数据要素市场的价格规律：来自上海数据交易中心的探索研究

尹文怡[1]　窦一凡[1]　汤奇峰[2]　黄丽华[1]

（1. 复旦大学管理学院，上海 200433；
2. 上海数据交易中心有限公司，上海 200436）

摘　要　数字经济时代，企业内快速积累的数据正在承担起新型生产要素的角色。然而，长期以来，数据在组织之间的流动缺乏必要市场支持，造成了普遍存在的“数据孤岛”现象，并制约了数据要素的有效配置。本文基于上海数据交易中心的前期经验和交易数据开展探索性研究。分析的核心结论是数据提供方的权威性和数据独特性对于成交价格具有决定性的影响；此外，不同场景下的数据价格在分布特征上存在显著不同。本文随后结合这些探索性的结论对于数据市场的健康发展进行了讨论，对我国当前各地兴建和完善数据市场的管理实践提供了一定的数据支持和方向指引。

关键词　数据交易，数据定价，生产要素，探索分析

中图分类号　F276.6

1　引言

在数字经济和人工智能时代，各类组织积累的海量数据已经成为科技创新和社会治理的核心动力和战略资源。然而，组织内“数据孤岛”现象的普遍存在，使得数据在组织之间的流动和融合往往受到很大限制，无法实现数据的市场化配置和价值挖掘。在这一背景下，我国于 2020 年 4 月提出《关于构建更加完善的要素市场化配置体制机制的意见》，将数据的战略作用提升到新型生产要素的高度，鼓励各级政府和组织积极探索和推动数据的市场化流通机制和交易方法。2021 年发布的“十四五”规划同样明确提出要“健全数据要素市场”。与土地、资本、劳动等传统生产要素相比，数据要素以其独特的技术、经济和市场特性，对数字经济时代国家经济社会发展具有更加关键的作用，也是我国提升数字经济全球竞争力的重要抓手。各地政府对此也高度重视，北京国际大数据交易所于 2021 年 3 月正式上线，其他交易所也在紧锣密鼓筹备，共同的目标是建立数据要素交易流通体系，推动建立市场化运作的数据交易市场，并为长期的数据资产化和数字化转型服务。

然而，自 2014 年贵阳大数据交易所建立以来，我国在数据交易市场的前期摸索过程举步维艰，交易规模的目标并未如期实现。在这当中，一个核心的难点在于市场供需双方对于数据价格仍未形成共识。例如，在本文研究团队前期的研究[1]调研中，18 个参与数据交易的企业中有 14 家都表示了类似于“目前数据价格不明确不成熟，更倾向于和另一方协商从而形成最终的价格结果”的态度。因此，如何总结经验并深入认识数据交易的特殊挑战并寻找应对策略，从而充分激活数据市场、激发潜在价值，是下一阶段我国健全数据要素市场中必须面对的重要议题。针对这一关键问题，本文得到了来自

通信作者：窦一凡，复旦大学管理学院教授、博士生导师，E-mail：yfdou@fudan.edu.cn。

国内当前处于领先水平的数据交易平台——上海数据交易中心的大力支持并开展相关研究。上海数据交易中心为本文提供了一套独特数据——2016 年 12 月至 2019 年 8 月的平台交易历史数据抽样样本——以开展探索研究。根据本文作者的不完全搜索，这也是在数据交易市场这一方向上，首篇基于实际交易数据所开展的研究，因此本文也主要致力于探索和寻找数据交易市场中的研究突破点——何种因素可能影响了数据市场的交易价格与交易量。

数据交易模式和运营机制的改进和提升，关键在于认识数据交易双方在从事交易过程中的主要顾虑并探索应对机制。在本文研究团队的前期访谈中[1]发现，数据买方普遍对于数据“质量”最为关注：18 位被访谈专家中共 9 位在访谈中提及“厂商权威性”，5 位提及“数据独特性”，一位来自物联网企业的买方专家坦言“除了考虑数据所处行业与应用场景与我们需求的匹配程度，如果数据卖方不够权威，我们不愿意与其合作或购买他们的数据”；另一位来自应用数据科学与人工智能提供解决方案企业的买方专家称“在购买数据时，数据的独家性（即数据独特性）会是除了数据维度、应用匹配度以外考虑的重要因素”。从访谈结果来看，本文认为对于实际的交易过程，数据“质量”首先体现在来源——即数据是否来自权威供方以及数据本身是否独特。因此，本文从这两个视角入手，探索数据权威性与数据独特性如何影响交易价格与交易量。在此基础上，本文试图从实证结果出发进一步探究数据交易市场目前存在的问题，并为数据交易平台的长期发展提供相应建议。

在研究内容的组织上，本文将首先对相关文献进行梳理与总结；在此基础之上，本文将进一步对数据供方权威性与数据独特性这两个关键变量的理论依据进行深入阐释；在明确理论依据的前提下，本文将对样本进行描述性统计，紧接着对数据进行实证检验，并开展一系列稳健性检验；最后，通过回归结果对数据交易市场“劣币驱逐良币”的演进过程做出了推测，并尝试为今后数据交易中如何激活数据市场、激发潜在价值提供相应管理启示。

虽然本文只是针对数据交易市场的探索性研究，但依然从不同的角度对于已有研究形成了拓展，并对于未来在这一领域开展更加深入的研究工作具有启发意义。首先，从理论角度看，数据交易和数据市场的研究依然非常稀缺，而已有经济学文献大多也是从数据的隐私保护[2, 3]等角度切入。相比之下，本文从数据的独特性（即数据的有限排他性）与数据的供方权威性（即数据的信息不对称程度）两个角度出发，探讨了这两点关键因素对数据交易价格的影响。从应用价值来看，本文结合来自上海数据交易中心的实际数据进行了较为全面的探索，揭示和证实了现有数据交易和数据市场中的种种困境，并对于未来数据市场的实践进行了较为深入的讨论，具有一定的实践指导意义。

2 文献综述与研究假设

2.1 文献综述

相比于传统商品，数据作为流通和交易的对象具有许多鲜明的差异，这在已有的文献当中已经有不少的讨论。例如，数据具有独特的规模经济与范围经济效应（economies of scale/economies of scope）[4]，这是因为企业投入建设基础数据设备的固定成本较高，但随后单位数据累积的边际成本较低，且随着数据的积累，企业能以更低的成本获得更大的收益与价值；同时，数据具有网络效应与互补性[5]：即在算法与技术的支持下，不同方面的数据集能合并补充，发挥网络效应并进一步增强人工智能算法，从而发挥更大的规模经济与范围经济效应；此外，数据具有非竞争性与有限排他性[6]：数据不同于石油等其他资源，不具有消耗性——即其能被无限需要的用户使用且不会对原有数据造成任何损失，但大部分企业也是顾虑于此，从而为保证自身数据资源的优势而囤积数据。数据在不同情境下具有不同的外

部性[7]：例如，在农业活动中对于肥料与驱虫剂效果的数据收集有利于所有种植业发展[8]（正外部性），但互联网企业对私人数据的采集会造成对个人隐私的侵犯（负外部性）。数据作为产品呈现出的形式多样：数据可以呈现为原始数据的形式——其价值最低、中间物的形式——即经过加工处理能够对企业经营活动产生效益的数据产品[9]，或仅仅是企业进行经济活动的副产品——这类数据的采集成本往往可以忽略不计[10]。

