

# 数据要素的界权、交易和定价研究进展<sup>\*</sup>

熊巧琴 汤珂

**摘要:**数据及其交易的价值日益凸显,但缺乏系统、跨学科的理论研究与支持。本文首先对数据界权进行综述,在场景性公正原则下,针对数据内容和数据产品细分数据权利,明确数据资产化的前提。本文继而总结数据资产的可复制性、价值不确定性等特点;并对数据的流通方式,以及数字产品、数据产品的交易和定价分别进行综述。相对于要求企业主动共享数据,若干数据交易技术(尤其是采取第三方交易平台的方式)更符合经济理性。数据产品的交易模式和卖方策略受应用场景和买方异质性的影响较大,差异化产品和定价是最主要的思想,区块链技术是完善数据交易的方案之一。最后,本文对数据资产的多种定价原则和方法的适用性和局限性进行综述。

**关键词:**数据界权 数据资产 数据交易 数据定价 数字经济

世界经济已从围绕物品和货币流动转变为围绕信息流动来组织(Drucker, 1992),“数据是新的石油”(Varian, 2018)成为共识。《中国互联网发展报告(2019)》指出,中国 2018 年的数字经济规模为 31.3 万亿元,占国内生产总值的 34.8%<sup>①</sup>;且中国的数据要素规模增速领先,预计 2025 年将成为世界最大的数据圈<sup>②</sup>。党的十九届四中全会决议将数据增列为生产要素,要求健全“由市场评价贡献、按贡献决定报酬的机制”<sup>③</sup>。2020 年 4 月 9 日,《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》明确了提升社会数据资源价值、健全要素市场运行机制的方向和重点改革任务,这对于引导各类要素协同向先进生产力集聚,加快完善社会主义市场经济体制具有重大意义<sup>④</sup>。

数据要素作为数字经济的微观基础具有战略性地位和创新引擎的作用。促进数据要素跨企业、跨产业的流通和交易有利于数据流向最有价值的地方,将数据对经济社会的乘数效应推到最大。但迄今为止,数据资产交易仍然面临大量亟待解决的问题,如数据资产化的前提、数据交易与个人隐私保护的争议、数据流通方式的选择、数据资产交易与传统商品交易的区别、数据交易所需要的技术支持、数据交易机制设计、数据资产的价值衡量和定价方式等。本文将通过对数据界权、交易和定价相关文献的梳理,从数据界权、数据资产特点、数据共享和交易机制设计、数据资产定价几方面进行综述,以回答上述问题。

<sup>\*</sup> 熊巧琴、汤珂,清华大学社会科学学院经济学研究所,邮政编码:100084,电子邮箱:xqq18@mails.tsinghua.edu.cn, ketang@mail.tsinghua.edu.cn。基金项目:国家自然科学基金(71973075);国家社会科学基金重大项目(16ZDA008);“四个一批人才”理论名家项目。感谢匿名审稿人的修改建议,文责自负。

① 中国互联网协会:《中国互联网发展报告(2019)》。

② Reinsel, D. et al(2019):《2025 年中国将拥有全球最大的数据圈》, <https://wenku.ofweek.com/show-41933.html>。

③ 中国政府网: [http://www.gov.cn/zhengce/2019-11/05/content\\_5449023.htm](http://www.gov.cn/zhengce/2019-11/05/content_5449023.htm)。

④ 中国政府网: [http://www.gov.cn/zhengce/2020-04/09/content\\_5500622.htm](http://www.gov.cn/zhengce/2020-04/09/content_5500622.htm)。

## 一、数据的权利界定和特点

### (一)数据的定义和权利界定

数据权属界定不清已成为数据要素化的最大制度障碍之一。根据科斯定理,当市场交易存在成本时,如果初始的权利界定不当,很可能影响资源的最终配置并带来社会福利损失(Coase,1960)。因此,数据权利归属于平台、个体或者政府的初始配置将影响数据市场的发展和社会福利水平。

数据的界权(delimitation of rights)和交易需要平衡数据市场发展和个人权利保护。为了厘清二者的争议,首先需要梳理数据的定义。根据国际标准化组织(ISO)的定义,信息(information)是关于事实、事件、事物、过程或思想等客体的知识,在特定语境中有特定的含义;数据(data)是信息的一种形式化方式的体现,以达到适合交流、解释或处理的目的<sup>①</sup>。信息可以在三个层面上被支配和控制,分别是物理层的数据载体(存储介质)、符号层的数据产品(data goods)以及内容层的数据信息。其中,数据产品是信息整合的表现形式和组织结果,数据信息是数据产品蕴含的信息内容。数据产品存储在计算机、云盘等数据载体上,数据信息则包括个人信息和非个人信息,个人信息又可分为敏感和非敏感的个人信息。数据产品和数据信息构成广义上的“数据”(纪海龙,2018)。因此,个人信息和数据产品并不等同,二者的界权形式、交易方式也当有所不同。个人信息属于数据主体的权利,数据产品则属于数据产品生产/持有者的权利。

但是,学界对个人信息中的非敏感信息的赋权结论并不完全一致。一方面,购买历史、出行方式等通常是个人行为的副产品(Veldkamp,2005;Ordoñez,2013),其本质上属于数据主体的权利。Miller & Tucker(2017)的研究也表明,给予消费者对个人数据的控制权有助于促进个人数据的分享与交易。另一方面,大多数研究认为个人或企业拥有数据所有权的优劣没有绝对的最优解,而是取决于一些关键系数,例如数据对企业的价值和消费者货币化个人数据的能力(Dosis & Sand-Zantman,2019),或者取决于消费者对隐私保护与数据交易收益的最优决策(Abowd & Schmutte,2019)。Jones & Tonetti(2019)认为若由消费者掌握数据所有权,数据交易量将接近社会最优化水平(但弱劣于最优);若由企业掌握数据所有权,数据交易量可以达到社会最优交易量,但交易量与数据带来的创新性破坏程度成反比,企业可能滥用数据交易,也可能交易不足。

根据 Nissenbaum(2010)提出的场景性公正(contextual integrity)原则,结合具体的情景、参与者、数据特点、流通原则以及场景中各方的合理预期,如果新的信息流不违背该场景下传统的民主公平规范或者能更好地实现该场景的价值,那么便应当确定相关主体的数据权益。由于数据要素的人格化禀赋效应(Thaler,1980;Radin,1982)<sup>②</sup>、低复制成本,且强势一方(比如企业)很容易通过合同、协议将个人的数据所有权低价甚至免费“交易”到自己手中,赋予个人绝对权利并不能有效保障个体的数据隐私和安全,且不利于数据要素的租赁和交易。为了解决个人和数据企业的激励问题(Admati & Pfleiderer,1990)和信息悖论问题(Arrow,1962),同时避免数据滥用和垄断,应当针对不同隐私和风险级别的个人信息,给予数据生产者(自然人)不同级别(可通过行使数据人格权需满足不同条件的方式来体现)的拒绝权、可携权、获取收益等数据控制权,赋予数据产品持有者(例如数据收集者、设备生产者等)有限的占有权(除所有权之外的收益权、使用权等权益集合)<sup>③</sup>来解决上述问题。美国在不同行业分别制定隐私法,如《财务隐私法》《公平信用报告法》《有线通信信息法》《健康

<sup>①</sup>ISO/IEC: Information Technology-Vocabulary, Online Browsing Platform, <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en:en>。

<sup>②</sup>禀赋效应指如果一项产权与持有者的人格密切相关,人们对其拥有的物品的价值评价会更高的现象(Thaler,1980;Radin,1982)。

<sup>③</sup>“占有权”借鉴自龙登高的《中国传统地权制度及其变迁》(2018)一书。

保险携带和责任法》等,并与数据市场自由交易相结合,实际上便体现了分类分级保护隐私和鼓励数据交易的思想。

对于具有较大经济价值且经脱敏(desensitization)的个人信息,消费者和数据搜集/中介商之间可采取租赁式交易,数据中介商每使用/销售一次个人数据,对该消费者支付租金;或者采取一次性买卖以及其他隐私补偿方法(Ghosh & Roth,2015;Easley et al,2018),从而刺激原始数据的生产和交易。另外,可借鉴美国《同意法案》(Consent Act),通过制定额外的规则和程序,促使公司承担利用消费者提供的个人数据的应尽责任,包括对内容层面的个人数据进行充分的脱敏(Agarwal et al,2019),以及其他必要的合法处理,例如匿名数据再识别化的风险评估、应急补救等。

