），

^① IT 桔子数据库收纳了中国历年泛互联网相关的企业，可以认定为本文所指的数据相关企业。

即可求得增值法下各省份 2003 年至 2020 年的数据资本存量。表 2 展示了估算结果，中国数据资本形成额从 2003 年的 3222.06 亿元增长至 2020 年的 44423.70 亿元，年均增长率为 16.69%；中国数据资本存量则从 2003 年的 14880.48 亿元增长至 2020 年的 174137.67 亿元，年均增长率为 15.57%。中国数据资本存量和 GDP 的比例从 2003 年的 0.11 逐年增加至 2020 年的 0.27，而数据资本存量和数字经济增加值的比例则从 2007 年的 1.96 倍增加至 2020 年的 3.23 倍。总体而言，增值法的估算下，数据资本形成额与数据资本存量的增长率均快于成本法的估算。

表 2 2003—2020 年基于增值法的数据资本估算结果 (2003 年不变价)

| 年份 | 数据资本形成额 (亿元) | 数据资本存量 (亿元) | 数据资本存量/ GDP | 数据资本存量/ 数字经济增加值 |
|------|-----------------|----------------|----------------|--------------------|
| 2003 | 3222.06 | 14880.48 | 0.11 | |
| 2004 | 3627.52 | 16473.14 | 0.11 | |
| 2005 | 4239.05 | 18448.10 | 0.10 | |
| 2006 | 5134.46 | 21027.59 | 0.10 | |
| 2007 | 6132.24 | 24220.21 | 0.10 | 1.96 |
| 2008 | 7136.30 | 27942.77 | 0.10 | 2.00 |
| 2009 | 8567.31 | 32528.73 | 0.11 | 2.23 |
| 2010 | 9875.91 | 37723.01 | 0.11 | 2.37 |
| 2011 | 11570.95 | 43805.49 | 0.12 | 2.58 |
| 2012 | 13750.71 | 51109.50 | 0.13 | 2.82 |
| 2013 | 17887.48 | 61311.39 | 0.14 | 2.88 |
| 2014 | 21203.20 | 73130.49 | 0.16 | 3.07 |
| 2015 | 25140.59 | 86890.79 | 0.18 | 3.26 |
| 2016 | 28584.53 | 101817.78 | 0.20 | 3.44 |
| 2017 | 32295.95 | 118013.23 | 0.21 | 3.23 |
| 2018 | 36186.75 | 135443.42 | 0.22 | |
| 2019 | 40536.03 | 154367.49 | 0.24 | |
| 2020 | 44423.70 | 174137.67 | 0.27 | |

资料来源：GDP 数据来自 2003—2019 年《中国统计年鉴》。数字经济增加值数据来自许宪春、张美慧《中国数字经济规模测算研究——基于国际比较的视角》（《中国工业经济》2020 年第 5 期）。其余数据为本文估算结果。

(三) 存储规模法估算

根据式 (13) 和式 (14)，需要首先估算历年的数据存储形成规模 d_t ，然后再通过永续盘存法进行数据存储规模的估算。具体的估算分为以下四个步骤。

第一步,估算 2003—2020 年以人民币计价的中国半导体存储器市场规模 M_t 。本文从中国海关、世界半导体贸易统计协会 (World Semiconductor Trade Statistics, WSTS) 等相关网站,获取中国半导体存储器市场规模数据。

第二步,估算 2003—2020 年新增半导体存储器存储规模 d_t 。本文从 WSTS 获取到了半导体存储器以美元每 GB 计价的价格 P_t ,从国家统计局获取到了美元汇率 ϵ_t 。同时,根据对华为、腾讯等中国计算服务行业企业的调研,一般存储介质的使用率在 60% 左右,因此可以直接求得历年新增的半导体存储器存储规模 $d_t = \frac{0.6 M_t}{P_t \epsilon_t}$ 。

第三步,使用永续盘存法估计 2003—2020 年中国的数据存储规模。同样基于对华为、腾讯等中国计算服务行业企业的调研,服务器存储介质的一般寿命在 5—8 年。根据平均年限法,取折旧率 δ_d 为 15%。不同于传统的固定资产,硬盘报废后数据均会转移至新买的硬盘之中,因此不存在数据意义上的残值。选定基年为 2003 年,根据每年市场投入的新增半导体存储器存储规模 d_t ,得到前十年 d_t 的几何年平均增长率 $\bar{g}_d = \sqrt[10]{d_{2013}/d_{2003}} - 1$ 。进一步可以根据式 (13) 得到 2003 年的数据存储规模,再根据式 (14) 得到 2004—2020 年历年的数据存储规模。