上述的各种特殊性导致数据无法像普通商品一样经由市场的自动调节形成价格，从而通过“看不见的手”来实现资源的有效配置。Falck 与 Koenen 指出了纯粹的市场机制无法保证数据资源的有效分配[7]——在此情况下各个国家针对不同情况对数据市场进行干预尤为重要。综合现有文献，现有数据市场交易举步维艰的原因[11]主要包括以下几点。其一，信息不对称：主要体现为数据质量的不确定性导致买卖双方无法建立信任，这一不确定性还体现在数据的价值往往在训练和处理之前无法确定，且其价值往往还取决于与其他数据的综合分析。其二，供应非透明化：目前的数据交易大多集中在场外交易或私下交易，没有统一的集中市场与数据中介来了解供需双方的潜在需求和潜在供应并促成交易。其三，交易成本对中小企业较高：数据交易不同于普通的商品交易，目前仍非标准化，因而对于中小企业而言签订契约的难度和成本尤其大，同时中小企业可能需要负担更大的监管成本，在资金压力下他们往往不会雇佣专业人士来观察市场潜在的需求并提供自身宝贵的数据资源。其四，负外部性的存在即已有市场力量对新进入者的阻碍：垂直整合的数据分析公司往往将数据购买者视为潜在的竞争者并在战略上阻碍其获得自己的数据，以此阻碍了更多的数据买方进入市场。

针对上述问题，Falck 与 Koenen 经过大量调研后指出，数据的市场价格无法自主形成，主要是因为数据本身仍是高度非标准化的[7]。因此，实际的数据交易价格往往取决于买方的支付意愿。后续的研究成果则集中于数据外部性与数据价格之间的联系，并以此为出发点对数据交易市场进行进一步的机制设计研究。例如，Acemoglu 等在最新发表的文献中建立了理论模型并证明了数据的某种负外部性（当一个用户在平台分享他的数据时会同时揭露其他用户的相关隐私数据）最终降低了数据的价格，而被压低的价格又进一步导致了更加过度的数据共享[12]。作者认为引入数据市场与平台之间的竞争并不能解决这一问题甚至可能进一步降低福利，因此提出了一种中介数据共享的方案以提高效率。又如，有文献从另一侧面考虑了企业在售卖数据给竞争对手时面临的负外部性[13]。这篇文章认为负外部性的存在增加了垄断性数据卖方的收入，并通过拍卖设计发现使得数据卖方福利和收入最大化的机制设计取决于其私人信息及外部性。

综合来看，现有文献对数据价格与市场机制的探讨局限于理论模型的建立与推导，仍没有对现实数据交易的实证性研究。Koutroumpis 等[14]在对现实数据市场的综述性文章中回顾了数据交易的历史发展与机构背景，比较了数据交易与想法交易的异同，并从 Roth 于 2002 年[15]和 2008 年[16]分别提出的机制设计视角出发，探讨了未来数据交易市场发展的两种可能性——少控制的大市场与全部控制的小市场。相比之下，本文从实证上拓展了现有研究的不足，用交易数据揭示与证实了现有数据交易过程中的种种困境并进一步探究了数据产品价格形成的机制。理论上，本文突破了已有文献对数据非竞争性与外部性这两大因素对数据价格与交易均衡形成的影响，从数据的独特性（即有限排他性）与数据的供方权威性（即信息不对称程度）两个角度出发，探讨了其对数据交易过程中价格形成机制的影响。

2.2 研究假设

2.2.1 数据独特性

当一份数据在市场当中重复出现并被不同的卖家售卖时，该数据的独特性会下降。对于那些在市

场上重复出现的数据，可能属于两种情况之一：一种是可能违背数据权属约定的二手倒卖；另一种是该数据公开且易得，如通过简单爬虫或者对政府公开数据搜集并进行不同的加工即可获得。其中，第二种情况更接近于相似产品的定价策略问题，利用传统的定价方法，如纳什谈判等手段可以求解[17]得到，数据供方的讨价还价能力越高且数据需方的最低支付意愿越高，数据价格越高。

然而，对于第一类二手倒卖的情况，我们依据 Coase 提出的二手市场交易理论构建如下的解释机制：大部分企业为保证自己的数据优势，往往会选择囤积数据造成数据孤岛，这是由于原始数据可复制性强，仅仅具备有限排他性，除非有法律保障，卖方在出售独有数据后很难防止买方在使用后复制并转让数据——因此在没有有效契约保障的情况下任何买家都可以购买数据后实现重新转卖，从而形成数据交易的二手市场[18]。Rust 于 1986 年提出，二手市场上的旧产品越多意味着市场上的替代品越多，垄断企业所面临的需求弹性越大，一手市场中垄断企业的利润空间将被压缩[19]。如果同一种数据产品在市场上存在更多的卖家，即意味着市场上对原始数据的替代品也越多，在原有垄断数据卖方无法控制交易成本的情况下，其利润将被二手市场卖方的竞争压缩。

另外，由于买方可以等待“二手数据”出现在市场上，数据的这种易复制性会导致交易价格的快速下跌。有趣的是，数据本身无形，并不会随着时间的推移而发生物理损耗，因此可被视作一类特殊的超级耐用品——而这与“科斯猜想”的假设颇为一致。Coase 认为对于耐用品而言，销售价格通常存在“时间不一致性”，其后果是若消费者具有理性预期，则会等待未来更低的市场价格，导致垄断厂商的收取价格也会最终降低到边际成本[18]。一旦出现一个数据买方，其他潜在买方将预期该买方在使用完毕后即会转手出售——换言之，即使市场上最初只有一份数据，买方也能够预期到最终可能出现多个卖方提供同一份数据，且他们相互存在替代关系。如果存在更多的卖方，理性买方总是预期能在未来等待到更低的价格，从而使得买方的支付意愿更低，进一步降低该数据产品的成交价格。另外，由于数据产品的复制成本极低且成本极小，即便考虑买方可能对数据时效性即需要尽快获取数据的情况，该数据产品的价格也能在短时间内经不断复制与重新出售降低到边际成本上。因此，越不容易复制即越独特的数据产品价格越高。

但值得注意的是，高价格与高独特性在该市场上可能并不意味着高质量数据，对于高度同质化（即容易被复制）的数据产品而言，由于数据产品的复制门槛低，被复制的数据少甚至无人复制往往意味着数据质量低或无法满足买方需求，因而无人愿意继续复制该数据并出售，因此我们猜测数据的独特性反而与销售总量与销售金额呈负相关关系。

由此我们可得以下三项假设：

H_{1a}：数据产品的独特性与其单价呈正相关关系；

H_{1b}：数据产品的独特性与其销售总量呈负相关关系；

H_{1c}：数据产品的独特性与其销售总金额呈负相关关系。

2.2.2 数据权威性

Shapiro[20]、Allen[21]与 Klein 和 Leffler[22]的理论研究表明，公司的个体声誉基于过去公司产出的产品质量，且会影响消费者对公司未来产品质量的预期，从而调整该产品的市场价格。由于数据产品在出售之前存在事前信息不对称，买方无法在使用数据之前明确知晓数据的真正价值，因此买方很可能依赖数据供方权威与否对数据产品的质量形成预期，从而影响其支付意愿，进一步影响市场交易价格。

