

数据要素经济学： 特征、确权、定价与交易

□张明 □路先锋 □吴雨桐

数据作为新型生产要素已经融入到生产、分配、流通和消费的各个环节,有助于生产力增强、企业创新和资源配置效率提升。相较其他传统要素,数据要素具有虚拟性与非消耗性、非竞争性、价值不确定性、非静态性、正外部性五大特征,并对包括规模经济、范围经济、边际报酬递减规律、私人物品与公共物品等基本经济学原理赋予新的意义。梳理和比较文献后发现,数据确权、数据定价和数据交易构成了数据要素市场的三大核心内容。从促进数据要素市场发展视角看,数据确权亟需明确权利性质和划分权利主体,数据定价需要结合具体的场景、定价策略和定价模型,数据交易则依赖数据参与者存在不同的商业模式和交易机制。展望未来,数据技术发展、数据产权分置和数据垄断问题将会成为影响数据市场发展不可忽视的重要因素。

关键词:数据要素;数据确权;数据定价;数据交易

中图分类号:F727 文献标识码:A 文章编号:1003—5656(2024)04—0035—10

引言

21世纪正在发生一场新的科技革命和产业变革,全球经济正在经历前所未有的巨大变化。据IDC发布的《数据时代2025》预测,全球数据量将从2018年的33ZB增加到2025年的175ZB。快速发展的数字化不仅导致交换的数据量急剧增加,而且数据经济规模也会空前增长。根据中国信通院的统计,2022年,我国数字经济规模达50.2万亿元,占GDP的比例达到41.5%。数据作为新型生产要素已成为数字化、网络化和智能化的基础,迅速地渗透到生产、分配、流通、消费和社会服务管理等各个领域,深刻地改变了我们的生产方式、生活方式以及社会治理方式。在此背景下,全国自2015年贵阳大数据交易中心设立后陆续诞生了50多个交易中心。截至2019年,有18个省市建立了省级大数据管理部门。2019年,十九届四中全会首次在中央层面确定数据可以作为生产要素参与分配,数据要素自此成为继土地、资本、劳动、技术之后的第五大生产要素。2022年6月,被称为“数据二十条”的《关于构建数据基础制度更好发挥数据要素作用的意见》在中央全面深化改革委员会第二十六次会议上审议通过。2023年3月7日,国家数据局宣告成立,负责协调推进数据基础制度建设,统筹数字资源共享开发和利用。

数据不仅引起了市场界和政策界的强烈关注,也在学术界产生了较大影响^[1-2]。一方面,数据作为生产要素被纳入到经济增长模型中^[3-4],数据要素被证明在企业创新、缓解信息不对称等领域对经济有显著的促进作用^[5];另一方面,围绕着数据的一系列问题引发了学术界的广泛讨论,例如为什么数据可以作为一种生产要素,数据与传统的生产要素有什么区别,数据对经济学的基本原理可能带来哪些改变等^[6]。

数据确权、数据定价和数据交易是数据要素流通和数据市场建设的三大核心内容。数据确权是指

作者简介:张明,中国社会科学院金融研究所研究员、博士生导师,国家金融与发展实验室研究员;路先锋(通讯作者),中国社会科学院大学应用经济学院博士研究生;吴雨桐,中国社会科学院大学应用经济学院博士研究生。

对数据所有权和使用权进行确认和授权的过程,是数据市场发展的前提。通过数据确权,可以明确和保护数据所有权,从而保护数据提供者的知识产权和隐私,防止滥用和不当使用,让数据提供者更加愿意开放和共享数据,并促进数字经济的发展和交易市场的稳定。数据定价是根据市场需求和供给、数据类型、数据质量等因素来确定数据价值,并将其转化为具体的货币金额的过程。数据定价为数据提供者和数据购买者之间建立公平和合理的交易机制,并发挥价格调节作用,引导市场开发和分享促进经济发展的数据资源,在数据市场发展起到关键作用。数据交易涉及到对数据资源进行买卖或租赁等方式的交换行为,在谈判和协商中确定数据价值、价格、使用条款等关键要素。数据交易与数据定价密不可分,两者相互影响。

本文的边际贡献有以下两点:第一,比较了数据要素与传统生产要素的特征差异,系统总结和阐述了数据要素具有虚拟性与非消耗性、非竞争性、价值不确定性、非静态性、正外部性五大特征,并且分析了这些特征对于规模经济、范围经济、边际报酬递减规律、私人物品与公共物品等基本经济学原理的影响;第二,本文探讨了数据市场的三大核心内容,即数据确权、数据定价和数据交易,结合学术理论和现实实践,深入阐述了数据确权的权利性质和权利主体划分,数据定价的对象、策略和模型,以及数据交易的商业模式和交易机制,归纳了数据要素市场研究的最新进展。