## (二)数据资产及其特点

基于上述对“数据产品”的赋权,数据产品交易的合法化前提已确立——数据是一种产权可界定、可交易的商品。从直观呈现的产品类型来区分,数据可分为数字产品(digital goods)和数据产品(Pei,2020)。前者是以数字形式存储、表现和使用的人类的思想、知识成果,如网易云歌曲、电子文献、在线课程等;后者是由网络、传感器和智能设备等记录的、可联结、可整合和可关联某特定对象的行为轨迹和关联信息,具有较强的分析价值,如各种机器生产和采集的内容。

数据要素化、数据资产化着重指的是数字化的数据,即数据产品。数据资产化的核心在于通过数据与具体业务融合,驱动、引导业务效率改善从而实现数据价值(何伟,2020)。一般而言,资产的核心特征主要包含三点:未来的收益性、所有者拥有对资产的控制权、由过往交易结果形成(Godfrey et al,2010)。因此,合法获取的由企业或个人产生的,预计会影响个人或企业未来的行为决策,并为个人或企业带来经济收益的各类数据资源都是数据资产<sup>①</sup>。大体量的数据产品集合又称作大数据资产。数据资产具有与传统资产、金融资产不同的特点。

1. 数据资产具有非竞争性且边际成本接近于零。数据资产可被无限分享和复制,且被分享和复制的数据资产一定程度上具有非竞争性,即使用者的增多不影响数据资产本身的价值(Moody & Walsh,1999)。然而,这给数据资产交易造成了困扰,只有少数情况下,数据产品的分享会给数据拥有者带来不利竞争(比如与其构成商业竞争关系)。当数据资产的复制既没有物理成本也不会损害个人或厂商的福利,甚至会给分享者创收时,即便理论上可以进行数据界权,也很难防止用户将数据资产进行二次转售,从而损害数据产品创作者的利益,这是数据交易需要克服的难题之一。

数据资产在成本、价格公开的影响方面也与普通资产不同。由于数据整合涉及对不同系统来源的数据信息进行大量的人工干预、翻译和融合(English,1993),数据产品首次创作成本高,但根据摩尔定律,随着大数据技术的发展,数据资产的整合和储存等成本将进一步降低,数据资产产品的首次创作成本也将下降(Squire et al,2009),而且数据资产的再生产边际成本接近于零(Lerner et al,2006)。此外,数据资产还存在价格外部性,数据价格的公开会泄漏数据的价值(Katz & Shapiro,1985)。

2. 数据资产的价值具有很大的不确定性。首先,数据资产具有事前不确定性(Kerber,2016)、协调性(Babaioff,2012)、自生性(self-generating)和网络外部性(Glazer,1993;Tomak & Keskin,2008)。买方如果交易前不了解该数据资产的详细信息,会较难明确该数据能带来的效用价值;但如果买方了解数据的全部信息,购买该数据对买方的价值将降低,这也就是前文所说的“信息悖论”。协调性是指不同的数据集组合可以带来不同的价值,这导致数据资产具有范围经济的特征。自生性指当同一组织或个人拥有的数据资产组合越多时,这些数据资产彼此之间越可能相互结合而产生新

<sup>①</sup>中国互联网金融协会互联网银行专业委员会:《数据资产、账户与要素交易流转机制研究——基于银行保险机构的实践》,https://mp.weixin.qq.com/s/uztBqvIeXf-pHGSNV8Rvpg。

的数据资产,从而带来更多的价值。网络外部性指的是数据产品的使用者越多,其价值越高,比如Google、微信等平台企业,使用个体越多,吸引的使用者越多,平台的数据资产价值越大。

其次,数据资产的价值与本身的体量、质量、时效性、整合程度之间存在一定的不确定性,与具体的应用场景相关。虽然大多数情况下数据资产具有规模报酬递增性,即随着数据产品中包含的有效数据内容的增多,该数据资产带来的价值越大。但是,部分运用数据进行企业产品需求预测(如亚马逊)的实证研究发现,数据量对预测和决策改善的价值达到顶峰之后可能下降(O'Reilly, 1986; Bajari et al, 2018)。一般情况下,数据准确度越高,价值越大(Wang & Strong, 1996),但如果数据的准确程度固定,而使用者知晓该准确度,此信息的纳入同样可以帮助使用者进行决策矫正(Haebich, 1997),从而产生更高的价值。在某些对时效性要求较强的应用场景中,只有最新的数据才有价值,比如消费者的住址、定位(Moody & Walsh, 1999)。但对于学者研究、行为预测等,历史数据和当前数据的重要性差别并不大,甚至早期的数据价值更大。另外,通常数据整合度越高,其价值越大,但Goodhue et al(1992)指出数据价值与整合度呈抛物线关系,20%的整合度可以达到80%的效用价值。Azevedo et al(2019)的研究则表明,互联网搜索中的A/B随机试验结果的分布可能是厚尾的:罕见的结果可能有非常高的回报,因而通过许多低质量和低统计能力的小型实验来测试大量创意反而更有利于发现大的创新。

最后,数据资产的价值与使用者的异质性密切相关(Bergemann et al, 2018)。这主要是因为数据资产只有被使用才会产生价值(没有被使用的数据资产事实上是企业的负债),数据资产的价值在于改变行动、改善数据资产持有者的决策和行为(Ackoff, 1967; 熊励等, 2018)。因此,使用者的目的、知识、能力、私有信息、已有的数据资产不同,会导致同样的数据资产对不同买方的价值差异很大。所以数据资产的价值评估很难作为一个标准品,由众多类似于股票交易市场上的买方共同定价。

## 二、数据资产交易

### (一)数据资产的流动方式

数据资本化有赖于数据要素跨企业、跨行业的流通和社会化配置。从数据资产到数据资本是数据要素化过程中一次“质的飞跃”,这类似于马克思提出的商品到货币“惊险一跃”(Mayer-Schönberger & Ramge, 2018)。数据的流动方式主要包括企业主动共享、自留使用和数据交易,企业对这三种方式的选择依赖于卖方是否与买方存在竞争、买卖双方的风险偏好水平等因素。

1. 数据共享、自留和交易的影响因素。如果数据产品持有者既销售数据又使用数据(与数据产品买方存在竞争),此时主动共享数据资产违背理性原则(Currarini & Feri, 1985; Jones & Tonetti, 2019)。Easley et al(2018)假设了一个古诺竞争、需求不确定、没有数据资产供应商的基准经济(benchmark economy)模型,此时数据共享能促进企业更好地适应消费者的需求,从而改善消费者剩余和社会总福利。但是,企业竞争使得分享数据后企业利润减少,因此数据资产的“主动共享”对于企业来说是一个囚徒困境,局中人的上策都是不共享自身拥有的数据资产。

数据主体主动共享数据的场景较少。在古诺竞争和需求确定的伯川德竞争中,只有企业面临的市场需求不相关、市场需求信号完美或者存在战略互补(strategic complements)时,企业才会共享信息。在需求不确定的伯川德竞争中,只有当数据共享后,所有企业知晓其他企业和自身利润函数之后的策略相关性有利于自身利润,企业才会共享数据(Raith, 1996; Gal-Or, 1985)。此外,对具有公共基础设施、公共价值性质的数据和主体,或者对伦理、安全等有高要求以至于法律限制交易的数据,各区域、各行业之间进行共享才会成为首选。为了在合规的条件下解决数据孤岛、整合数据以产生更大的价值,联邦学习(federated learning)等分布式机器学习方案受到推崇,极大推动了数据价值

