

数据生产要素研究进展*

徐翔 厉克奥博 田晓轩

摘要:在数字经济时代,数据被广泛地视为一种新的生产要素,在社会生产活动中发挥着越来越重要的作用。本文首先探讨了数据生产要素的经济学概念及其主要特征,强调了数据生产要素的虚拟性和非竞争性特征,以及由此产生的边际产出递增、强正外部性、产权模糊与衍生性等特点。进一步,本文梳理了数据生产要素价值的估计方法以及将数据生产要素纳入经济增长框架的宏观经济理论。之后,本文相继分析了数据生产要素对微观企业决策与生产效率的影响,与数据要素相关的产业组织问题,以及金融市场中数据所发挥的降低信息摩擦的作用。最后,本文介绍了数据生产要素相关的隐私问题及其可能的交易机制,提出了当前数据生产要素相关研究存在的一些主要不足与未来拓展研究的方向。

关键词:数据生产要素 数据价值 经济增长 企业决策 市场机制

进入21世纪以后,数据对个人行为、企业决策、产业升级以及经济增长的影响与日俱增。《经济学者》2017年5月的封面文章大胆预言:“数据是新的‘石油’,也是当今世界最宝贵、同时也是最需要加强监管的资源。”现如今,数据已经出现在经济社会的各个角落,大量的数据被生成、记录与整理。大数据技术和人工智能技术的发展使得对于数据的使用贯穿于整个社会生产过程之中。强调数据所发挥作用的数字经济(digital economy)已经逐渐形成。2017年12月8日,习近平总书记在主持中共中央政治局就实施国家大数据战略进行的第二次集体学习时指出:“要构建以数据为关键要素的数字经济。建设现代化经济体系离不开大数据发展和应用”^①。数据的生产要素地位得到进一步明确。

2010年后,各主要经济体开始将数据相关技术与产业的发展问题上升到国家战略层面。一国拥有的数据规模与数据分析处理能力,已经成为其国家竞争力的重要组成部分。2012年3月29日,美国政府宣布投资2亿美元启动“大数据研究和发展计划”,以应对大数据革命带来的机遇,推进相关研究机构进一步进行科学发现、创新研究和商业转化^②。2019年12月,美国政府发布《联邦数据战略与2020年行动计划》,明确将数据作为一种战略性资源进行开发。2020年2月19日,欧盟委员会公布了一系列围绕数据资源的发展规划,包括《欧洲数据战略白皮书》《人工智能白皮书》等多份文件,详细概述了欧盟未来五年实现数字经济所需的政策措施和投资策略,以及构建一个真正的欧洲数据统一市场的发展目标^③。

在我国,随着近年来数字经济的高速发展,尤其是社会生产过程的广泛网络化、数字化与智能

* 徐翔、田晓轩,中央财经大学经济学院,邮政编码:100081,电子邮箱:seanxuxiang@126.com;hinstien@163.com;厉克奥博,清华大学中国经济思想与实践研究院,邮政编码:100084,电子邮箱:likeaobo@126.com。基金项目:国家社会科学基金重点项目(18AZD007);国家社会科学基金后期资助项目(19FJYB006);北京高等学校卓越青年科学家计划(BJJWZYJH01201910034034)。感谢匿名审稿人的修改建议,文责自负。

① 中国政府网:http://www.gov.cn/xinwen/2017-12/09/content_5245520.htm。

② 美国白宫新闻网:<https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200>。

③ 欧盟委员会网站:<https://digital-strategy.ec.europa.eu/en/policies/strategy-data>。

化,数据作为一种生产要素所发挥的作用在社会经济中也已经充分凸显。2019年10月31日,中国共产党第十九届中央委员会第四次全体会议通过《中共中央关于坚持和完善中国特色社会主义制度推进国家治理体系和治理能力现代化若干重大问题的决定》^①。文件中提出,“健全劳动、资本、土地、知识、技术、管理、数据等生产要素由市场评价贡献、按贡献决定报酬的机制”。2020年3月30日,中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》,明确提出要加快培育数据要素市场,为进一步发挥数据生产要素的作用指明了方向。^②上述政府文件与发展规划均充分表明,数据生产要素已成为中国经济在高质量发展时期的重点。

事实上,对于数据经济的探索已经成为近年来经济学研究的热点与前沿。美国社会科学联合会(ASSA)在其2020年年会上专门举办了“大数据,国民账户和公共政策”子论坛,数十位学者深入讨论了数据在宏观经济层面的核算和应用等问题。在2021年1月的美国经济学联合会(AEA)年会上,数十场分论坛以大数据技术在经济中的使用等相关话题为讨论主题。从经济学的整体发展趋势上看,数据生产要素的研究热点已经逐步形成,一批重要的理论与实证文献正如雨后春笋般不断出现。

一、数据生产要素的概念与主要特征

(一)数据生产要素的概念

在最新版的《牛津英语词典》中,数据(data)被定义为“被用于形成决策或者发现新知的事实或信息”。根据国际标准化组织(ISO)的定义,数据是对事实、概念或指令的一种特殊表达方式,用数据形式表现的信息能够更好地被用于交流、解释或处理。在《现代汉语词典》(第七版)中,对于数据的解释是:“进行各种统计、计算、科学研究或技术设计等所依赖的数值”。从经济活动的角度,加拿大统计局(2018)将数据定义为“已经转化成数字形式的对于现实世界的观察”。采取数字形式的数据能够被储存、传输以及加工处理,数据的持有者也能够从中提取新的知识与信息。Farboodi & Veldkamp(2020a)将数据理解为“可以被编码为一系列0和1组成的二进制序列的信息”。按照这一定义的数据既包括数字化的音乐、影像资料和专利等,也包括统计数据 and 交易记录。前者可以被看作是以数据形式存在的 product 与服务,而后者则更多地被看作是为了生产知识而进行的投入。从上述定义中可以看出,数据的概念与信息密切相关,数据在很多时候被视为信息的一种表现形式或者数字化载体。

Jones & Tonetti(2020)研究了数据进入生产过程的基本模式,进而定义了作为生产要素的数据。他们认为,“数据”可以被视为“信息”(information)中不属于“创意”(idea)和“知识”(knowledge)的部分。从整体上看,对于信息的理解是由二进制位串(bit strings)表示的经济物品(economic goods)。创意和知识是信息的子类,遵循Romer(1990)等的定义,被Jones & Tonetti(2020)视为一组能够被用于生产经济物品的指令。在他们看来,所有除了创意和知识之外的其他信息都是数据。作为生产要素的数据,其本身不能被直接用于生产经济物品,但是却能在生产过程中发挥作用——如创造新的知识或者形成对于未来的预测,进而指导经济物品的生产。

(二)数据生产要素的主要特征

数据生产要素首先是一种虚拟的、存在于数据库与互联网空间中的资源。因此,大多数已有研究都将数据的虚拟性(virtuality)视为该生产要素的一项核心特征(Jones & Tonetti, 2020)。虚拟性的存在就意味着数据必须以其他生产要素作为载体才能发挥作用。在当前的技术条件下,数据在大多数时候是存在于信息与通信技术(ICT)产品中的,二者的有效结合被视为目前全球经济增长的主要动力之一(Jorgenson & Wu, 2016)。虚拟性是数据与其他传统生产要素如劳动力、资本和土地

^① 中国政府网:http://www.gov.cn/zhengce/2019-11/05/content_5449023.htm。

^② 中国政府网:http://www.gov.cn/zhengce/2020-04/09/content_5500622.htm。

的最主要差异,也是知识、技术、管理和数据等新生产要素的主要特点。对于数据等虚拟生产要素的依赖,是数字经济的主要特点之一,也是数字经济与传统经济的主要区别(Mueller & Grindal, 2019)。除了虚拟性之外,数据还具有非竞争性、排他性、规模报酬递增、强正外部性、产权模糊以及衍生性等特征。

1. 非竞争性。除虚拟性之外,数据生产要素的另外一项核心特征是非竞争性。非竞争性一般指一个使用者对该物品的消费并不减少它对其他使用者的供应。同一组数据可以同时被多个企业或个人使用,一个额外的使用者不会减少其他现存数据使用者的效用,数据也因此具有非竞争性(Acquisti et al, 2016)。Veldkamp & Chung(2019)也强调,对于数据的额外使用的边际成本为0,这是数据与其他生产要素之间的一个主要差异。由于非竞争性的存在,任何数量的企业、个人或机器学习算法都可以同时使用同一组数据,而又不会减少其他人可用的数据量,这就决定了数据的高使用效率与巨大的潜在经济价值(Jones & Tonetti, 2020)。