第四步,估算中国各省份 2003—2020 年的数据存储规模。本文计算增值法下,历年各省数据资本存量占全国数据资本存量的比例,并将其等同为各省数据存储规模占全国数据存储规模的比例,最终得到了每个省份 2003—2020 年的数据存储规模,2020 年中国市场的数据存储规模达到了 423.41EB。

(四) 估算结果分析

下面基于前述对中国 2003—2020 年的数据资本存量的估算,分析中国数据资本形成、演变和分布的基本特征。

1. 中国数据资本总体情况

图 2 展示了成本法和增值法下我国的数据资本存量估算结果,可以发现,在 2003—2011 年之间,成本法和增值法的估算结果几乎没有差异;在 2011 年之后,增值法估算结果开始逐渐高于成本法估算结果,且差异逐年增加。2011 年为中国移动互联网元年,智能手机开始普及,以微信为代表的智能手机 APP 开始大量发布上线。2012 年,中国手机网民数量首次超过了电脑网民数量。2013 年,中国移动互联网爆发式发展,中国政府颁发 4G 牌照,大数据产业规模高速增长。2014 年,大数据首次被写入了政府工作报告,以大数据、云计算为代表的行业开始在中国迎来快速增长期,而算法的进步也让数据的价值得到进一步的开发。因此,结合中国实践,图 2 中的增值法所展现出的数据资本存量更为合理。

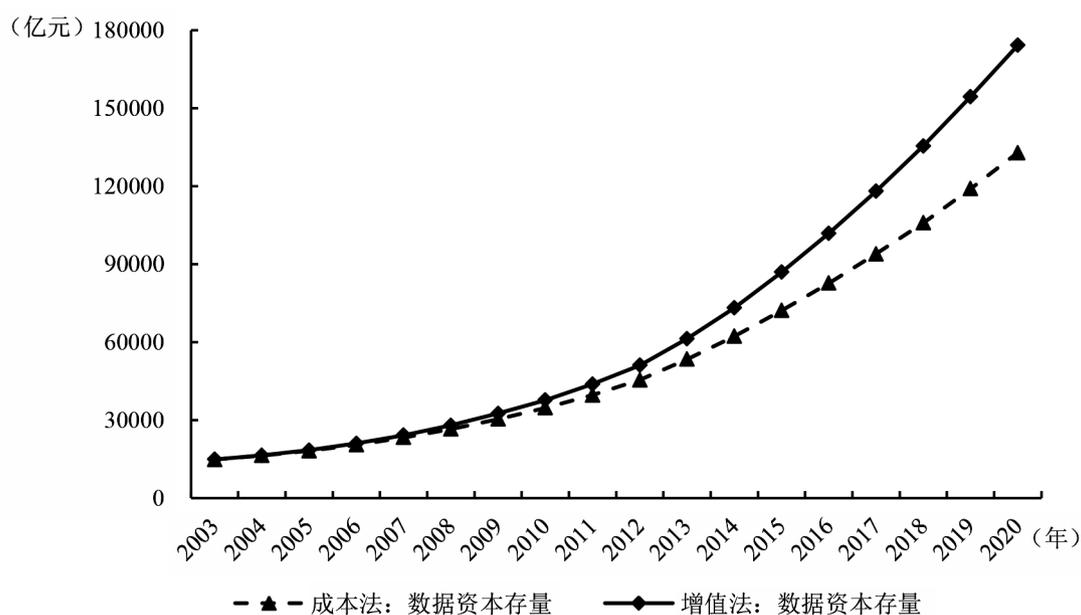


图 2 2003—2020 年成本法和增值法估算的中国数据资本存量 (2003 年不变价)

表 3 展示了 2003—2020 年基于增值法的数据资本存量估算结果，中国人均数据资本存量从 2003 年的 0.12 万元/人增长至 2020 年的 1.23 万元/人，增长至 10.25 倍。在此期间，中国人均 GDP 增长至 3.54 倍。人均数据资本存量和人均 GDP 的比例从 2003 年的 0.11 增长到 2020 年的 0.32，增长至 2.91 倍。可见，数字经济背景下中国数据资本存量高速发展，数据资本在经济增长中的地位不断攀升。

