

# 电商农户大数据使用：驱动因素与增收效应\*

曾亿武<sup>1</sup> 张增辉<sup>1</sup> 方湖柳<sup>1</sup> 郭红东<sup>2</sup>

**摘要：**随着新技术革命的推进，大数据逐渐与“三农”融合。本文聚焦电商农户使用大数据产品辅助网店经营决策的现象，在理论分析的基础上，利用山东曹县、江苏沭阳和浙江临安三地电商农户的问卷调查数据，采用二元 Probit 模型、熵均衡法和分位数回归模型，对电商农户大数据使用行为的驱动因素与增收效应进行了实证研究。研究发现，在农村电子商务发展较好的地区，已有近三成的电商农户在其网店经营过程中使用了大数据产品；以培训为主要形式的知识转移和以内群体交往为核心的知识溢出对驱动电商农户使用大数据发挥了重要的积极作用；大数据使用显著提升了电商农户的收入水平，并促进了电商农户内部收入差距的缩小。

**关键词：**电商农户 大数据 培训经历 内群体交往 收入效应

**中图分类号：**F325 F724.6 **文献标识码：**A

## 一、引言

大数据是指一种规模大到在采集、存储、管理、分析等方面明显超出了传统数据库软件工具能力范围的数据集合，具有数据容量大、数据传递快速、数据类型多样、数据更新及时、数据质量高等特征，通过合理运用大数据，可实现以低成本创造高价值（Chi et al., 2016; Rodriguez et al., 2017）。随着现代社会的快速发展、信息化的快速膨胀和互联网的迅猛传播，各种海量的数据化信息被不停地生产、收集、存储、处理与利用，大数据时代随之来临，带来了全方位的社会变革，深刻地影响着人们的生产和生活。目前大数据分析技术正被加速应用于不同的行业，例如金融、保险、网络营销以及科学研究（Cooper et al., 2013; Waga and Rabah, 2014）。政府组织也在使用大数据分析技术增强服务公民的能力，以更好地应对来自经济、医疗、创造就业、自然灾害和恐怖袭击等方面的挑战（Kim et al., 2014）。

---

\*本文研究受到浙江省自然科学基金项目“电商农户大数据产品使用：属性偏好与增收效应”（编号：LQ20G030016）、国家自然科学基金项目“中国淘宝村形成机理及其‘三农’影响效应研究”（编号：71673244）、杭州市属高校第二轮优秀创新团队扶持计划、2019年杭州市属高校优秀青年博士扶持计划资助。感谢匿名评审专家提出的宝贵意见，但笔者文责自负。本文通讯作者：郭红东。

近十年来，中国农村电子商务发展十分迅速，越来越多的农户<sup>①</sup>借助第三方电商平台开展网络营销。2017年，全国农村网店达到985.6万家<sup>②</sup>；截至2019年8月，全国已发现“淘宝村”共计4310个，“淘宝镇”1118个，过去一年全国“淘宝村”和“淘宝镇”网店年销售额合计超过7000亿元，在全国农村网络零售额中占比接近50%，活跃网店数达到244万个<sup>③</sup>。

电子商务平台汇聚各方主体，沉淀交易记录、评论信息、搜寻痕迹等海量数据，通过对这些大数据的挖掘和分析，能够为网商的线下生产和线上经营提供重要指引（曾亿武等，2018）。现实中，已有部分电商农户使用“生意参谋”<sup>④</sup>等大数据产品辅助网店的经营决策，但仍有很多电商农户尚未选择应用大数据。究竟存在什么关键因素显著影响着电商农户的大数据使用行为，并且，大数据的使用会对电商农户的收入水平和收入差距产生怎样的影响效应，对这些问题进行研究，具有重要的理论价值和现实意义。一方面，有助于加深对电商农户大数据使用行为规律和效应机理的理解，以及形成对大数据使用实际效果的客观评价和正确认识；另一方面，可以为政府部门促进农村电子商务高质量发展以及推动大数据广泛应用和深度服务于农村地区的相关政策制定提供科学依据和决策参考。