一、数据要素的特征与经济学原理

数据要素是继资本、劳动、土地和技术之外的第五大生产要素,本部分旨在概括其独特的经济学特征,并分析其对经济学基本原理的影响。

(一)数据要素的特征

第一,虚拟性与非消耗性。区别于资本、劳动、土地等传统要素,数据在形态上呈现虚拟性,且不具有损耗性,数据不像传统资本,如机器、建筑物或自然资源,会自然地衰减或耗尽。相反在使用数据的过程中又会产生新的数据,使用的频率越高,产生的新数据越多,数据的体量越大。第二,非竞争性。物质资本和劳动力具有竞争性,意味着一方使用它们会降低其他人的可用性,并且是排他性的,这意味着可以排除其他人使用它们。数据则是非竞争的,一个人的位置历史、医疗记录和驾驶数据可以同时被许多公司使用。第三,价值不确定性。数据价值受数据准确性和完整性影响。而数据的准确性和完整性依赖于数据生态系统,即各种数据产生、收集、传输、存储、分析和应用等环节组成的全局性数据网络体系,包括数据处理平台、数据服务提供商、数据使用者等多个参与方。在这个生态系统中,数据的质量和值受到整个环境和各个参与方的影响,需要通过协同合作和共享开放等方式,不断完善和优化数据生态系统的运作机制和治理模式。第四,非静态性。非静态性主要体现在动态转换和时效性两个方面:其一,与物质资本和劳动力不同,数据可以通过创新和新技术来创建和转换,物质资本和劳动力相对固定,不易改造或提高;其二,数据是时刻更新的,尽管交易中可以约定所交易的数据是实时更新的或是历史产生的,但时效性是影响数据价值的重要特征之一,数据的价值可能随着时间的推移而贬值,贬值的速度取决于数据的类型和市场需求。第五,正外部性。当产品或服务的价值随着越来越多的人使用而增加时,就会产生网络效应。网络效应是指一种产品或服务的价值随着越来越多的用户采用而增加的现象,数据网络效应可以带来数据使用的良性循环,使所有用户受益。数据是网络效应的关键推动因素,尤其是在数字经济中,大量数据可用于改进算法以及个性化产品和服务,会产生积极的正外部性。

(二)数据要素的经济学原理

数据要素是否具有规模经济和范围经济效应?是否依然满足边际报酬递减规律?以及数据要素究竟属于公共物品还是私人物品?本部分试图回答这些新问题。

1. 规模经济与范围经济

首先,数据的虚拟性和非消耗性对规模经济有深远的影响。数据存储成本随着时间的推移逐渐降低,使企业能够积累更多的数据。数据可以被企业用于市场细分、客户定位和个性化营销等,从而提高企业的效率和生产力。此外,数据的非消耗性还为企业提供了更多的商业机会,如将数据作为一种商品出售给其他公司或研究机构以创造新的收入来源。同时,企业还可以与其他公司合作,共享数据并开发新的产品和服务,从而提高整个行业的效率 and 创新能力。

其次,数据的非竞争性也给规模经济效应带来了新的影响。一方面,数据的非竞争性使得数据复制的成本几乎为零,企业的数据驱动决策模式依赖数据的规模,随着数据的不断积累,数据驱动决策的效率更高。因此,数据规模的扩大可以通过促进企业创新和产品质量的提高,进一步降低成本。另一方面,规模经济在数据驱动的背景下可能会有所不同。数据驱动行业的规模经济是由利用数据和算法提供更加个性化和差异化的产品和服务的能力驱动的。在这种情况下,产出与成本之间的传统关系可能仍然存在,但规模经济的来源可能会从传统的物理投入转向数据和算法^[7]。

最后,数据要素的正外部性带来了“网络效应”,推动范围经济的出现。平台从收集的用户数据中了解得越多,平台对每个用户的价值就越大,那么就表现出数据网络效应。此外,拥有更多的数据也可以帮助企业实现产品和服务的差异化,使用数据分析可以让公司更好地识别和利用不同产品或服务之间的互补性,从而进一步实现范围经济。因此,即使不生产大量不同的产品,企业也可能实现范围经济。这是因为数据分析可以帮助公司确定组合不同产品或服务的最有效方式,从而为客户提供更大的价值。

2. 边际报酬

数据要素是否符合边际报酬递减规律在学术界存在较大的争议。部分学者认为数据要素是一种高级生产要素,具有网络效应、规模经济和正外部性,数据越多边际报酬越高^[8]。而其他学者则从数据质量较差、可模仿、易被替代等角度进行思考,认为数据要素边际报酬递减。实际上,争议的根源在于数据具有极强的场景依赖性,数据在不同的情境下很可能表现出完全相反的经济学规律。王超贤等^[9]从状态依赖视角分析了数据边际报酬规律,认为数据质量、规模报酬、外部性、学习效应和质量阶梯是影响数据要素边际报酬的影响因素,并从数据生产端和数据交易端分析了数据的报酬性质。例如,自动驾驶、智能诊疗、图像识别在生产端拥有强大的规模经济、正外部性、生产力阶跃效应,在交易端这些数据密集型产品或服务的价值容易体现,具有明确的规模报酬递增性质;相反,像智能家居和经济学研究此类产品式活动在生产端不具有规模经济,也不具有需求端正外部性,具有明确的规模报酬递减性质。