进一步地，Tirole 在 1996 年提出了集体声誉的概念[23]。集体声誉指消费者对所有单个卖方个体声誉感知的加总。在数据交易行业中，买家对某一数据质量的预期也可能基于数据交易行业整体过去提

供数据产品的平均质量情况。持有资源基础理论认为声誉是能为企业提供可持续优势的宝贵财富[24]。基于数据平台交易的相对匿名性，买卖双方不可接触性，在对市场平均情况毫不知情的情况下首次购买数据时，买方将优先选择权威供方提供的数据产品。也因此，数据供方的权威性在平台交易市场中尤为重要。与此同时，也有研究表明卖家的绝对声誉水平与买家对该卖家商品的支付意愿之间存在正向关系[25]。这为产品提供了溢价的可能性。因此，数据供方是否对于该数据具有足够权威很可能影响数据产品的成交价格与成交量。

H_{2a}：数据卖方的权威性与该数据在交易中的成交价格呈正相关关系；

H_{2b}：数据卖方的权威性与该数据在交易中的销售总量呈正相关关系；

H_{2c}：数据卖方的权威性与该数据在交易中的销售总金额呈正相关关系。

3 数据介绍与描述性统计

3.1 数据来源

本文得到了来自上海数据交易中心的大力支持。作为数据的提供方，上海数据交易中心有限公司于 2016 年 4 月 1 日正式成立，是一家经上海市人民政府批准，上海市经济和信息化委员会、上海市商务委员会联合批复成立的国有控股混合所有制企业。上海数据交易中心的发起成立单位包括上海市信息投资股份有限公司、中国联合网络通信集团有限公司、中国电子信息产业集团有限公司、申能集团有限公司、上海仪电控股集团公司、上海晶赞科技发展有限公司、万得信息技术股份有限公司、万达信息股份有限公司、上海联新投资管理有限公司等多家国资和民营企业。目前，上海数据交易中心流通的数据覆盖各行业的各类结构化和非结构化数据，中心的主要收入来源则是从数据供方的数据销售收入中提成。由于交易中心自身背景所赋予的公共属性，几年来交易中心所设定的提成的比例较小，并且在许多数据交易培育期的行业推行了减少和免除提成等优惠，因此可以认为平台在提成方面的收费比例设定对于市场最终的成交价格影响可以忽略不计。

由于采用开放市场形式进行的数据交易在全球范围内目前都属于起步阶段，市场的交易数量和频率无法与其他成熟的电子市场直接对比，因此我们选择了 3 年作为时间窗口，自 2016 年 12 月至 2019 年 8 月从上海数据交易中心的交易记录中抽取了 6 222 条交易记录。在随机抽取的方法中，我们的依据是随机抽取 38 家卖方（约为市场卖家数量的 1/10），并遍历这些卖家在市场中提供的总计 113 个数据产品，并提取在时间窗口内所有购买 113 项产品的买家共 64 家。每条交易记录包含了交易编号、结算日期、数据类型、交易品名称、数据单价、供需方名称、计费方式、计费数量、佣金比例、需方结算、供方金额等信息。针对 6 222 条数据，我们进行了必要的数据清洗工作，去除了计费数量为零或单价为零的单条记录，最终得到了 6 042 条交易记录。根据上海数据交易中心的说明，这些数据记录的价格是最终成交价格。以下，我们将围绕这些数据来进行描述性统计，并尝试结合上述的理论假设来构建回归模型加以验证。无疑这一数据规模依然较小，但对于针对这一新兴市场的研究依然具有探索意义。

3.2 描述性统计与回归模型构建

受限于数据较为稀疏，本文并未构建专门的时间序列数据，而是将三年时间窗口中的所有交易数据视作截面数据进行回归分析。为了尽量考虑时间带来的影响，我们在后续的稳健性分析中增加了时间固定效应。另外，受限于我们所能够获得的数据产品样本量较小，仅有 113 个。考虑到这些数

据方面的约束，为了尽可能挖掘这批数据的价值，我们选择增加研究的颗粒度，以单笔交易作为回归分析的对象。

首先，我们对主要变量（表 1）进行变量描述[①]，并在表 2 中进行相关性矩阵分析。结果表明，各个变量数值合理，并且除数据买方性质与数据行业之间存在一定相关性外（购买大量数据的大买方往往购买的是广告类数据），其他自变量之间不存在高度相关。

表 1　变量描述

变量名称		变量标记	变量含义	变量类型	变量说明
因变量	数据产品价格	Price	某一数据产品的价格	数值变量	数据产品交易价格/元
	数据产品销量	Volume	某一数据产品的销量	数值变量	数据产品交易个数/条
	数据产品销售额	Sales	某一数据产品的销售额	数值变量	数据产品销售金额/元
自变量	数据行业	Industry	数据产品所处行业	0~1 变量	征信类数据为 1，广告类数据为 0
	数据独特性	Uniqueness	某一数据产品独特与否	0~1 变量	若某一数据产品的供方个数只有一个则定义为 1，而存在多个供方则定义为 0
	数据权威性[1)]	Authority	某一数据产品权威与否	0~1 变量	数据产品只存在一手权威卖家则原始卖家定义为 1，若存在转手卖家则定义为 0
控制变量	数据买方性质	Dummy_Hoarder	某一数据产品是否存在大买方	0~1 变量	大买方定义为平均购买金额远超其他[2)]且购买数据产品个数远超其他的买方，存在这样买方的数据产品定义为 1，不存在的定义为 0
	数据需求	Dummy_Demand	某一数据产品单次交易的需求量	0~1 变量	每一笔交易的需求量，大于中位数的定义为 1，反之为 0
	供方固定效应	Supplier	某一数据产品的供应方	类别变量	某一具体数据产品的供应方
	数据查询方式	Dummy_Search_Type	某一数据产品的查询方式	0~1 变量	按查询收费的定义为 1，按查得收费的定义为 0
	时间因素	Year_Online	某一数据产品每一笔交易的发生时间	类别变量	每一笔交易的发生时间，2017 年定义为 0，2018 年定义为 1，2019 年定义为 2

1）我们首先聘请了 5 位上海数据交易中心专家独立为我们编码每个产品的权威供方，若该产品对应的供方是原始数据提供者则标记权威性为 1，若为二手转卖供方则标记权威性为 0，我们核对了 5 位专家的编码结果并聘请相关数据买方确认该编码的稳健性

2）其中三家企业的平均购买金额与个数在数量级上远超其他买家约 10^6 倍

表 2　变量相关系数矩阵

变量	样本量	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）	（9）
（1）数据产品价格	6 042	1.000								
（2）数据产品销量	6 042	−0.144***	1.000							
（3）数据产品销售额	6 042	−0.061***	0.696***	1.000						
（4）数据行业	6 042	0.251***	−0.519***	−0.236***	1.000					
（5）数据独特性	6 042	0.029*	0.132***	0.071***	−0.254***	1.000				