表 3 2003—2020 年基于增值法的人均数据资本存量估算结果 (2003 年不变价)

| 年份 | 人均数据资本存量 (万元/人) | 人均 GDP (万元/人) | 人均数据资本存量/ 人均 GDP | 数据资本 存量增速 (%) | GDP 增速 (%) |
|------|--------------------|------------------|---------------------|------------------|---------------|
| 2003 | 0.12 | 1.09 | 0.11 | | |
| 2004 | 0.12 | 1.08 | 0.11 | 10.70 | 10.10 |
| 2005 | 0.14 | 1.25 | 0.11 | 11.99 | 11.30 |
| 2006 | 0.16 | 1.42 | 0.11 | 13.98 | 12.70 |
| 2007 | 0.19 | 1.67 | 0.11 | 15.18 | 14.20 |
| 2008 | 0.21 | 1.77 | 0.12 | 15.37 | 9.60 |
| 2009 | 0.24 | 1.86 | 0.13 | 16.41 | 9.20 |
| 2010 | 0.28 | 2.14 | 0.13 | 15.97 | 10.60 |
| 2011 | 0.32 | 2.33 | 0.14 | 16.12 | 9.50 |
| 2012 | 0.38 | 2.56 | 0.15 | 16.67 | 7.70 |
| 2013 | 0.45 | 2.72 | 0.17 | 19.96 | 7.70 |
| 2014 | 0.53 | 2.90 | 0.18 | 19.28 | 7.30 |
| 2015 | 0.63 | 3.00 | 0.21 | 18.82 | 7.00 |

续表 3

| 年份 | 人均数据资本存量 (万元/人) | 人均 GDP (万元/人) | 人均数据资本存量/ 人均 GDP | 数据资本 存量增速 (%) | GDP 增速 (%) |
|------|--------------------|------------------|---------------------|------------------|---------------|
| 2016 | 0.73 | 3.20 | 0.23 | 17.18 | 6.80 |
| 2017 | 0.84 | 3.44 | 0.24 | 15.91 | 6.90 |
| 2018 | 0.96 | 3.69 | 0.26 | 14.77 | 6.70 |
| 2019 | 1.10 | 3.84 | 0.29 | 13.97 | 5.53 |
| 2020 | 1.23 | 3.86 | 0.32 | 12.81 | 2.75 |
| 年均 | | | | 15.59 | 8.56 |

注：本文使用的数据资本存量、人均 GDP 等数据均已经折算成以 2003 年为基年的实际值，其中数据资本存量使用的是增值法的估算结果。

2. 分行业数据资本

图 3 分别展示了我国 2003 年和 2020 年数据资本形成额排名前十的行业。可以发现公共管理和社会组织的数据资本形成额始终排名靠前，从 2003 年的 585.30 亿元排名第一至 2020 年的 7372.81 亿元排名第二，增加至近 12 倍。同时，尽管 2003 年和 2020 年排名前十的行业中有九个行业相同，但部分行业的排名发生了较大的变化。其中信息传输、软件和信息技术服务行业在 2003 年的数据资本形成额仅为 354.22 亿元，排名第五，在 2020 年则已排名第一，达到 9006.60 亿元。

