

# 数据要素定价机制研究进展<sup>\*</sup>

欧阳日辉 杜青青

**摘要:**数据要素定价机制是数据要素市场建设的重要内容,是买卖双方在制度、场景和技术等约束条件下进行数据要素交易价格确定的制度安排。本文首先在明确数据要素定价的客体是具备生产要素属性的数据产品和服务的基础上,探讨了影响数据要素定价的成本、价值、场景等因素,归纳了数据要素定价的一般性原则和特定性原则。其中,制度设计是数据要素定价的关键,场景对数据要素定价影响很大。然后,本文梳理了数据要素的交易参与者、市场结构、交易机制和交易规则在交易制度设计中的作用,评述了交易场景和应用场景对数据要素定价的影响,综述了不同交易场景下数据要素定价的主要方法与模型。最后,本文介绍了机器学习、区块链、智能合约和密码技术等数字技术在数据要素定价中的应用。

**关键词:**数据要素 数据产品 定价机制 交易制度 数字技术

数据是数字经济的关键生产要素与核心资源,它与其他生产要素深度融合,赋能传统产业,对经济增长产生乘数倍增作用。在此过程中,定价是数据要素市场化配置的关键环节,党和政府高度重视数据要素定价机制的建立健全。2020年4月,《关于构建更加完善的要素市场化配置体制机制的意见》中强调,“丰富数据产品”“健全生产要素由市场评价贡献、按贡献决定报酬的机制”“完善要素交易规则和服务”<sup>①</sup>。此后,相关部门落实党中央和国务院的部署,2021年11月,工业和信息化部印发《“十四五”大数据产业发展规划》提出,到2025年初步建立数据要素价值评估体系,推动建立市场定价、政府监管的数据要素市场机制<sup>②</sup>。2022年1月,国务院印发的《“十四五”数字经济发展规划》进一步明确提出,鼓励市场主体探索数据资产定价机制,逐步完善数据定价体系<sup>③</sup>。地方政府也积极探索建立数据要素定价机制,比如,《广东省数据要素市场化配置改革行动方案》提出健全数据市场定价机制<sup>④</sup>;《上海市数据条例》提出,市场主体可以依法自主定价,但要求相关主管部门组织相关行业协会等制订数据交易价格评估导则,构建交易价格评估指标<sup>⑤</sup>。由此可见,政府对建立数据要素定价机制尚处于探索阶段。

目前,国内外数据交易机构和理论界都在探索数据要素定价的方法、模型和策略。在实践中,数据资产价值评估主要采用市场法、收益法及成本法等传统方法,或者基于统一费用、溢价和线性定价

<sup>\*</sup> 欧阳日辉,中央财经大学中国互联网经济研究院,邮政编码:100081,电子邮箱:ouyangcass@163.com;杜青青,北京语言大学经济研究院,邮政编码:100083,电子邮箱:duqing\_du@163.com。基金项目:教育部哲学社会科学重大课题攻关项目“互联网金融的风险防控与多元化监管模式研究”(15JZD022);研究阐释党的十九届五中全会精神国家社会科学基金重大项目“数据要素参与收入分配的机制与策略研究”(20ZDA047);北京高校卓越青年科学家计划项目(BJJWZYJH01201910034034)。

① 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)。

② 中华人民共和国工业和信息化部:[https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2021/art\\_c4a16fae377f47519036b26b474123cb.html](https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2021/art_c4a16fae377f47519036b26b474123cb.html)。

③ 中国政府网:[http://www.gov.cn/zhengce/content/2022-01/12/content\\_5667817.htm](http://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm)。

④ 广东省人民政府:[http://www.gd.gov.cn/zwgk/gongbao/2021/20/content/post\\_3369676.html](http://www.gd.gov.cn/zwgk/gongbao/2021/20/content/post_3369676.html)。

⑤ 上海市人民政府:<https://www.shanghai.gov.cn/nw12344/20211129/a1a38c3dfe8b4f8f8fcb5e79fbc9251.html>。

等简单的定价方法(Zhang & Beltran,2020)。已有学术文献多数讨论数据要素定价方法和模型,对数据要素定价机制的研究尚处于起步阶段。数据作为生产要素必须基于场景考虑数据要素定价,比土地、劳动力、资本、技术等传统生产要素的定价机制更为复杂。此外,数字技术也对数据要素定价产生影响。据此,本文从数据要素定价机制的视角,对数据要素定价的影响因素和原则、交易制度、定价方法和模型,以及数字技术在数据定价中的应用等文献进行评述。

## 一、数据要素定价的对象、影响因素与主要原则

### (一)数据要素定价对象与机制

讨论数据要素定价应该区分哪种形式的数据可以作为生产要素。Pei(2020)认为,经过加工的数据主要分为两类:电子书和在线音乐等数字产品;数据集和数据报告等数据产品。其中,数字产品作为最终商品直接用于消费,不是生产要素。只有数据产品才是数据要素定价的对象,原因有二:一是数据要素加工后形成数据产品进行交易,能够给数据提供方、数据需求方和数据经纪人带来收益和效用(Meierhofer,2019);二是数据产品投入生产过程与其他生产要素融合应用能够提升最终产品或服务的性能或生产效率(Müller et al,2018)。

Yu & Zhang(2017)定义了数据产品,指经过抓取、重新格式化、清洗、加密等处理后的数据产品和服务,如数据集和由数据集衍生的信息服务。在数据要素市场中,根据加工精细程度和传输技术手段的不同,数据产品主要包括数据包、数据 API、数据报告和数据服务。另外,服务商可以提供个性化的数据产品和服务。随着数据要素的应用以及大数据交易平台的发展,将“数据要素作为一种商品”进行定价的思路得到了认可。所以,数据要素定价是指对数据资源通过加工形成的、可以作为生产要素的数据产品和服务进行定价。

不同于实物商品和金融产品,数据要素具有外部性、异质性、价值溢出、交易场景多元等特征(Short & Todd,2017),分析数据产品定价时需要充分考虑这些特征,进而对传统价格理论进行创新。具体而言,数据要素定价机制需要考虑三个因素:第一,数据的价值具有高度情景相关性,定价必须基于场景(Short & Todd,2017),但传统的价格理论没有考虑场景因素,难以解释数据要素定价。第二,数据要素市场结构比较复杂,存在单边市场交易双方博弈(Liang et al,2018),也存在多边市场的“柠檬市场”(Heckman et al,2015),这会影响定价策略。第三,不同于商品交易必须是所有权的转移,数据要素交易既可以是数据使用权,也可以是数据所有权<sup>①</sup>,需要针对不同的交易权利设计不同的交易合同。基于以上分析,本文认为,数据要素的定价机制包括但不限于由供求决定价格的定价方法、策略和模型,是买卖双方对制度、场景和技术等多种约束条件下进行数据交易价格确定的制度安排。

### (二)数据要素定价的影响因素

1. 成本是卖方确定数据产品价格的关键因素。数据的成本结构与实物商品不同,数据的总成本是重置成本减去贬值损失。数据产品的重置成本分为三类:数据的采集、确认和描述等建设成本,数据存储和整合的运维成本,人力成本、间接成本以及服务外包成本等管理成本<sup>②</sup>。Fontana et al(2020)认为可以采用成本法估计英国的国家医疗服务体系(NHS)收集的医疗数据的价值。但是,很多研究指出了成本法给数据要素定价造成的困扰:一是成本量化困难,作为生产经营生产物的数据资产没有对应的直接成本,且不易进行间接成本的分摊(Közgazdász,2017);二是传统成本法的定价效果不佳,成本法是对历史价值的评估,忽略了数据产品的增值潜力,需要随着数据的使用不断调

<sup>①</sup>国家工业信息安全发展研究中心:《2020—2021年中国数据要素市场发展报告》,https://http://www.199it.com/archives/1241135.html。

<sup>②</sup>中国资产评估协会:《资产评估专家指引第9号——数据资产评估》,http://www.bicpaedu.com/hyzz/20200114/48001.html。

整初始估值(Adler et al,2016)。

更为重要的是,数据产品具有很高的固定成本和几乎为零的边际成本(Shapiro & Varian,1998),后者导致买方可以很容易地生成副本并以较低价格转售,从而引发数据盗版问题(Jones & Tonetti,2020),损害数据所有者或控制者的权益。基于以上特征,Adler et al(2016)认为,数据定价无法采用传统的边际成本定价法,而需要考虑潜在价值、顾客感知等其他因素,将成本法用于设定价格区间的下限是可行的。

2. 数据价值是影响交易双方对数据产品定价的主要因素。构建数据资产价值评估指标体系,是建立数据资产价值评估模型的基础。Gartner 和中关村数海数据资产评估中心提出的价值评估指标体系涵盖数据的内在价值、业务价值、绩效价值、成本价值、市场价值以及经济价值,包含数据的数量、范围、质量、颗粒度、关联性、时效性、来源、稀缺性、行业性质、权益性质、交易性质、预期效益<sup>①</sup>。在确定指标体系之后,通常结合模糊综合评价法进行指标量化<sup>②</sup>,即运用层次分析法,请专家针对数据的各评价指标进行打分,然后根据打分情况计算出每个影响因素的权重,将定性评价转化为定量指标。而且在不同使用场景,各指标的权重不同或对指标有所取舍(Short & Todd,2017)。