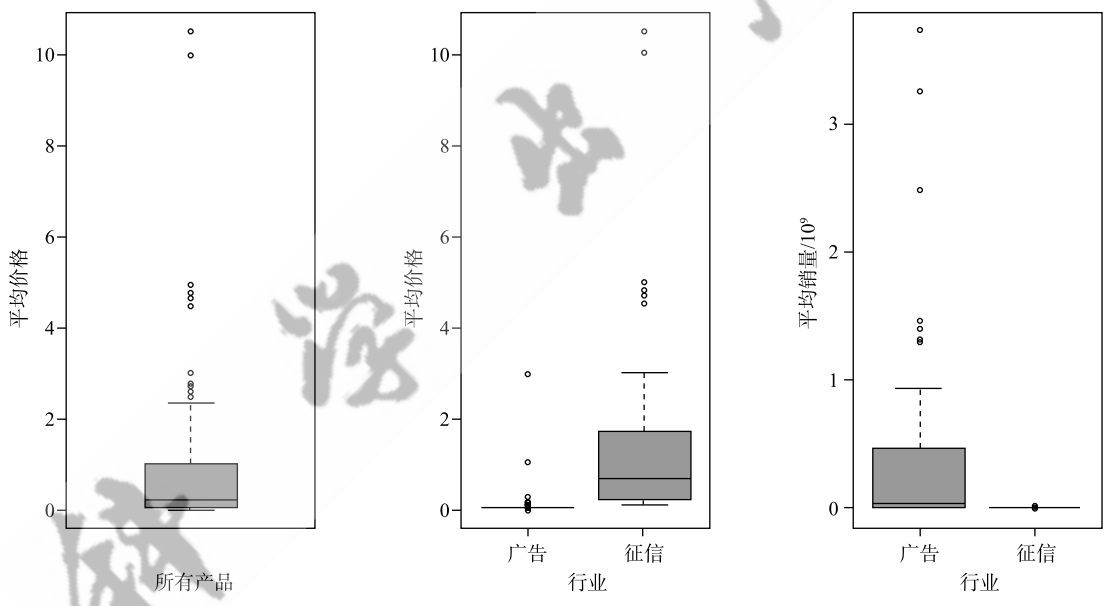
① 由于公司销售数据需要保密，描述性统计具体内容可联系作者经公司授权后获取。

续表

变量	样本量	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）	（9）
（6）数据权威性	6 042	−0.026*	−0.104***	−0.029*	0.199***	0.109***	1.000			
（7）数据买方性质	6 042	−0.156***	0.422***	0.171***	−0.564***	0.113***	−0.130**	1.000		
（8）数据查询方式	6 042	−0.093***	−0.134***	−0.078***	0.305	−0.004***	−0.258***	−0.155***	1.000	
（9）时间因素	6 042	−0.101***	0.006	−0.016	−0.069***	0.103***	0.213***	0.087***	−0.125***	1.000

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

其次，我们对该样本中不同行业类型的数据进行了进一步详细的描述性统计，以直观的形式探究数据的价格与销量如何受到行业分布的影响，不同行业数据产品的重要变量分布如图 1 所示。第一类征信类数据的需求主要来自金融领域；第二类广告场景类数据需求主要源于新媒体场景中的精准营销。上述两类数据交易虽然在市场当中都普遍存在，然而二者在数据特性上却存在显著的差异。征信类数据大多直接与个人数据绑定（如身份证二维核验、个人投资与任职查询等），并且价值也较大程度上受到分析模型的设定影响（不同征信企业所选择的评分模型可能存在差异），因此数据价值对于不同交易主体的异质性较高。广告类数据更加碎片化与标准化（如设备类型、消费类网站访问时长等），同时用以处理这些数据的主要模型思路也较为成熟（各类推荐算法等）。



（a）数据产品平均价格分布　（b）数据产品分行业平均价格分布　（c）产品分行业平均销量分布

图 1　数据产品平均价格与分行业平均价格销量分布

从图 1 中可以看出，广告类数据存在基准定价，约为 0.05 元/条，而征信类数据定价空间相对较大。从平均销量来看，广告类数据销量远大于征信类数据销量，因此两类产品市场总交易价值出现显著差异。由此可以看出，相较于广告类数据，征信类数据产品无论平均价格还是单笔交易价格都更高，但相应地，其销量也有显著下降，导致最终征信类数据的销售额显著小于广告类数据。为进一步支持行业异质性对于数据价格与销量的影响，我们选取了样本中 113 个产品的平均价格、销量与销售额进行 ANOVA 检验。结果（表 3）显示不同行业应用场景下数据产品价格的均值，方差与分布均都存在

显著不同（p<0.01），同时其销量分布也显著不同（p<0.01）。

表 3　数据价格行业特征 ANOVA 方差分析

自变量	平方和	自由度	均方和	F 比值	p 值
模型 1：平均价格为因变量					
行业因素	44.39	1	44.39	17.77	$5.13\times10^{-5***}$
误差	277.46	111	2.50		
模型 2：平均销量为因变量					
行业因素	6.574×10^{18}	1	6.574×10^{18}	16.8	$7.93\times10^{-5***}$
误差	4.344×10^{19}	111	3.913×10^{17}		
模型 3：平均销售额为因变量					
行业因素	2.023×10^{16}	1	2.023×10^{16}	17.11	$6.88\times10^{-5***}$
误差	1.312×10^{17}	111	1.182×10^{15}		

***表示 p<0.01

基于上述分析，本文在控制了行业异质性、是否存在数据大买方、数据交易需求、数据查询方式、时间因素等控制变量和数据供方的固定效应后，根据我们对于数据供方权威性和数据独特性的研究假设，我们对每笔发生的交易 k，分别建立如下模型：

$$\ln(\text{Price}_k)=\beta_0+\beta_1\text{Industry}_k+\beta_2\text{Uniqueness}_k+\beta_3\text{Authority}_k+\alpha_1\text{Dummy\_Search\_Type}_k+\alpha_2\text{Dummy\_Hoarder}_k+\alpha_3\text{Dummy\_Demand}_k+\sum_{T=2018,2019}\alpha_T\text{Year\_Online}_{k,T}+\sum_{q=1,2,\ldots,37}\alpha_q\text{Supplier}_{k,q}+\varepsilon_k$$

$$\ln(\text{Volume}_k)=\beta_0+\beta_1\text{Industry}_k+\beta_2\text{Uniqueness}_k+\beta_3\text{Authority}_k+\alpha_1\text{Dummy\_Search\_Type}_k+\alpha_2\text{Dummy\_Hoarder}_k+\alpha_3\text{Dummy\_Demand}_k+\sum_{T=2018,2019}\alpha_T\text{Year\_Online}_{k,T}+\sum_{q=1,2,\ldots,37}\alpha_q\text{Supplier}_{k,q}+\varepsilon_k$$

$$\ln(\text{Sales}_k)=\beta_0+\beta_1\text{Industry}_k+\beta_2\text{Uniqueness}_k+\beta_3\text{Authority}_k+\alpha_1\text{Dummy\_Search\_Type}_k+\alpha_2\text{Dummy\_Hoarder}_k+\alpha_3\text{Dummy\_Demand}_k+\sum_{T=2018,2019}\alpha_T\text{Year\_Online}_{k,T}+\sum_{q=1,2,\ldots,37}\alpha_q\text{Supplier}_{k,q}+\varepsilon_k$$