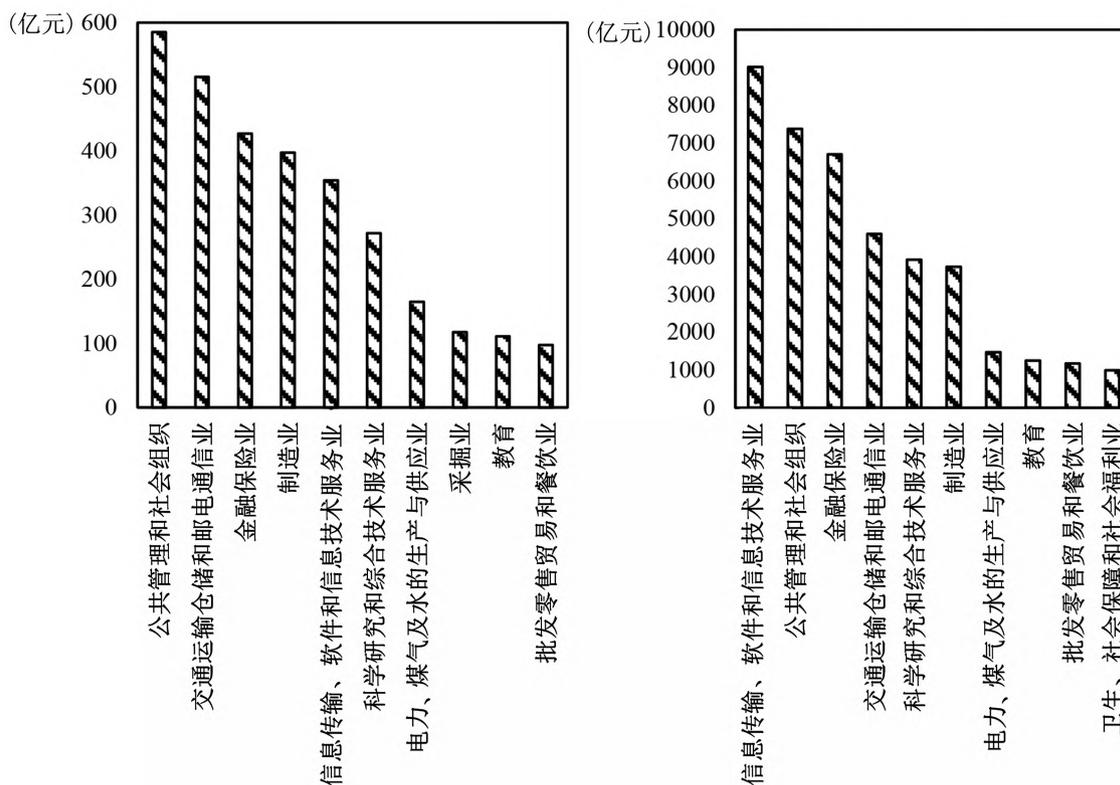


图 3 2003 年和 2020 年排名前十行业的数据资本形成额 (2003 年不变价)

三、数据资本对中国经济增长的贡献

本节探索数据资本在中国经济增长中所扮演的角色。研究发现,一方面,通过探究不同估算方法下数据资本对中国经济增长的贡献,增值法估算结果更具科学性;另一方面,实证结果表明数据资本确实对中国经济增长作出了重要贡献,对比物资资本、人力资本等生产要素,数据资本已成为中国经济增长重要动能。

(一) 数据资本对经济增长的贡献

基于柯布—道格拉斯生产函数,引入数据资本后假定生产函数的形式如下:

$$Y = A D^{\alpha} K^{\beta} H^{\gamma} L^{\delta} \quad (19)$$

其中, Y 表示总产出, D 、 K 、 H 和 L 表示数据资本存量、物质资本存量、人力资本存量和劳动力数量。 α 、 β 、 γ 和 δ 分别表示对应的产出弹性。假定生产的规模报酬不变,即 $\alpha + \beta + \gamma + \delta = 1$ 。利用劳均产出形式,得到如下计量经济方程:

$$\ln y_{it} = \text{Cons} + \alpha \ln d_{it} + \beta \ln k_{it} + \gamma \ln h_{it} + \epsilon_{it} \quad (20)$$

其中, y_{it} 表示省份 i 在第 t 年的劳均产出, d_{it} 、 k_{it} 、 h_{it} 分别表示省份 i 在第 t 年的劳均数据资本、劳均物质资本和劳均人力资本。

本文基于已有研究估算了 2003—2019 年^①省际物质资本存量,估算过程中各个省份的固定资产投资和固定资产投资价格指数来源于《中国经济统计年鉴》,折旧率选取为 8.5%。^② 劳动力数据来自各个省份的统计年鉴中年末就业人口数量;人力资本数据使用中央财经大学中国人力资本与劳动经济研究中心基于 J—F 方法所得到的省级层面人力资本总量。