数据价值的影响因素主要包括数据要素的完整性、准确性、层次性、协调性和异质性等。首先,数据要素的完整性和准确性与数据要素价值成正比(Heckman et al,2015;Immonen,2015)。完整性是指数据要尽可能涵盖被记录对象的属性,包括数据体量、数据采集时间连贯、数据关系完备等(Pérez-Pons et al,2019)。准确性表示数据被记录的精准程度,是数据质量的核心指标(Sukumar et al,2015)。数据量与数据价值成正比,数据集包含的信息量通过信息熵衡量(Li et al,2017;Mao et al,2019)。但是,一旦信息量没有达到或者超过某个最佳点,决策绩效就会下降(Moody & Walsh,1999)。Bajari et al(2019)利用亚马逊的零售数据证实了这个结论。其次,数据产品的层次性包含技术含量、稀缺性和数据维度。数据产品和服务的技术含量越高,其价值也越高(Yu et al,2011)。稀缺性表示数据被所有者独占的程度,如果某类数据仅由一个机构掌握,其所蕴含的商业信息价值很高。此外,数据维度越多,适用的范围也越广,应用价值就越高。Yu & Zhang(2017)基于数据质量多维度以及多维度之间的相互作用建立了数据定价双层编程模型。当处于同一维度的数据质量标准提升时,另一维度的数据质量标准下降,数据卖方对数据处理上的投入增加时,会将此作为约束条件纳入数据产品定价模型。所以,考虑数据质量的多维度、多版本策略能够实现更好的市场细分。再次,数据要素具有协调性或协同性。不同类型的数据、数据集或数据产品的组合会产生不同的增量价值(Zheng et al,2017)。最后,异质性源于数据结构不同(Koutris et al,2013)、采集主体不同、价值高度依赖使用场景(Short & Todd,2017)、市场分割以及买方异质性(Bergemann et al,2018)等,很难给出统一的定价公式。然而,数据质量指标之间的复杂互动也会影响数据质量,比如,提高一个特定数据集的准确性可能会以牺牲其完整性为代价。

此外,卖方追求长期利润最大化,考虑客户感知价值而不是基于成本的定价,从而会设计客户感知价值定价模型(Luong et al,2016)。根据客户感知设定数据价格需要确定一组价值驱动因素,这些因素主要包括买方对数据产品成本和效用的认知、供应商社会声誉带来的声誉价值、心理动机以及对数据产品潜在价值的认知等(Wu et al,2019)。

3. 场景影响数据效用,进而影响数据产品定价。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。一方面,数据价值与具体的应用场景相关,数据要素只有被使用才会产生价值,同样的数据对不同买方的价值差异很大,卖方会根据买方异质性实行价

①Laney, D. (2015), "Why and how to measure the value of your information assets", <https://www.gartner.com/en/documents/3106719>。

② 普华永道:《开放数据资产估值白皮书》,CSDN,[https://blog.csdn.net/weixin\\_58337139/article/details/118696079](https://blog.csdn.net/weixin_58337139/article/details/118696079)。

格歧视策略。另一方面,因交易场景不同数据定价方法而异,比如,收益现值法适合基于项目数量和用户数量确定租赁费用的订阅方式,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。基于场景的定价特点是数据要素定价与其他要素定价最大的不同(熊巧琴、汤珂,2021)。

### (三)数据要素定价的主要原则

数据要素定价的基本原则是选择定价方法和模型的重要依据。Pei(2020)提出,数据定价遵循真实性、收益最大化、避免套利、公平性、保护隐私和高效匹配等原则,但在不同的使用场景和定价模型中有所取舍。本文将数据要素定价的原则分为一般性原则和特定性原则两类。其中,一般性原则与产品的定价原则相仿,但具体内涵有所不同;在数据要素特定的交易场景和定价模型中,重视坚持真实性、避免套利和保护隐私等特定性原则。

1. 一般性原则。数据要素定价也遵循商品定价的基本原则,比如,以价值为依据、成本为基础、市场竞争为导向。收益最大化、公平性和高效匹配被认为是数据产品定价的一般性原则。因为数据产品的复制成本很低,数据定价模型普遍遵循收入最大化而非利润最大的原则。比如在拍卖模型中,卖方以收入最大化为原则确定拍卖的数量(Goldberg et al,1999),基于查询的数据定价追求无套利和收入最大化目标,据此建立定价算法(Koutris et al,2015)。

公平性原则不仅指买卖双方的公平定价,还需要考虑利益相关者的公平分配。Shapley(1953)提出,公平分配应具备平衡性、对称性、零要素、可加性四个条件,值是满足所有要求的唯一分配方式。受这个思路的影响,Jia et al(2019)认为,Shapley 值是用于衡量数据产品收益公平分配的最佳工具。Agarwal et al(2018)提出,由于数据产品的复制成本低、再生产边际成本接近零,卖方可以低成本复制相同的数据产品,获取更多的 Shapley 值,从而获得不合理的收益。这对数据要素市场的实际公平提出了挑战。

高效匹配原则指定价模型必须以适当的价格来匹配买卖双方,提高计算效率实现高效匹配。数据产品价值因应用场景而异,有效地计算众多交易参与者的市场报价是对数据交易平台的基本要求。如果计算效率过低,则会影响数据价值和交易效率。密码学、区块链等数字技术的结合应用是解决计算效率问题的方向,例如,Liu et al(2021)将“盲目多项式评估”的密码学技术与区块链范式相结合,设定明确的协议并将其实例化,构建了具有不依赖第三方、数据处理规模大等优势的计算函数。

2. 特定性原则。真实性是市场有效的保障,可以促使卖家提供真实效用价值最大化的数据产品。真实性原则是拍卖机制的核心原则(Agarwal et al,2018),买家只愿意支付真实效用价值最大化的价格(Pei,2020)。Jiao et al(2018)提出了一个数据服务提供商和买家之间基于拍卖的数据交易模型,同时坚持了真实性、收入最大化和高效匹配原则。

无套利性指参与者无法通过不同市场的价格差异获利,是基于查询的数据定价的核心原则,可以分解为无信息套利和无捆绑套利。Koutris et al(2015)为查询定价设计了一种满足卖方收入最大化、无套利和公平分配的定价算法,如果预先的设定价格不存在套利情况,就存在唯一的定价函数,能满足无套利和无折扣条件。Zheng et al(2017)对移动人群感应数据进行查询定价,同时遵循捕捉数据的不确定性、无套利和收入最大化三个原则。Li & Raghunathan(2014)则认为,在对个人敏感数据交易提供查询服务时,卖方必须接受一些套利的风险,以便制定合理的价格。

保护隐私原则在隐私含量高的数据交易场景中被重点考虑。网络平台用户的个人信息、数据提供方的经营信息以及第三方交易平台的信息很容易在交易中泄漏。例如,训练机器学习模型的样本通常来自存储在云服务器上的用户内容,在提取过程中存在隐私泄露风险(Koushanfar et al,2019)。因此,理论界积极探索保护数据产品隐私的方法,包括不得出售未经脱敏的原始数据,建立去中心化和可信的数据交易平台,使用区块链技术保护隐私,采取买卖双方直接交易方式等(Hynes et al,2018)。

## 二、交易制度设计与数据要素定价

在数据要素市场中,制度设计是数据要素定价的关键,数据要素交易机构的运行由一系列的制度安排维系,交易制度设计的目的是减少交易成本、降低交易风险、实现资源最优配置。

### (一)数据要素的交易参与者

Spiekermann(2019)认为,数据要素交易的参与者主要包括三类市场主体:数据提供商;数据买家;第三方服务提供商和数据交易中介。

1. 数据提供商提供数据整合服务,将数据加工成可以交易的数据产品。按照传统的产权理论,商品交易后所有权发生转移,然而现实中大多数的数据要素产权模糊,数据提供者也不一定要向买方让渡所有权。当前,数据提供商尤其是大型互联网企业,包括网络搜索引擎所有者、新闻媒体、在线购物平台等,拥有数据、流量、算法等方面的优势,对数据的商业价值认知最深,整合能力最强,在数据交易市场上占比最大(Li & Gao, 2021)。平台企业还会通过收购等方式强化优势。Li et al(2019)的研究表明,亚马逊、eBay 和谷歌的多次并购都是以获取目标公司的数据资源和提升数据分析能力为目的。平台也会采取提供优惠、积分或者服务等资源互换模式换取对消费者的购买历史记录、网站浏览历史记录、Web 搜索历史记录等个人数据的使用权(Li et al, 2017),再凭借算法等技术优势对用户数据进行处理和整合,进而出售数据产品。数据提供商对数据要素定价有话语权,因此政府对数据提供商定价进行有效监管有利于数据要素市场形成公允的价格。