2. 排他性。当数据的规模足够庞大、内容足够复杂和广泛时,数据生产要素就表现出高度的排他性(exclusive),拥有数据的企业和机构会选择“窖藏”而非分享数据。在现实中,大多数私营机构都不会随意公开自己产生、收集和拥有的各项数据,即使公开这些数据能够创造巨大的经济社会价值(Gaessler & Wagner, 2019)。数据的排他性还体现在,其他虚拟生产要素(如知识和技术)会随着人事变动或劳动力迁移而公开或扩散,但数据的复杂性和广泛性使其无法被包含在人力资本当中(Jones & Tonetti, 2020)。以机器学习为例,机器学习产生的知识是公开的,而输入到机器学习算法的数据却是保密的——每一家企业都在收集自己的数据,形成训练集之后交由人工智能进行训练、学习。虽然有一些公司将自己的训练数据集公之于众以鼓励研发,然而大多数企业都将数据视为自己的一项核心竞争力而极少公开。Varian(2018)强调,数据公司可能会将一些数据整理、编制起来形成数据集和数据库,并根据某些条款向第三方出售访问和使用许可,利用排他性获得收入。例如,有几家美国的数据公司将美国人口普查数据和其他种类的地理数据合并,然后向第三方出售这些数据的使用许可权,同时禁止转售或重新许可,利用数据的排他性攫取利润。

3. 规模报酬递增。数据生产要素的非竞争性进一步产生了规模报酬递增的效应。从企业的角度,其所拥有的数据可以被每一名员工使用,因而数据规模越大、种类越丰富,产生的信息和知识就越多,进而呈现出规模报酬递增的特点。如果数据对于整个行业乃至经济体的参与者开放,数据规模扩大带来的经济价值就将更为可观(Jones & Tonetti, 2020)。Romer(1990)提出,知识的非竞争性意味着劳动力和知识的结合将形成规模报酬递增的效果。与知识相同,数据生产要素的规模报酬递增也需要与劳动力结合才能体现出来。在Veldkamp & Chung(2019)建立的包含数据生产要素的生产者模型中,商品的产量(Y)取决于生产中投入的数据量(Ω)和劳动力数量(L)。在只考虑竞争性投入(劳动力)的情况下,数据是规模报酬不变的;但若同时考虑劳动力和数据,数据将会是规模报酬递增的,进而带来生产效率的提升。

4. 正外部性。数据的正外部性首先体现在数据收集型企业生产效率的提升上。Schaefer & Sapi(2020)发现,使用雅虎搜索引擎的用户,其搜索数据显著改进了该引擎的搜索质量,进而产生了很强的正外部性。一方面,当许多用户输入同一个特定的关键词时,为该关键词收集的反馈数据量就会增加。搜索引擎会从用户在搜索结果页面上的点击行为了解用户期望得到的搜索结果,从而提高搜索质量,形成直接的网络效应。另一方面,用户使用搜索引擎的历史越长,搜索引擎的学习速度就越快,这就强化了用户数量增多所带来的正外部性。此外,数据生产要素还通过改善运营、促进创新和优化资源配置的方法,在提升企业组织效率和用户体验上实现正外部性。通过结合和分析来自样车试驾、车间报告和其他来源的数据,宝马公司可以快速发现潜在的问题和漏洞,并在新车型推出之前消除这些问题(Kshetri, 2014)。IBM等拥有较强数据分析能力的科技公司,更是能在几天内完成“分析数据—发现漏洞—解决问题”的全处理过程,从而有效地优化产品维修和养护服务,提高客户满意度。

5. 产权模糊性。数据生产要素在产权归属上存在一定的模糊性,其所有权和产生的各项产出在企业与消费者之间的分配尚不清晰。消费者在使用互联网公司等企业提供的各项 ICT 产品和服务的过程中会产生大量数据。这些数据往往由企业直接收集和整理,消费者在客观上没有处置和使用这些数据的机会。Varian(2018)认为,相较更适用于竞争性物品的“所有权”(ownership)概念,更应当关注非竞争性数据的“访问权”(data access),理由是数据很少以出售私人物品的方式进行“出售”,而是多被许可用于特定用途。例如,出于安全性的考虑,应当允许多方同时访问自动驾驶汽车的数据,包括引擎、导航系统和乘客的手机等。然而,数据的所有权也会影响数据访问权,数据产权的不同分配方式会对消费者剩余与总社会福利造成影响(Jones & Tonetti, 2020)。

6. 衍生性。各类经济活动的参与者在生产、销售和消费产品和服务的过程中,很早就开始记录各种交易、市场与用户数据,数据在大多数时候是经济活动衍生出的一种副产品,经济活动产生的数据会被金融市场的参与者深入分析,从而影响金融市场投资,进而影响实体经济。资产市场中的投资者会在经济状况良好时进行更多的投资活动,这些投资活动会产生有关当前经济状态的公共数据(Veldkamp, 2005)。如果经济状况在形势良好且数据量充足时出现波动,资产价格就会快速调整,从而导致金融市场崩盘;当经济形势不好时,数据的匮乏和高度的不确定性就会降低投资者的反应速度,反过来减少金融市场波动对于实体经济的负面影响。Ordóñez(2013)提出,更大的社会投资规模会产生更多反映经济总体状况的数据,进而影响市场信贷行为。当经济整体下行时,创业投资的成功率较低,贷款者从大量创业活动失败的数据中推断出经济在下行,并迅速提高利率以应对更高的违约可能性。当经济不景气时,市场中的利率较高且借贷活动较少;而当经济状况好转时,数量有限的现存创业投资活动也只能产生相应少量的有关经济状况转换的数据,贷款人逐渐了解到这一点,并逐渐调低贷款利率,这又促进了企业投资和产出扩张。由于存在通过分析数据以协助投资决策的现实激励,生产企业和金融机构会自发地记录、整理和存储各项经济活动的有关数据,即使其主营业务与这些数据无关,而数据生产要素的衍生性就体现于此。

二、数据生产要素与经济增长

(一)GDP 统计中缺失的数据生产要素

国内生产总值(GDP)衡量了一国生产的所有最终产品的货币价值,反映了一个经济体的经济实力、生产能力与市场规模。然而,在当前的 GDP 核算体系中,数据生产要素的规模及其创造的经济价值均未得到充分体现。其主要原因在于,现有 GDP 测算大多基于人们为商品和服务支付的费用,包括数据在内的、没有货币价格的商品(或称零价商品)在统计上不会对 GDP 产生任何贡献(Brynjolfsson & Collis, 2019)。然而,这些零价商品充斥于整个国民经济和人们的生活中,创造了大量的经济价值。从这个角度来说,现有的国民经济核算方法已不再适用于以数据作为核心生产要素的数字经济,根据此方法得出的 GDP 指标不能很好地反映一国经济的真实情况(金星晔等,2020)。

一些已有研究开始尝试测度各类虚拟生产要素对于总产出的贡献。Hulten & Nakamura(2017)扩展了传统的核算模型,允许技术直接影响消费者福利,使技术进步除了在传统意义上以提高全要素生产率的方式“节约资源”(resource-saving)之外,还产生了“节约产出”(output-saving)的效应。例如,ICT 技术推动的电子商务等新经济模式的发展减少了对于传统商品和服务(如交通运输)的需求,而电子商务本身又因为缺乏明确价格没有被统计进 GDP 指标中,因此可以说作为虚拟生产要素的电子商务绕开 GDP 直接创造了消费者剩余。通过估计此类产出节约型技术的价值并计入 GDP,可以实现对传统 GDP 核算框架的扩展。数据生产要素也体现出类似的性质,对于大数据的使用减少了企业对于传统商品服务的需求。例如,金融企业使用大数据建立的用户图像能够帮助其发掘潜在借款人,而无须客户经理主动联系客户(Begenau et al, 2018)。因此,与技术类似,数据生产要素也产生了额外的生产者剩余和消费者剩余,应被纳入 GDP 核算体系之中。

Brynjolfsson et al(2019a)考虑到数字经济中新商品的频繁引入和零价商品的不断增加,在传统

GDP 的基础上提出了一个新的度量标准——“GDP-B”，其中包含了具有隐含价格的免费数字商品，通过量化和捕捉这些商品对福利的贡献，改善了传统 GDP 核算中对于数据生产要素的遗漏和误测。Byrne & Corrado(2020)认为，家庭是数字革命的主要阵地，在内容上不断创新的数字服务作为一种免费商品，通过 ICT 技术传输给连接的用户，成为 ICT 资本所产生的服务。他们建立了一个框架来衡量 ICT 资本数字服务创新和增长的价值，作为现有 GDP 中个人消费部分的有效补充。

上述研究所描述的现有 GDP 统计对数据生产要素的低估或忽视，有助于我们更好地理解近年来统计数据所显示的生产率增速的放缓。自 2005 年起，发达经济体的劳动生产率增长速度开始下降，此后一直保持较低水平。2005—2016 年间，美国的总体劳动生产率平均年增长仅为 1.3%，远低于 1995—2004 年间 2.8% 的年平均增长率。在经合组织编制的 29 个其他国家的生产率数据库中，有 28 个国家的增长率数据出现类似的下降。1995—2004 年间，这些国家的未加权平均年劳动生产率增长为 2.3%，而 2005—2015 年间的这一数据仅为 1.1% (Brynjolfsson et al, 2019a)。这些统计数据反映的生产率下滑与人们在经济生活中感受到的技术进步与生产率提升并不相符，被定义为索洛提出的生产率悖论的新版本，即“现代生产率悖论”(Beaudreau, 2010)。