回归中分别采用三种数据资本存量数据:(1)简单成本法。这类简单成本法没有考虑数据生产区别于其他中间投入品生产的特征,没有对数据价值链的三个环节进行分解,也没有使用机器学习对数据生产的劳动力成本进行细致估算。直接假定各行业总劳动力成本中,数据相关劳动力成本的比例等同于这些行业 ICT 中间投入占总中间投入的比例。^③ 具体估算结果不在文中展示,仅作回归对比。(2)成本法。本文的成本法对数据价值链三个环节进行分解,并重点采用机器学习较为精确地估算了各个环节的劳动力成本。(3)本文提出的增值法。增值法在成本法的基础上,针对数据要素的特征,考虑了数据资本形成过程中算法进步对数据的增值。

① 为了排除疫情带来的影响,本文利用 2003—2019 年的面板数据进行分析,利用 2003—2020 年的面板数据得出的结论与此类似。

② C. A. Holz and Y. Sun, "Physical Capital Estimates for China's Provinces, 1952-2015 and Beyond," *China Economic Review*, vol. 51, 2018, pp. 342-357.

③ 参见徐翔、赵墨非:《数据资本与经济增长路径》,《经济研究》2020 年第 10 期。

本文基于 31 省份 \times 16 年的面板数据得到回归结果,如表 4 所示。其中,第(1)和(2)列所使用的数据资本存量基于简单成本法,第(3)和(4)列则基于成本法,第(5)和(6)列基于增值法。第(1)、(3)和(5)列采用随机效应进行估计,第(2)列、第(4)列和第(6)列采用固定效应,Hausman 检验支持固定效应回归。采用简单成本法、成本法和增值法估算的数据资本的产出弹性分别为 0.08、0.11 和 0.11。在 2003—2019 年期间,数据资本对中国经济增长的产出弹性低于物质资本和人力资本。数据资本每增加 1%,可以带动产出增长约 0.11%。

表 4 各类要素对 GDP 产出弹性的回归结果

| 变量 | (1) lngdp | (2) lngdp | (3) lngdp | (4) lngdp | (5) lngdp | (6) lngdp |
|------------------|----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| lnd ₀ | 0.0798*** (3.43) | 0.0803*** (3.33) | | | | |
| lnd ₁ | | | 0.1122*** (3.11) | 0.1140*** (2.95) | | |
| lnd ₂ | | | | | 0.1040*** (3.80) | 0.1057*** (3.64) |
| lnk | 0.2746*** (15.64) | 0.2784*** (8.01) | 0.2699*** (7.90) | 0.2740*** (7.77) | 0.2763*** (8.26) | 0.2807*** (7.99) |
| lnh | 0.5077*** (12.46) | 0.4976*** (4.76) | 0.4687*** (5.81) | 0.4559*** (4.88) | 0.4645*** (6.32) | 0.4511*** (5.33) |
| Observations | 527 | 527 | 527 | 527 | 527 | 527 |
| R ² | 0.972 | 0.972 | 0.972 | 0.972 | 0.973 | 0.973 |
| Specification | RE | FE | RE | FE | RE | FE |

注:***、**和*分别表示在 1%、5%和 10%水平上显著,括号内为 t 统计量。其中 d₀、d₁、d₂分别为简单成本法、成本法和增值法所得到的劳均数据资本存量。

基于表 4 中的回归结果,可进一步计算各要素对中国经济增长率的平均贡献率,^①结果如表 5 所示。2003—2019 年期间,对中国经济增长率的平均贡献率最大的要素主要还是物质资本和人力资本。就数据资本而言,增值法下数据资本的贡献率最高,为 16.30%;成本法次之,为 15.52%;简单成本法最低,仅为 11.84%。可以认为,数据资本对中国过去十多年的经济增长作出了不可忽视的贡献。结果显示简单成本法低估了数据资本的产出弹性和对 GDP 增长率的平均贡献率。尽管成本法下数据资本的产出弹性与增值法类似,但是下面的计算将进一步说明成本法仍然低估了数据资本对 GDP 增长率的贡献。

^① 本文计算的各要素对经济增长率的平均贡献率为各年各省份贡献率的平均值。

表 5 2003—2019 年我国各要素对 GDP 增长的平均贡献率

| 面板 1：简单成本法 | | | | | |
|------------|---|--------|----------------|--------|-------|
| 对经济增长率的贡献 | K | H | D ₀ | L | TFP |
| | | 39.86% | 44.01% | 11.84% | 1.29% |
| 面板 2：成本法 | | | | | |
| 对经济增长率的贡献 | K | H | D ₁ | L | TFP |
| | | 39.23% | 40.32% | 15.52% | 1.40% |
| 面板 3：增值法 | | | | | |
| 对经济增长率的贡献 | K | H | D ₂ | L | TFP |
| | | 40.19% | 39.90% | 16.30% | 1.46% |