2. 数据服务商从事数据的采集、整理、聚合、分析等加工业务。数据服务商即第三方服务供应商可以在数据采集、数据存储、数据分析、数据流通中提供多项服务,提升专业化分工程度,降低数据产品的生产成本,因而提高了数据价值形成的效率,为买卖双方节省了成本。根据 Muschalle et al(2013)的调研,应用程序供应商将数据集转化为一连串预先编译好的数据提供查询服务,买方只需根据具体需求进行查询;数据相关算法开发商提供将源数据转化为集成数据的算法等技术,如数据挖掘、匹配、清理、相关性计算和沿袭跟踪,并将这些算法作为用户定义的黑盒函数上传到数据市场。

3. 数据交易中介在数据要素市场的建设中不可或缺。它们不仅提供了操作规范、安全性高的交易场所,而且承担建立健全数据资产评估、登记结算、交易撮合、争议仲裁等市场运营的职责。数据交易中介打造存储、搜索、交换和托管数据及相关算法的通用平台(Muschalle et al, 2013; Spiekermann, 2019),提供针对不同品类的数据产品的定价模式和相应的技术支持。按照营业范围和模式,数据交易中介分为三种类型:一是交易所以中间代理人身份提供数据交易撮合服务。交易的数据产品类型和定价方式由买卖双方自主决定,此时交易中介只是买卖双方的桥梁,不提供数据处理等其他服务。二是以数据经纪人身份收集数据、处理并出售各类数据产品和服务,即作为掌握大宗数据资源的聚合平台,集中开展一对多的数据供需匹配服务。三是数据交易中介吸收政府、企业、社会多方参与,组成交易联盟,联盟内的成员可以共享数据。数据联盟形式的优点是形成数据资产的利益捆绑,有助于保护投资者利益,且数据质量较高,但是该方法门槛较高,数据流通仅限于联盟内部成员。

### (二)数据要素市场的结构类型

市场结构对实物商品定价影响深远,对数据要素市场也是如此。在数据要素市场中,价格制定者设计数据产品定价模型必然要考虑数据市场结构(Liang et al, 2018),根据买卖双方的数量和相对市场力量,遵循定价原则制定合适的定价策略。

1. 数据交易的单边市场和双边市场。已有研究将数据要素市场类型的分析作为数据定价模型设定的基础,对数据市场类型进行归纳。Zhang & Beltran(2020)认为,数据要素市场主要包括单边

市场、集中式双边市场和分散式双边市场三种类型。

(1)单边市场。单边市场包括买方市场和卖方市场,其中一方在定价中占主导优势。比如卖方市场上,卖方可以根据买方竞争的程度选择定价方法(Adler et al,2016)。当买方竞争程度高、有强烈的动机去独占数据资产时,卖方可以进行拍卖或基于买方的预算来制定销售价格;当买方竞争程度很低时,卖方可以使用成本法、收益法、市场法等传统法定价。数据交易市场上优势方的数量会影响定价策略, Ma et al(2019)以卖方市场建立的理论模型发现,当市场有两个卖家时,一个卖家为要求高的买家服务,另一个为要求低的买家服务,市场实现价格均衡。与二元竞争相比,如果市场上仅有一个卖家,卖方有动力为不同质量要求的买家提供可选择的数据产品方案,以此获取更多的利益。不少研究为简化模型或分析过程,在理论模型中假设要素市场是单边市场,给出了只有单个买家或卖家时的定价策略和模型(Ghorbani & Zou,2019),但是现实中,数据交易市场更偏向于双边市场类型。

(2)依赖数据中介机构的集中式双边市场。数据经纪人拥有较强的数据采集和数据分析能力,能提供整合程度和数据质量更高、种类更丰富的数据产品和服务(Kitchin & Lauriault,2015)。此外,数据的排他性以及中小企业数据利用能力的不足会加剧行业垄断,而数据经纪人有利于缓解这一难题(Li et al,2019)。更为重要的是,数据经纪人有动力采用多种数据保护技术,比如,使用密码学和隐私计算等技术实现数据加密,提供限制手段或规定数据的重复使用次数,推动私人数据转换为可交易的数据产品;或者在不影响数据所有权的前提下交易数据使用权(Hynes et al,2018),既能保护数据所有者权益,又能在合法情况下最大限度地开发数据要素价值。

数据经纪人之间也存在竞争与合作,进而影响社会福利水平。对于具有竞争性和排他性的私人数据,数据经纪人倾向于向数据拥有者提供更多补偿来获取更多数据量;对于非竞争性的数据则相反(Belleflamme,2018),因为高补偿会导致数据所有者与多家数据经纪人分享他们的数据,降低买方意愿价格并损害数据经纪人的利益。但是,大多数的数据要素具有非竞争性,数据经纪人的竞争无法使消费者获得更多社会福利(Ichihashi,2021)。如果数据集的价值具有较高互补性,数据经纪人可能选择合并数据集,进行联合销售(Gu et al,2021)。选择竞争还是合作,受到合并数据集价值的互补性和替代性的影响:当数据是“次加性”时,即扣除合并成本后的合并价值低于单个数据集的价值之和时,数据经纪人倾向于联合销售;当数据是“超级加法”时,即合并后的价值大于单个数据集的价值之和,竞争就会加剧。此外,采用 Shapley 值法验证数据互补性是一种比较好的思路,可以测定成员在参与合作中获得的回报。

(3)买卖双方直接交易的分散式双边市场。分散式双边市场的典型是点对点交易,买卖双方直接沟通,就交易对象、价格、时间、交割方式等内容签订意向协议。考虑到隐私和数据要素的权益归属,分散式双边市场包括两种交易模式:第一种是数据所有权属于消费者,数据收集企业向消费者购买数据。比如,Coggeshall et al(2011)的研究为个人用户开发了一个数据交易平台,用户可以通过平台对与自身相关的数据进行存储、控制、维护所有权和货币化,买方可以在平台上购买经过匿名化处理的个人数据访问权。第二种是数据产权属于企业,企业和数据购买者都应“为隐私付费”,对数据所有者的隐私损失提供补偿。Jones & Tonetti(2020)认为,第一种模式可以带来接近最优水平的分配,产生的社会福利高于第二种模式。第二种交易模式虽然会带来更好的交易量,但是企业可能会过度使用数据且不充分尊重数据所有者的隐私,而且企业有囤积数据的倾向。

Koutroumpis et al(2020)参照 Roth 的市场设计方法,提出了数据要素市场的四种类型,见表 1。这与 Zhang & Beltran(2020)的观点基本一致。其中,双边市场的流动性较低,交易成本高昂;如果不考虑数据独占性,多对一或一对多市场可能运行良好,无须在交易后对数据进行严格保护。多对多市场的数据平台所有者可以利用定价策略提高买卖双方的参与度,并在收集大量数据的基础上,

利用网络外部性实现盈利。Koutroumpis et al(2020)对比研究了四种匹配机制,认为有效的数据要素市场应该满足交易成本低、制度安排完善、匹配算法稳定等条件。实际上,市场设计的有效性与监管机制密切相关,监管机构直接在分散或多边市场中引入技术或契约制度来激励交易极为困难,可行办法是通过完善监管体系引导多边平台形成规范的大规模数据交易系统,从而提升定价效率和交易匹配效率。

表1 按匹配机制划分的数据市场类型与市场特征

匹配机制	市场设计	交换条件	实例	资产变现能力	交易成本	安全性
一对一	双边	协商	个人数据经纪人;Acxiom	低	高	高
一对多	发散	标准化	Twitter API;Facebook API	高	低	低
多对一	集中	隐性易货	谷歌 Waze;谷歌搜索	高	低	不确定
多对多	多边	标准化或协商	无	高	低	低

注:Acxiom 是一家位于加利福尼亚州旧金山的 SaaS 公司,提供数据连接平台,其服务包括数据上载,用于营销目的的在线离线数据传输。Waze 是一个基于 GPS 的导航移动软件应用程序,提供免费的地图导航和基于位置数据的广告投放服务,于 2013 年被谷歌收购。

资料来源:Koutroumpis et al(2020)。

2. 数据市场的垄断和竞争。市场结构可以分为竞争市场、垄断市场和寡头垄断市场等,一些研究认为数据交易市场的集中度较高,容易形成卖方垄断<sup>①</sup>。掌握大规模数据或高信息含量的数据时,数据控制者倾向于囤积数据,而不是分享,此时数据要素具有高度排他性。Varian(2018)也指出,优势企业可以建立数据库,向第三方机构出售数据使用权,利用排他性获得收入。也有研究反对这一观点,认为在数据市场上,单个主体的市场份额无法达到 50%,难以形成垄断(Tucker & Hill, 2014)。一方面,数据要素具有非竞争性和产权模糊性,优势企业无法完全掌握数据的控制权从而防止数据复制和传播;另一方面,特定场景下所需数据具有一定的可替代性,以信贷场景为例,信贷机构既可以使用央行征信中心的数据,也可以从互联网企业获取客户消费信息和信用记录,两种数据都可以达到风险控制的目的。总体来看,缺乏有效监管的数据市场很容易形成垄断,数据控制者会采取价格歧视的定价策略细化需求模型和价格函数,攫取数据商品的超额利润。