3. 私人物品与公共物品

除了规模经济与范围经济、边际报酬递减规律外,数字平台的兴起和数据驱动的决策制定模糊了私人物品和公共物品之间的区别。根据经济学经典定义,私人物品具有竞争性和排他性,而公共物品则具有非竞争性和非排他性。然而,数据创造了非竞争性但具有排他性的新商品,例如数字音乐。近年来,学者们提出了对传统经济理论的各种修改,以解释这些新型商品:一种方法是将“俱乐部商品”纳入经济理论^[10]。“俱乐部商品”是非竞争性但具有排他性的商品,只有支付会员费才能获得;另一种方法是关注数据作为新生产要素的作用,数据可以被视为一种新型资本,可用于生产私人 and 公共产品^[11]。还有学者提出了“半公共产品”(quasi-public goods)的概念,即同时具有公共产品和私人产品特征的产品。例如,使用者可以自主选择是否分享个人信息,从而在一定程度上实现数据的排他性。半公共产品的出现对经典经济学中私人产品和公共产品的理论框架进行了一定的修正和拓展。

4. 数据要素特征与经济学原理的内在联系

通过以上分析,我们发现数据要素与传统要素不同的五大特征给数据要素的基本经济原理赋予了

一些新的内容和争议。如图1所示,本文试图总结数据要素新特征与经济学基本原理之间的内在联系。首先,数据的虚拟性与非消耗性、非竞争性共同推动了数据要素的规模经济效应;数据的正外部性则通过企业产品差异化、“网络效应”等为企业带来范围经济。其次,影响数据要素边际报酬的因素很多,数据要素的价值不确定性和非静态性导致数据具有极强的场景依赖性,导致规模经济效应和正外部性带来的范围经济也表现出场景异质性,以及学术界对于数据边际报酬的争议。因此,关于数据要素的边际报酬递减规律,我们需要根据具体的数据应用场景具体分析。最后,数据要素的非竞争性模糊了私人物品和公共物品的界限,学者们提出“俱乐部商品”和“半公共物品”的概念对经济学原理进行拓展。

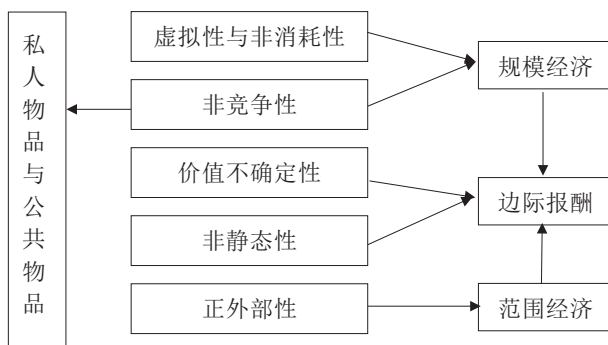


图1 数据要素的典型特征与经济学原理

资料来源:笔者绘制。

二、数据确权

数据确权是数据作为生产要素顺畅流转的前提,在法学、公共管理和经济学等领域中均有广泛探讨。法学者主要关注制度约束,包括行为控制模式和权利规范模式;公共管理学者主要关注数据确权与数据开发的公共治理;经济学者主要关注机制设计,以确保数据权利分配贴近于社会最优。但总的来说,数据确权主要涉及两大主题:一是明确数据权利性质,二是划分数据权利主体。

(一)明确数据权利性质

首先,数据能否作为客体被确权是明确数据权利性质的前提。有学者认为数据的无形性使其缺乏民事客体所要求的独立性,因此无法被权利化且本身不具备经济价值^[12]。但与此相对应的是,作品、发明、商标等无形物也是知识产权客体,却已被纳入民事权利范畴。这表明物质或无形并不是判断是否为客体的必要条件。从法律实践的角度来看,物权法、合同法等规范已经在一定程度上涉及到数据确权问题。例如,商业数据库和数据集等具有商业价值的数据可以通过物权法进行确权;而针对数据交易,则可使用合同法进行确权。

其次,按照权利属性,数据确权包括所有权与使用权。数据所有权是指数据主体对其数据拥有的所有权利,包括决定其数据的收集、使用和共享方式,并有权控制其数据的使用和传播。数据所有权的概念在法律和政策领域中得到了广泛关注和探讨。欧盟《通用数据保护条例》(GDPR)规定了数据主体对于其个人数据的访问、更正、删除、限制使用和移植的权利。美国加利福尼亚州通过《消费者隐私法》(CCPA)规定了数据主体对于其个人数据的访问、删除和限制使用的权利。数据使用权是指企业或其他机构在获得数据主体授权后可以使用其个人数据的权利。数据使用权的合法性需要遵守相关的数据保护法规和道德准则,以确保数据使用的公正性和透明性。欧盟GDPR规定了数据处理的合法性原则,即必须经过数据主体的明确同意或符合法律规定的其他合法依据才能使用个人数据。