(二) 不同阶段数据资本对经济增长的贡献

本文在第三部分结合中国数字经济发展历程，对数据资本形成额的增长趋势进行了简要分析，并说明经历了 2011 年移动互联网元年后，基于增值法的中国数据资本形成额迎来快速增长。因此下面直接采用基于增值法的数据资本存量，以 2011 年为节点，分别对 2003—2011 年和 2011—2019 年进行分阶段回归，结果如表 6 所示。在 2003—2011 年间，数据资本的产出弹性仅为 0.14，明显小于物质资本和人力资本的 0.30 和 0.44；而在 2011—2019 年，数据资本的产出弹性达到了 0.19，与 2003—2011 年相比得到显著提升，弹性大小与人力资本类似。可见，在 2011—2019 年我国数字经济快速发展背景下，数据资本每增长 1%，可以带动 GDP 增长 0.19%，进一步说明数据资本已成为中国经济增长的新动能之一。

表 6 各要素对 GDP 产出弹性的异质性回归结果

| 变量 | (1) | (2) |
|-----------------------------|---------------------|---------------------|
| | 2003—2011 | 2011—2019 |
| | lngdp | lngdp |
| ln _{d₂} | 0.1437*** (3.14) | 0.1921*** (6.75) |
| lnk | 0.3018*** (5.03) | 0.2509*** (6.37) |
| lnh | 0.4373*** (4.05) | 0.2046*** (3.37) |
| Observations | 279 | 279 |
| R ² | 0.963 | 0.938 |
| Specification | FE | FE |

注：***、**和* 分别表示在 1%、5%和 10%水平上显著，括号内为 t 统计量。d₂为增值法估算得到的人均数据资本存量。

基于表 6 的回归结果, 分别计算 2003—2011 年和 2011—2019 年各要素对经济增长率的平均贡献率如表 7 所示。在 2003—2011 年, 数据资本对 GDP 增长率的平均贡献率仅为 13.65%, 其重要性远不如物质资本和人力资本; 而在 2011—2019 年, 数据资本对 GDP 增长率的平均贡献率达到了 34.46%, 与 2003—2011 年相比显著提升。

表 7 不同时期各要素对 GDP 增长的贡献率

| 面板 1: 2003—2011 年 | | | | | |
|-------------------|--------|--------|----------------|-------|-------|
| 对经济增长率的贡献 | K | H | D ₂ | L | TFP |
| | 46.11% | 35.41% | 13.65% | 2.50% | 2.33% |
| 面板 2: 2011—2019 年 | | | | | |
| 对经济增长率的贡献 | K | H | D ₂ | L | TFP |
| | 37.44% | 21.50% | 34.46% | 2.09% | 4.51% |

结 论

本文从数据价值链视角出发, 基于数据采集、数据清洗与存储、数据加工三个环节重构了成本法, 并通过考虑数据分析技术带来的数据价值增值提出增值法。利用成本法、增值法及存储规模法三种方法, 从价值层面和物理层面对我国全国及省级数据资本存量做出系统估算。尤其采用机器学习的方法, 基于我国《职业大典》中各类职业的详细描述, 估算了数据价值链三个环节中各个职业从事数据相关工作的时间占比, 对中国数据生产三个环节中的劳动力成本进行了详细的估算。估算结果表明, 2003—2020 年, 我国数据资本存量在总量和人均上均实现大幅度增长, 人均数据资本存量和人均 GDP 的比例从 2003 年的 0.11 增长到 2020 年的 0.32, 增长至近 3 倍。

基于数据资本存量估算结果, 本文利用面板模型估计数据资本存量对中国经济增长的贡献。可以发现, 近年来数据资本对增长的贡献相对于以前明显加大, 已成为中国经济增长的重要动能。2003—2019 年, 数据资本每增加 1%, 可以带动 GDP 增长约 0.11%, 对经济增长率的平均贡献率为 16.30%; 2011—2019 年期间数据资本的产出弹性达到 0.19, 数据资本对经济增长率的平均贡献率达到 34.46%。