一些实证研究提出，如果能够将 ICT 技术与数据等新生产要素带来的潜在经济价值计入 GDP 中，实际生产率就会高于以当前统计数据为基础计算出的水平，“现代生产率悖论”(modern productivity paradox)将不攻自破(Acemoglu et al, 2014; Brynjolfsson et al, 2019a)。也有研究反对这一观点，认为数字产品所制造的生产者剩余和消费者剩余的规模不足以解释大部分的生产率增速放缓，主要有三个方面的理由(Syverson, 2017):其一，即使 ICT 技术和数据分析带来的消费者剩余都以某种方式计入 GDP，未必能够弥补生产率增速下降所造成的 GDP 损失。其二，销售 ICT 产品的企业会将零价商品(如数据服务和相关软件)的价值计入互补性商品的价格中。其三，消费者在非工作时间使用 ICT 服务并从中获得剩余，但非工作时间的收益增加并不能表明生产过程中的生产率被低估。上述三个原因在现实中确实存在，但是其中涉及的主要是数字化零价产品带来的消费者剩余。考虑到企业使用零价数据和知识获得的大量生产者剩余，劳动生产率仍有可能是被低估的，将数据等零价产品引入 GDP 核算仍能在一定程度上解决“现代生产率悖论”。

(二)数据生产要素的规模估计

为了克服 GDP 统计中存在的不足，Brynjolfsson et al(2019a)采取大规模线上选择实验来估计零价商品(如 Facebook 提供的免费社交服务)所创造的经济价值，并将其作为 GDP-B 的组成部分。在该项实验中，受试者需要在两个选项里做出“二选一”：保留某种数字产品的使用权，或者放弃某种商品以换取特定金额的货币补偿。该实验只要求每位消费者回答这一个问题，并通过每次实验的数千名消费者赋予不同的货币补偿，从而估计出消费者对于数字产品的支付意愿，并作为该产品的价值估计。根据这一实验，Facebook 用户样本愿意接受的单月补偿金额中位数为 48 美元，可以被视为 Facebook 提供的数字化服务的隐性价格。Brynjolfsson & Collis(2019)在上述结果的基础上估计出了 Facebook 提供的免费社交服务每年创造的经济价值：2004—2017 年间美国消费者通过使用 Facebook 获取了 2310 亿美元的剩余。如果将这部分消费者剩余加入 GDP，会使美国的 GDP 增长平均每年增加 0.11%。与之相对，美国在此期间(2004—2018 年)年平均 GDP 增速仅为 1.83%，这就意味着经济增长中的 6% 来源于 Facebook 提供的免费服务。除了 Facebook 提供的社交服务外，Brynjolfsson et al(2019b)还利用谷歌市场调研工具(Google Surveys)对美国互联网用户的代表性样本进行了更大规模的在线实验，估计了用户对于不同类型的数字化产品和服务的接受意愿(WTA)。实验结果显示，搜索引擎是最有价值的数字商品(平均价值为 17530 美元/年)，其次是电子邮件(8414 美元/年)和数字地图(3648 美元/年)。

从企业生产的角度，一些研究将具有虚拟性的数据生产要素视为企业的一种无形资产(Erickson & Rothberg, 2015)。为了研究无形资产对企业市场价值的影响，Belo et al(2019)建立了一个广义的新古典主义投资模型，其中包含有四种类型的准固定投入(quasi-fixed input)，两种有形

资本:实物资本(机器、厂房)和劳动力(工人),以及两种无形资本:知识资本(创新活动的累计投资)和品牌资本(提升品牌知名度的累计投资)。计算结果表明,在过去的几十年里,对于整体经济尤其是高科技产业来说,物质资本对企业价值的重要性不断下降,而知识资本的重要性逐渐上升,在总资产价值中的占比从1970年的24.9%大幅增加到2010年的44.8%。数据生产要素被定义为信息中非知识的部分,因此也可以采取类似的方法予以估计。也有一些研究将数字经济中的“零价商品”看成是互联网用户的ICT资本所产生的一种资本服务(capital services),其价值取决于用户使用ICT资本消费网络供应内容的强度,其规模取决于ICT资本的质量(Byrne & Corrado, 2020)。根据Byrne & Corrado(2020)的估计,1987—2017年间,数字服务内容的创新使美国每个互联网用户每年的消费者剩余增加了近2000美元。2008—2017年间,数字服务内容的创新使得美国的每年实际GDP增速提高了0.3~0.6个百分点。

Reinsdorf & Ribarsky(2019)总结了从宏观视角衡量数据经济价值的三种可能方法:(1)市场法。数据资产的价值由市场上可比产品的市场价格来决定。例如,企业收集的关于所在行业的数据的价值,取决于企业直接购买相应的数据或信息所需要支付的市场价格。(2)成本法。数据资产的价值取决于生产信息时的成本是多少,这一方法直接度量企业获取、收集、整理、分析与应用数据的成本。(3)收入法。数据资产的价值取决于对未来能够从数据中获取的现金流数额的估计。这种方法将数据资产类比为金融资产,用贴现的办法度量其价值。

加拿大统计局(Statistics Canada, 2020)采用成本法的思路,用劳动力成本的数据来估计数据资产的价值。该方法首先在加拿大国家职业分类(National Occupation Classification)中筛选出与数据生产有关的职业,并分别假设各职业在生产数据资产上耗费的工作时间占比,作为各自生产数据的机会成本占直接劳动力成本的权重。在估计出各职业在数据生产上的机会成本后,设定总工资成本的50%为间接劳动力成本和其他成本之和,并附加一个3%的增值(markup)作为资本形成的额外价值,进而得到对数据资产的价值估计。其估计的结果是,加拿大在数据资产上的投入自2005年起年均增长5.5%,占全国固定资本形成总额的5.9%~8.0%。从存量上看,加拿大统计局估计的2018年该国数据净资本存量为1570亿~2170亿美元,占非住宅建筑、机器设备以及知识产权总量的6.1%~8.4%,占知识产权净资产存量的68.9%~95.2%。按照这一估计结果,数据生产要素已经成为加拿大拥有的知识产权的核心内容。

政府是宏观经济数据和金融数据的主要生产者,商业部门可以利用政府数据进行生产和投资决策、营销和库存管理以及长期战略规划,但政府数据的价值却难以准确衡量。美国经济统计局(2014)提出“政府数据密集型部门”(government data-intensive sector,以下简称GDIS)的概念,以代表在生产过程中严重依赖政府数据的企业,并估计出2012年GDIS型企业的总产出为2208亿美元。与此同时,政府资金支持的下降可能会导致公共数据的质量及其为企业提供的价值受损。美国政府在2017财年的13个主要统计机构的预算总额为22.57亿美元,与2004—2013年间的平均预算相比,实际美元价值下降了8.7%,这就可能导致美国政府数据的质量下降,不利于企业决策与美国经济发展(Hughes-Cromwick & Coronado, 2019)。

上述研究充分表明,数据生产要素的经济价值已经非常可观,且其相关投资与存量规模均呈持续增长趋势,对于经济增长的贡献亦十分显著。需要注意的一点是,上述结果是建立在数据生产要素对于经济增长的间接贡献(如促进其他生产要素的结合效率)大多是不可测度的前提之下,企业、政府和个人在数据上的实际投入也可能被低估。因此,数据生产要素对于经济增长的实际贡献可能要比上述文献中估计的更高。

(三)包含数据生产要素的经济增长理论

大多数理论研究认为,数据生产要素主要通过驱动知识生产影响经济增长。Agrawal et al(2018)构建了一个“基于组合”(combinatorial-based)的知识生产函数,结合Jones(1995)的半内生增长模型,研究知识的产生过程及其对经济增长的影响。从数据生产要素的角度,大数据分析技术的

进步有效提升了算法预测有用知识组合的准确度,进而提高了新知识的发现率。新知识的不断发现提高了全社会的生产效率,促进了经济增长速度的提升。这一理论研究明确了数据影响知识生产、进而影响经济增长潜力的核心机制。

Aghion et al(2019)在Agrawal et al(2018)的研究基础上引入了企业异质性,提出了大数据等新技术的发展以及数据等新生产要素的出现导致增长下滑的潜在机制。在这一模型中,企业异质性的来源主要有二:产品质量(product quality)和加工效率(process efficiency)。ICT、大数据等技术的加速发展使得企业的成本降低,因此加工效率更高的企业(其附加值更高)会迅速扩张自己的生产线范围,进入新的市场并取得利润;而由于这种效率难以被模仿,低效的企业则很难进入市场,并会因此而减少创新,阻碍创造性毁灭的速度。他们认为,创新活力的降低导致了企业活力的衰退以及国民收入中劳动收入份额的下降,进而对经济增长产生了负面作用。