最后,按照生成场景,数据确权涵盖了个人数据、企业数据和政府数据。首先是个人数据,包括与个人相关的各种信息,如姓名、地址、电话号码、个人健康记录、金融交易记录和社交媒体活动。这些数据对个人至关重要,缺乏适当的保护可能会被不法分子利用,从而涉及诈骗、身份盗窃等违法行为。但对于无法识别的自然人非个人数据,法律不应在数据收集、转让和使用上施加过多限制^[13]。其次是企业数据,指企业拥有的各种信息,如财务报表、销售统计、客户数据和产品研发信息。企业数据具有极高的商

业价值,可以帮助企业做出有根据的决策,提高效率和竞争力。然而,企业也需要保护这些数据,以防泄漏、滥用或受到黑客攻击等威胁。尽管《非个人数据自由流动框架》明确排除了个人数据的保护范围,但随着数字化的发展,一些国家已开始扩展对个人数据定义的范围,将企业等其他类型的数据纳入相关保护范围^[14]。最后是政府数据,指政府机构收集和持有的各种信息,如税务数据、人口普查数据和公共安全数据。政府数据对于制定政策和规划社会发展至关重要,若处理不当,可能对公众造成伤害,例如个人隐私泄露和数据滥用等问题。

(二)数据权利主体的划分

数据作为法律客体被确权的合理性越来越得到认可,下面本文基于数据所有权和数据使用权的框架具体分析目前关于数据权利主体的划分情况。

第一,数据所有权归属的争议。关于数据所有权归属问题,当前存在三种代表性观点。第一种观点认为数据所有权应该划归给消费者。因为数据具有非竞争性,只有在市场上所有厂商都能够使用数据时,数据价值才能最大化。如果将数据所有权分配给企业,则企业可能会选择保护和独享数据,从而拒绝与竞争对手共享数据。相反,如果将数据所有权分配给消费者,则消费者会选择与所有厂商共享数据,从而最大化数据价值^[15]。第二种观点认为数据所有权应该划归给企业。因为企业拥有数据产生规模经济和范围经济的优势,同时数据可替代性和隐私悖论等问题不太可能引发严重的隐私问题和数据垄断^[16]。第三种观点则认为不存在明确的权利主体。因为数据的应用场景多样多变,且数据具有复制性和非排他性,同一数据在不同应用场景下会产生不同的价值,数据所有权归属于消费者还是企业应该取决于数据的价值^[17]。数据的所有权归属于企业时,可能导致数据被过度使用并对隐私造成过度侵害。反过来,如果数据所有权归属于消费者,则可能使企业缺乏在数据处理上进行投入的动力。因此,在数据价值相对较低且处理效率不足时,将数据所有权划归给消费者是合适的选择;但在高价值数据的情况下,需要采用不同的策略来进行权利配置。

第二,从数据所有权转向数据使用权。数据的使用权逐渐成为人们更为关注的焦点,尤其是在数据所有权争议和确权困难的背景下。申卫星^[18]提出了一种二元权利结构,即将数据原发者拥有数据所有权与数据处理者拥有数据用益权分开考虑,以实现数据财产权益分配的均衡。数据用益权可以通过数据所有权人授权、数据采集、加工等事实行为取得,也可以通过共享、交易等方式继受取得。此外,根据杨竺松等^[19]的研究,在不完全契约理论的视角下,可以将数据开发过程分为开发环节和使用环节两个阶段,阐释数据开发合约中要素投入无效率的来源,从社会最优视角认为公共数据应由公共部门掌握,而专属性较强的数据则应由市场调节。因此,数据所有权转向数据使用权的趋势日益明显。

第三,数据使用权的分级授权模式。分级授权模式以市场化的授权协议来合理、合法地使用数据,而不考虑平台上数据衍生出的复杂权利及相关权属问题。刘涛雄等^[20]认为数据可以被视作生成品,其初始产权应当归参与生成过程的主体所有,并通过参与主体之间分散的市场化契约来确立初始产权。基于“按贡献分配”和“事前确权”的原则,可以根据数据的生成场景提出数据确权的分级授权体系。用户和数字平台可以根据数据的不同生成场景,以市场化的方式达成不同层级的数据授权协议,让平台可以基于这些协议不同程度地使用数据要素,并开展数字经济相关的生产活动。

三、数据定价

总体来看,数据要素定价首先要明确定价对象,然后根据数据场景专用性设定不同的定价策略,再进一步和经济学模型的定价思路结合,最终构成数据市场上看到的数据报价。

(一)数据定价对象和应用场景

数据定价的对象经常会涉及两个概念：一是数字产品，二是数据产品。数字产品是指那些无形的商品，但可以通过电子产品消费，例如电子书、可下载的音乐、在线广告和互联网优惠券，许多数字产品以某种方式具有物理对应关系，但并非绝对必要；数据产品是指数据集作为产品和从数据集衍生出来的信息服务。虽然在某些场景下，数字产品和数据产品的界限也很模糊，但是本文所阐述的数据要素定价主要是指数据产品的定价。