的共享。

通过法令强制要求企业主动共享数据的方案则会产生额外的社会成本。除去行政成本,企业还将通过各种各样的策略行为,如给数据加“噪音”、降低数据质量等,减少社会福利。美国证券交易委员会(SEC)2000年曾颁布《金融衍生工具法》(FD),规定上市公司必须向公众披露重大信息,但是,Bushee et al(2004)和Cohen et al(2014)发现该法案的颁布促使公司们通过各种方式进行合谋,从而控制向公众发布的数据。Bailey et al(2003)则指出自FD法案颁发后,企业分享的数据质量显著降低。

当卖方同样是数据资产的消费者时,数据卖方没有动机主动共享数据,但选择数据交易或自留使用则受到买卖双方风险偏好程度的影响。假设一个将数据资产金融化的场景,卖方将数据产品加工为数据基金进行销售。如果卖方是风险厌恶的,则卖方倾向于只销售数据基金,不在二级市场中(例如利用数据资产进一步交易或者投入创新开发)和数据买方进一步竞争(Bhattacharya & Pfleiderer,1985)。当数据无法被二次销售(如数据时效性较强)且数据持有者是风险中性时,其不会销售数据产品,而是倾向自留使用以减少竞争;当数据持有者是风险厌恶的且其他数据交易买方是风险中性时,则持有者倾向于销售数据产品以平衡风险分摊和竞争加强的影响,而不是自留使用;否则,将选择销售数据和自留数据参与后续竞争的混合策略(Grinblatt & Ross,1985;Admati & Pfleiderer,1988)。

2. 第三方“数据销售商”的作用和问题。在平台或消费者拥有数据资产的寡头竞争市场中,独立收集和处理的第三方“数据销售商”可以有效促进消费者、企业之间的数据流通;且相比于单个数据提供者,第三方“数据销售商”会从多种渠道、多个企业处收集和整合数据产品,从而提供更多元、综合的数据产品(Chiang et al,2003;Yan,2008)。因此当数据卖方不与买方形成竞争时,卖方通常需要权衡独自销售可减少的交易成本和通过第三方销售可增加的销量,进而做出最优的销售渠道选择。如果数据产品的需求存在网络外部性,那么原始数据提供者的最优策略为同时采用自我销售和第三方销售渠道,并提供差异化数据产品(versioning):在自我销售渠道上销售高质量数据产品,在第三方销售渠道上销售低质量数据产品(Liu et al,2015;Khouja & Wang,2010)。

另外,第三方数据机构还可以帮助数据买方中和风险、增加激励。在一个两企业参与、赢者通吃、成功不确定的探索创新模型中(比如专利开发),由于“死胡同”低效性(dead-end inefficiency)和过早出局低效性(early-switching inefficiency)普遍存在,由一个独立的第三方数据销售商来负责补偿创新失败企业的损失,才能弥补企业激励不足带来的社会福利损失(Akcigit & Liu,2016)。

但目前我国的第三方数据交易平台普遍存在活跃度低、交易数量不足的问题<sup>①</sup>,原因是多方面的:(1)基于生命周期视角,现阶段我国大数据交易产业处于成长期。(2)从制度因素来看,现阶段存在数据权属不清(申卫星,2020),交易标准和规范不明晰等问题,同时《网络安全法》等法律法规的实施导致中小型数据供需主体的市场合规成本增高等。(3)在技术方面,现有数据交易技术机制尚不成熟,难以兼顾隐私保护与流动。(4)就市场因素而言,一方面互联网巨头垄断数据,“数据自留”行为十分普遍,另一方面现有第三方数据交易平台仅具备信息撮合功能,尚未培养成熟的数据交易市场。数据的价值往往来自对数据的加工和挖掘所传达的有效信息,大量的中小供需主体缺乏这项能力。因而,第三方数据交易平台可重点撮合专业性的大宗数据资源聚合平台参与交易。

## (二)数据资产的交易设计

与数据产品相比,数字产品的交易和定价市场的发展已较为成熟。数据产品的交易可借鉴数字

<sup>①</sup>中国大数据产业生态联盟:《2017中国大数据产业生态地图暨中国大数据产业发展白皮书》,https://www.bdinchina.com/Article/info/id/609.html。

产品,但由于数据的可复制性、低边际成本、可整合性和价值不确定性等特点更为明显,数据交易模式和卖方策略受应用场景、买方异质性和市场结构的影响更大。

1. 数字产品交易。数字产品的交易手段主要可分为:捆绑销售(bundling)、订阅和租赁式(subscription and renting)和拍卖(auction)。由于数字产品的低复制成本和买方异质性,对数字产品或服务进行捆绑销售十分普遍(Shapiro et al, 1998; Aguiar & Waldfogel, 2018)。Daskalakis et al (2017)研究发现,单独对每单位的数字产品进行定价和大捆绑(grand bundle)都能达到收益最大化。Haghpanah & Hartline(2020)研究表明,如果价格敏感型买家认为产品更具互补性,那么大捆绑是最佳选择。订阅指消费者支付固定价格之后可以免费使用一段时间平台内的数字产品服务,租赁指对单独每一次使用付费(Alaei et al, 2019)。根据客户的产品使用率和价值的差异,平台的最大化收入决策有所不同。大多数平台同时提供订阅和租赁,并谨慎制定二者的价格,以使得使用率低的客户尽可能订阅数字产品,而使用率高的客户尽可能租赁数字产品(Pei, 2020)。

数字产品主要适用的拍卖方式包括赞助搜索拍卖(Edelman et al, 2007; Varian, 2009; Che et al, 2017)、无限供给拍卖(Goldberg et al, 2001; Goldberg & Hartline, 2001)和在线拍卖(Lavi & Nisan, 2000)等。其中,“赞助搜索拍卖”指通过对搜索引擎中的关键词进行竞价,从而为广告商等竞得展示内容的位置(Lahaie et al, 2007)。“随机抽样拍卖”(random sample auction)是无限供应的数字产品实现真实反映竞拍者估值的典型方法,类似于“第二价格拍卖”(Vickrey, 1961)。这是因为数字产品的边际成本很低,“第二价格”很可能接近零(Pei, 2020)。随机抽样拍卖将买方随机分为两部分,分别进行最优拍卖定价,各部分的买方最终能否成功竞得数据取决于其投标价格是否高于另一部分买方拍卖所定的最优价格(Goldberg & Hartline, 2001)。此时,买方的出价与是否获得标的无关,只有按照真实支付意愿出价才可能获得标的。数字产品在线上进行交易时,不同客户往往在不同时间出价,“在线拍卖”系统将在每个出价到达时做出竞拍成功与否的决定。Lavi & Nisan(2000)和 Bar-Yossef et al(2002)分别对有限供给和无限供给条件下的数字产品的在线拍卖提出了激励相容的在线拍卖机制。

2. 数据产品交易。数据交易按照应用场景可分为营销、风险规避和人员搜索三大类<sup>①</sup>。其中营销是指运用数据列表(data lists)和数据追加集(data appends)<sup>②</sup>等对消费人群进行分割、匹配、定位、营销分析和消费预测。风险规避主要应用于个人和机构等进行信用等级构建和诈骗检测。人力资源数据则多用于人员搜索应用场景。

尽管应用场景、具体的交易机制可能不同,数据交易卖方策略的核心思想主要是差异化产品和价格,或只选择部分买方进行交易(Fudenberg & Villas-Boas, 2012; Shapiro & Varian, 1998),其核心在于区分买方异质性,真实反映买方的效用。如果交易方的风险容忍度不同而卖方可以完全歧视,卖方倾向于对风险容忍度更高的买方收取更高的价格;如果卖方不能实现完全歧视,则会选择二级和三级价格歧视,促使买方自我选择(Admati & Pfleiderer, 1986)。

根据卖方对数据资产加工整合的精细程度,数据交易可分为直接交易和间接交易(Admati & Pfleiderer, 1986)。直接交易指卖方直接提供未经加工的原始数据,如消费者的年龄、收入等数据,这是大多数潜在客户开发公司和一些金融数据销售公司(如 Bloomberg、Wind)采用的方式。间接交易指的是卖方通过对数据的整合再加工形成一定程度的标准品或数据资产组合。直接交易和间接交

<sup>①</sup>Federal Trade Commission(2014), “Data brokers: A call for transparency and accountability”, <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>.