当然, 本文的估算结果依赖于对数据资本折旧率、数据相关企业增长率等参数的前置假定。若参数设置不合理, 例如设定了过低的数据资本折旧率或者过高的数据相关企业增长率, 可能会导致对数据资本存量的高估。今后随着相关统计资料的不断完善, 可以不断改进参数设置, 进而提高估算的准确性。

基于上述研究结果, 本文提出如下建议。

第一, 立足数据资本形成的价值链规律, 积极发展人工智能 (AI) 等相关数据

分析技术，提升数据资本存量。本文利用增值法估算发现，数据分析技术的提升会对已有数据带来价值增值。例如，ChatGPT 的横空出世不仅在很大程度上依赖于海量的数据，也反过来提升了这些数据的价值。因此，在不断积累新数据的同时，也应积极发展以 AI 为代表的的分析技术，提升数据价值，助力数据资本积累。

第二，基于数据价值链的三大环节，优化数据产业空间布局，发挥各自比较优势，实现区域间更好的协同互补。本文估算结果表明，数据资本存量地区分布不均衡。经济发展水平高、互联网经济发达的东部地区省份数据资源丰富，数据资本存量相对较高。而东部地区的比较优势在于高素质的数据分析人才，中西部地区则相对在发展算力所需的能源成本等方面有比较优势。应优化数据产业空间布局，大力推进“东数西算”工程，将东部地区产生的丰富数据更多布局在中西部地区进行存储和运算，而东部地区更多发挥人才优势做大做强数据分析产业。

第三，充分发挥数据价值链不同环节尤其是第一环节在实现高质量就业中的作用。本文估算了不同阶段劳动力成本的情况，发现总体看来第一阶段劳动力成本是最主要的，占到总劳动成本的一半左右。可见第一阶段是整个数据资本积累的基础，是数据产业吸收劳动力就业的主阵地，且第一阶段的劳动力相对后续的环节来讲，所需要的专业技能相对容易掌握，就业形式灵活，对于在中国经济不断数字化的进程中创造高质量就业有很大空间和潜力。

〔责任编辑：梁 华〕

carries out, the better will be its innovation performance, because the interconnected innovation behaviors will form the basic characteristics of innovation cost-saving and knowledge preference connections. Interconnected innovation provides a new path for Chinese technological latecomers to search for knowledge, make breakthroughs in core technologies, and regulate innovation.

Estimation of Data Capital and Its Contribution to China's Economic Growth: From the Perspective of Data Value Chain

Liu Taoxiong, Rong Ke and Zhang Yadi • 44 •

With the vigorous development of the digital economy, the significance of data capital in contributing to China's economic growth is increasingly evident. Data undergoes various stages including collection, cleansing, storage, and processing, which cumulatively adds value to itself, thereby generating data capital. This study adopts the lens of the data value chain, utilizing cost and value-added methods to estimate the value-level data capital formation and data capital stock while employing the storage scale method to ascertain the physical scale of data. Based on the results of data capital estimation, provincial panel models are applied to gauge its contribution to economic growth. The findings indicate that between 2003 and 2020, both the formation and stock of data capital in China have surged on both a total and per capita basis, with a growth rate way faster than that of GDP. Following 2011, the output elasticity and the contribution of data capital to the economic growth rate have both significantly surpassed the previous stages, establishing data capital as an important driver of China's economic growth.

The Narrative of Modernization in Chinese Urban Literature since the New Period

Lu Zhen • 65 •

Since the new period, urban literature has developed in parallel with China's modernization process. Urban literature's narrative of modernization has had an important impact on contemporary literature. Writers delve into the physical landscape of the city and the inner world of urban dwellers, writing about the improvement of the modern functions of the city, the reform of the economic system, the integrated development of urban and rural areas, and the shaping of the cultural personality, imagining an ideal form of the city and the urbanites, and establishing a two-way interactive narrative between the "modernization of the city" and the "modernization of humanity." As the value kernel of the narrative of modernization in urban literature, the "modernization of humanity" takes the modernization experience of man in the city as a link, highlights the multidirectional qualities of the mind of city dwellers, and explores the ideal civic ethics and morality. Entering the new era, the narrative of modernization in urban literature pursues the double deepening of "poetry" and "thought." Based on the task of building a modern socialist country in all aspects, it aims at enriching the people's inner world, discovering and paying

• 205 •