与传统的生产要素相比,数据在企业中发挥的作用更为灵活多样,通过单纯的宏观视角仍然难以有效刻画。Farboodi & Veldkamp(2020a)建立了一个类索洛增长框架的数据经济增长模型,讨论具有非竞争性及衍生性的数据要素积累对于数据经济与宏观经济的整体影响。在这一框架下,数据积累根据其绝对水平的高低,会产生收益递减和收益递增两个相反方向的影响:企业在利用大数据进行预测时,预测误差最多只能被降低到0,这一自然界限加上经营过程中不可预测的随机性,使得数据在规模够大时必然具有递减的收益;而在数据量不够充足时,数据则具有递增的收益;更多的数据使企业更具生产力,从而产生更多的生产和交易,而这又创造出更多的数据,并进一步提高生产力和数据生成,形成“数据反馈循环”(data feedback loop)的过程。这种递增的收益会反过来造成所谓的“数据贫困陷阱”(data poverty trap),使得那些数据拥有量低的企业、行业或国家长期局限于少量的生产和交易,从而无法取得进一步发展。与Aghion et al(2019)的判断类似,Farboodi & Veldkamp(2020a)认为数据生产要素的发展未必会提升长期经济增长率。

虽然数据经济的增长不同于传统的经济增长模式,在短期内数据的积累过程可能会获得递增的回报,但从长期来看,数据生产要素的回报率仍可能呈递减的趋势(Farboodi & Veldkamp, 2020a; Jones & Tonetti, 2020)。数据积累能够提高产出,同时也能够作为技术创新的投入要素,但在长期,这二者本身都不能独自维持无限的增长,对于创新的支持与鼓励仍有必要。Nordhaus(2015)认为,虽然计算机和人工智能可以完成许多常规任务从而释放劳动力,但那些随着经济环境而变化的非常规任务却很难进行编程。Nordhaus(2015)就信息投入与传统投入之间可替代性进行了7项“替代检验”(alternative tests),结果表明,世界经济仍未达到持续自发增长的“经济奇点”,新生产要素仍符合边际产出递减的基本规律。因此,对于数据生产要素所驱动的经济增长也不应盲目乐观,应充分重视技术进步与知识生产等新经济增长理论强调的内生增长机制(Romer, 1990)。

三、数据生产要素影响经济的微观机制

(一)数据驱动的决策过程

大数据技术的迅速发展与广泛应用,促使越来越多的企业开始从管理者主导的经验型决策转向高度依赖数据分析结果的科学决策模式,即所谓的“数据驱动型决策”(data-driven decision making,以下简称DDD模式)。Brynjolfsson et al(2011)构建了基于企业微观数据估计DDD模式的使用情况与贡献度的分析框架。基于美国179家上市公司的调查数据及公开信息,他们预测DDD模式可以解释2005—2009年间美国企业5%~6%的生产率增长。Provost & Fawcett(2013)从概念上区分了数据科学(data science)和数据驱动决策(DDD模式),强调了企业在决策方式转型中引入数据科学的重要性。他们将数据科学定义为通过自动化数据分析来理解各种现象的原理、过程和技术,其最终目标是改进决策质量;而数据驱动决策则是基于数据分析而非纯粹直觉的决策实践,是数据科学在企业决策中的应用方法。数据科学使大规模的自动决策成为可能,同时也依赖于企业对大数据的存储、分析和处理技术水平。与之相对的,数据驱动决策能够提高企业决策的科学性,进而提高

资源配置效率和企业表现,是数据科学在企业管理上的应用形式。McAfee et al(2012)提出,相较于传统的经验型决策,DDD模式优势十分明显:充分利用DDD模式的管理者,能够基于经验证据(empirical evidence)而非商业直觉(business intuition)来做出企业决策,提高了决策的科学性与准确性。他们对北美330家企业公共管理实践和业绩数据进行调查研究发现,DDD模式使用程度更高的企业在财务和运营状况上有着更好的表现。进一步的实证分析显示,在一个行业中使用DDD模式占比最高的前三名企业,其平均生产效率和利润率分别会比其他竞争对手高5%和6%左右。

大数据分析带来的竞争优势促使越来越多的企业尤其是制造业企业开始转向DDD模式。在2005—2010年间,美国制造业中使用DDD模式的企业占比从11%增长到30%,预计到2020年这一比例将超过50%,DDD模式已然成为美国制造业的“新常态”(Brynjolfsson & McElheran, 2016a)。Brynjolfsson & McElheran(2016b)基于2005年和2010年美国人口普查局(USCB)的年度制造业调查(ASM)以及管理和组织实践调查(MOPS)数据,实证研究了美国制造业企业对于DDD模式的使用情况及其经济影响。回归结果显示,属于多生产单位型企业(multi-unit firms)的大型工厂更多且更早地使用DDD模式,并通过这一转变提升了企业绩效。随着时间推移,DDD模式的先行者和后来者之间的绩效差距会减小。此外,DDD模式和ICT资本规模以及技术工人数量等变量之间还存在相互促进的互补关系。

虽然许多企业都在通过投资于ICT和大数据来转向DDD模式,但该模式存在的一些客观不足也开始逐渐显现。Brynjolfsson & Mitchell(2017)通过一项决策实验比较了机器学习领域中的“学徒系统”(learning apprentice)与一般企业管理者决策之间的效率差异。“学徒系统”是一种特殊的DDD模式。在这一系统中,人工智能程序充当学徒协助人类工作者,通过观察人类的决策进行学习,并将这些人类决策作为训练自己的示例。机器从它所协助的多个个体人类的组合数据中进行学习,按照大数据与人工智能的基本理论,机器将会做出比训练它的团队中的每一个人更优的决策。然而,实验结果显示,机器学到的专业知识仍可能会受到团队技术水平与相关决策变量的在线可用性的限制,在现实中做出的决策未必会比管理者基于经验做出的结论更好。Bajari et al(2019)对于亚马逊的产品销售量预测系统的实证分析表明,产品在市场上销售时段的增加产生的额外数据会提升预测表现,同一产品种类中的产品数量增长产生的额外数据则无法改进预测质量,说明不同类型的数据对于数据驱动的决策和预测的影响存在异质性。上述研究表明,简单地加大在数据生产要素上的投资力度未必会提升企业的决策质量,需要进一步分析数据驱动决策的作用机制,并意识到数据驱动型决策的局限性,人类和机器在决策上的表现各有所长(Abis, 2020)。

企业转向数据驱动型决策的过程并不是一蹴而就的,需要面对诸多管理上的挑战(McAfee et al, 2012)。为了采用DDD模式,企业需要对管理模式进行深入改革。企业高层决策者必须要逐渐接受数据驱动的、基于实证证据的决策方式,而企业为此需要额外雇用能够发现数据中的模式规律并将其转化成可用商业信息的数据科学家(data scientists)或数据策略师(data strategists)。整个企业的组织结构都必须围绕DDD模式进行调整。Janssen et al(2017)也指出,影响数据驱动决策质量的相关因素还有很多,大数据和大数据分析未必能带来更好的企业决策。一方面,大数据的多样性和大数据分析的广泛使用,增强了企业发现欺诈行为的能力,从而有利于防止决策错误;但另一方面,它同样也可能加剧对于用户和消费者的歧视,带来额外的社会福利问题。

事实上,不仅仅是企业决策,数据生产要素也具有改进政府政策决策的可能性。Höchtel et al(2016)通过构建一个政策周期模型,证明大数据分析能够改进政府的政策决策过程。在大数据分析的支持下,政府在设计 and 实施政策计划中不应一味遵从分阶段连续执行的传统模式,而应该在每一个阶段对政策措施的当前效果进行持续评估,通过采用大数据情景分析实时制定替代方案。在必要时应该提前放弃先前计划的政策,通过实时决策提高政策制定的效率。此外,由于能够更快、更好地处理收集到的大量未结构化信息,大数据分析能够在政策周期的每一个阶段都实现公众的广泛参与,从而能够更充分地集思广益。随着更多数据通过预测分析技术转化为可用信息并指导决策,社

会部门可以更合理地将社会支出分配到需要改善的领域,以此提高社会福利水平(Coulton et al, 2015)。

(二)数据生产要素与生产效率

除了驱动企业决策外,数据还能直接提升企业的生产效率。Müller et al(2018)使用计量方法对大数据分析和企业生产效率之间的关系进行了实证研究,通过将2008—2014年间814家美国企业对于大数据分析的使用情况与Compustat数据库中的财务绩效数据相结合,结果发现,对于样本中的所有企业来说,拥有大数据分析会使平均生产率提高4.1%;具体到行业层面,企业生产率的大幅提升与大数据分析高度相关;大数据分析使得信息技术密集型行业的生产率提高6.7%,竞争性行业的生产率提高5.7%。这无疑为大数据分析的商业价值提供了有力的实证证据。具体来看,已有研究归纳了数据提升企业生产效率的三种主要实现机制:信息挖掘、协同创新和产品质量提升。