<sup>②</sup>买方在数据交易前拥有消费者的某些数据,所购买的数据产品可作为已有数据的补充。

易都可以实现价格歧视。

(1)直接交易。当数据产品的价值可预期、部分可知时(Akcigit & Liu, 2016; Riley & Samuelson, 1981),可以采取直接交易方式。此外,当数据的负外部性相对较小而买方的异质性过大,例如私有信息和需求不同导致买方对数据的组合要求迥异,导致卖方无法设计足够多的数据产品来满足所有买方的要求(Admati & Pfleiderer, 1990),也可考虑直接交易。数字产品的捆绑销售、订阅租赁以及各种拍卖方式均可一定程度上运用于数据产品的直接交易(Pei, 2020; Akcigit & Liu, 2016)。例如,卖方可将不同质量的数据产品捆绑销售(Baake & Boom, 2001),通过低价销售或免费赠送低质量的数据产品,将支付意愿和质量需求高的买方引流到高质量付费产品上,从而获取更高的利润。卖方也可通过租赁或销售“部分数据”(partial information)的方式来识别高价值和低价值买方,因为高价值买方对部分数据的价值评估很低,倾向于高价订阅完全数据(Jones & Tonetti, 2019)。

直接交易还可通过交互式协议(interactive protocol)进行动态机制设计(Babaioff et al, 2012)。假设市场仅存在一个垄断的数据产品卖方、一个购买数据追加集来优化营销和避险决策的买方,双方的私有信息分别是数据(关于世界状态的信息)和买方的估值。卖方主要有三种协议方式可选择:其一,通过密封信封机制(sealed envelope mechanism)向买方透露产品价格,买方决定是否购买,此定价方式下的数据产品被视为一般商品。其二,买方支付一定的价格可以获得卖方提供的数据随机样本,再决定是否购买(pricing mapping)。其三,卖方先向买方提供一部分随机数据样本,买方基于这部分样本判断整体数据的价值,再决定是否购买(pricing outcomes)。这种方式使得卖方可以根据买方的事后选择来判断买方类型并进行价格歧视。

(2)间接交易。间接交易主要适用于数据的网络外部性为负、外部性较强、买方相对风险宽容时。此外,当卖方处于寡头或完全竞争市场时,由于卖方之间对彼此拥有的数据产品的信息不完全,没有卖方可以通过直接交易攫取更大的剩余,此时买方可采取众包、众筹等模式,卖方可在竞标成功后进行间接交易设计(Admati & Pfleiderer, 1990)。根据买方的异质性需求,卖方可以通过设计不同成本—收益的数据产品来筛选买方类型。以买方购买数据追加集进行风险规避和市场营销活动(比如是否对某个消费者借贷或推送广告)为例, Bergemann et al(2018)指出,在卖方垄断、买方异质的市场中,由于买方交易前拥有不同的数据资产,且对销售数据产品质量的预判不一,买方对卖方手中的数据追加集的价值评估不同。卖方可以将这些数据追加集设计成不同的统计实验(假设检验)<sup>①</sup>来筛选买方,买方是贝叶斯决策者,由此卖方可以通过买方选择的“一类错误、二类错误组合水平”来识别买方的私有信息,本质上是二级价格歧视。

结合数据产品价格和使用的负外部性,即数据资产价格对数据价值的泄漏、数据资产的使用价值随使用者增多而降低的特点,数据供应商可通过设立基金、制定两部定价法(每股价格和固定费用相结合的方式)来实现任意数据资产的组合销售。数据价格的外部性越高,每股价格更高,两部定价法获利越大(Bhattacharya & Pfleiderer, 1985)。如果买方是异质的,且可能和卖方构成竞争,那么具有一定市场力量的卖方将同时参与数据基金的销售和二级市场的交易(Grinblatt & Ross, 1985; Admati & Pfleiderer, 1988)。此时,卖方的利润不仅源于数据基金的销售额,也包括通过数据基金参与二级市场的投资收益。

在双边市场中,可由统一的第三方机构或市场机制来提供间接交易的数据产品。此时,多个买方需要不同准确度程度的训练数据集,多个数据卖方共同出售各自的数据,价格真实反映买方估值、卖方收益最大化、收入公平分配给合作提供数据集的不同卖家(Shapley, 1953; Agarwal et al, 2019)、

<sup>①</sup>比如,原假设  $H_0$  为“该消费者风险低,宜借贷/推送”,备择假设  $H_1$  为“该消费者风险高,不宜借贷/推送广告”,给定某个固定的阈值,卖方提供不同的数据产品会带来不同的“一类错误、二类错误组合水平”。

无套利(Li et al,2015)等目标非常重要。Agarwal et al(2019)提出由统一的在线市场设计不同的数据产品,并基于特定数据组合产品的零后悔拍卖机制来激励买方透露真实的价值评估。继而根据以上结果对买方销售不同的数据产品和收取价格,同时按照卖方的边际贡献来公平分配数据产品的总销售收入。在充分竞争的双边市场中,卖方没有足够大的市场力量来控制市场,只能提供真实的数据,而非增加噪音或者复制数据滥竽充数。

在直接和间接交易中,卖方亦可能只选择买方中的一部分人进行交易。Bergemann et al(2018)指出,当满足一定条件时(比如卖方提供给高、低类型买方的差异化产品的边际收益相等),卖方将仅向部分买方销售数据产品。考虑到数据的可复制性、易于转售等问题,数据交易最可能采取的方式是“价高者得”,也就是最终只有使用价值最大、支付意愿最高的企业能够使用该数据资产(Akcigit & Liu,2016)。但“价高者得”是一个次优决策,并不能实现利润最大化。另外,随机抽样拍卖方式也是通过随机选择一部分买方的方式来促使买方反映自我的真实价值。

(3)数字产品和数据产品的差异。虽然数字产品和数据产品的交易模式存在相似之处,但二者也存在显著区别(Pei,2020)。首先,数字产品的最小一单位,如一部电影、一篇电子文献本身是有价值的,而数据产品的最小一单位,例如一条匿名化的消费者购买数据记录,其本身的价值却微乎其微。因而数字产品可仅出售最小单位的量,而数据虽可以按条数计价,但大都以数据集出售。其次,数据产品的可整合性极高,与不同的数据集整合可形成不同的数据产品,因而防止数据套利是数据交易需要考虑的因素。数字产品虽然也因低复制成本而将不同数量、类型捆绑销售(Bakos & Brynjolfsson,1999,2000),但捆绑与数据的整合有本质差异。最后,数字产品可以折旧后再消费、转售,且基本不影响使用价值,而数据产品的转售则至少存在两种情形:一方面,数据内容本身可以被复制、整合成完全不一样的数据产品进行再消费和转售;另一方面,间接交易、“一对一”制定的差异化数据产品的转售却可能受到极大限制。

(4)数据交易技术和区块链的应用。针对数据交易存在的信息悖论、二次转售、道德风险等问题,传统交易技术和方法在一定程度上可以帮助解决部分问题。例如,运用预览、专家评估、声誉建设等方法,可以解决数据资产的信息悖论问题。运用密封信封(cryptographic envelopes)技术和专利权,通过审计、统计跟踪系统,可以帮助解决数据资产的可复制性问题。另外还可以接受数据产品的可复制性,将数据商品与销售者希望广泛传播的其他数据进行捆绑销售,如投放广告来赚取额外收益(Varian,1999)。但是,这些传统手段并不能从根本上解决问题,随着区块链和智能合约等技术的发展,上述问题得到了更有效的解决。