实际上,发展阶段不同,市场结构也可能不同。Koutroumpis & Aija(2013)认为,在数据市场发展的初级阶段,一些数据平台所有者会在数据获取和定价等方面占据优势。因此,一些关于数据定价的研究都明确或隐含地假设了一个垄断的市场结构,数据卖方(数据买方)并不关心与他人的竞争。例如,Bergemann et al(2018)给出了卖方垄断市场的数据定价模型。模型假设数据卖方是垄断者,为实现利润最大化目标,数据卖方提供不同质量的数据产品菜单,购买方根据使用意愿进行选择。随着数据要素市场的发展,数据卖方可能会出现寡头垄断的局面。Balasubramanian et al(2015)研究了双寡头垄断市场结构下免费和按使用量付费两种定价策略对数据卖方收益的影响。

### (三)数据要素的交易机制和交易规则

1. 市场交易机制的设计和选择。交易机制设计为参与者提供了满足利益最大化、真实披露等既定目标的交易互动方式。如果数据产品的排他性较强,可以设计拍卖机制进行定价。拍卖制度可以激励代理人披露真实估值。拍卖机制遵循公开、公平、公正的原则,具有快速和批量的特征,缩短交易时间,提高交易效率(Jin et al,2015)。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。其中,贵阳大数据交易所对拍卖机制的适用场景进行了

<sup>①</sup>OECD(2019),“Measuring the digital transformation: A roadmap for the future”, [https://www.oecd-ilibrary.org/science-and-technology/measuring-the-digital-transformation\\_1b9f3165-en](https://www.oecd-ilibrary.org/science-and-technology/measuring-the-digital-transformation_1b9f3165-en)。

描述:一是买方仅仅为了短期需求行为而不愿意支付年度数据采购费用;二是买方希望一次性垄断数据而不愿意其他买方再共享此类数据。

拍卖机制的现有研究成果如表 2 所示,包括双边拍卖、反向拍卖和 VCG 拍卖等多种拍卖类型(Liang et al,2018),选择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,选定的定价原则等因素。

表 2 拍卖机制在数据交易和定价中的应用

拍卖方式	市场结构			描述	定价原则	参考文献
	卖家	中介机构	买家			
维克里拍卖	多个	无	多个	隐私是数据的主要成本,卖方会根据隐私提供要价;当买方对数据有一个固定的准确度目标时,维克里拍卖(即第二价格密封拍卖)的应用保证付款最小化的同时实现了准确度目标;当买方有固定预算时,拍卖机制可以最大限度地提高最终估计的准确性,同时确保最终付款总额不超过预算	真实性;隐私补偿;买方总付款最小化	Ghosh & Roth (2015)
维克里拍卖/序贯拍卖	多个	有	多个	修改了传统维克里拍卖和序贯拍卖模型,为卖家决定拍卖数据产品的数量提供解决办法;提出修改的序贯拍卖定价模型,用以确定当卖家将同一数据产品卖给多个竞拍者时,对获胜竞拍者应提供多少补偿	社会福利最大化	Chen et al(2018)
维克里拍卖/VCG 拍卖	一个或多个	有	多个	设计了频谱交易的拍卖机制,通过修正维克里拍卖来解决多赢家拍卖策略下,传统 VCG 拍卖存在的损害卖家收益和受到“合谋”攻击的问题;拍卖商根据约束条件和买方的满意程度来选择最终赢家,该机制可以提升买方使用效率和卖方收入	真实性;社会福利最大化	Sofia & Edward (2020)
单轮密封投标拍卖	多个	有	多个	模拟了无限供应市场上的密封投标拍卖,将拍卖机制的总效用与固定定价的总效用做对比,发现拍卖定价机制的效果更好	真实性;社会福利最大化	Goldberg et al (1999)
反向拍卖	多个	无	多个	为移动传感数据设计了一个反向拍卖模型,假设买方(平台)有一个 $\Gamma = \{\tau_1, \dots, \tau_n\}$ 的传感任务,每个任务 $\tau_i$ 会产生价值 $v_i$ 。每个卖家选择一个任务子集 $\Gamma_i \subseteq \Gamma$ ,执行这一任务需要花费的成本为 $c_i$ 。卖方 $s_i$ 为传感数据设置的价格为 $b_i$ ,并将投标 $(\Gamma_i, b_i)$ 提交给买家,在收集所有的出价后,买方选择一个卖家子集 $S$ 作为赢家,并确定对每个赢家 $s_i$ 的付款 $p_i$ 。在拟定的拍卖机制下,从 $S = \emptyset$ 开始迭代选择带来最大非负边际净利润的卖家	真实性;个体理性;平台收益最大化	Yang et al(2012)
反向拍卖/VCG 拍卖	多个	无	多个	当卖家对一组传感任务进行竞价时,必须满足每个传感任务的数据都符合质量要求,采用贪婪算法选择赢家,可以实现最佳总效用的近似值;当卖家对多组传感任务进行竞价时,采用迭代下降算法选择赢家,可以实现接近最优总效用的目标	真实性;社会福利最大化	Jin et al(2015)



拍卖方式	市场结构			描述	定价原则	参考文献
	卖家	中介机构	买家			
双边拍卖	多个	无	多个	设计了一个迭代拍卖机制,目标是社会福利最大化,买卖双方直接交易原始数据。为了根据提交的出价设定下一轮的规则,拍卖者的问题被表述为一个受限的优化问题,并得出最佳规则,然后开始新一轮的拍卖,直到观察到社会福利的收敛	个体理性;社会福利最大化	Cao et al(2017)
	多个	有	多个	中间商收集卖家数据,对卖家进行隐私补偿,并对收集的原始数据进行处理,根据买家需求出售数据服务;采用贝叶斯利润最大化拍卖计算最优价格和数量,实现中间商收益最大化	真实性;中间商收益最大化;计算效率	Jiao et al(2018)

针对排他性弱、潜在买家多的数据产品,卖家采取直接交易的方式,以保底价格定价,并设计产品菜单供消费者自主选择产品。例如,Cummings et al(2015)假设市场有两类数据提供者,第一类数据提供者都是单独的个体,价格方差主要受差异化隐私处理的影响;第二类数据提供者是有能力从其控制的子人群(如学生、教授等)中收集不同规模的随机样本的组织,价格方差主要与样本中包含的个体数量有关。通过选择方差来模拟消费者对感兴趣的数据产品的选择,实现总体统计的无偏估计以及买方支付成本最小化的目标。

Wei & Nault(2014)认为,在卖方制定产品菜单和价格、买方直接购买的交易方式中,卖方采取了基于买方异质性的差异化定价策略。Moriarty(2021)认为这种定价策略会不公平地剥夺交易所创造的社会剩余,在线销售商应该披露他们正在进行差异化定价或者停止这一策略。也有研究认为社会总体福利标准关注的是福利的无谓损失而非福利转移损失,如 Li et al(2019)指出,虽然价格歧视通过减少消费者的支付意愿和购买价格之间的差额来减少消费者剩余,但是减少的消费者剩余作为增加的利润转移给了卖方,在一般均衡中,卖方增加的利润又作为收入分配给家庭,促进了社会的公平分配。

2. 合同设计的影响因素。拍卖机制和直接交易机制中的差异化定价是应对信息不对称和买方异质性的选择,是数据交易市场的主要交易机制。但是交易机制具有多样性,合同设计能够规范交易流程,为不同的交易机制提供保障,进而提高交易成功率。交易合同设定的主要影响因素如下:

(1)数据使用量。如果买方不确定对数据的需求量,可以设计两部收费合同(Li & Raghunathan, 2014),即数据消费者先为一定数量的数据集或查询支付固定费用,当使用量超过固定额度后,再按单位价格购买数据。如果消费者对数据产品能否满足其效用存疑,可以提供“前期免费,后期付费”合同,或者先提供数据产品的低阶测试版(Choudhary et al,1998)。这至少会给数据经纪人带来两个好处:一是先建立网络外部性效应,然后再推出价格更高的高阶版本,获取更多市场份额;二是通过提供免费产品或服务换取用户有价值的敏感信息(Li et al,2017),利用这些信息改善决策或者直接将信息出售给第三方机构获利。分层定价可以根据购买数量的变化设定不同的单价,且数据集的价格和数量不是简单的线性关系。Zhang & Beltran(2020)认为,分层定价至少包含两种策略:一是随购买数量增加降低单价;二是根据边际支付意愿对消费者分类,然后分别定价。

(2)数据所有权和使用权。数据所有权交易主要是对数据包和数据报告的产权转让,交易合同以数据包为定价单位,由卖方上传数据产品并定价,然后数据经纪人整理成产品菜单,买方根据需求发起交易要约。数据使用权交易主要指 API 数据,指在保证所有者权利的基础上给予买方访问权限,交易合同以数据调用次数为定价单位,由于 API 数据需要持续更新,合同还规定了查询 API 的衍生产品和服务的定价内容(Balazinska et al,2011),买方可以在云端发出查询请求,然后购买查询 API 接口。在 Balasubramanian et al(2015)的研究中,如果市场上只存在一个卖家,出售数据使用权比出售所有权获得的收益更高;如果市场上存在两个卖家,分别交易数据所有权和使用权,此时交易