个体和企业竞争激烈的环境中生存有赖于在正确的时间掌握正确的信息(Kubina, 2015; Veldkamp & Chung, 2019),数据生产要素正是正确信息的重要来源。一方面,企业通过大数据分析可以提高组织内部信息的透明度,从而产生更广泛、更深入和更准确的观察,进而改进决策的质量。另一方面,企业能够利用大数据刻画出更复杂、更完整的客户画像,从而有针对性地提供更准确的定制产品和服务,进而提升生产效率。从这个角度来看,数据生产要素可以被视为一种能够提高企业生产效率的信息资产(Farboodi et al, 2019)。能有效利用数据(指从每单位生产中获取更多数据)的企业,在开始时可能会因为建立自己的数据库而产生亏损。然而,如果企业能够在这一阶段顶住财务压力而对数据类投资予以持续的资金支持,就能够迅速超越规模更大但数据利用效率低的企业。

数据生产要素提升企业生产效率的第二个机制是鼓励企业间的协同创新。创新本身是一个试错的过程,研发过程中的失败尝试所形成的大量数据和信息对企业也具有重要意义(Akcigit & Liu, 2016)。如果企业之间不能进行信息共享,那么研究的时间和资源就很可能浪费在其他企业已经发现毫无结果的项目上,从而导致无效率的均衡。Akcigit & Liu(2016)建立了一个“赢家通吃”(winner-takes-all)的动态博弈模型研究这一问题,发现竞争性企业的博弈结果是:要么重复进行高失败风险的实验,希望通过某次幸运的成功带来高额回报;要么提前放弃风险研究,中止创新进程。这两种选择均会因为信息外部性的存在引发显著的效率损失。反之,如果企业能够交流、共享研发过程中的各种数据,便能够显著提高创新活动的效率,进而实现生产效率提升。

除了信息挖掘与协同创新之外,企业还能利用数据预测并选用最优的生产技术,进而改善产品质量。Farboodi & Veldkamp(2018)仿照索洛(Solow, 1956)的经典新古典增长模型中对资本流入和流出的分析,对数据的“流入”(经济活动产生的新数据)和“流出”(数据的折旧)进行了理论建模。模型结果显示,当数据存量非常低时,新产生的数据量远远大于数据的折旧量,数据生产要素会快速积累,产品的质量和价值也会随之快速提高;而随着数据存量的增加,数据流入的速度会逐渐放缓,新产生的数据量与折旧量之间的差距变得越来越小。最终,数据流入和流出交于一点,数据经济达到稳态均衡水平,此时数据存量、产品质量以及GDP都将保持不变。由此可见,当数据量积累到一定程度时,收益递减的力量就会占据主导地位。而在数据量不够充足时,“数据反馈循环”将会产生递增的回报——拥有更多数据的企业会生产更高质量的产品,这又会促使这些企业增加投资、生产和销售,进而生成更多数据。Veldkamp & Chung(2019)进一步指出,数据帮助企业选择更好的生产技术的渠道有两种:一方面,交易记录会揭示消费者的偏好,帮助企业了解消费者更加偏好何种颜色的鞋子、何种动力的汽车等,从而可以进行针对性的生产;另一方面,消费者的偏好是在持续不断地变化当中,企业必须始终保证自己能够适应这种变化。通过对生产和销售过程中产生的数据进行分析,企业可以获得指导其最优化自身运营实践的有用信息。

(三)数据生产要素与产业组织

产业组织理论关注一个行业内不同企业间的组织或者市场关系。企业拥有的数据生产要素规

模与其在 ICT 技术上的投入密切相关。这一投入与企业规模之间又相互影响,进而影响企业动态(firm dynamics)与产业组织方式。因此,对于数据生产要素与产业组织之间关系的研究,就应该首先关注不同规模企业的 ICT 投入。鉴于以往研究企业层面 ICT 资本生产率的文献的数据来源相对单一,Tambe & Hitt(2012)构建了一个新的面板数据集,基于约 1800 家企业在 1987—2006 年间的 ICT 从业人员数量以及相匹配的生产投入,研究企业 ICT 投资与企业规模之间的关系。研究发现,中型企业的 ICT 投资回报显著低于名列《财富》世界 500 强的大型企业,ICT 投入的边际产出在 2000—2006 年间比以往任何时期都高,这表明企业(尤其是大型企业)对于 ICT 技术的使用效率得到了显著提升,这很有可能是规模更大、种类更加丰富的数据生产要素带来的结果。

Begenau et al(2018)提出,现代经济的两大发展趋势是企业规模的扩大与 ICT 技术的进步,数据生产要素对于这两种趋势的形成均做出了一定贡献。作者建立了一个企业重复静态博弈模型,得出两个核心结论:首先,大企业比小企业更擅长利用金融市场中的大数据来降低资本成本;其次,大型企业因为有更多的经济活动和更长的经营历史,从而产生了更多可供处理的数据。随着计算机性能的提高,丰富的数据也支持了更多的财务分析。数据分析改善了投资者的预测质量,减少了股票投资的不确定性,降低了企业的资本成本。当投资者能够处理更多的数据时,大企业的投资成本会下降得更多,从而使其规模变得更加庞大。Farboodi et al(2019)采取了类似的建模方法,在重复静态博弈模型中引入了一个名为“数据精通”(data-savvy)的新变量用于描绘企业对于数据使用的精通程度,并通过求解这一模型他们发现:首先,数据有助于企业提高生产率,更高的生产率使得企业进行更多投资,规模变得更大,从而生产出更多的数据,构成了一个“数据反馈循环”;其次,由于额外的生产过程会产生更多数据,企业为了获得更多、更好的数据会进一步增加投资,数据质量也因此得到显著改善。

数据生产要素的积聚还在一定程度上导致了“超级明星公司”(superstar firms)的出现。Autor et al(2020)认为,在 ICT 技术和包括数据在内的无形资产上的竞争优势催生了超级明星公司,这些公司具有高附加值和低劳动力份额的特点,造成产品市场集中度的显著上升,以及宏观意义上劳动收入份额的下降。Tambe et al(2020)提出了“数字资本”的概念,用于指代数字技术密集型企业对实现新技术价值所需的无形资产进行的投入(如员工 ICT 技能培训、企业决策结构和软件定制等方面的累计投资等),通过创建一个关于 ICT 相关劳动力投入的企业面板数据库进行研究发现,在大多数“超级明星公司”中积聚了大量数字资本,进而导致了一定程度的垄断。Ciuriak(2019)分析了过去三百多年的经济史中要素租金的分配方式,指出数据生产要素的集聚产生了大规模的租金,催生了超级明星企业,并为战略性的贸易和投资政策产生了强有力的激励。

(四)数据生产要素与信息摩擦

对于金融市场来说,数据生产要素是投资决策的核心参考变量,发挥了降低信息摩擦(information frictions)的重要作用。计算机技术的进步使科技公司能够收集实时、精确的基本面指标,并将其出售给专业投资人士。这些数据通过降低信息获取成本而提高了金融产品中的价格信息含量(price informativeness),这对投资者产生了两个主要影响(Zhu,2019):一方面,当价格迅速而全面地反映未来收益时,经理人就很难有机会利用其内部信息优势获取个人交易收益;另一方面,关于基本面的数据揭示了企业当前业务的衰退趋势或在未来实现增长的机会,可以指导投资者在状况恶化时减少投资,在机会扩大时增加投资,从而提高投资效率。

随着金融科技技术的不断创新与应用,金融交易的价格信息含量总体上呈现上升趋势。Farboodi & Veldkamp(2020b)探讨了金融部门信息处理效率的确定性增长所带来的结果。由于其他投资者的需求数据能够协助投资者抵御需求冲击,因此当金融部门的信息处理效率提高时,获取未来股息(基本面)与需求(对价格的非基本面冲击)信息的动机就会发生改变,企业会选择处理越来越多关于其他投资者需求的数据,而不是企业的基本面数据。因此,信息处理效率的提升也就可以解释为什么金融分析从一个主要调查企业基本面盈利能力的部门,转变成了一个做少量基本面分析而主

要集中于获取和处理客户需求的部门。虽然对于基本面数据的分析在减少,但是金融产品的价格信息含量仍然会上升,投资者可以借此更好地预测资产回报并规避投资风险。

根据金融资源配置的基本逻辑,随着越来越多的技术被用于处理和传输金融数据,资本确实会得到更有效的配置,从而增加社会收益。Bai et al(2016)建立了一个衡量股票价格中信息含量的新指标,并通过计算该指标发现,1960年以来标准普尔500指数的股票价格在预测公司未来收益上的能力已经提高。Farboodi et al(2018)提出了另一种观点:大数据技术的进步引发了股票市场上的结构效应(composition effect),即以标准普尔500指数中的企业为代表的历史悠久的大企业的股票价格信息含量显著提高,其他中小企业的股票价格信息含量却有所下降。换言之,金融大数据的发展对于股价信息含量的影响具有规模异质性。受益于大数据技术,大企业股票的定价变得更为准确,但中小企业股票价格的实际信息含量并没有实质性的提升,因而对于总体市场来说,大数据技术的进步对于价格信息含量的提升有限。