区块链(blockchain)技术方案可以通过赋权、“零知识证明”(zero knowledge proof)(Feige et al,1988)、完整记录交易过程、提供可信的执行环境等来缓解上述交易问题。区块链系统中的任意多个节点可以通过密码学算法,将一段时间内发生在系统中全部信息交流的数据计算和记录到一个数据块(block)上,并且生成该数据块的指纹用于链接(chain)下个数据块和校验。此时,系统所有参与节点将共同认定该记录是否为真,同时,每个参与区块链的节点的合法性可以得到认可,保证了交易节点的真实性和合法性。通过区块链上私钥和公钥的双认证技术,可以有效验证数据交易方的身份,确认交易环节是否如约进行(Wang et al,2018)。另外,区块链技术生成数字时间戳(timestamp)可以对数据资产进行界权,同时有效地记录交易的时序。由于每个数据资产可以拥有独特的哈希值(hash value)和时间戳,数据资产一旦在区块链上被界权,后续的每一步操作都会被实时记录,很难被更改和删除,这也为解决数据资产二次转售无迹可寻的难题提供了思路。

区块链和智能合约<sup>①</sup>的结合对解决数据交易的“信息悖论”问题和行为规范方面起到了不可忽

<sup>①</sup>智能合约是用代码编写的协议,一旦预定义的条件得到满足,它就会自动执行(Szabo,1994)。



视的作用。当智能合同通过区块链实施时,一旦交易前定下的条件得到满足,协议将自动执行,减少数据交易的欺诈可能性,规范交易秩序(Chen et al,2019)。Hörner & Skrzypacz(2016)提出,假设卖方销售有关某状态好或者差的信号数据,买方对此数据产品的价值评估是公开信息,卖方可以设计自我执行合同来解决买方的失信问题。Su et al(2020)提出了一个基于区块链网络和可信执行交易环境(trusted execution environment)来保证数据商品和支付的可信交易框架,避免了传统第三方交易平台截留数据进行转卖、买卖双方彼此欺瞒等道德风险问题。Sabbagh(2019)则针对数据音乐作品的版权保护问题提出了利用 The Music Modernization Act(MMA)<sup>①</sup>建立的音乐作品数据库,结合区块链和智能合同来实施强制许可方案。Banerjee & Ruj(2019)提供了一个较详细的交易流程的理论方案,同时纳入法规来保证该设计具有可监管性。

当然,区块链技术和智能合约本身亦存在数据隐私、可扩展性和互操作性等各类挑战(Underwood,2016)。区块链和智能合约目前尚无法完全解决数据资产的转售等问题。比如,数据持有人仍可以对原有数据稍做改变生成新的数据,虽然数据有所失真,但此时哈希值和时间戳并不能完全消除未经授权的转售等道德风险问题。

### 三、数据资产定价

#### (一)数字产品的定价方法

数字产品的收入主要有三个来源,包括直接的货币收入、个人数据/隐私收入(Bergemann & Bonatti,2015)和时间/注意力收入(Lambrecht et al,2014),这三种收入流往往存在此消彼长的关系(Pauwels & Weiss,2008)。向客户销售数字产品和服务,如电子书、网易云音乐等,可以获取货币收入,也可以获取采集、使用消费者的个人数据的权益,还可以向广告商等出售其在数字产品中的空间和位置(Athey et al,2018)来获取消费者的时间/注意力收入。数字产品销售商需要在定价方案中考虑它们的最佳组合方式(Gallaughier et al,2001)。例如,视频或新闻平台通过投放广告可获得时间/注意力收入,若消费者为避免广告选择充值,则平台现金收入增多(Prasad et al,2003),但价格敏感型的消费者也可能直接离开该平台(Chiou & Tucker,2013)。过度采集或者不当使用个人数据可能会带来客户的流失,而适度放弃现金流,例如采取低端产品或样品免费模式则可能会培育起一个长期客户池(Boom,2004)。

数字产品常见的定价方式主要包括按离散单位计价、按使用量和时长计价和混合定价(Pei,2020)等类。根据市场结构(市场竞争的强弱、是否为垄断市场等)、产品粘性、消费者价格敏感性和需求弹性等方面的差异,使用量的边际定价趋势有所区别。例如,百度网盘的普通会员收费为10元/月,比非会员额外增加3T的容量空间,超级会员收费为25元/月,比非会员额外增加5T,边际支付是递增的;而大多数电子游戏的定价则随着使用时长而递减。订阅和租赁属于典型的多因素混合计价,例如亚马逊Prime影音服务,单个数字产品的费用随基础订阅费的等级有所差异,但均会低于实体替代产品的价格(Brynjolfsson & Smith,2000)。

#### (二)数据产品的定价原则和方法

数字产品的定价方式大多数可以直接借鉴到数据产品上,这主要是因为二者都拥有低复制成本、较大的价格歧视空间,而且在定价对象方面存在重合之处。例如,数字产品销售商通过提供数字产品或服务的方式获取消费者个人数据的授权,换一个角度便是商家对个人数据内容支付的对价(Pei,2020)。因此,数据产品通常也可通过提供部分免费数据来获取数据买方流量(Lerner et al,2006);或者根据使用量计费,例如根据使用条数计价;又或者收取订阅和租赁费的混合价格(Wu &

<sup>①</sup>这个数据库的目的是存储音乐作品信息,例如版权所有者的身份和位置,并根据实时搜索情况更新。

Banker, 2010)。由于数据价值的不确定性,对数据资产的价格进行统一规定几乎是不现实的,应该在界权基础上由市场去发现数据资产的价格。同时,数据要素的定价离不开具体交易场景,需要根据典型应用场景有针对性地核算数据要素价值。

数据资产定价的主要原则包括:价格可以真实地反映买家的效用、卖方收入最大化、收入公平分配给合作提供数据集的不同卖家(Shapley, 1953; Agarwal et al, 2019)、无套利(Li et al, 2015)、隐私保护和计算效率(Goldberg et al, 2001; Pei, 2020)。具体的定价方法基本体现了以上原则的取舍和融合。例如, Koutris et al(2013)对查询式数据(query based data)制定出一套线性规划方案,可以同时满足卖方收入最大化、无套利、公平分配原则。随机抽样拍卖定价方式能有效促进价格真实地反映买家效用,但不一定能使得卖方整体收入最大化。而一些能最大限度满足无套利原则的定价方法可能需要较大的计算量(Koutris et al, 2015)。以下将对主要的定价方法及其适用场景进行介绍。

1. 传统会计学定价。收益法、成本法和市场法等可以适用于不同类型的数据。收益法关注商品的效用价值或现值,其收益可以依靠合同定期支付。对以原始数据直接交易的数据资产的定价,可以通过收益现值法,根据买方的实际收益所得、使用次数或时间等,按比例支付给卖方,但选择合适的折现率比较困难(Henderson et al, 2015)。收益法的典型应用场景包括基于项目数量和用户数量制定比例租赁费用的订阅方式,根据买方的质询、模型训练精度定价(Koutris et al, 2013; Tang et al, 2013; Agarwal et al, 2019)等方式。Bergemann & Bonatti(2015)曾提出,对于用以广告精准推送的消费者行为数据,可通过广告的收益和成本的线性定价来获取最大收入,且该价格与市场集中度负相关,与该数据集的协同性(不同信息产品之间的合作产生的增量价值)正相关。

成本法易于操作且定价相对直观,但如果仅依靠成本法则忽略了买方异质性和数据特点所产生的价值,很可能会低估数据的价值。因此,成本法比较适用于买方差异不大、制作成本几乎是公开信息、供给竞争激烈的数据产品,同时也适用于对个人数据的隐私补偿定价(Ghosh & Roth, 2015)。Fleischer & Lyu(2012)针对个人的隐私数据定价提出了“公布价格”机制,卖方以接受或拒绝合同的方式来获取数据的隐私成本收入。对于搜集和加工边际成本递减的数据类型,成本法给出的数据资产的价格应该比平均成本更低。

市场价格法则强调数据资产的交易价格,主要考虑重置成本(用新资产替换已有资产的成本)、当前成本(用类似用途的资产替换资产的成本)或可变现净值(资产可以出售的金额减去出售成本)。大多类似知识产权的数据都可适用市场法,但其运用限制也较为明显,比成本法更费时和昂贵(Henderson et al, 2015),且还要求市场上已有类似的数据交易作为参照。