数据所有权的收益更高,而使用权交易的成本上升会使两个卖家的收益同时上升。

(3)买家偏好。数据经纪人可以根据买方偏好,通过合同设计提供数据产品的不同版本和价格搭配供代理人选择。例如,Harmon et al(2009)认为,关注搜索成本的消费者不愿意花时间去搜寻与购买相关的信息,并倾向于将高价格与高质量联系起来,或者随机购买高价格的数据产品。对此,数据经纪人需要提供不同数据质量的合同菜单(Bergemann et al,2018),以激励消费者显示其具体需求。考虑到可操作性,数据经纪人希望能够限制菜单长度,Mehta et al(2021)提供了一种近似最优收入的合同菜单设计,即菜单长度为 $L$ 时(最多有 $L$ 项产品-价格组合),满足数据经纪人对实际收入和最优收入进行权衡和选择的需求。

Li & Raghunathan(2014)根据对数据敏感度的偏好将买方分为两种类型:一种关注信息总数;另一种关注私人信息含量。数据经纪人可以分别为这两类买方提供低敏感度和高敏感度的数据。此外,关注预期价格的买家不急于使用数据产品,可以提供数据期货式交易和现货式交易。当数据买方对某一数据产品存在较高预期且不一定以使用该数据产品为目的时,可以采取期货式交易方式(Gundepudi et al,2001)。此时,买方认为该数据产品的潜在价值很高,提前以某一价格支付数据产品的购买费用,并且可以在未来任意时间点将这笔期货转让出去,可购买的期货产品包括远期、近期和即时合约。另一种情况是买方期待折扣等低价购买时机,等待现货购买。折扣和津贴的类型主要包括提前付款、淡季打折、批量购买、零售折扣、现金折扣和以旧换新津贴六种(Wu et al,2019),如Balasubramanian et al(2015)提出为信息产品的买方提供基于使用频率的折扣。

### 三、基于场景的数据定价方法和模型

#### (一)基于应用场景和交易场景的定价

数据的交易场景非常广泛,难以设定一个具有普遍适用性的数据定价标准。各行业的数字化程度、数据丰裕度和交易场景等存在明显差异,一些学者根据不同场景研究不同行业的数据定价问题。例如,Bourreau et al(2018)提出了一个网络广告行业的数据交易模型,建议采用拍卖法进行定价,数据交易平台控制交易中出售多少消费者数据。Son et al(2021)研究了非公共组织医疗数据的定价,认为初始价格由成本和数据价值决定,然后根据交易中的实时偏好和供求变化进行动态调整。FTC(2014)将应用场景分为市场营销产品、风险控制产品和人员搜索产品三类<sup>①</sup>。例如,上海数据交易中心提供的中国受众画像库(CAP产品)通过增补企业缺失的用户画像来帮助企业开展客户洞察、客户运营和后续的市场营销活动。

数据有应用场景,但没有交易场景的情形也大量存在。这种情景主要指政府免费开放数据(Arribas-Bel et al,2021)、企业共享数据(Moody & Walsh,1999)以及并购、诉讼等非交易场景。在非交易场景之下,数据的价格本质上是一种对于数据价值的评估。开放数据可以根据成本和消费者的支付意愿定价,或者采取“免费+增值”的模式,提供免费的基础版本和作为商业产品的增强版本(Arribas-Bel et al,2021)。企业之间的数据共享可以采取俱乐部制度,或者数据联盟形式。并购场景下,参与者关注的是未来经营状况,可以采用收益法定价、实物期权等定价方案;为了突出数据价值,在必要时可以采用Shapley值定价。诉讼场景下,可以选择成本法和比较法进行定价。如果找不到可比对象,可以采用知识产权领域对标准必要专利许可费定价中的Georgia-Pacific方法(Epstein & Marcus,2003),这样就可以根据各种具体情形,在可供选择的基准的基础上进行定价。

#### (二)交易场景中数据产品的定价方法和模型

1. 订阅和租赁。批量廉价数据具有获得性和可复制性强的特点,可以采取固定定价和按次计价

<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>。

等订阅和租赁方式交易(Liang et al,2018)。按次计价即按照对数据的使用次数收费,主要用于对数据使用权的交易。根据Fruhirth et al(2020)对美国现有20个数据交易平台的调查结果,超过半数使用了固定定价的方法,即采取事前定价的方式,根据对数据产品的价值评估确定其价格并挂牌销售。对买方来说,固定定价是与其使用限制有关的每月或每年的订阅费用,在这种模式下,时间是决定他们需要支付多少费用的唯一因素。例如,Infochimps收取每月的订阅费,并允许数据消费者调用一定数量的查询。

2. 协议定价与博弈模型。协议定价是目前使用最广泛的定价方式,主要包含两种使用场景:一是当价格意见不统一时,买卖双方在数据交易平台的撮合下进行商议,双方讨价还价之后确定成交价;二是没有中介机构,买卖双方直接进行交涉。如果对某数据产品的估值定价意见不统一,买卖双方可以采取有中介机构撮合的协议定价方式,从而获得更大的定价自主权和商议空间。协议定价的过程由博弈模型刻画,包括基于非合作博弈的定价模型、基于斯塔克伯格博弈的定价模型以及基于讨价还价博弈的定价模型。

(1)非合作博弈。参与者之间不会建立合作关系,每个卖家都是自私的,并独立设定使自身利润最大化的价格(Luong et al,2016)。在非合作博弈模型中,所有参与者都必须发布一个透明的定价策略,这在真实的数据要素市场中难以实现,因为参与者无法计算纳什均衡,导致该模型的实用性不高(Liang et al,2018)。

(2)斯塔克伯格博弈。Liu et al(2019)搭建了一个两阶段的斯塔克伯格博弈模型:第一阶段,数据经纪人是价格领导者,公布数据所有者给出的所有价格策略;第二阶段,数据需求者根据领导者公布的信息出价并发出购买请求;最后,数据经纪人进行交易撮合。Xu et al(2020)构建了一个三阶段的斯塔克伯格博弈模型:在第一阶段,数据所有者公布定价策略,服务提供商根据公布的价格做出对原始数据的购买决策;在第二阶段,服务提供商对原始数据进行处理后向买方提供包含一定数量不同版本的数据产品的菜单;在第三阶段,买方根据菜单做出购买决定。然而,在斯塔克伯格博弈中,数据拥有者需要确定一个价格领导者,并在其宣布定价策略后制定价格,在实践中如何找到价格领导者存在困难。

(3)讨价还价博弈。讨价还价博弈是解决复杂谈判条件的适当方案,因此它通常被用于数据拍卖。数据供应商和需求方分别提出自身的最优定价策略,如果需求方出价大于供应商出价,交易达成,反之则谈判继续(Mao et al,2017)。Jung & Park(2019)考虑隐私保护的重要性,将差分隐私应用于数据市场,构造鲁宾斯坦讨价还价模型来确定隐私损失的价值(噪声参数 $\epsilon$ ),以社会福利水平最大化为目标确定数据产品的定价公式。该模型也存在缺陷,一是供需双方的谈判过程耗时又浪费资源,实施困难;二是在鲁宾斯坦讨价还价模型中,如何确定出价区间没有统一定论,并且如果对方无法接受出价,理论上可以进行无限轮次讨价还价,实际中则难以达成交易。

3. 隐私定价。一些学者提出了隐私补偿定价方法,主要适用于隐私含量较高的个人数据交易场景,购买数据的成本主要是提供给数据主体的隐私补偿。在买方有准确性要求的假设下,Ghosh & Roth(2015)认为买方必须购买隐私,买价即隐私成本,每个数据提供者的隐私成本函数表示为 $c_i(\epsilon) = v_i \cdot \epsilon$ , $v_i$ 是个人 $i$ 的单位隐私成本, $\epsilon$ 为数量单位;买方采用经典的维克里拍卖方式购买数据,即选择出价最低的 $m$ 个卖方,并为拍卖中的每个赢家提供统一补偿 $b \cdot \epsilon$ , $b$ 是第 $m+1$ 个最低出价。Yang et al(2019)进一步说明了隐私成本取决于数据提供者的隐私态度。每个数据提供者 $i \in K$ 用一个非递减函数(即补偿方案) $C_i: \tau \rightarrow R^+$ 来建模,代表数据平台和数据提供者之间就数据提供者的隐私损失 $\tau$ 应得到多少补偿的承诺, $\tau$ 由差分隐私定义。然后,作者通过一个双层编程模型模拟消费者的自我选择过程和垄断者(数据交易平台)的决策行为。数据交易平台的总利润 $G(c, p, x) = \sum_{j=1}^M \sum_{i=1}^N (p_i - c_i) x_{ij}$ ,其中 $x_{ij}$ 度量消费者的自我选择过程, $i$ 表示隐私数据的敏感性水平, $x_{ij} = 1$ 表示消费者愿意购买隐私敏感度为 $i$ 的数据产品。不同的隐私敏感性数据提供的效用不同,价格 $p_i$ 随