4　分析结果

4.1　数据独特性对数据成交价与成交额的影响

对该横截面数据集的单笔交易而言，在控制了所有供应方的固定效应后，表 4 的模型 1、模型 2 证实了我们的假设 H_{1a}：即对于同一个商家提供的不同数据产品，数据产品越独特（即提供该商品的厂家越少），数据产品的价格越高。表 4 中模型 3~模型 7 显示，在未控制供应商固定效应的情况下，数据越独特，单笔交易的成交额与成交量越大；但在控制供应商固定效应的情况下，该相关性并不显著，这可能是因为供应商的异质性吸收了数据产品的独特性：有部分供应商只提供一个商品且供应商数量一共只有 38 家。为进一步探索数据独特性与数据成交量与销售额之间的关系，我们将在 5.1.1 稳健性分析

中进一步对征信类数据和广告类数据分别进行回归分析并提供相应的解释。

表 4 数据独特性对单笔交易成交量与成交价格的影响（OLS 回归）

自变量	ln（Price）		ln（Volume）		ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7
行业（Industry）	2.52*** （0.047）	2.39*** （0.065）	−9.64*** （0.655）	−8.96*** （0.128）	−7.45*** （0.09）	−7.11*** （0.09）	−6.57*** （0.13）
数据独特性（Uniqueness）	0.42*** （0.026）	0.49*** （0.033）	0.29*** （0.052）	−0.40*** （0.064）		0.71*** （0.054）	0.09 （0.06）
常数（Constant）	−2.37*** （0.056）	−3.81*** （0.400）	11.71*** （0.111）	1.20*** （0.791）	9.71*** （0.111）	9.34*** （0.11）	−2.60*** （0.825）
观察个数	6 042	6 042	6 042	6 042	6 042	6 042	6 042
调整后 R^2	0.480	0.656	0.874	0.919	0.818	0.823	0.880
控制变量	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	N	Y	N	Y	N	N	Y

***表示 p<0.01

注：括号内为 OLS 回归的标准误

4.2 数据权威性对数据成交价与成交额的影响

对于假设 H_2，从表 5 的回归结果不难发现，在控制相关变量后，数据权威性对价格（表 5 中模型 1~模型 3）和销售额（表 5 中模型 7~模型 9）有显著的正向影响，但对成交量的影响并不显著。

表 5 数据权威性对单笔交易成交量与成交价格的影响（OLS 回归）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
行业（Industry）	2.26*** （0.046）	2.41*** （0.038）	2.43*** （0.047）	−12.5*** （0.107）	−12.6*** （0.111）	−9.63*** （0.09）	−10.2*** （0.103）	−10.2*** （0.107）	−7.19*** （0.097）
数据权威性（Authority）	0.38*** （0.026）	0.14*** （0.026）	0.34*** （0.025）	0.07 （0.074）	0.10 （0.08）	−0.04 （−0.052）	0.265*** （0.071）	0.237** （0.072）	0.30*** （0.053）
数据独特性（Uniqueness）		0.36*** （0.028）	0.37*** （0.026）		0.10*** （0.075）	0.30*** （0.05）		0.182* （0.072）	0.67*** （0.054）
常数（Constant）	−2.02*** （0.055）	−3.14*** （0.037）	−2.23*** （0.056）	15.65*** （0.099）	15.74*** （0.108）	11.70*** （0.113）	12.7*** （0.095）	12.61*** （0.104）	9.47*** （0.112）
观察个数	6 042	6 042	6 042	6 042	6 042	6 042	6 042	6 042	6 042
调整后 R^2	0.478	0.422	0.495	0.702	0.702	0.874	0.627	0.730	0.823
控制变量	Y	N	Y	N	N	Y	N	N	Y
厂商固定效应	N	N	N	N	N	N	N	N	N

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

注：括号内为 OLS 回归的标准误

5 讨论与总结

5.1 稳健性检验

5.1.1 分行业稳健性检验

不同的样本对于所得的结果具有不同的敏感性，因此在稳健性检验时，也常常先进行分样本回归，常见的分类方法有按照人口规模分类、按照地理位置分类、按照城乡分类、按照性别不同分类等。

例如，刘怡等在研究婚姻匹配对代际流动性的影响时提出婚姻匹配是中国代际传递的重要机制[26]，在稳健性检验中，根据子代的城乡分布，将子代样本划分为城镇和乡村样本，比较分析城镇和乡村地区的代际流动性及其婚姻匹配机制在代际传递中的影响。

在本文描述性统计部分，已经论述了不同的数据行业应用场景对其价格与销量的分布影响十分显著，因此可将样本分为广告类数据与征信类数据两个子样本并对假设 H_1 与 H_2 分别进行分样本回归，从表 6 和表 7 中可以得到如下结论：从各检验的回归系数及其显著性中可以发现，对于征信行业而言数据独特性对数据价格起到的作用更显著且正相关性更大（系数为 0.73，p 值接近 0）；相反地，对于广告类数据，数据独特性对数据价格起到的作用不显著且有负相关性（系数为−0.1，p 值为 0.046），我们推测这可能是由于广告类数据产品更为标准化——每一条数据所包含的信息往往较为固定，如浏览网页次数、点击量、停留时间等，且更易于复制——如能够方便地通过网络爬虫获得，因此该数据集无法在独特性上体现价值，价格也往往较为稳定——通过描述性统计可发现广告类数据集的价格聚集在 0.05 元/条的市场定价。

表 6　数据独特性与权威性对单笔交易成交量与成交价格的影响（征信类数据）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
数据权威性（Authority）		0.45***（0.026）	0.39***（0.026）		0.15**（0.078）	0.12*（0.047）		0.60***（0.049）	0.50***（0.049）
数据独特性（Uniqueness）	0.73***（0.042）		0.50***（0.029）	−0.59***（0.075）		0.24***（0.052）	0.14*（0.079）		2.65***（0.080）
常数（Constant）	−2.49***（0.253）	0.49***（0.044）	0.48***（0.043）	4.68***（0.457）	2.17***（0.078）	2.16***（0.078）	2.18***（0.483）	2.66***（0.082）	2.65***（0.080）
观察个数	5 352	5 352	5 352	5 352	5 352	5 352	5 352	5 352	5 352
调整后 R^2	0.437	0.167	0.211	0.752	0.652	0.654	0.702	0.587	0.600
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

注：括号内为 OLS 回归的标准误

表 7　数据独特性与权威性对单笔交易成交量与成交价格的影响（广告类数据）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
数据权威性（Authority）		0.34**（0.112）	0.33***（0.119）		−3.36**（0.501）	−3.33***（0.500）		−3.02***（0.464）	−3.00***（0.464）
数据独特性（Uniqueness）	0.02（0.035）		−0.1*（0.046）	−0.84（1.18）		0.33（0.195）	−0.07*（0.146）		0.23（0.181）
常数（Constant）	−3.98***（0.281）	−4.42***（0.319）	−4.36***（0.320）	−0.838（1.187）	4.17**（1.35）	3.94**（1.35）	−4.8***（1.163）	−0.25（1.246）	−0.41（1.25）
观察个数	690	690	690	690	690	690	690	690	690
调整后 R^2	0.486	0.063	0.067	0.593	0.264	0.266	0.567	0.302	0.302
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