数据的使用具有极强的场景依赖性，主要分为数据订阅、数据租赁、数据查询、定制数据和数据托管等几种场景。数据订阅是指用户按照一定的时间周期或数量购买数据，适用于需要频繁获取特定类型数据的用户，如市场分析师和投资者。数据查询是指用户可以按需随时获取数据，适用于需要即时访问特定数据的用户，如移动应用程序和数据查询。访问的优势在于随时获取数据，无需长期承诺或额外费用。定制数据是指用户需要特定数据，但市场上无现成数据可满足需求，用户可以选择委托数据提供商进行特定数据的定制，适用于对数据准确性和特殊要求较高的用户。尽管这种场景需要付出高昂费用并进行定制开发，但对特定需求的满足是其优势所在。数据托管是指用户将数据存储在第三方数据中心而非本地设备上，适用于需要安全、可靠、高性能数据存储和处理的用户，如企业级应用程序和金融机构。托管数据的优势在于用户无需购买昂贵硬件设备和维护复杂系统，同时能获得更高的数据可用性和安全性。

（二）数据定价策略

根据数据产品的定价对象，以及订阅、租赁、查询、定制数据、数据托管等应用场景，数据定价策略可分为使用量、使用时间、使用内容和组合搭配四种定价策略。

第一，数据使用量定价策略。这一定价通常基于数据使用量的阶梯式计费，即客户使用的数据量越多，单价越低。数据使用量定价策略在中国的云计算和SaaS市场中很常见。例如，阿里云、腾讯云等云计算服务商提供了不同类型的云存储和数据库服务，其定价策略往往采用按使用量计费的方式。以阿里云的对象存储服务为例，其定价分为标准存储、低频访问存储、归档存储等不同类型，每种类型都采用了按使用量计费的策略，即按照客户使用的数据存储量计算费用。此外，SaaS行业的在线会议、视频直播等服务也采用了数据使用量定价策略，即根据客户使用的带宽和观看时长等因素来计费。

第二，数据使用时间定价策略。客户需要按照时间段来支付费用，例如每月或每年的固定费用，这种定价策略适用于需要长期使用数据的客户，例如高校、科研院所、政府机构等。以财新数据通为例，它会根据用户所选择的数据类型和使用时间来计算价格。例如，对于某个数据集，如果用户选择使用一个月，则需要支付相应的费用。如果用户选择使用更长时间，费用也会相应增加。

第三，数据使用价值的定价策略。这一定价策略主要考虑数据的稀缺性、时效性、精度以及数据的应用场景和客户需求等因素。对于稀缺的数据，定价会更高，比如某些地质数据，对于石油公司而言非常重要。数据的时效性很重要，定价也会更高，比如准确的天气预报数据对于农业和交通等行业非常关键。另外，如果数据的精度很高，定价也会更高，比如新能源汽车使用的高精度地图。

第四，数据组合定价策略。将多个数据集合在一起作为一个整体进行销售，常用于需要多个数据源进行交叉验证和分析的场景。定价方式通常根据数据集的价值、数量、时效性、覆盖范围等因素来评估和确定。我国部分数据交易中心和平台也提供数据组合定价策略，如天眼查数据中心、易观千帆数据中心等。例如，天眼查数据中心提供了多种数据组合方案，如“企业品牌大数据分析套餐”“企业信用数据分析套餐”等，根据不同的客户需求和场景提供多种组合选择，定价方式也根据数据集的价值和数量等因素进行评估和定价。

（三）定价模型

数据定价策略实际上只是微观实践中的一种定价思路,学术界还提出了基于数据的定价模型,以测算符合数据交易者利益最大化或数据市场均衡的价格。

第一,无套利定价模型。Lin 和 Kifer^[21]研究了数据访问的无套利定价,初始阶段客户仅需提供查询包,定价则根据查询包而定。但具体价格并非基于客户查询内容确定,在客户正式购买查询答案之前,他们知晓计费方式但不了解具体价格,一旦同意购买,客户将在获得答案时付费。Chen 等^[22]则为多个版本的机器学习模型开发了无套利定价设计,通过在最优模型中引入随机高斯噪声为不同购买者生成不同版本。第二,收入最大化定价模型。Xia 和 Muthukrishnan^[23]考虑了指定买家购买的用户数量的最大和最小情况,并提供了一个近似算法来最大化收入。Chawla 等^[24]则在假设所有买家都是单一的且供应无限的情况下,通过考虑基于查询和视图的定价,以实现仲裁者的收入最大化。第三,博弈模型。博弈论在数据商品定价领域是有效的方法,包括非合作博弈、斯塔克伯格博弈和讨价还价博弈三种方案。非合作博弈是指所有参与者都不合作,以竞争方式销售他们的数据。Luong 等^[25]设计了一个定价模型来评估物联网传感数据,供应商扮演参与者的角色,决定定价策略,所有参与者必须透明地发布其定价策略。Haddadi 和 Ghasemi^[26]提出了一个斯塔克伯格模型,以保护公开其定价策略的玩家,并证明了领导者可以获得比使用其他模型更好的回报。最后是讨价还价博弈,供应商和消费者需要就价格进行协商,只有在达成一致时才会发生交易。第四,隐私补偿定价模型。当用户与他人共享数据时,用户可能会在某种程度上披露自己的隐私。Ghosh 和 Roth^[27]设计了真实的市场,数据买家想要购买数据来估计统计数据,而卖家想要赔偿他们的隐私损失,在设计中,只有一个查询,并且对其数据的个人评估是私有的,数据所有者被要求报告使用其数据的成本。