敏感性上升而提高。实验结果表明,多层次的隐私划分可以实现总效用最大化。

4. 基于查询服务的定价。查询定价是对在线数据库定价的优化,因为查询要求更细的颗粒度(Lin & Kifer, 2014),即不用购买整个数据库,而是选择想要的覆盖范围或聚合水平。具体操作为,卖方首先在数据库中几个视图上设置明确的价格,买方输入想要购买的任意查询,系统就会根据已经设定价格的视图自动派生出查询价格(Koutris et al, 2015)。Koutris et al(2013)研究了查询需要多个卖家的数据库联合运算的情况,通过在卖家之间引入一个收入分享政策,每个卖家都能得到一份查询价格,该价格与卖家在整数线性规划程序定价中所能得到的最大收入成正比。Deep & Koutris(2017)设计了名为 Qirana 的查询定价系统,该查询定价模型可以在无套利原则下对广泛的 SQL 查询(聚合查询)进行实时定价。由在线数据库提供数据的交易场景并不少见,查询定价可以使交易过程和结果更符合消费者实际需求,但是仍存在算法程度高、选择用于预先定价的视图难度大、重复求解、对存储空间要求高以及更新数据库的工作量大等问题(Koutris et al, 2015)。

5. 为机器学习模型定价。随着机器学习服务(machine learning as a service, MLaaS)行业不断发展,消费者可以直接购买机器学习模型而不是原始数据集。原始数据集和机器学习模型的定价主要有四个区别(Cong et al, 2021):首先,后者主要作为一个整体进行定价和销售,前者可以根据粒度进行调整;其次,由于机器学习模型训练的技术难度更高,对其实施差异化定价策略难度更大;再次,机器学习模型通常出于特定目的而设计,所以价值确定难度小于原始数据集;最后,模型市场中防止套利的难度更大。

Chen et al(2019)设计了机器学习模型的定价模型,认为模型准确性是价格的主要影响因素,卖方首先在整个原始数据集上训练一个最佳模型,然后通过向最佳模型的参数添加不同方差的高斯噪声来产生不同版本的机器学习模型,噪声方差的数值与模型的预期错误率成正比;当且仅当定价函数是单调的且相对于噪声方差的倒数具有次可加性时,满足无套利条件。Liu et al(2021)提出了一个更为完整的方案,包含数据所有者、模型购买者和经纪商。首先,数据所有者根据隐私敏感度和 Shapley 值指定他们想要的补偿函数;然后,模型购买者根据 Shapley 值的覆盖率和为保护隐私而添加到模型中的噪声来衡量模型的相对效用,提供他们愿意支付的价格函数;最后,经纪商同时考虑隐私补偿和模型噪声来建立具有不同 Shapley 值覆盖率和 DP 参数的模型定价机制。

## 四、数字技术在数据要素定价中的应用

### (一)机器学习对数据定价模型的优化

在数据要素的交易和定价中,机器学习可用于处理快速变化、大型复杂的数据集,通过在机器学习模型中输入真实数据来检验定价模型的有效性还能极大提高模型计算效率,实现动态定价。在算力支持下,机器学习对客户进行画像,优化数据定价机制。

1. 验证定价模型的有效性。首先,机器学习不仅能使用降维技术缩小定价模型中影响因子的范围,还能对因子有效性进行检验。例如,Harvey et al(2016)利用机器学习领域中的自助法检验了1967年以来发表的实证论文中金融资产定价因子的有效性,结果表明很多文献的研究结果存在错误。其次,机器学习可以处理复杂的函数,并重点关注函数对现实的模拟程度。考虑到影响因素之间的相互作用, Li et al(2013)认为数据定价应该采用非线性函数形式,而机器学习可以自动辨识因子之间的非线性定价结构,并给出预测结果。进一步地, Yang et al(2019)认为数据质量是影响数据产品定价的关键要素,为此构建了基于质量等级的效用函数,使用神经网络模型(ANNs)对真实数据集进行分类训练,证明了该复杂函数的有效性。最后,数据资产定价可以借鉴机器学习在金融资产定价领域中的应用,检验预期对资产价格的影响。Jegadeesh & Wu(2013)使用朴素贝叶斯法衡量了1995—2010年间美国金融公司的45860份公司年报文件中的情绪词汇,指出其中的积极和消极词汇会影响市场反应,进而影响金融资产价格。当使用收益法为数据资产

定价时,可以借鉴上述方法,比如通过朴素贝叶斯法检验市场预期是否会影响折现因子的大小,进而提高定价模型的准确性。

2. 实现定价模型动态调整。动态数据定价也被称为智能数据定价模型(SDP),是差异化定价模式的一个特例(Liang et al,2018)。将机器学习技术嵌入数据要素定价模型可以使产品价格根据数据价格影响指标的变化得到快速而准确的调整。例如,Tsai et al(2017)利用机器学习构建智能数据定价模型(TDP)应用了 TDP-TR 和 TDP-KNN 两种定价算法:前者采用动态定价方案,根据用户过去的支付意愿,引入惩罚函数与附加收益函数分别表示损失和收益,应用优化算法最大化消费者的目标效用函数来确定价格;后者采用 K 近邻算法,根据用户过去的网络使用情况来估计未来的数据使用量。Balasubramanian et al(2015)进一步指出,利用机器学习算法可以控制消费者的需求量,比如在使用量超过预先设定的固定数据配额时,系统会自动实施额外费用等补充计划。动态定价也存在不足,比如,动态定价造成的价格波动更大,影响消费者购买体验;而且更为精准的差异化定价可能引起顾客“不公平”等负面感知价值,导致购买意愿降低(Isabel et al,2021)。

## (二)区块链、智能合约和密码学技术的应用

目前的数据交换和共享都是基于中心化服务器的设计理念,存在数据所有权界定不清、数据所有者隐私泄露、交易透明度低等问题,这无疑加大了数据定价的难度,而区块链、智能合约和密码学技术的应用可以在一定程度上解决以上难题。

1. 区块链技术生成的时间戳。区块链将包含数据资产的区块的哈希值传递给时间戳服务器,时间戳服务器对哈希值和时间记录进行签名,生成时间戳,哈希值和时间戳成为数据资产存在的凭证,有效防止了数据资产的所有权界定不清的问题(Fallucchi et al,2021)。并且,前一个时间节点生成的时间戳和后一个时间节点生成的时间戳环环相扣,认证程度不断加深,有利于防止二次转售等对数据资产权利的不正当使用行为。时间戳可以实现数据资产尤其是社交媒体等产生的私人数据的可视化,为私人数据的交易和定价提供了凭证(Misue,2016)。

2. 区块链和智能合约。数据拥有者将数据发布和存储在区块中,使数据具有不可篡改性、可追溯性和安全性等特征(Wang et al,2018),可以保证自身权利和数据的安全,有利于提高平台交易的安全性和合规性。买方在访问区块链平台后,向数据提供者发出交易请求,双方签订智能合约,约定数据处理规则并协商同态加密密钥,完成数据产品的交易(Yue et al,2017)。区块链能够自动执行智能合约中的数据拍卖协议,从而防止数据竞价中的串通行为,减少交易争议。例如,Wei & Xiong (2021)提出一种基于区块链和智能合约的反串通数据拍卖机制,在智能合约中设计反合谋数据拍卖算法,通过密封的竞价、竞价掩码和竞价揭示实现反串通数据拍卖,并设置惩罚机制用于惩罚数据拍卖中的不当行为。Ke et al(2021)建立了一个双向匿名拍卖协议,以政策驱动的变色龙哈希和修订的可链接和可编辑的环形签名作为构建模块,保证竞标者出价的竞争性和匿名性,并允许每个参与者在拍卖协议结束时验证出价证明的有效性,进而减少交易中的争议。此外,在去中心化和匿名的交易环境中,智能合约对加密数据的访问和执行权限进行管理,买方在加密数据使用结束后通过智能合约将其返还提供方,智能合约对数据进行销毁,实现了数据所有权和使用权的分离。

3. 密码学技术。密码学领域的同态加密和非对称加密(Ma et al,2021)有利于解决数据交易中的隐私泄露问题。同态加密技术对隐私数据进行加密,智能合约对加密数据进行处理,处理后的结果解密后可以得出与原数据相同的分析结果,既满足了需求方的查询要求,又避免了原始数据的泄露。非对称加密技术可以对数据交易方的身份进行验证,明确了数据来源,又能防止数据伪造和数据篡改,保证了交易往来的合法性(Wang et al,2018)。

此外,智能化采集、云计算和物联网技术解决了海量数据的采集、存储和分析等面临的技术难题,降低了数据要素的重置成本,进而对数据要素定价产生间接影响。数据采集方面,企业内部经营数据和生产数据主要通过高性能内存计算设备上运行的数据库和数据仓库以及传感器等物联网设

备进行采集(Faroukhi et al,2020);外部数据主要通过爬虫技术、埋点检测技术、用户调研等方法进行采集(Xu et al,2017)。离线数据构建在开源的 HDFS 文件系统和 MapReduce 运算框架上,在线数据分析系统构建在云计算平台的 NoSQL 系统上(Yang,2021)。Balazinska et al(2011)提出,基于云的数据市场为企业、终端用户和应用开发者提供了“一站式购物”的机会,可以帮助用户节约搜寻和获取所需数据的成本。