除了影响价格信息含量外,大数据分析技术还降低了贷款信用风险管理中的信号传递和信息搜索成本,减少了借贷过程中存在的信息不对称性问题。在对金融市场中的信息摩擦现象进行分析时,Yan et al(2015)以P2P(peer-to-peer)贷款平台为对象,发现了大数据分析在贷款行业中对减少信息摩擦所起到的积极作用。由于借贷过程中存在信息不对称问题,借款人会发出信号并传达有关其本人和投资项目特征的信息,而贷款人则需要搜索信用记录等信息并筛选贷款申请人。在传统的银行体系中,银行进行贷款决策并负责信用信息的收集和评估。而随着ICT技术的不断突破,基于大数据的金融科技已经成为贷款行业中的一个颠覆性驱动力,企业收集、呈现和评估信息的手段变得更加先进和丰富,信用信息的检索成本大幅度降低,信用数据的收集也从被动的信息检索转变为主动的信息收集。

大数据、ICT与人工智能等新技术的发展改变了金融行业的基本经营方式与盈利逻辑,推动了金融行业的全面转型升级。Vives(2019)调查了银行业的技术革新,并考察了其对于行业竞争的影响以及效率提高和客户福利的潜力,认为银行业正在进行从依赖实体分支行开展业务向依赖ICT技术、大数据以及高度专业化的人力资本的系统性变革。随着数字化水平的提升,银行在支付和咨询服务等核心业务上面临着相较于其他中介机构更大的竞争威胁。新近崛起的金融科技部门开始在金融服务中使用创新的信息和自动化技术,使得数字技术的采用速度以及相关用户的获取速度都显著加快。

四、数据生产要素的隐私问题与交易机制

(一)数据生产要素的隐私问题

虽然企业能够利用数据生产要素创造巨大的经济利益,然而这些利益的产生往往是以侵犯消费者隐私为代价的,消费者得到的补偿却微乎其微,这也是企业与消费者间数据产权归属矛盾的主要症结所在。Kshetri(2014)指出,大数据分析技术在很大程度上都需要利用高速数据(high-velocity data),其中比较典型的就是从移动设备中获取的点击流数据(click-stream data)^①以及GPS定位数据。企业可以通过收集此类敏感数据进行精准的短期预测,但这同时也损害了消费者的隐私和安全利益。Kshetri(2014)还总结了多个国际咨询机构对于消费者如何看待大数据的调研结果,提出数据安全和隐私问题已经在全球范围内引起了广泛的担忧,消费者们越来越关注企业的数据收集方法,特别是诸如网站Cookies和GPS追踪器等追踪技术的使用。Acquisti et al(2016)从隐私的经济属性(the economics of privacy)的视角,探讨了个人信息的经济价值,以及消费者在个人数据共享和隐私之间的权衡取舍等问题,认为现实经济中的个人数据既有私人价值也具有商业价值:一方面,数据

^①这是指用户在网站上持续访问的轨迹。用户对网站的每次访问包含了一系列的点击动作行为,这些点击行为数据就构成了点击流数据,它代表了用户浏览网站的整个流程。

共享可以减少市场摩擦,促进交易;另一方面,对数据商业价值的利用往往会导致私人效用的降低,有时甚至是整体社会福利的减少。具体例子包括,零售市场中的价格歧视、保险和信贷市场中的数量歧视、垃圾邮件以及身份盗用的风险等。

随着信息技术的快速发展,企业获取用户数据的手段和能力都在增强,消费者虽然从基于大数据分析的针对性产品推荐中获益,但也会承受个人隐私被侵犯所带来的货币成本和负效用。新的数字技术已经针对注重隐私的消费者提供了个性化的控制选项,这似乎将隐私处理的选择权交给了消费者,但大多数消费者实际上缺乏保护、管理其个人数据的意识与运用技术的成熟度。而侵犯隐私的技术服务在日常通信、求职和一般消费中已经屡见不鲜。无论是企业的商业数据还是个人数据,其所有权、访问权以及交易的法律框架仍未得到完全确定(Duch-Brown et al,2017)。此外,虽然法律上的数据所有权有所缺失,但数据市场中的议价却产生了实际上的数据所有权,这往往会导致数据所有权的碎片化或反公共化。拒绝共享个人数据可能会提高消费者的隐私利益,但这种反公共化的做法会阻碍数据聚合的范围经济效应,使数据的非竞争性所能带来的经济利益无法得到充分实现,造成数据市场的失灵。

然而,使社会福利最大化的数据产权分配方式很难进行事先预测(Duch-Brown et al,2017),数据经济收益与个人隐私权益间的取舍也并不能一刀切(Acquisti et al,2016),而是应当根据具体情况进行适当安排。Tene & Polonetsky(2012)提出了一个平衡企业及研究人员数据利益和个体隐私权益的模型设想,要求政策制定者首先确定需要征得用户同意方可使用的“个人可识别”(personally identifiable)^①的数据类型范围,并将这些个人可识别的数据置于法规监管框架中;在此基础上,兼顾隐私数据保护原则与其他诸如公共卫生、国家安全和环境保护等社会价值的平衡,当预期数据使用的收益明显大于隐私风险时,即使个体拒绝同意,也仍然认定数据的使用是合法的。

Jones & Tonetti(2020)建立了一个数据经济的理论框架,专门研究数据产权的不同分配方式对消费者福利的影响。他们的基本判断是,当企业拥有数据的产权时,就不会充分尊重消费者的隐私。此时政府可能出于对隐私权的关注而大幅限制企业对消费者数据的使用,然而,这在产生隐私收益的同时却也使非竞争性的数据生产要素不能以最优的规模投入使用,继而导致无效率的情况出现。倘若不存在政府的该类限制,从理论上讲,非竞争性就能使数据在企业间广泛使用,从而产生巨大的社会经济收益,但如果出售数据会加快“创造性毁灭”(creative destruction)的速度,企业就会有囤积数据并阻止其他企业使用的动机,这同样会造成无效率的均衡结果。基于上述分析,Jones & Tonetti(2020)提出,消费者数据所有权产生的消费和福利远高于企业数据所有权,将数据产权赋予消费者可以带来接近最优的分配。如果这一判断是准确的,数据的市场机制就不符合科斯定理。

如果按照 Jones & Tonetti(2020)的建议,将数据所有权赋予消费者,是否就能够实现对于数据价值的充分保护与合理分配?一些实验研究表明,消费者可能会轻易地将自己和亲友的数据交给企业,有时仅仅是为了换取非常低的回报。Athey et al(2017)将这种表现出关注数据隐私的态度但却在现实中轻易披露数据和信息的消费者行为定义为“数字隐私悖论”(digital privacy paradox),并通过两个实验验证了数字隐私悖论确实存在,并且研究发现,无论是受试者的个人特征(如性别、年龄等),还是受试者所表达出的关于隐私的偏好,对于其在数据隐私披露上做出的实际选择都影响甚微。Liu et al(2020)的研究也给出了类似的结论。他们首先在支付宝中添加一个第三方小程序,然后通过一项调查获得用户对于数字隐私的偏好,最后再让用户选择是否使用这一必须提供个人数据才能运行的小程序。实验结果显示,那些声称非常在意数据隐私的用户实际上大多都与第三方小程序共享个人数据。随着授权的数据越来越多,用户对数据隐私的担忧也在逐步增加,但是这种担忧仅仅止步于“担忧”,并没有改变其将个人数据共享给第三方平台的行为。

^①即同真实身份和隐私信息相匹配的数据,与匿名化的数据相反。

(二)数据生产要素的交易机制

针对数据生产要素的隐私与权益问题,一些学者试图通过设计数据生产要素的交易机制来提升社会的总福利水平。Elvy(2017)探讨了两种极端化的数据交易模式对于经济福利的影响。第一种名为“为隐私付费”(pay-for-privacy)的模式将数据初始产权交给了企业,要求消费者在购买产品或服务时支付一笔额外的费用来保证其数据不被收集和挖掘并用作广告用途。第二种名为“私人数据经济”(personal data economy)的模式将数据产权放还消费者,在该模式中,类似于 Datacoup 这样的数据收集企业需要直接向个人购买数据。两种模式都促进了隐私数据向可交易产品的转化,但同时也可能产生或加剧隐私数据获取的不平等,并进一步导致掠夺或歧视性行为的发生。