四、数据交易

数据权利的明确为数据交易市场提供了合法的参与者,形成了不同的数据交易商业模式,数据定价则为数据市场提供了交易的基础条件,加上数据买卖的合同设计,一起构成了数据市场交易的核心内容。

(一)数据交易的参与者

根据数据生成阶段、加工阶段和消费阶段的划分,数据交易市场包括了三类参与者,即数据拥有者、数据消费者和数据经纪人。数据所有者是拥有数据并愿意将其数据货币化的人,例如拥有自己的社交、财务、位置、健康数据的个人,以及收集有关其用户数据的公司;数据消费者是寻求外部数据以改进他们的决策、产品设计、服务和客户管理的人或组织,例如广告商、软件开发商、零售商、制造商、电信服务提供商;数据经纪人是一个中介,从数据所有者那里获取数据,整合后开放数据,然后将数据出售给数据消费者,并从数据交易中获利。

(二)数据交易的商业模式

根据互联网上数据产品公司的分析,并以数据交易实体的角色为依据,数据交易模式可以分为“单边数据提供商模式”和“双边交易平台模式”。

第一,单边数据提供商模式。此类模式下的卖方有数据产品提供者和数据服务提供者两种类型。数据产品提供者将其持有的数据卖给数据消费者,如万得、国泰安数据库,将整理好的数据库打包供消费者下载使用。而数据服务提供者则将其持有的数据整合为可以向买方提供的服务,典型的例子是提供人脸识别服务的 Clearview AI,该公司将其从公开互联网等渠道收集的人脸照片整合为人脸识别服务,提供给执法机构等客户。单边数据提供商模式可以分为买方市场和卖方市场,其中一方在定价中具有主导优势,而另一方则相对弱势。以卖方市场为例,卖方可以根据买方竞争的程度来选择定价方法。在

这种情况下,卖方可以通过限制竞争、控制供应量和提高价格等手段来实现垄断地位。

第二,双边交易平台模式。双边交易市场有两种类型,一是依赖数据中介机构的集中式双边市场,二是买卖双方直接交易的分散式双边市场。前者产生的原因在于,数据中介机构拥有较强的数据采集和分析能力,可以提供整合程度更高、数据质量更好、种类更丰富的数据产品和服务。以数据堂为例,它拥有1000余家合作伙伴,遍及50多个国家和地区,装备专业采集设备和加密工具。客户提出需求后,数据堂能够为其量身定制采集方案并确保数据的安全。这种更市场化、门槛较低的交易模式调动了供需双方的积极性,促进了数据的汇聚和再利用。分散式数据交易平台不依赖中介机构,具有去中心化的特点。卖家可以在数据交易平台上联系潜在的买家,之后由数据交易平台管理他们之间数据交易的流程。大部分的数据交易平台上可以交易各种类型的数据,如我国贵阳大数据交易中心、北京国际大数据交易中心等。但也有少数专注于汽车、能源等某些行业和物联网传感器数据等特定类型数据的交易平台。例如,专门为物联网传感器数据提供数据交易平台的美国公司Terbine。

(三)数据市场交易

数据市场交易既包括数据提供者和数据购买者的交易形式,也包括数据交易后数据的转移模式。

第一,数据交易机制。数据交易主要采用直接交易机制和拍卖机制两种方式。对于排他性较弱、潜在买家众多的数据产品,卖家通常采取直接交易方式以保底价格定价并设计产品菜单供消费者自主选择。单边交易、集中式双边交易和数据管理系统等也多采用直接交易方法。而如果数据产品具有较强的排他性,可以采用拍卖机制进行定价。拍卖机制遵循公开、公平、公正的原则,具有快速和批量的特点,可以激励代理人披露真实估值,从而提高交易效率^[28]。目前已有许多研究探讨了双边拍卖、反向拍卖和VGG拍卖等多种拍卖类型。选定何种拍卖机制需要考虑买方和卖方的相对市场力量、是否有中介机构充当拍卖商、定价原则等因素。我国的贵阳大数据交易所、华中数据交易所、上海数据交易中心等多家数据交易平台都采用拍卖机制。