## 五、结论与展望

数据是发展数字经济的关键生产要素,高效的数据要素配置可以加速数据成为经济增长的新动力、新引擎。本文在梳理数据要素定价的影响因素、定价原则和交易制度设计的基础上,对数据要素定价机制进行了评述。数据要素定价机制有三个显著特点:第一,数据所有权的概念没有定论,数据的隐私及其保护会影响数据所有者提供数据的意愿,进而影响数据定价,数据要素定价需要对数据的所有权和使用权进行拆分。第二,数据要素定价与场景高度相关是数据要素定价机制的最大特征,交易参与者可以利用市场结构和交易规则,从策略上影响竞争者的行为或估值;在制造、电商、金融等应用场景和交易场景中,优化定价模型、定价策略与设计更好的激励机制,可以提高市场化配置效率。所以,优化交易制度设计创造更多的交易场景是建立高效的数据要素定价机制的关键所在。第三,数字技术不仅改善和丰富数据的来源和安全性,还可以降低交易成本,促成交易,对数据要素定价的影响越来越大。

本文认为数据要素的定价涉及多个学科,不仅运用到价格理论、价值理论和运筹学等理论,还涉及电子商务、大数据管理、数据挖掘和机器学习等领域,应该采用跨学科研究方法。国内专门研究数据要素定价的文献凤毛麟角,已有的研究大多介绍数据定价方法和模型,对于不同定价方法的适用场景缺乏论证,很少有研究将理论模型与数据定价的实践紧密联系起来。数字技术在数据要素定价中的作用越来越重要,但国内外已有文献对这个问题的研究明显不足,这可能与交易机构对所用数字技术采取保密措施有关系。

数据要素估值和定价对于数据要素市场建设和市场化配置至关重要,是数字经济研究中的一个热点问题和重点问题。未来需要进一步强化以下方面:第一,加强数据要素定价的基础理论研究,探索构建基于场景的数据定价理论体系,加强数据要素交易模式、交易机制、产业链、定价指标的研究,从理论上建立涵盖数据确权、算法定价、收益分配的数据交易全生命周期的价格体系。第二,数据要素的定价方法应该与数字技术的应用更好地结合起来,深入分析数字技术在定价中的作用机制,利用新技术手段构建自动定价和动态定价模型。第三,完善大宗数据资源交易平台的交易规则,探索根据使用场景和数据购买者设定个性化的交易合同,细化研究不同层次市场和不同交易场所的数据要素价格形成机制。第四,加强数据要素市场的会计和审计研究,提升数据定价的透明度和数据市场的效率。

### 参考文献:

- 熊巧琴 汤珂,2021:《数据要素的界权、交易和定价研究进展》,《经济动态》第2期。
- Adler, R. et al(2016), “The valuation and pricing of information assets”, *Pacific Accounting Review* 28(4): 419—430.
- Agarwal, A. et al(2018), “A marketplace for data: An algorithmic solution”, ArXiv Working Paper, No.1805.08125.
- Arribas-Bel, D. et al(2021), “Open data products: A framework for creating valuable analysis ready data”, *Journal of Geographical Systems* 23(10):1—18.
- Bajari, P. et al(2019), “The impact of big data on firm performance: An empirical investigation”, *AEA Papers and Proceedings* 109:33—37.
- Balasubramanian, S. et al(2015), “Pricing information goods: A strategic analysis of the selling and on-demand pricing mechanisms”, *Marketing Science* 34(2):218—234.
- Balazinska, M. et al(2011), “Data markets in the cloud: An opportunity for the database community”, *Proceedings*

of the VLDB Endow 4(12):1482–1485.

- Belleflamme, P. (2018), “Economic models of data: A complex relation between supply and demand”, *Economics of Privacy*, Working Paper, Aix-Marseille School of Economics.
- Bergemann, D. et al(2018), “The design and price of information”, *American Economic Review* 108(1):1–48.
- Bourreau, M. et al(2018), “The value of consumer data in online advertising”, *Review of Network Economics* 16(3):269–289.
- Cao, X. et al(2017), “An iterative auction mechanism for data trading”, *2017 IEEE International Conference on Acoustics*, pp. 5850–5854.
- Chen, L. et al(2019), “Towards model-based pricing for machine learning in a data marketplace”, *Proceedings of the 2019 International Conference on Management of Data*, pp. 1535–1552.
- Chen, Z. et al(2018), “Research on the auction strategies and pricing of big data”, *Journal of University of Science and Technology of China* 48(6):486–494.
- Choudhary, V. et al(1998), “Economic benefits of renting software”, *Journal of Organizational Computing and Electronic Commerce* 8(4):277–305.
- Coggeshall, J. et al(2011), “System and method for an individual data marketplace and monetization”, Patent, No. US20110295694 A1.
- Cong, Z. et al(2021), “Data pricing in machine learning pipelines”, ArXiv Working Paper, No. 2108.07915.
- Cummings, R. et al(2015), “Accuracy for sale: Aggregating data with a variance constraint”, *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pp. 317–324.
- Deep, S. & P. Koutris(2017), “QIRANA: A framework for scalable query pricing”, *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 699–713.
- Epstein, J. & A. J. Marcus(2003), “Economic analysis of the reasonable royalty: Simplification and extension of the Georgia-Pacific factors”, *Journal of the Patent & Trademark Office Society* 85:555–583.
- Fallucchi, F. et al(2021), “Blockchain framework in digital government for the certification of authenticity, time stamping and data property”, *Proceedings of the Hawaii International Conference on System Sciences*, pp. 285–295.
- Faroukhi, A. Z. et al(2020), “Big data monetization throughout Big Data Value Chain: A comprehensive review”, *Journal of Big Data* 7(1):1–22.
- Fontana, G. et al(2020), “Ensuring that the NHS realises fair financial value from its data”, *Lancet Digital Health* 2(1):10–12.
- Fruhvirth, M. et al(2020), “Discovering business models of data marketplaces”, *Proceedings of the 53rd Hawaii International Conference on System Sciences*, pp. 5738–5747.
- Ghosh, A. & A. Roth(2015), “Selling privacy at auction”, *Games and Economic Behavior* 91:334–346.
- Ghorbani A. & J. Y. Zou(2019), “Data Shapley: Equitable valuation of data for machine learning”, ArXiv Working Paper, No. 1904.02868.
- Goldberg, A. et al(1999), “Competitive auctions and digital goods”, *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 735–744.
- Gu, Y. et al(2021), “Data brokers co-opetition”, Social Science Research Network Working Paper, No. 3308384.
- Gundepudi, P. et al(2001), “Forward versus spot buying of information goods”, *Journal of Management Information Systems* 18(2):107–131.
- Günther, W. A. et al(2017), “Debating big data: A literature review on realizing value from big data”, *Journal of Strategic Information Systems* 26(3):191–209.
- Harmon, R. et al(2009), “Pricing strategies for information technology services: A value-based approach”, *Proceedings of the 42nd Hawaii International Conference on System Sciences*, pp. 1–10.
- Harvey, R. et al(2016), “... and the cross-section of expected returns”, *Review of Financial Studies* 29(1):5–68.
- Heckman, J. R. et al(2015), “A pricing model for data markets”, *Proceedings of the iConference 2015*, pp. 1–12.
- Hynes, N. et al(2018), “A demonstration of sterling: A privacy-preserving data marketplace”, *Proceedings of the VLDB Endowment* 11(12):2086–2089.