但是,传统会计评估法可能会低估数据集的价格,这是因为拥有数据资产的企业一般会进行相机决策。如果现有数据集质量不佳或市场需求疲软等,企业可能会放弃或延迟开发数据集。这也意味着当企业计划将数据要素纳入生产环节中时,数据资产具有了隐含期权的特征。因而,可考虑将实物期权理论(Myers, 1977)融入数据资产定价,但此方面的研究仍有待进一步发展。

2. 基于“信息熵”定价。根据 Shannon(1948)的信息论,“信息熵”表示信息中排除冗余后的平均信息量,是与买家关注的某事件发生的概率相关的相对数量。信息熵越大,某事件发生的不确定性越小,正确估计它的概率越高。因而,熵越大,信息内容的有效性越大,交易价格越高。“信息熵”定价在传统金融、期权领域运用广泛,主要通过不同时间的历史数据来预测未来时期的期权价值(Rubinstein, 1994; Buchen & Kelly, 1996)。因此,通过对数据元组(组成数据集的小单位)的隐私含量、被引用次数、供给价格、权重等因素的结合,可以对数据资产的信息熵进行动态定价。Shen et al(2019)提出了对个人数据的信息熵定价方法。信息熵定价法充分考虑了数据资产的稀缺性,且相对于数据的内容和质量,更关注数据的有效数量和分布(Li et al, 2017)。

3. 数据资产价值的多维度定价。根据数据资产价值的多个维度进行定价可以兼顾卖方、买方和

数据资产本身的核心关注点。数据资产价值的评估要素主要应考虑数据成本、数据质量、数据产品的层次和协同性、买方的异质性等(Wang & Strong, 1996)。数据资产的采集、存证、传输、加工、营销均会产生成本。数据质量的指标主要包括完整性、独特性、时效性、有效性、准确性和一致性<sup>①</sup>。数据产品的层次主要指其技术含量、稀缺性等,协同性则指的是不同信息产品之间的合作产生的增量价值。一般以上指标与数据价值成正比,而买方异质性则使得数据价值的方差很大。不同的买方拥有不同的风险厌恶程度、数据偏好、信息使用成本和变现能力,即便是相同的数据,价值差异也很大。因此,如果买方异质性较强,企业一般会先筛选买方类型,再进行价值评估和差异化定价。

Sajko et al(2006)提出可对数据的现值、成本、数据本身的特征和质量等多个维度的重要性和价值展开定量评级,再结合群组决策和价值矩阵等定性分析方法,得到数据在每个维度的价值,最后得到综合价值。同时,可以结合 Zeithaml(1988)提出的价格价值、竞争价值、情感价值、功能价值和社会价值五个维度<sup>②</sup>,设计客户感知价值定价模型。另外,可以借助人工智能提高大规模数据产品定价的计算效率。人工智能定价在医学图像数据(Litjens et al, 2017)、自然语言处理(Young et al, 2018)等对决策模型的精度和准确度有特殊要求的领域的应用将越来越普遍。

综上,在对数据资产内容的定价时,为了保护数据内容,防止内容泄漏造成数据价值降低,卖方可利用上述各种资产估值方法对随机抽样部分的内容进行价值评估,进而对整体数据资产内容进行定价。对大数据型的数据资产,则可以不按内容计价,而是在数据标准化的基础上,以“盒”为单位,通过一般性价值评估对标准化的单位“盒”数据估值,进而按件计价交易。

#### 四、结论

数据要素是数字经济的微观基础和创新引擎,促进数据要素的流通和交易是未来产业发展的大趋势。数据界权、交易和定价机制的明晰有利于保护数据主体权益并维护数据安全,同时引导数据要素协同向先进生产力集聚,从而促进完善社会主义市场经济体制。本文综合经济学、计算机科学、法学等跨专业学科文献,对数据资产界权、交易和定价进行综述。

本文首先通过区分数据内容和数据产品,提出数据自然人享有个人数据内容的分级分类的数据控制权,数据文件持有者则享有数据产品有限制的占有权(经济所有权或用益数权),这有利于解决个人数据权益和数据流通的争议。接着,本文总结了数据资产的可复制性、价值不确定性等特征,数据资产的交易和定价与传统资产、金融资产不同。继而通过对数据交易、自我保留、数据共享三种数据流通方式的比较总结,本文指出了数据交易的合理性、第三方数据售卖商对数据流通的重要性和现阶段数据交易不足的原因。随后,本文比较了数字产品和数据产品的交易和定价,着重说明了数据产品直接交易和间接交易的具体方式、卖方策略,并指出其受应用场景、买方异质性和市场结构的影响。差异化产品和定价是数据产品交易的核心思想,可以通过线性定价、两部定价法、拍卖、机制设计等多种方式进行。同时,本文还分析了区块链和智能合约等技术方案在解决数据交易二次转售、“信息悖论”和行为规范方面的作用和不足。最后,本文比较分析了数据资产的定价原则,以及包括传统会计定价、多维度综合定价、信息熵定价等在内的多种方法,并将其运用于数据资产内容和大数据型资产的价值评估中。

<sup>①</sup>英国 DAMA 工作组(2020):《数据质量评估的六个主要维度》,CSDN, <https://blog.csdn.net/bigdatapang/article/details/104157430>。

<sup>②</sup>价格价值与制作数据产品的成本相关;竞争价值包括该数据产品的稀缺性等;情感价值主要指买卖双方在交易过程中的感受、情绪和情感效用(Rabin, 1993);功能价值主要指买方在使用数据产品中所获取的效用价值;社会价值指的是需求方通过数据应用提高社会自我概念而带来的效用。

通过上文的综述和分析,本文针对数据交易产业发展和政策制定提供以下建议。第一,完善数据要素界权,在场景性公正原则下针对数据内容和数据产品细分数据权利,加强数据安全和隐私保护,充分挖掘和释放数据要素的价值。第二,以应用需求为导向建设第三方大数据交易平台,提升其协调、服务数据买卖方的技术和能力,鼓励专业性的数据聚合企业通过直接交易和间接交易方式,“一对多”与其他数据需求企业开展匹配交易。第三,重视区块链技术在数据交易中的应用,制定统一的技术标准及规则,规范数据交易技术和市场。第四,构建数字生态,坚持市场主导原则,完善市场监督。

#### 参考文献:

- 何伟,2020:《激发数据要素价值的机制、问题和对策》,《信息通信技术与政策》第6期。
- 凌斌,2010:《界权成本问题:科斯定理及其推论的澄清与反思》,《中外法学》第22期。
- 纪海龙,2018:《数据的私法定位与保护》,《法学研究》第6期。
- 龙登高,2018:《中国传统地权制度及其变迁》,中国社会科学出版社。
- 申卫星,2020:《论数据用益权》,《中国社会科学》第11期。
- Abowd, J. M. & I. M. Schmutte(2019), “An economic analysis of privacy protection and statistical accuracy as social choices”, *American Economic Review* 109(1):171—202.
- Ackoff, R. L. (1967), “Management misinformation systems”, *Management Science* 14(4):B147—B156.
- Admati, A. R. & P. Pfleiderer(1986), “A monopolistic market for information”, *Journal of Economic Theory* 39(2):400—438.
- Admati, A. R. & P. Pfleiderer(1988), “Selling and trading on information in financial markets”, *American Economic Review* 78(2):96—103.
- Admati, A. R. & P. Pfleiderer(1990), “Direct and indirect sale of information”, *Econometrica* 58(4):901—928.
- Agarwal, A. et al(2019), “A marketplace for data: An algorithmic solution”, *Proceedings of the 2019 ACM Conference on Economics and Computation*, 701—726.
- Aguiar, L. & J. Waldfogel(2018), “As streaming reaches flood stage, does it stimulate or depress music sales?”, *International Journal of Industrial Organization* 57:278—307.
- Akcigit, U. & Q. Liu(2016), “The role of information in innovation and competition”, *Journal of the European Economic Association* 14(4):828—870.
- Alaei, S. et al(2019), “Optimal subscription planning for digital goods”, Social Science Research Network Working Paper, No. 3476296.
- Athey, S. et al(2018), “The impact of consumer multi-homing on advertising markets and media competition”, *Management Science* 64(4):1574—1590.
- Azevedo, E. M. et al(2020), “A/B testing with fat tails”, *Journal of Political Economy* 128(12):4614—5000.
- Baake, P. & A. Boom(2001), “Vertical product differentiation, network externalities, and compatibility”, *International Journal of Industrial Organization* 19(1—2):267—284.
- Babaioff, M. et al(2012), “Optimal mechanisms for selling information”, *Proceedings of the 13th ACM Conference on Electronic Commerce*, 92—109.
- Bailey, W. et al(2003), “Regulation fair disclosure and earnings information: Market, analyst, and corporate responses”, *Journal of Finance* 58(6):2487—2514.
- Bajari, P. et al(2018), “The impact of big data on firm performance: An empirical investigation”, *AEA Papers and Proceedings* 109:33—37.
- Bakos, Y. & E. Brynjolfsson(1999), “Bundling information goods: Pricing, profits, and efficiency”, *Management Science* 45(12):1613—1630.
- Bakos, Y. & E. Brynjolfsson(2000), “Bundling and competition on the internet”, *Marketing Science* 19(1):63—82.
- Banerjee, P. & S. Ruj(2019), “Blockchain enabled data marketplace-design and challenges”, *Computing Research*