如果将数据产权直接赋予企业或消费者均不能实现最好的福利结果,那么能否通过引入数据中介来改善数据生产要素的配置? Bergemann & Bonatti(2019)详细分析了以数据公司形式存在的数据中介(data intermediary)在信息市场中的作用及其与经济福利的关系。数据中介从个人消费者处收集信息并销售给企业,而企业利用这些信息来改进价格和数量决策。单个企业可以利用得到的数据划分市场,并进行三级价格歧视(third-degree price discrimination)^①,从而获得超额利润。在这种情况下,商品价格与消费者支付意愿的相关性越大,企业获得的利润就越大,而消费者剩余和社会总福利则会越低。虽然传统理论表明三级价格歧视所造成的社会福利降低会使数据中介无利可图,但 Bergemann & Bonatti(2019)指出,如果给定消费者需求不确定性和消费者数据信息含量的范围,以及假设数据中介聚合了单个消费者的信号并根据总需求的后验估计将信息传递给企业,那么市场中总会存在一个临界值 \bar{n} ,使得当且仅当消费者的数量 N 满足 $N > \bar{n}$ 时,数据中介能够进入市场并获取正的利润,并对消费者福利造成损害。

既然如此,数据中介对于消费者福利的负面影响能够通过竞争化解吗? Ichihashi(2020)给出了否定的答案。在其数据市场模型中,包含一名消费者、多个竞争性的数据中介以及一家下游企业。在该模型中,购买数据的下游企业会通过价格歧视和骚扰广告侵吞消费者权益,因此中介可能会对收集消费者数据的行为进行补偿。这种补偿既可能是货币转移,也可能是诸如免费的在线网络地图服务等非货币的回报。由于数据生产要素具有非竞争性,如果多个中介机构都提供较高的补偿,那么消费者就会与所有中介机构共享同一套数据,并接受所有平台的货币和服务补偿,这无疑降低了数据的下游价格,损害了数据中介的收益。而中介平台也预测到了这一问题,因此最终所有中介都不会有提供高补偿的动机。因此,消费者从数字市场的竞争性中得到的福利要少于传统的商品市场。最后,Ichihashi(2020)还指出,在中介收集的数据集互斥且需向消费者提供补偿的条件下,数据集集中度越高、越接近数据垄断的均衡,中介机构的利润就越高,而消费者的福利越低。

五、小结与研究展望

截至目前,对于数据生产要素的研究已经取得了一些关键性的重要突破。对于数据生产要素影响企业表现的微观机制,以及数据生产要素影响经济增长的路径,均形成了一系列具有重要意义理论与实证成果。当然,由于这一研究领域仍处于发展早期,尚存以下几点比较明显的不足之处有待改进:

第一,在数据生产要素驱动经济增长的研究中,对于溢出效应的讨论相对不足。与大数据相关的 ICT 技术和数据分析技术均表现出一定的通用目的技术(general purpose technology)特征,能够作为经济社会活动的基础性技术间接推动生产效率的提升。数据生产要素本身具有的非竞争性也导致其具有较强的正外部性,如被广泛使用的政府公开数据显著提升了企业的决策效率。然而,在目前研究数据生产要素对经济增长影响的文献中,大多缺少对于上述两种溢出效应的理论建模,这就可能导致数据生产要素对于总产出的综合影响被低估。因此,数据生产要素的宏观研究也应更加

^①即对于同一商品,完全垄断厂商根据不同市场上的需求价格弹性不同,对不同类型的消费者实施不同的价格。

深入地探讨数据生产要素的溢出效应。

第二,无论是基于微观数据还是宏观数据,数据生产要素相关的实证文献较之理论文献仍相对薄弱。这主要是因为目前对于数据生产要素的测度与估计仍缺少共识性方法,相关数据要么不可得,要么不可比,从而导致对于数据生产要素的实证研究相对分散、割裂,缺少一致性和系统性。绝大多数实证研究都以自行调研、收集的企业数据为基础,其采用的研究方法与分析方法也各不相同,导致相关实证研究的体系性较差,也不利于数据生产要素基础理论的提炼与升华。

第三,对于数据生产要素的跨国分析严重不足。想要准确把握数据生产要素对于经济增长与经济发展的综合影响,就需要以国家层面的宏观测度与对应的生产率估计为基础。目前除加拿大统计局等少数统计机构之外,无论是各国的统计机构还是研究机构,均没有对数据生产要素的测度与估计予以足够重视。虽然一些国际组织如 IMF 和 OECD 发布了一些相关的研究报告,但是这些研究均尚未落实到指标测算与跨国分析的层面,这就限制了相关宏观研究的进一步发展和深化。

第四,对于数据生产要素如何影响社会福利的探讨仍不够深入。数据作为一种生产要素,在进入生产活动后自然会产生要素收入的分配问题。虽然现有研究已经开始关注数据相关的产权和分配问题,然而大多数讨论是在理论层面上进行的,缺少有说服力的经济证据。此外,绝大多数相关文献建议将数据产权转交给消费者,这虽然在理论上能获得最高的消费者福利,但在实际推进的过程中却面临极大的现实困难。因此,为了防止数据生产要素的发展导致收入不平等进一步恶化,需要对于数据生产要素的福利经济学含义进行更加深入的探讨,通过更好的市场机制设计与政策手段,实现收入的均等化与社会福利的最大化。

有理由相信,随着数字经济的进一步发展以及数据生产要素核心地位的进一步确立,相关研究也将向纵深发展。中国的数字经济发展在全球占据优势地位,中国也是最早在官方规划性文件中将数据纳入生产要素范畴的国家之一。在 ICT 基础设施和大数据产业的相关投入规模方面,中国也均居全球前列。此外,中国还是全世界生产数据最多的国家之一。据国际数据公司(IDC)统计,中国 14 亿多人口创造的数据超过整个世界数据量的 20%。我们坚信,数据生产要素将在中国下一阶段的经济发展中发挥举足轻重的作用,基于中国视角的相关理论和实证研究也会如雨后春笋般涌现,值得经济学界予以加倍重视。

参考文献:

- 金星晔等,2020:《数字经济规模核算的框架、方法与特点》,《经济社会体制比较》第4期。
- Abis, S. (2020), “Man vs. machine; Quantitative and discretionary equity management”, SSRN Working Paper, No. 3717371.
- Acemoglu, D. et al(2014), “Return of the Solow paradox? IT, productivity, and employment in US manufacturing”, *American Economic Review* 104(5):394—399.
- Acquisti, A. et al(2016), “The economics of privacy”, *Journal of Economic Literature* 54(2): 442—492.
- Agrawal, A. et al(2018), “Finding needles in haystacks: Artificial intelligence and recombinant growth”, NBER Working Paper, No. 24541.
- Aghion, P. et al(2019), “A theory of falling growth and rising rents”, NBER Working Paper, No. 26448.
- Akcigit, U. & Q. Liu(2016), “The role of information in innovation and competition”, *Journal of the European Economic Association* 14(4):828—870.
- Arrieta-Ibarra, I. et al(2018), “Should we treat data as labor? Moving beyond ‘free’”, *AEA Papers and Proceedings* 108:38—42.
- Athey, S. et al(2017), “The digital privacy paradox: Small money, small costs, small talk”, NBER Working Paper, No. 23488.
- Autor, D. et al(2020), “The fall of the labor share and the rise of superstar firms”, *Quarterly Journal of Economics* 135(2):645—709.
- Bai, J. et al(2016), “Have financial markets become more informative”, *Journal of Financial Economics* 122(3):625—654.
- Bajari, P. et al(2019), “The impact of big data on firm performance: An empirical investigation”, *AEA Papers and Proceedings* 109:33—37.