注：括号内为 OLS 回归的标准误

另外值得注意的是，对于这些容易获取的广告类数据，一方面，复制份数越少即独特性越小的数据往往意味着其质量和可用价值越低，因而无人选择复制并出售，而这可能导致的结果是，数据独特性越高反而意味着总销售额和成交量越低；另一方面，广告类数据的数据独特性和数据权威性对于数

据质量的影响较小，由于广告类数据原始商家的数据权属不够明确，部分由权威商家提供的数据可能遭受买方复制修改二手倒卖后降价，导致广告类的数据权威性对于最终销量的影响反而是负向的。与之相反，我们不难观察到，在获取数据更加困难的征信类行业，数据权威性对于价格、销量和销售额的正向影响都更显著。

5.1.2 去除大买家后的稳健性检验

陈强远等在研究中国技术创新主要激励政策对企业技术创新质量和数量的影响中提到，高新技术企业认定等技术创新激励政策可能存在自选择问题，即整体绩效较好的企业更容易享受优惠政策，这可能导致估计结果存在偏误[27]。

同样地，本文认为购买数据金额与产品个数远超其他的大买方可能存在自选择问题：即大买方比普通买方更容易影响产品的市场定价与总销售额。因此本文在交易记录中去除 3 个大买家后重新对剩余交易记录进行回归并进行了稳健性检验。我们从表 8 中不难发现去除大买家后数据权威性对数据成交量的影响结果有所变化（数据权威性对成交量影响不够显著），而这可能与大买家会购买更多数据的行为相关。

表 8 数据独特性与权威性对单笔交易成交量与成交价格的影响（去除大买家）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
行业（Industry）	2.43*** （0.072）	2.28*** （0.047）	2.49*** （0.048）	−9.07*** （0.138）	−9.8*** （0.092）	−9.63*** （0.096）	−6.6*** （0.145）	−7.5*** （0.095）	−7.14*** （0.098）
数据权威性（Authority）		0.39*** （0.026）	0.34*** （0.026）		0.013 （0.051）	−0.03 （0.052）		0.40*** （0.053）	0.31*** （0.053）
数据独特性（Uniqueness）	0.54*** （0.035）		0.41*** （0.027）	−0.49*** （0.068）		0.23*** （0.055）	0.06 （0.071）		0.71*** （0.056）
常数（Constant）	−3.80*** （0.409）	−1.99*** （0.056）	−2.22*** （0.057）	1.105 （0.788）	11.8*** （0.11）	11.71*** （0.114）	−2.7** （0.826）	9.88*** （0.114）	9.49*** （0.116）
观察个数	5 804	5 804	5 804	5 804	5 804	5 804	5 804	5 804	5 804
调整后 R^2	0.606	0.405	0.423	0.893	0.833	0.833	0.847	0.765	0.772
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，**表示 p<0.05

注：括号内为 OLS 回归的标准误

5.1.3 调整时间窗口后的稳健性检验

本文数据样本期为 2016 年 12 月至 2019 年 8 月，在建立回归模型时本文为控制时间效应选择了 2017 年至 2019 年所有交易记录作为回归样本。表 9~表 11 的稳健性检验通过调整时间窗口的方式验证了结果的稳健性。

表 9 数据独特性与权威性对单笔交易成交量与成交价格的影响（2017 年）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
行业（Industry）	2.31*** （0.560）	2.85*** （0.380）	3.31*** （0.379）	−10.4*** （1.220）	−9.5*** （0.872）	−9.87*** （0.089 2）	−8.1*** （1.188）	−6.6*** （0.866）	−6.57*** （0.890）
数据权威性（Authority）		−0.55 （0.443）	−0.40 （0.43）		−3.6*** （1.019）	−1.0* （0.052）		−4.19*** （1.012）	−4.17*** （1.01）
数据独特性（Uniqueness）	1.15*** （0.210）		1.08*** （0.200）	−0.55 （0.456）		−3.78*** （1.017）	0.60 （0.443）		0.07 （0.471）

续表

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
常数（Constant）	−2.43***（0.608）	−1.41*（0.594）	−2.43***（0.608）	−4.3*（1.810）	6.80***（1.36）	7.75***（1.429）	−7.84**（1.76）	5.39***（1.352）	5.32***（1.423）
观察个数	471	471	471	471	471	471	471	471	471
调整后 R^2	0.557	0.522	0.606	0.779	0.736	0.737	0.694	0.619	0.618
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

注：括号内为 OLS 回归的标准误

表 10　数据独特性与权威性对单笔交易成交量与成交价格的影响（2018 年）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
行业（Industry）	2.78***（0.153）	2.30***（0.057）	2.40***（0.058）	−10.2***（0.263）	−10***（0.099）	−10.8***（0.102）	−7.4***（0.297）	−8.48***（0.104）	−8.33***（0.107）
数据权威性（Authority）		0.10**（0.035）	0.07（0.035）		0.12*（0.062）	0.107***（0.062）		0.22***（0.065）	0.17***（0.065）
数据独特性（Uniqueness）	0.51***（0.04）		0.25***（0.035）	−0.39***（0.071）		0.16*（0.062）	0.11（0.081）		0.41**（0.065）
常数（Constant）	−0.04（0.49）	−2.41***（0.060）	−2.54***（0.062）	7.19***（0.848）	12.02**（0.104）	12.02**（0.104）	7.14**（0.960）	9.61***（0.110）	9.39***（0.114）
观察个数	3 029	3 029	3 029	3 029	3 029	3 029	3 029	3 029	3 029
调整后 R^2	0.604	0.454	0.463	0.931	0.900	0.901	0.880	0.851	0.853
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

注：括号内为 OLS 回归的标准误

表 11　数据独特性与权威性对单笔交易成交量与成交价格的影响（2019 年）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
行业（Industry）	3.02***（0.380）	2.11***（0.052）	2.33***（0.053）	−3.6***（0.888）	−11***（0.106）	−10.8***（0.111）	−0.5（0.954）	−9.0***（0.111）	−8.5***（0.114）
数据权威性（Authority）		0.78***（0.033）	0.72***（0.033）		0.23***（0.068）	−0.14*（0.069）		1.02***（0.072）	0.87***（0.07）
数据独特性（Uniqueness）	0.28***（0.043）		0.42***（0.035）	−0.13（0.100）		0.62***（0.072）	0.15（0.107）		1.04***（0.074）
常数（Constant）	−0.07（0.29）	−2.16***（0.053）	−2.43***（0.057）	6.25***（0.68）	12.2***（0.109）	11.8***（0.118）	6.17***（0.732）	10.1***（0.115）	9.38***（0.120）
观察个数	2 542	2 542	2 542	2 542	2 542	2 542	2 542	2 542	2 542
调整后 R^2	0.823	0.583	0.605	0.945	0.900	0.902	0.912	0.852	0.863
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，*表示 p<0.10