- Repository Working Paper, No. 1811.11462.
- Bar-Yossef, Z. et al(2002), “Incentive-compatible online auctions for digital goods”, *SODA'02: Proceedings of the Thirteenth Annual ACM/IEEE Symposium on Discrete Algorithms*, 964–970.
- Bergemann, D. & A. Bonatti(2015), “Selling cookies”, *American Economic Journal: Microeconomics* 7(3):259–294.
- Bergemann, D. & A. Bonatti(2019), “Markets for information: An introduction”, *Annual Review of Economics* 11: 85–107.
- Bergemann, D. et al(2018), “The design and price of information”, *American Economic Review* 108(1):1–48.
- Bhattacharya, S. & P. Pfleiderer(1985), “Delegated portfolio management”, *Journal of Economic Theory* 36(1):1–25.
- Boom, A. (2004), “‘Download for free’: When do providers of digital goods offer free samples?”, Discussion Papers 2004/28, School of Business & Economics, Free University Berlin.
- Brynjolfsson, E. & M. D. Smith(2000), “Frictionless commerce? A comparison of internet and conventional retailers”, *Management Science* 46(4):563–585.
- Buchen, P. W. & M. Kelly(1996), “The maximum entropy distribution of an asset inferred from option prices”, *Journal of Financial and Quantitative Analysis* 31(1):143–159.
- Bushee, B. J. et al(2004), “Managerial and investor responses to disclosure regulation: The case of Reg. FD and conference calls”, *Accounting Review* 79(3):617–643.
- Che, Y. K. et al(2017), “An experimental study of sponsored-search auctions”, *Games and Economic Behavior* 102: 20–43.
- Chen, Y. et al(2019), “FaDe: A blockchain-based fair data exchange scheme for big data sharing”, *Future Internet* 11(11):225–225.
- Chiang, W. Y. K. et al(2003), “Direct marketing, indirect profits: A strategic analysis of dual-channel supply-chain design”, *Management science* 49(1):1–20.
- Chiou L. & C. Tucker(2013), “Paywalls and the demand for news”, *Information Economics and Policy* 25(2):61–69.
- Coase, R. H. (1960), “The problem of social cost”, *Journal of Law and Economics* 3:1–44.
- Cohen, L. et al(2014), “Playing favorites: How firms prevent the revelation of bad news”, NBER Working Paper, No. w19429.
- Currarini, S. & F. Feri(2018), “Information sharing in oligopoly”, In: L. C. Corchón & M. A. Marini(eds), *Handbook of Game Theory and Industrial Organization*, Edward Elgar.
- Daskalakis, C. et al(2017), “Strong duality for a multiple-good monopolist”, *Econometrica* 85(3):735–767.
- Dosis, A. & W. Sand-Zantman(2019), “The ownership of data”, Social Science Research Network Working Paper, No. 3420680.
- Drucker, P. F. (1992), “The new society of organizations”, *Harvard Business Review* 20:281–293.
- Easley, D. et al(2018), “The economics of data”, Social Science Research Network Working Paper, No. 3252870.
- Edelman, B. et al(2007), “Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords”, *American Economic Review* 97(1):242–259.
- English, L. P. (1993), “Accountability to the rescue”, *Database Programming and Design* 6:54–59.
- Feige, U. et al(1988), “Zero-knowledge proofs of identity”, *Journal of Cryptology* 1(2):77–94.
- Fleischer, L. K. & Y. H. Lyu(2012), “Approximately optimal auctions for selling privacy when costs are correlated with data”, *Proceedings of the 13th ACM Conference on Electronic Commerce*, 568–585.
- Fudenberg, D. & J. M. Villas-Boas(2012), “Price discrimination in the digital economy”, In: M. Peitz & J. Waldfoegel(eds), *The Oxford Handbook of the Digital Economy*, Oxford University Press.
- Gallaughan, J. M. et al(2001), “Revenue streams and digital content providers: An empirical investigation”, *Information & Management* 38(7):473–485.
- Gal-Or, E. (1985), “Information sharing in oligopoly”, *Econometrica* 53(2):329–343.
- Ghosh, A. & A. Roth(2015), “Selling privacy at auction”, *Games and Economic Behavior* 91:334–346.

- Glazer, R. (1993), "Measuring the value of information: The information intensive organization", *IBM Systems Journal* 32(1):99-110.
- Godfrey, J. et al(2010), *Accounting*, John Wiley & Sons.
- Goldberg, A. V. & J. D. Hartline(2001), "Competitive auctions for multiple digital goods", *ESA'01: Proceedings of the 9th Annual European Symposium on Algorithms*, 416-427.
- Goldberg, A. V. et al(2001), "Competitive auctions and digital goods", *SODA'01: Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, 735-744.
- Goodhue, D. L. et al(1992), "The impact of data integration on the costs and benefits of information systems", *MIS Quarterly* 16(3):293-311.
- Grinblatt, M. S. & S. A. Ross(1985), "Market power in a securities market with endogenous information", *Quarterly Journal of Economics* 100(4):1143-1167.
- Haebich, W. (1997), "A methodology for data quality management", *Proceedings of the 1997 Conference on Information Quality*, 194-208.
- Haghpanah, N. & J. Hartline(2020), "When is pure bundling optimal?", Technical Report, Working Paper, The Pennsylvania State University.
- Henderson, S. et al(2015), *Issues in Financial Accounting*, Pearson Higher Education Australia.
- Hörner, J. & A. Skrzypacz(2016), "Selling information", *Journal of Political Economy* 124(6):1515-1562.
- Jones, C. I. & C. Tonetti(2019), "Nonrivalry and the economics of data", *American Economic Review* 110(9):2819-2858.
- Katz, M. L. & C. Shapiro(1985), "Network externalities, competition, and compatibility", *American Economic Review* 75(3):424-440.
- Kerber, W. (2016), "A new, intellectual, property right for non-personal data? An economic analysis", *Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil* 11:989-998.
- Khouja, M. & Y. Wang(2010), "The impact of digital channel distribution on the experience goods industry", *European Journal of Operational Research* 207(1):481-491.
- Koutris, P. et al(2013), "Toward practical query pricing with QueryMarket", *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 613-624.
- Koutris, P. et al(2015), "Query-based data pricing", *Journal of the ACM* 62(5):1-44.
- Lahaie, S. et al(2007), "Sponsored search auctions", In: N. Nisan et al(eds), *Algorithmic Game Theory*, Cambridge University Press.
- Lambrecht, A. et al(2014), "How do firms make money selling digital goods online?", *Marketing Letters* 25(3):331-341.
- Lavi, R. & N. Nisan(2004), "Competitive analysis of incentive compatible online auctions", *Theoretical Computer Science* 310(1-3):159-180.
- Lerner, J. et al(2006), "The dynamics of open-source contributors", *American Economic Review* 96(2):114-118.
- Li, C. et al(2012), "A theory of pricing private data", *ACM Transactions on Database Systems* 39(4):1-28.
- Li, X. et al(2017), "A first look at information entropy-based data pricing", *2017 IEEE 37th International Conference on Distributed Computing Systems*, 2053-2060.
- Ligett, K. & A. Roth(2012), "Take it or leave it: Running a survey when privacy comes at a cost", In: P. W. Goldberg & M. Guo(eds), *WINE'12: Proceedings of the Eighth International Workshop on Internet and Network Economics*, Springer.
- Liu, Z. et al(2015), "Selling information products: Sale channel selection and versioning strategy with network externality", *International Journal of Production Economics* 166:1-10.
- Mayer-Schönberger, V. & T. Ramge(2018), *Reinventing Capitalism in the Age of Big Data*, Basic Books.
- Miller, A. R. & C. Tucker(2017), "Privacy protection, personalized medicine and genetic testing", *Management Science* 64(10):4648-4668.
- Moody, D. & P. Walsh(1999), "Measuring the value of information: An asset valuation approach", *ECIS'99: Proceedings of Seventh European Conference on Information Systems*, 1-17.