- Beaudreau, B. C. (2010), "The dynamo and the computer: An engineering perspective on the modern productivity paradox", *International Journal of Productivity and Performance Management* 59(1):7-17.
- Begenau, J. et al(2018), "Big data in finance and the growth of large firms", *Journal of Monetary Economics* 97(8): 71-87.
- Belo, F. et al(2019), "Decomposing firm value", NBER Working Paper, No. 26112.
- Bergemann, D. & A. Bonatti(2019), "Markets for information: An introduction", *Annual Review of Economics* 11: 85-107.
- Brynjolfsson, E. & K. McElheran(2016a), "Data in action: Data-driven decision making in US manufacturing", US Census Bureau Center for Economic Studies Paper, No. CES-WP-16-06.
- Brynjolfsson, E. & K. McElheran(2016b), "The rapid adoption of data-driven decision-making", *American Economic Review* 106(5):133-139.
- Brynjolfsson, E. & T. Mitchell (2017), "What can machine learning do? Workforce implications", *Science* 358 (6370):1530-1534.
- Brynjolfsson, E. & A. Collis(2019), "How should we measure the digital economy", *Harvard Business Review* 97 (6):140-148.
- Brynjolfsson, E. et al(2019a), "GDP-B: Accounting for the value of new and free goods in the digital economy", NBER Working Paper, No. 25695.
- Brynjolfsson, E. et al(2019b), "Using massive online choice experiments to measure changes in well-being", *Proceedings of the National Academy of Sciences* 116(15):7250-7255.
- Brynjolfsson, E. et al(2011), "Strength in numbers: How does data-driven decision-making affect firm performance", SSRN Working Paper, No. 1819486.
- Brynjolfsson, E. et al(2019), "Does machine translation affect international trade? Evidence from a large digital platform", *Management Science* 65(12):5449-5460.
- Bughin, J. (2017), "The best response to digital disruption", *MIT Sloan Management Review* 58(4):80-86.
- Bulger, M. et al(2014), "Data-driven business models: Challenges and opportunities of big data", *Research Councils UK: New Economic Models in the Digital Economy*, Oxford Internet Institute.
- Byrne, D. & C. Corrado(2020), "Accounting for innovations in consumer digital services: IT still matters", In: C. Corrado et al(eds), *Measuring and Accounting for Innovation in the 21st Century*, University of Chicago Press.
- Carriere-Swallow, M. Y. et al(2019), "The economics and implications of data: An integrated perspective", International Monetary Fund Department Paper Series, No. 9/16.
- Ciuriak, D. (2019), "Economic rents and the contours of conflict in the data-driven economy", Policy Brief, Centre for International Governance Innovation.
- Coulton, C. J. et al(2015), "Harnessing big data for social good: A grand challenge for social work", Grand Challenges for Social Work Initiative Working Paper, No. 11.
- Duch-Brown, N. et al(2017), "The economics of ownership, access and trade in digital data", JRC Digital Economy Working Paper 2017-01.
- Einav, L. & J. Levin(2014a), "Economics in the age of big data", *Science* 346(6210):1-6.
- Einav, L. & J. Levin(2014b), "The data revolution and economic analysis", *Innovation Policy and the Economy* 14 (1):1-24.
- Erickson, S. & H. Rothberg(2015), "Big data and knowledge management: Establishing a conceptual foundation", *Electronic Journal of Knowledge Management* 12(2):108-116.
- Elvy, S. A. (2017), "Paying for privacy and the personal data economy", *Columbia Law Review* 117(6):1369-1460.
- Farboodi, M. et al(2018), "Where has all the big data gone?", SSRN Working Paper, No. 3164360.
- Farboodi, M. & L. Veldkamp(2020a), "A growth model of the data economy", NBER Working Paper, No. 28427.
- Farboodi, M. & L. Veldkamp(2020b), "Long run growth of financial data technology", *American Economic Review* 110(8):2485-2523.
- Farboodi, M. et al(2019), "Big data and firm dynamics", *AEA Papers and Proceedings* 109:38-42.
- Gaessler, F. & S. Wagner(2019), "Patents, data exclusivity, and the development of new drugs", *Review of Eco-*

- nomics and Statistics*, forthcoming.
- Henderson, J. V. et al(2012), “Measuring economic growth from outer space”, *American Economic Review* 102(2): 994—1028.
- Hillebrand, K. & L. Hornuf(2021), “The social dilemma of big data: Donating personal data to promote social welfare”, CESifo Working Paper, No 8926
- Höchtel, J. et al(2016), “Big data in the policy cycle: Policy decision making in the digital era”, *Journal of Organizational Computing and Electronic Commerce* 26(1—2):147—169.
- Hoeren, T. (2014), “Big data and the ownership in data: Recent developments in Europe”, *European Intellectual Property Review* 36(12):751—754.
- Hughes-Cromwick, E. & J. Coronado(2019), “The value of US government data to US business decisions”, *Journal of Economic Perspectives* 33(1):131—146.
- Hulten, C. & L. Nakamura(2017), “Accounting for growth in the age of the internet: The importance of output-saving technical change”, NBER Working Paper, No 23315.
- Ichihashi, S. (2020), “Non-competing data intermediaries”, Bank of Canada Staff Working Paper, No 28.
- Janssen, M. et al(2017), “Factors influencing big data decision-making quality”, *Journal of Business Research* 70(1):338—345.
- Jetzek, T. et al(2014), “Data-driven innovation through open government data”, *Journal of Theoretical and Applied Electronic Commerce Research* 9(2):100—120.
- Jones, C. I. & C. Tonetti(2020), “Nonrivalry and the economics of data”, NBER Working Paper, No 26260.
- Jorgenson, D. W. & K. M. Vu(2016), “The ICT revolution, world economic growth, and policy issues”, *Telecommunications Policy* 40(5):383—397.
- Kshetri, N. (2014), “Big data’s impact on privacy, security and consumer welfare”, *Telecommunications Policy* 38(11):1134—1145.
- Koh, D. et al(2020), “Labor share decline and intellectual property products capital”, *Econometrica* 88(6):2609—2628.
- Kubina, M. et al(2015), “Use of big data for competitive advantage of company”, *Procedia Economics and Finance*, 26(1):561—565.
- Liu, Z. et al(2020), “Data privacy and temptation”, NBER Working Paper, No 27653.
- Lohr, S. (2012), “The age of big data”, *New York Times*, Feb 11.
- McAfee, A. et al(2012), “Big data: The management revolution”, *Harvard Business Review* 90(10):60—68.
- Mueller, M. & K. Grindal(2019), “Data flows and the digital economy: Information as a mobile factor of production”, *Digital Policy, Regulation and Governance* 21(1):71—87.
- Müller, O. et al(2018), “The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics”, *Journal of Management Information Systems* 35(2):488—509.
- Morris, S. & H. S. Shin(2002), “Social value of public information”, *American Economic Review* 92(5):1521—1534.
- Nickerson, D. W. & T. Rogers(2014), “Political campaigns and big data”, *Journal of Economic Perspectives* 28(2): 51—74.
- Nordhaus, W. D. (2015), “Are we approaching an economic singularity? Information technology and the future of economic growth”, NBER Working Paper, No 21547.
- Ordóñez, G. (2013), “The Asymmetric effects of financial frictions”, *Journal of Political Economy* 121(5):844—195.
- Provost, F. & T. Fawcett(2013), “Data science and its relationship to big data and data-driven decision making”, *Big Data* 1(1):51—59.
- Reinsdorf, M. & J. Ribarsky(2019), “Measuring the digital economy in macroeconomic statistics: The role of data”, International Monetary Fund Working Paper.
- Romer, P. M. (1990), “Endogenous technological change”, *Journal of Political Economy* 98(5): 71—102.
- Rothberg, H. N. & G. S. Erickson(2017), “Big data systems: Knowledge transfer or intelligence insights?”, *Journal of Knowledge Management* 21(1):92—112.
- Solow, R. M. (1956), “A contribution to the theory of economic growth”, *Quarterly Journal of Economics* 70(1):65—94.
- Statistics Canada(2020), “The value of data in Canada: Experimental estimates”, *Latest Developments in the Canadi-*

- an Economic Accounts* (Working Paper Series), No. 9.
- Syverson, C. (2017), "Challenges to mismeasurement explanations for the US productivity slowdown", *Journal of Economic Perspectives* 31(2):165–186.
- Schaefer, M. & G. Sapi(2020), "Learning from data and network effects: The example of internet search", SSRN Working Paper, No. 3688819.
- Tambe, P. & L. Hitt(2012), "The productivity of information technology investments: New evidence from IT labor data", *Information Systems Research* 23(3):599–617.
- Tambe, P. et al(2020), "Digital capital and superstar firms", NBER Working Paper, No. 28285.
- Tene, O. & J. Polonetsky(2012), "Privacy in the age of big data: A time for big decisions", *Stanford Law Review Online* 64:63–69.
- Varian, H. R. (2014), "Beyond big data", *Business Economics* 49(1):27–31.
- Varian, H. (2018), "Artificial intelligence, economics, and industrial organization", NBER Working Paper, No. 24839.
- Veldkamp, L. (2005), "Slow boom, sudden crash", *Journal of Economic Theory* 124(2):230–257.
- Veldkamp, L. & C. Chung(2019), "Data and the aggregate economy", *Journal of Economic Literature*, forthcoming.
- Vives, X. (2019), "Digital disruption in banking", *Annual Review of Financial Economics* 11: 243–272.
- Yan, J. et al(2015), "How signaling and search costs affect information asymmetry in P2P lending: the economics of big data", *Financial Innovation* 1(19):1–11.
- Zhu, C. (2019), "Big data as a governance mechanism", *Review of Financial Studies* 32(5): 2021–2061.

Research Progress on Data as a Factor of Production

XU Xiang¹ LI Keabo² TIAN Xiaoxuan¹

(1. Central University of Finance and Economics, Beijing, China;

2. Tsinghua University, Beijing, China)

Abstract: In the digital economy era, data is widely seen as a factor of production, and plays a key role in social economic activities. In this paper, we first study the economic concepts and main characteristics of data, and emphasize the key features of data on the basis of virtuality and non-rivalry; increasing marginal returns, strong positive externalities, vague property and derivativeness. We then introduce major estimation methods for estimating the value of the data factor of production, and underlining macro theories of economic growth. After that, we investigate the impact of the data factor of production on micro decision-making and productivity, related industrial organization issues, and how data reduced information frictions in financial markets. At last, we introduce data-related privacy issues, and potential data market mechanisms.

Keywords: Data Factor of Production; Data Values; Economic Growth; Firm Decision-making; Market Mechanisms

(责任编辑:刘洪愧)

(校对:刘新波)