注：括号内为 OLS 回归的标准误

缩短时间窗口可以排除其他政策的影响，如王雄元和卜落凡在研究“一带一路”如何影响企业创新行为的研究中提到，中国于 2013 年正式提出“一带一路”倡议，因此仅保留 2013 年及以后开通“中欧班列”的样本，有助于将研究统一置于“一带一路”倡议的背景下，排除可能的其他政策干扰[28]。同样地，本文通过缩短时间窗口至每年从而排除了2017~2019年有关数据隐私保护等政策变化对交易量与成交价格带来的影响。本文发现 2017 年的子样本无法支持假设 H_2 即数据权威性正向影响数据价格、成交量与销售额的假设，这一方面可能是因为样本量较小（471 条交易记录）所导致的；另一方面可能是因为数据交易所刚刚成立，买方仍处于模糊的摸索阶段，在购买过程中对于数据供方是否权威不够了解。

5.1.4 去除尝试性交易后的稳健性检验

本文考虑到在很多交易中，买方只购买了 10 条以内的数据，支付金额小于 1 元，我们猜测该类交易为买家的尝试性交易，因此在表 12 的稳健性检验中我们排除了这些交易记录并再次进行回归分析，有趣的是我们发现在排除了尝试性交易的情况下，数据独特性的相关结果稳健且对销售额与销售量起到正向影响；但数据权威性对成交额与成交量都起到了负向影响，即数据由越权威的原始商家提供，其成交金额最后却越少，我们猜测这是因为经过了几次尝试性交易后，由于原始商家的数据权属不够明确，部分由权威商家提供的数据遭到了买方二手倒卖后降价造成的。

表 12 数据独特性与权威性对单笔交易成交量与成交价格的影响（去除尝试性交易）

自变量	ln（Price）			ln（Volume）			ln（Sales）		
	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
行业（Industry）	1.89*** （0.068）	2.15*** （0.045）	2.26*** （0.047）	−8.03*** （0.151）	−9.9*** （0.121）	−9.46*** （0.124）	−6.1*** （0.153）	−7.79*** （0.118）	−7.2*** （0.119）
数据权威性（Authority）		0.43*** （0.033）	0.39*** （0.033）		−0.3*** （0.088）	−0.48*** （0.087）		0.12*** （0.08）	−0.08 （0.084）
数据独特性（Uniqueness）	0.29*** （0.042）		0.26*** （0.034）	−0.21* （0.094）		1.12*** （0.090）	0.07 （0.096）		0.71*** （0.056）
常数（Constant）	−3.79*** （0.613）	−2.63*** （0.053）	−2.77*** （0.056）	4.10** （1.359）	16.6*** （0.14）	15.97*** （0.147）	0.310 （1.386）	13.9*** （0.139）	13.21*** （0.142）
观察个数	3 014	3 014	3 014	3 014	3 014	3 014	3 014	3 014	3 014
调整后 R^2	0.729	0.586	0.593	0.893	0.827	0.836	0.885	0.765	0.783
控制变量	Y	Y	Y	Y	Y	Y	Y	Y	Y
厂商固定效应	Y	N	N	Y	N	N	Y	N	N

***表示 p<0.01，**表示 p<0.05，*表示 p<0.10

注：括号内为 OLS 回归的标准误

5.2 对于数据交易市场的管理启示

虽然由于数据规模有限，本文研究依然属于探索类工作，但通过对于来自数据交易市场的交易数据展开的上述分析，已经可以帮助从业者和学术界对于数据交易市场这一新兴市场的规律形成初步认识。我们把本文的研究结论和与之有关的启示共同用图 2 来表示。

在图 2 中，浅色背景文本框代表了本文假设 H_1 和 H_2 的显著结果，而深色背景文本框则是与之相对应的市场后果，整体连起来看，就形成了数据交易市场“劣币驱逐良币”的演进过程。我们根据回归结果推测的具体解释如下：首先，对于任何一个新兴的数据交易市场而言，市场所有者和发起者会积极动员具有一定权威性的数据提供方进入市场，从而吸引更多数据的买家。这一点与大多数的平台市

场的初期策略是一致的。但本文的假设 H_2 结果表明，权威卖家的市场价格也相对较高，从而才能满足权威卖家参与该市场的收益预期；然而，由于数据的特殊性，任何购买获得数据的买方如果对这一利润感兴趣，也可以转售手中的数据，并且这个“仿冒”的过程难以控制和界定，如通过对于原数据的重抽样形成的新数据集，是否对原始数据构成侵权是没有成熟标准的。一旦高价数据出现替代品和二手交易，本文的假设 H_2 研究结果表明，价格会相应更低，从而导致数据的权威卖方无法获得理想的利润，最终缺乏参与动力而离开市场。由此可见，简单地以电商平台的逻辑来销售数据是行不通的——数据在易于复制、非标准化、无法判断是否仿制等方面的特性，使得数据交易市场更加容易陷入过度竞争，从而影响到数据交易市场的健康发展。

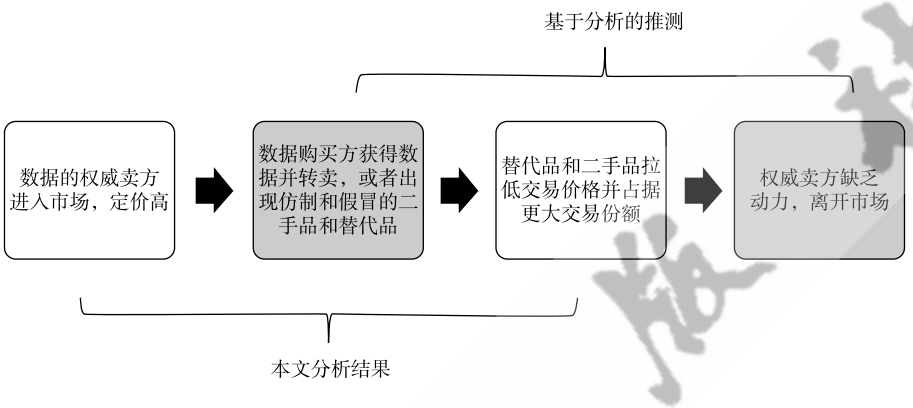


图 2　基于本文分析结果进行的数据市场发展趋势

如何破解数据市场“劣币驱逐良币”的趋势？本文研究虽然无法直接给予解答，但是可以对解决方案提出一定的思考和启发——图 2 的逻辑显示，核心的问题出现在第二个环节，就是需要在数据交易过程中，对于数据权属的确定和转移过程给予充分的重视并采用必要的技术手段来合理设计。具体而言，如果需要对于数据的所有权进行交易，则需要通过技术和制度手段来保证数据的所有权透明可追踪，如利用区块链技术来进行标记；如果数据交易对于数据本身的所有权不进行交易，只是对于数据的使用权进行交易，则需要对于数据本身的产品和服务形态进行更加完善的封装——一言以蔽之，没有封装、无法追踪权属的原始数据不应当参与数据市场交易。例如，在北京国际大数据交易所的官方描述当中就特别强调“北京国际大数据交易所将以数据技术为支撑，采用隐私计算、区块链等手段分离数据所有权、使用权、隐私权”，而本文的研究结论对于这一设计思路的必要性从实证数据的角度提供了直接的支持。