- Myers, S. C. (1977), "Determinants of corporate borrowing", *Journal of Financial Economics* 5(2): 147–175.
- Nissenbaum, H. (2010), *Privacy in Context: Technology, Policy, and Social Life*, Stanford University Press.
- O'Reilly III, C. A. (1986), "Individuals and information overload in organizations: Is more necessarily better?", *Academy of Management Journal* 23(4):684–696.
- Ordóñez, G. (2013), "The asymmetric effects of financial frictions", *Journal of Political Economy* 121(5):844–895.
- Pauwels, K. & A. Weiss(2008), "Moving from free to fee: How online firms market to change their business model successfully", *Journal of Marketing* 72(3):14–31.
- Pei, J. (2020), "A survey on data pricing: From economics to data science", ArXiv Working Paper, No. 2009.04462.
- Prasad, A. et al(2003), "Advertising versus pay-per-view in electronic media", *International Journal of Research in Marketing* 20(1):13–30.
- Rabin, M. (1993), "Incorporating fairness into game theory and economics", *American Economic Review* 83(5):1281–1302.
- Radin, M. J. (1982), "Property and personhood", *Stanford Law Review* 34:957–1015.
- Raith, M. (1996), "A general model of information sharing in oligopoly", *Journal of Economic Theory* 71(1):260–288.
- Riley, J. G. & W. F. Samuelson(1981), "Optimal auctions", *American Economic Review* 71(3):381–392.
- Rubinstein, M. (1994), "Implied binomial trees", *Journal of Finance* 49(3):771–818.
- Sabbagh, C. R. (2019), "Envisioning a compulsory-licensing system for digital samples through emergent technologies", *Duke Law Journal* 69(1):231–265.
- Sajko, M. et al(2006), "How to calculate information value for effective security risk assessment", *Journal of Information and Organizational Sciences* 30(2):263–278.
- Shannon, C. E. (1948), "A mathematical theory of communication", *Bell System Technical Journal* 27(3):379–423.
- Shapiro, C. et al(1998), *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business Press.
- Shapiro, C. & H. R. Varian(1998), "Versioning: The smart way to sell information", *Harvard Business Review* 6: 106–106.
- Shapley, L. S. (1953), "A value for n-person games", In: H. W. Kuhn & A. W. Tucker (eds), *Contributions to the Theory of Games*, Princeton University Press.
- Shen, Y. et al(2019), "Pricing personal data based on information entropy", *Proceedings of the 2nd International Conference on Software Engineering and Information Management*, 143–146.
- Squire, B. et al(2009), "The impact of mass customisation on manufacturing trade-offs", *Production and Operations Management* 15(1):10–21.
- Su, G. et al(2020), "BDTF: A blockchain-based data trading framework with trusted execution environment", ArXiv Working Paper, No. 2007.06813.
- Szabo, N. (1994), "Smart contracts", Unpublished manuscript.
- Tang, R. et al(2013), "The price is right", In: H. Decker et al(eds), *Database and Expert Systems Applications*, Springer.
- Thaler, R. H. (1980), "Toward a positive theory of consumer choice", *Journal of Economic Behavior and Organization* 1(1):39–60.
- Tomak, K. & T. Keskin(2008), "Exploring the trade-off between immediate gratification and delayed network externalities in the consumption of information goods", *European Journal of Operational Research* 187(3):887–902.
- Underwood, S. (2016), "Blockchain beyond bitcoin", *Communications of the ACM* 59(11):15–17.
- Varian, H. R. (1999), "Markets for information goods", In: K. Okina & T. Inoue(eds), *Monetary Policy in a World of Knowledge Based Growth, Quality Change and Uncertain Measurement*, Macmillan Press.
- Varian, H. R. (2009), "Online ad auctions", *American Economic Review* 99(2):430–434.
- Varian, H. (2018), "Artificial intelligence, economics, and industrial organization", In: A. Agrawal et al(eds), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.
- Veldkamp, L. L. (2005), "Slow boom, sudden crash", *Journal of Economic Theory* 124(2): 230–257.
- Vickrey, W. (1961), "Counterspeculation, auctions, and competitive sealed tenders", *Journal of Finance* 16(1):8–37.

- Vives, X. (1988), "Aggregation of information in large Cournot markets", *Econometrica* 56(4): 851—876.
- Vives, X. (1990), "Trade association disclosure rules, incentives to share information, and welfare", *RAND Journal of Economics* 21(3):409—430.
- Wang, D. et al(2018), "A novel digital rights management in P2P networks based on bitcoin system", In: F. Li et al (eds), *Frontiers in Cyber Security*, Springer.
- Wang, R. Y. & D. M. Strong(1996), "Beyond accuracy: What data quality means to data consumers", *Journal of Management Information Systems* 12(4):5—34.
- Wu, C. et al(2019), "Cloud pricing models: Taxonomy, survey, and interdisciplinary challenges", *ACM Computing Surveys* 52(6):1—36.
- Wu, S. & R. Banker(2010), "Best pricing strategy for information services", *Journal of the Association of Information Systems* 11(6):339—366.
- Yan, R. (2008), "Profit sharing and firm performance in the manufacturer-retailer dual-channel supply chain", *Electronic Commerce Research* 8(3):155—172.
- Young, T. et al(2018), "Recent trends in deep learning based natural language processing", *IEEE Computational Intelligence Magazine* 13(3):55—75.
- Yu, H. & M. Zhang(2017), "Data pricing strategy based on data quality", *Computers & Industrial Engineering* 112:1—10.
- Zeithaml, V. A. (1988), "Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence", *Journal of Marketing* 52(3):2—22.

## Research Progress on the Right Delimitation, Exchange and Pricing of Data

XIONG Qiaoqin TANG Ke  
(Tsinghua University, Beijing, China)

**Abstract:** Data and its exchange are two increasingly important pillars in research and practice, which lack systematic and interdisciplinary theoretical research, however. This article first reviews the delimitation of data rights. It proposes separating data rights for data products and data information with contextual integrity, which clarifies the premise of data monetization. Then this article summarizes the characteristics of data asset, including low replication costs and high value heterogeneity. Next, it synthesizes the circulation mode of data, exchange mechanisms and pricing of digital goods and data goods. Data exchange, primarily through third-party data vendors, is more in line with market participations' economic rationality than requiring firms to share data in most scenarios. It then points out that data assets' business patterns and seller strategies are greatly affected by application scenarios and buyers' heterogeneity. Differentiated products and pricing are essential strategies. Blockchain transaction technology solutions could efficiently regulate data exchange. Finally, it presents a unified overview of the applicability and limitations of data assets' pricing principles and methods.

**Keywords:** Delimitation of Data Rights; Data Asset; Data Exchange; Data Pricing; Digital Economy

(责任编辑:刘洪愧)

(校对:李仁贵)