电商农户作为互联网时代农村地区的一个新兴群体，正在不断受到学者们的关注和研究。与传统农民相比，电商农户群体呈现年轻化、拥有较强的计算机使用技能、互联网使用频率高、重视提升产品附加值、注重消费者体验、具有开放和分享精神等特征（曾亿武、郭红东，2018）。电商农户成长遵循浮现孕育、立足生存、步入崛起、拓展集聚的演进路径，其健康发展需要从品牌化建设、农村电商物流渠道下沉建设、农村电商产业园建设、农村电商人才体系建设等方面进行合理规划（王昕宇、黄海峰，2016）。部分电商农户经历了简单复制模仿、经营规模扩张、品牌形象塑造的持续成长演化路径，地方电商生态系统的建设状况显著地影响着电商农户由小到大、由弱变强的转型升级过程（张庆民等，2019）。实践表明，电子商务的兴起赋予了农民新的发展机会以及更多的可能性。电子商务有助于增加农户的订单量、开发新的客户，尤其是促进其在国内市场的销售（Galloway et al., 2011; Zapata et al., 2013; Fan and Garcia, 2018）。电子商务拓宽农民的就业渠道，在农村以及偏远

<sup>①</sup>在本文中，农户是一个区位概念，指居住在农村地区的家庭，不局限于从事农业经营活动；并且按照是否有开网店，将农户分为电商农户和非电商农户。换言之，本文将电商农户定义为利用电商平台开展网络销售的农村家庭。

<sup>②</sup>数据来源：中国国际电子商务中心研究院发布的《中国农村电子商务发展报告（2017—2018）》，<http://www.199it.com/archives/805043.html>。

<sup>③</sup>数据来源：阿里研究院发布的《中国淘宝村研究报告（2009—2019）》，<http://www.100ec.cn/detail--6525509.html>。

<sup>④</sup>“生意参谋”诞生于2011年，是阿里巴巴开发的商家端统一数据产品平台，覆盖淘宝、天猫等阿里系所有平台和PC、无线等全部终端，集数据作战室、市场行情、装修分析、来源分析、竞争情报等数据产品于一体，涉及数据指标上千个，是大数据时代赋能电商卖家的重要平台。据天下网商披露，2016年“生意参谋”累计服务商家超2000万，月服务商家超500万；月成交额30万元以上的商家中，逾90%在使用“生意参谋”；月成交额100万元以上的商家中，逾90%每月登录“生意参谋”天次达20天次以上。详见<http://www.iwshang.com/Post/Default/Index/pid/244213.html>。

地区的作用更为明显 (Atasoy, 2013), 电子商务还对农村创业活动具有促进作用 (芮正云、方聪龙, 2018; 王金杰等, 2019)。电子商务引发了农村配送渠道的重构以及“村民—中间商”相互依存关系的重置, 并有助于提升产品的适应性 (Leong et al., 2016)。更重要的是, 电子商务能显著地促进农户收入增长 (Cho and Tobias, 2010; 鲁钊阳、廖杉杉, 2016; 曾亿武等, 2018; 李琪等, 2019)。尽管如此, 在经营上, 电商农户总体上存在专业素质缺乏、本地同业竞争激烈、物流成本高、融资渠道和营销策略单一、政策普惠性低等问题 (路征等, 2015)。电商农户在网络空间表意上呈现网店装饰不足、朴素乡土风和单纯卖货的特点, 与城市电商存在较大差距 (邵占鹏, 2019)。

综上, 关注电商农户的已有文献主要涉及了电商农户的特征、电商农户的成长路径、电子商务对于农户产品销售、就业创业和收入等方面的影响以及电商农户的网店经营问题。现实中, 部分电商农户通过使用大数据成功克服经营短板并保持着竞争优势, 对大数据的利用构成这些电商农户成长以及网店经营绩效的重要影响因素, 但是已有研究尚未关注电商农户使用大数据产品辅助制定经营决策的现象。国外有部分相近文献关注农场主采纳农场管理信息系统的影响因素 (Carrer et al., 2017; Pivoto et al., 2019)、将农场数据共享到大数据平台的意愿 (Turland and Slade, 2019) 以及使用遥感数据指导农场开展精准农业生产的行为和效益 (Coble et al., 2018; Toscano et al., 2019), 但并不是针对电商农户这个群体, 也没有涉及源自销售端的大数据。