第二,数据转移模式。数据转移通常涉及三种模式。第一种是托管交易模式,每个数据所有者将其数据存储于数据交易中心的数据库中,并通过数据交易中心与数据购买者进行交互与交易。在此模式下,数据所有者对于交易过程中的对象、数量等内容没有任何决定权,而其权益完全取决于数据交易中心的信用。第二种是聚合交易模式,在该模式下,数据交易中心通过API接口链接数据所有者,数据所有者无需将其数据事先存储至数据交易中心,而是由其自行管理。当数据购买者有需要购买数据时,数据拥有者可以通过数据交易中心与其进行实时交互,并将信息返回给数据购买者。虽然表面上看来,数据在聚合交易模式下被数据所有者所控制,但仔细分析API接口数据获取机制后会发现,采用聚合交易模式的数据交易中心具有保留交易数据的能力和机会。因此,随着交易的不断进行和数据的积累,一个数据聚合中心逐渐变成了一个数据托管中心。第三种是基于区块链技术的数据转移模式,利用了区块链的不变性、防篡改性、可追溯性以及智能合约的可编程性,实现去中心化,降低传统的数据交易模式的安全风险。Xiong和Xiong^[29]提出了一种基于区块链的数据转移方案,包括验证数据的可用性时使用相似性学习方法,建立数据购买者和数据所有者间的挑战响应机制。

五、结论与展望

随着数字化时代的到来,数据作为新型生产要素已经成为当今社会中的重要资源。数据可以促进各个领域的生产力增强、创新与发展,同时也能够提高资源配置效率。相比传统要素,数据要素具有虚拟性与非消耗性、非竞争性、价值不确定性、非静态性和正外部性五大特征。这些特征使得数据要素对包括规模经济、范围经济、边际报酬递减规律、私人物品与公共物品等基本经济学原理产生新的影响。

数据要素市场上的三大核心内容是数据确权、数据定价和数据交易。在数据确权方面,首先要明确权利性质,其次根据数据所有权和使用权分析权利主体,其中数据分级授权模式更加符合实践。在数据定价方面,存在着订阅、租赁、访问、数据定制和数据托管等应用场景,根据数据的应用场景可以制定基于使用量、使用时间、使用价值和组合定价四种策略,结合无套利模型、收入最大化模型、博弈模型和隐私补偿等定价模型,可以更好地实现数据市场均衡价格。在数据交易方面,则需要根据数据参与者包括数据拥有者、数据消费者和数据经纪人,根据数据参与者数据交易模式可以分为“单边数据提供商模式”和“双边交易平台模式”,数据交易机制可以根据数据的排他性分为直接交易和拍卖交易两类,数据转移有数据托管、数据聚合和区块链去中心化三种模式。

展望未来,在数据要素市场化的发展过程中,数据技术发展、数据产权分置和数据市场结构演变三种情景会对数据确权、数据定价和数据交易带来较强影响。

第一,数据技术的进步将深刻影响数据市场。区块链技术在数据确权方面发挥关键作用,通过去中心化和建设数据管理平台,确保数据确权更安全、透明、可靠。大数据和机器学习的结合提高了数据定价的准确性,更好地预测市场需求和价值,制定更合理的数据定价策略。技术如区块链降低了数据交易成本,提高了数据流通效率,推动了数据定价的公正和透明。然而,技术发展带来了一些潜在的挑战,包括成本上升、存储和传输问题以及合规性难题。在数据确权、数据定价和数据交易中,必须关注知识产权、隐私保护等法律法规,确保合规性成为维护数据市场可持续发展的必要条件。第二,数据产权的分置需求日益增强,公共数据、企业数据和个人数据具有主体差异性。公共数据趋向于共享和开发,企业数据宜建立确权授权机制,个人信息数据需要规范处理活动、强化保护,推动匿名化处理和创新技术手段。在数据定价方面,公共数据可向公众无偿开放,企业数据则适合根据维护成本定价。个人数据的隐私性导致采用隐私补偿定价模型,未来的发展将取决于隐私计算的创新。第三,数据厂商垄断值得警惕。数据具有规模经济和范围经济的优势,掌握大规模或高信息含量的数据的企业倾向于囤积数据而不是分享,导致数据要素具有高度排他性。一些优势企业会通过建立数据库并出售数据使用权以获得收入。然而,现有文献和社会观点普遍认为企业往往出于遏制竞争对手的考虑,不愿意将消费者数据分享或出售给竞争对手,进而造成所谓的“数据垄断”。一旦垄断形成,数据控制者会采取价格歧视的定价策略以获取超额利润,并通过细化需求模型和价格函数来实现这一目标。未来,反数据垄断可能成为促进数据市场健康发展的重要任务。

参考文献:

- [1]徐翔,厉克奥博,田晓轩.数据生产要素研究进展[J].经济学动态,2021(4):142-158.
- [2]欧阳日辉,杜青青.数据要素定价机制研究进展[J].经济学动态,2022(2):124-141.
- [3]徐翔,赵墨非.数据资本与经济增长路径[J].经济研究,2020,55(10):38-54.
- [4]杨俊,李小明,黄守军.大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论[J].经济研究,2022,57(4):103-119.
- [5]徐翔,赵墨非,李涛,等.数据要素与企业创新:基于研发竞争的视角[J].经济研究,2023,58(2):39-56.
- [6]黄阳华.基于多场景的数字经济微观理论及其应用[J].中国社会科学,2023(2):4-24.
- [7]NEETHIRAJAN S.The role of sensors, big data and machine learning in modern animal farming[J].Sensing and bio-sensing research,2020(29):100367.
- [8]蔡跃洲,马文君.数据要素对高质量发展影响与数据流动制约[J].数量经济技术经济研究,2021,38(3):64-83.
- [9]王超贤,张伟东,颜蒙.数据越多越好吗——对数据要素报酬性质的跨学科分析[J].中国工业经济,2022(7):44-64.
- [10]CHOHAN U W,D SOUZA A.Club theory: a contemporary economic review[J].Available at SSRN 3512557, 2020.
- [11]BRYNJOLFSSON E,MCAFEE, A.The second machine age: work, progress, and prosperity in a time of brilliant technolo-

gies[M].WW Norton & Company, 2014.

[12]梅夏英.数据的法律属性及其民法定位[J].中国社会科学,2016(9):164-183.

[13]程 啸.论大数据时代的个人数据权利[J].中国社会科学,2018(3):102-122+207.

[14]韩旭至.数据确权的困境及破解之道[J].东方法学,2020(1):97-107.

[15]JONES C I, TONETTI C. Nonrivalry and the economics of data[J]. American economic review, 2020, 110(9): 2819-2858.

[16]费方域, 闫自信, 陈永伟, 等. 数字经济时代数据性质、产权和竞争[J]. 财经问题研究, 2018(2): 3-21.

[17]DOSIS A, SAND-ZANTMAN W. The ownership of data[J]. Journal of law, economics and organizations, 2019.

[18]申卫星.论数据用益权[J].中国社会科学,2020(11):110-131.

[19]杨竺松, 黄京磊, 鲜逸峰. 数据价值链中的不完全契约与数据确权[J]. 社会科学研究, 2023(1): 85-93.

[20]刘涛雄, 李若菲, 戎 珂. 基于生成场景的数据确权理论与分级授权[J]. 管理世界, 2023, 39(2): 22-39.

[21]LIN B R, KIFER D. On arbitrage-free pricing for general data queries[J]. Proceedings of the VLDB Endowment, 2014, 7(9): 757-768.

[22]CHEN L, KOUTRIS P, KUMAR A. Model-based pricing for machine learning in a data marketplace[C]. <https://doi.org/10.48550/arXiv.1805.11450>.

[23]XIA C, MUTHUKRISHNAN S. Arbitrage-free pricing in user-based markets[C]. // Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, 2018: 327-335.

[24]CHAWLA S, DEEP S, KOUTRIS P, TENG, Y. Revenue maximization for query pricing[J]. ArXiv preprint arXiv: 1909.00845, 2019.

[25]LUONG N C, HOANG D T, WANG P, et al. Data collection and wireless communication in Internet of Things (IoT) using economic analysis and pricing models: A survey[J]. IEEE communications surveys & tutorials, 2016, 18(4): 2546-2590.

[26]HADDADI S, GHASEMI A. Pricing-based Stackelberg game for spectrum trading in self-organised heterogeneous networks[J]. IET communications, 2016, 10(11): 1374-1383.

[27]GHOSH A, ROTH A. Selling privacy at auction[C]. // Proceedings of the 12th ACM Conference on Electronic Commerce, 2011: 199-208.

[28]JIN H, SU L, CHEN D, NAHRSTEDT K, et al. Quality of information aware incentive mechanisms for mobile crowd sensing systems[C]. // Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2015: 167-176.

[29]XIONG W, XIONG L. Smart contract based data trading mode using blockchain and machine learning[J]. IEEE access, 2019(7): 102331-102344.

(收稿日期: 2023—10—30 责任编辑: 赵爱清)

Data Factor Economics: Characteristics, Right Confirmation, Pricing and Transaction

Zhang Ming, Lu Xian-feng, Wu Yu-tong

Abstract: As a new factor of production, data are integrated into all aspects of production, distribution, circulation and consumption, and they are helpful to enhance productivity, innovation and resource allocation efficiency. Compared with other traditional factors, data have five characteristics: virtuality and non-consumption, non-competitiveness, value uncertainty, non-staticity, and positive externalities. Data economics will make some changes to the basic economic principles such as economies of scale, economies of scope, the law of diminishing marginal returns, private goods and public goods. Combining and comparing the literature reveals data rights confirmation, data pricing and data transactions constitute three core elements of data market. From the perspective of promoting the development of data factor markets, data rights confirmation needs clarifying the nature of rights and dividing rights subjects, data pricing should combine specific scenarios, pricing strategies and pricing models, while data transactions relies on the existence of different business models and trading mechanisms for data participants. Looking forward, the development of data technology, the division of data property rights and data monopoly will become three important forces shaping the evolvement of the data market.

Key Words: Data Factor; Data Right Confirmation; Data Pricing; Data Transaction