6　总结

数据要素市场的建立和发展已经成为我国“十四五”规划中“数字中国”建设工作所明确的基础任务之一，而数据要素市场的健康发展将扮演着打破“数据孤岛”和发挥数据的“乘数效应”的重要作用。对于这一新型的市场形态，本文基于一组来自国内领先数据交易市场的真实交易数据展开探索性研究，对数据市场当中的价格如何受到供方权威性和数据独特性的影响进行了重点分析，并通过图 2 的讨论，对我国数据交易市场未来的发展进行了展望和讨论。

囿于数据的规模有限，本文目前只是探索性的研究工作，但对于数据交易市场而言，有大量的研究工作有待开展，如基于数据的各类独特性质来探索其定价策略和商业模式等，对于我国未来的数据市场机制设计和发展都将具有重要的引导和启发意义。

参 考 文 献

[1] Tang Q，Shao Z，Huang L，et al. Identifying influencing factors for data transactions：a case study from shanghai data exchange[J]. Journal of Systems Science and Systems Engineering，2020，29（6）：697-708.

[2] Acquisti A，Taylor C，Wagman L. The economics of privacy[J]. Journal of Economic Literature，2016，54（2）：442-492.

[3] Chen L，Huang Y，Ouyang S，et al. The data privacy paradox and digital demand[Z]. NBER Working Papers，2021.

[4] Duch-Brown N，Martens B，Mueller-Langer F. The economics of ownership，access and trade in digital data[Z]. Digital Economy Working Paper，JRC Technical Reports，2017.

[5] Goldfarb A，Trefler D. The Economics of Artificial Intelligence：An Agenda[M]. Chicago：University of Chicago Press，2019：463-492.

[6] Carrière-Swallow Y，Haksar V. The economics and implications of data[Z]. IMF Working Papers，2019.

[7] Falck O，Koenen J. Resource“data”：economic benefits of data provision[C]. CESifo Forum. München：ifo Institut-Leibniz-Institut für Wirtschaftsforschung an der Universität München，2020，21（3）：31-41.

[8] Wolfert S，Ge L，Verdouw C，et al. Big data in smart farming – a review[J]. Agricultural Systems，2017，153：69-80.

[9] Jones C I，Tonetti C. Nonrivalry and the economics of data[J]. American Economic Review，2020，110（9）：2819-2858.

[10] Farboodi M，Mihet R，Philippon T，et al. Big data and firm dynamics[J]. AEA Papers and Proceedings，2019，109：38-42.

[11] Koutroumpis P，Leiponen A，Thomas L D W. The（unfulfilled）potential of data marketplaces[Z]. ETLA Working Papers，2017.

[12] Acemoglu D，Makhdoumi A，Malekian A，et al. Too much data：prices and inefficiencies in data markets[Z]. NBER Working Papers，2019.

[13] Agarwal A，Dahleh M，Horel T，et al. Towards data auctions with externalities[EB/OL]. https://arxiv.org/abs/2003.08345，2020.

[14] Koutroumpis P，Leiponen A，Thomas L D W. Markets for data[J]. Industrial and Corporate Change，2020，29（3）：645-660.

[15] Roth A E. The economist as engineer：game theory，experimentation，and computation as tools for design economics[J]. Econometrica，2002，70（4）：1341-1378.

[16] Roth A E. What have we learned from market design?[J]. The Economic Journal，2008，118（527）：285-310.

[17] Horn H，Wolinsky A. Bilateral monopolies and incentives for merger[J]. The RAND Journal of Economics，1988：408-419.

[18] Coase R H. The coase theorem and the empty core：a comment[J]. The Journal of Law and Economics，1981，24（1）：183-187.

[19] Rust J. When is it optimal to kill off the market for used durable goods? [J]. Econometrica，1986，54：65-86.

[20] Shapiro C. Premiums for high quality products as returns to reputations[J]. The Quarterly Journal of Economics，1983，98（4）：659-679.

[21] Allen F. Reputation and product quality[J]. The RAND Journal of Economics，1984：311-327.

[22] Klein B，Leffler K B. The role of market forces in assuring contractual performance[J]. Journal of Political Economy，

1981，89（4）：615-641.

[23] Tirole J. A theory of collective reputations（with applications to the persistence of corruption and to firm quality）[J]. The Review of Economic Studies，1996，63（1）：1-22.

[24] Rindova V P，Williamson I O，Petkova A P. Reputation as an intangible asset：reflections on theory and methods in two empirical studies of business school reputations[J]. Journal of Management，2010，36（3）：610-619.

[25] Obloj T，Capron L. Role of resource gap and value appropriation：effect of reputation gap on price premium in online auctions[J]. Strategic Management Journal，2011，32（4）：447-456.

[26] 刘怡，李智慧，耿志祥. 婚姻匹配、代际流动与家庭模式的个税改革[J]. 管理世界，2017，（9）：60-72.

[27] 陈强远，林思彤，张醒. 中国技术创新激励政策：激励了数量还是质量[J]. 中国工业经济，2020，385（4）：81-98.

[28] 王雄元，卜落凡. 国际出口贸易与企业创新——基于“中欧班列”开通的准自然实验研究[J]. 中国工业经济，2019，379（10）：82-100.

Understanding Pricing in the Data Factor Market: An Exploration Study at Shanghai Data Exchange

YIN Wenyi[1]，DOU Yifan[1]，TANG Qifeng[2]，HUANG Lihua[1]

（1. School of Management，Fudan University，Shanghai 200433，China；
2. Shanghai Data Exchange，Shanghai 200436，China）

Abstract In the era of the digital economy，firms are accumulating a large amount of data which performs as an increasingly important factor in the production process. However，since the effective data markets are largely missing，the data generated at firms are mostly isolated，which undermines the process of effective reallocation of data among organizations. This paper conducts an exploratory study with a novel real-world dataset provided by one of the leading data exchanges in China，Shanghai Data Exchange. The results suggest that the data prices differ significantly in distribution characteristics，and the authority and uniqueness of data providers are significantly associated with the transaction price. These findings shed light on the policy-making and infrastructure improvement of the data market.

Key words Data transaction，Data pricing，Factors of production，Exploratory analysis

作者简介

尹文怡（1999—），女，复旦大学管理学院统计系在读学生。E-mail：18307100087@fudan.edu.cn。

窦一凡（1985—），男，复旦大学管理学院信息管理与信息系统系教授、博士生导师。E-mail：yfdou@fudan.edu.cn。

汤奇峰（1968—），男，上海数据交易中心有限公司 CEO，人工智能正高级工程师，现任中欧数字经济专家组中方专家，上海大数据联盟理事长、上海市信息化专家委员会大数据专业委员会专家。E-mail：keven@chinadep.com。

黄丽华（1965—），女，复旦大学管理学院信息管理与信息系统系教授、博士生导师。E-mail：lhhuang@fudan.edu.cn。