- Ichihashi, S. (2021), “Competing data intermediaries”, *RAND Journal of Economics* 52(3):515–537.
- Immonen, A. et al(2015), “Evaluating the quality of social media data in big data architecture”, *IEEE Access* 3:2028–2043.
- Isabel, P. R. et al(2021), “Does it matter who gets a better price? Antecedents and consequences of online price unfairness for advantaged and disadvantaged consumers”, *Tourism Management Perspectives* 40(1):1–14.
- Jegadeesh, N. & D. Wu(2013), “Word power: A new approach for content analysis”, *Journal of Financial Economics* 110(3):712–729.
- Jia, R. et al(2019), “Towards efficient data valuation based on the Shapley value”, ArXiv Working Paper, No. 1902.10275.
- Jiao, Y. T. et al(2018), “Profit maximization mechanism and data management for data analytics services”, *IEEE Internet of Things Journal* 5(3):2001–2014.
- Jin, H. et al(2015), “Quality of information aware incentive mechanisms for mobile crowd sensing systems”, *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 167–176.
- Jones, C. I. & C. Tonetti (2020), “Nonrivalry and the economics of data”, *American Economic Review* 110(9):2819–2858.
- Jung, K. & S. Park(2019), “Privacy bargaining with fairness: Privacy-price negotiation system for applying differential privacy in data market environments”, 2019 *IEEE International Conference on Big Data*, pp. 1389–1394.
- Ke, H. (2021), “BA<sup>2</sup>P: Bidirectional and anonymous auction protocol with dispute-freeness”, *Security and Communication Network*, pp. 1–12.
- Kitchin, R. & T. P. Lauriault(2015), “Small data in the era of big data”, *Geo Journal* 80(4):463–475.
- Koutris, P. et al(2013), “Toward practical query pricing with query market”, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 613–624.
- Koushanfar, F. et al(2019), “Deep learning on private data”, *IEEE Security & Privacy* 17(6): 54–63.
- Koutris, P. et al(2015), “Query-based data pricing”, *Journal of the ACM* 62(5):1–44.
- Koutroumpis P. & L. Aija(2013), “Understanding the value of (big) data”, 2013 *IEEE International Conference on Big Data*, pp. 38–42.
- Koutroumpis, P. et al(2020), “Markets for data”, *Industrial and Corporate Change* 29(3):1–16.
- Közgazdász, F. (2017), “Valuing methods for information assets: Literature review”, *Forum on Economics and Business* 20(130): 54–78.
- Li, C. et al(2017), “A theory of pricing private data”, *Communications of the ACM* 60(12):79–86.
- Li, J. & S. Gao(2021), “Platform economy and internet platform monopoly”, *BCP Business & Management* 13:160–163.
- Li, M. et al(2013), “Optimal versioning strategy for information products with behavior-based utility function of heterogeneous customers”, *Computers & Operations Research* 40(10): 2374–2386.
- Li, X. B. & S. Raghunathan(2014), “Pricing and disseminating customer data with privacy awareness”, *Decision Support Systems* 59(3):63–73.
- Li, W. et al(2019), “Value of data: There’s no such thing as a free lunch in the digital economy”, Research Institute of Economy, Trade and Industry Working Paper, No. 19022.
- Liang, F. et al(2018), “A survey on big data market: Pricing, trading and protection”, *IEEE Access* 6:15132–15154.
- Lin, B. R. & D. Kifer(2014), “On arbitrage-free pricing for general data queries”, *Proceedings of the VLDB Endowment* 7(9):757–768.
- Liu, J. (2021), “Dealer: An end-to-end model marketplace with differential privacy”, ArXiv Working Paper, No. 2003.13103.
- Liu, K. et al(2019), “Optimal pricing mechanism for data market in blockchain-enhanced Internet of things”, *IEEE Internet of Things Journal* 6(6):9748–9761.
- Liu, Y. et al(2021), “Blind polynomial evaluation and data trading”, in: K. Sako & N. O. Tippenhauer(eds), *Applied Cryptography and Network Security*, Springer.
- Luong, N. C. et al(2016), “Data collection and wireless communication in internet of things (LoT) using economic analysis and pricing models: A survey”, *IEEE Communications Surveys & Tutorials* 18(4):2546–2590.
- Ma, J. et al(2021), “Privacy-preserving federated learning based on multi-key homomorphic encryption”, ArXiv Working Paper, No. 2104.06824.



- Ma, Y. (2019), "Monopoly and competition in the markets for information", Social Science Research Network Working Paper, No. 3505101.
- Mao, W. et al(2019), "Pricing for revenue maximization in IoT data markets: An information design perspective", 2019 *IEEE Conference on Computer Communications*, pp. 1837—1845.
- Mao, Y. et al(2017), "A strategic bargaining game for a spectrum sharing scheme in cognitive radio-based heterogeneous wireless sensor networks", *Sensors* 17(12):27—37.
- Mehta, S. et al(2021), "An approximation scheme for data monetization", Social Science Research Network Working Paper, No. 3808875.
- Meierhofer, J. et al(2019), "Data products", in: M. Bruschler et al(eds), *Applied Data Science*, Springer.
- Misue, K. (2016), "Visualization of social data with timestamps", *Journal of the Visualization Society of Japan* 36(141):2—8.
- Moody, D. & P. Walsh(1999), "Measuring the value of information: An asset valuation approach", *ECIS'99: Proceedings of Seventh European Conference on Information Systems*, pp. 1—17.
- Moriarty, J. (2021), "Why online personalized pricing is unfair", *Ethics and Information Technology* 23(3):1—9.
- 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.
- Muschalle, A. et al(2013), "Pricing approaches for data markets", in: M. Castellanos et al(eds), *Enabling Real-Time Business Intelligence*, Springer.
- Pei, J. (2020), "A survey on data pricing: From economics to data science", ArXiv Working Paper, No. 2009.04462.
- Pérez-Pons, M. E. et al(2019), "Towards financial valuation in data-driven companies", *Journal of Computer Science and Technology* 12(2):28—33.
- Shapiro, C. & H. R. Varian(1998), "Versioning: The smart way to sell information", *Harvard Business Review* 76(6):106—114.
- Shapley, L. S. (1953), "A value for n-person games", in: H. W. Kuhn & A. W. Tucker (eds), *Contributions to the Theory of Games*, vol. 2, Princeton University Press.
- Short, J. E. & S. Todd(2017), "What's your data worth?", *MIT Sloan Management Review* 58(3):17—19.
- Sofia, D. S. & A. S. Edward(2020), "Auction based game theory in cognitive radio networks for dynamic spectrum allocation", *Computers & Electrical Engineering* 86:106734.
- Son, J. C. et al(2021), "A study on the trading price estimation algorithm for healthcare transaction data", *Research Square*, forthcoming.
- Spiekermann, M. (2019), "Data marketplaces: Trends and monetisation of data goods", *Intereconomics* 54(4):208—216.
- Sukumar, S. R. et al(2015), "Big data in health care: How good is it?", *International Journal of Health Care Quality Assurance* 28(6):1—9.
- Tsai, Y. C. et al(2017), "Time-dependent smart data pricing based on machine learning", in: M. Mouhoub & P. Langlais(eds), *Advances in Artificial Intelligence*, Springer.
- Tucker, D. S. & W. Hill(2014), "Big mistakes regarding big data", Social Science Research Network Working Paper, No. 2549044.
- Varian, H. (2018), "Artificial intelligence, economics, and industrial organization", NBER Working Paper, No. 24839.
- 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, Communications in Computer and Information Science*, Springer.
- Wang, H. et al(2018), "Blockchain challenges and opportunities: A survey", *International Journal of Web and Grid Services* 14(4):352—375.
- Wei, X. & B. R. Nault(2014), "Monopoly versioning of information goods when consumers have group tastes", *Production and Operations Management* 23(6):1067—1081.
- Wei, X. & L. Xiong(2021), "Anti-collusion data auction mechanism based on smart contract", *Information Sciences* 555(1):386—409.

- Wu, C. et al(2019), “Cloud pricing models: Taxonomy, survey, and interdisciplinary challenges”, *ACM Computing Surveys* 52(6):1–36.
- Xu, C. Z. et al(2020), “Data pricing for blockchain-based car sharing: A Stackelberg game approach”, 2020 *IEEE Global Communications Conference*, pp. 1–5.
- Xu, L. et al(2017), “Dynamic privacy pricing: A multi-armed bandit approach with time-variant rewards”, *IEEE Transactions on Information Forensics & Security* 12(2):271–285.
- Yang, D. et al(2012), “Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing”, *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, pp. 173–184.
- Yang, J. (2021), “Big data privacy protection technology”, 2021 *International Conference on Artificial Intelligence and Information Technology*, pp. 1–5.
- Yang, J. et al(2019), “Big data market optimization pricing model based on data quality”, *Complexity*, 2019(2):1–10.
- Yu, G. W. et al(2011), “The assessment of product information quality in digitization manufacturing processes”, *Advanced Materials Research* 411:419–423.
- Yu, H. & M. Zhang(2017), “Data pricing strategy based on data quality”, *Computers & Industrial Engineering* 112:1–10.
- Yue, L. et al(2017), “Big data model of security sharing based on blockchain”, *Proceedings of the 2017 3rd International Conference on Big Data Computing and Communications*, pp. 117–121.
- Zhang, M. & F. Beltran(2020), “A survey of data pricing methods”, *Social Science Research Network Working Paper*, No. 3609120.
- Zheng, Z. et al(2017), “An online pricing mechanism for mobile crowdsensing data markets”, *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 1–10.

## Research Progress on the Pricing Mechanisms of Data

OUYANG Rihui<sup>1</sup> DU Qingqing<sup>2</sup>

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

2. Beijing Language and Culture University, Beijing, China)

**Abstract:** The pricing mechanism of data elements is an important part of the construction of the data element market. It is an institutional arrangement for buyers and sellers to determine the price of data transactions under the constraints of systems, scenarios, and technologies. This article firstly clarifies that the object of data element pricing is data products and services with the attributes of production factors. Then it summarizes the factors that affect the pricing of data products, such as cost, value and scenario, and summarizes the general and specific principles of data factor pricing. In the pricing mechanism of data elements, institutional design is the key to the pricing of data elements, and different pricing methods have been derived from scenarios. Then this article composes the roles of transaction participants, market structure, transaction mechanism and transaction rules in the design of transaction system, reviews the impact of data transaction scenarios and application scenarios on pricing, and summarizes the main methods and models of pricing in different transaction scenarios. Finally, it presents an overview of the literature on the impact of digital technologies such as machine learning, blockchain, smart contracts, and cryptography on data pricing.

**Keywords:** Data Elements; Data Products; Pricing Mechanism; Trading System; Digital Technology

(责任编辑:李仁贵)

(校对:刘洪愧)