本文在理论分析的基础上, 提出若干研究假说, 利用浙江临安、山东曹县和江苏沭阳三个抽样区域 397 个电商农户的问卷调查数据, 采用二元 Probit 模型、熵均衡法和分位数回归模型, 实证分析电商农户大数据使用行为的驱动因素与增收效应。驱动因素方面, 本文重点聚焦培训经历和内群体交往两个因素, 它们构成电商农户知识获取的两种基本渠道, 通过影响他们对新技术的主观感知, 进而影响其技术采纳行为; 增收效应方面, 本文同时关注大数据使用对电商农户整体收入水平和内部收入差距的影响。本文的主要贡献在于: 第一, 积极响应实践中需要研究的新现象、新问题, 拓展了农村电子商务领域的研究视角, 丰富了电商农户微观层面的定量研究; 第二, 为电商农户内部大数据使用行为的差异性提供一个知识获取渠道视角的解释, 证实了以培训为主要形式的知识转移和以内群体交往为核心的知识溢出对于驱动电商农户使用大数据的积极作用; 第三, 阐释大数据使用的增收逻辑, 并证实了大数据使用有助于提升电商农户的收入水平, 同时促进了电商农户内部收入差距的缩小。

## 二、理论分析与研究假说

### (一) 培训经历、内群体交往与电商农户大数据使用

电商农户对大数据的使用本质上是一种新技术采纳行为。在技术采纳行为理论看来, 个体对新技术的主观感知是影响其技术采纳意向和行为的关键因素。例如计划行为理论 (TPB) 认为, 个人感知到完成某一技术采纳的难易程度是影响个人技术采纳行为的重要解释变量, 个人认为所拥有的资源与机会越多而预期阻碍越小, 其感知行为控制也就越强, 并促成技术采纳行为的发生 (Ajzen, 1991)。再如技术接受模型 (TAM) 提出, 感知有用性和感知易用性分别反映技术采纳者对于技术

优越程度和操作难度的主观评价，共同决定着个体对于新技术的行为态度（Davis, 1989）。主观感知固然影响技术采纳行为，但个体对新技术的主观感知存在差异的深层次原因更为重要，即技术接受模型所提到的“外部变量”。对于电商农户而言，知识获取渠道（access to knowledge）是影响他们对新技术主观感知的一个重要因素。借鉴 Fallah and Ibrahim（2004）的有关观点，电商农户的知识获取渠道可以划分为知识转移（knowledge transfer）和知识溢出（knowledge spillover）两种基本途径。知识转移是有意识地开展正式的学习与交流，目前其主要的实践形式是参与电商技能培训；知识溢出即电商农户之间日常交往过程中无意识的知识共享和信息传播。

在现代人力资本理论框架中，职业培训与医疗保健、正规教育、迁移活动、“干中学”等一道构成人力资本积累的重要投资方式（Schultz, 1961; Becker, 1964; Grossman, 1972; Lucas, 1988）。职业培训是提升人力资本中技术素质和专业能力的重要途径，具有较强的针对性和实用性，能够在短时期内快速有效传递专用性知识和技能。已有研究表明，职业技能培训对提升农户人力资本要素的作用显著（张银、李燕萍, 2010; 屈小博, 2013），对正规教育具有一定的替代效应，在正规教育不足的情况下，农户通过投资技能培训也能获得较高的回报（展进涛、黄宏伟, 2016）。参与电商培训是电商农户提升自身电子商务专业知识和技能的重要途径。部分村庄的实践显示，政府举办的各类电商培训，有助于提高电商农户的网店经营水平（魏延安, 2017; 曾亿武等, 2019），并推动部分电商农户开始使用甚至是深度使用平台企业提供的大数据产品。电商培训经历有助于增进电商农户对电子商务和大数据分析技术的了解程度，提升其相应的人力资本，使电商农户对大数据分析技术的主观感知结构发生调整，即具备了较强的感知行为控制以及较高水平的感知有用性和感知易用性。随着电商培训经历的不断积累，电商农户将愈发深刻认识到应用大数据分析技术的重要性，从而获得更强的技术警觉性，这种创业者群体应有的警觉性的增强将提升他们对技术背后商业机会的识别（郭红东、周惠珺, 2013），并推动其大数据使用行为的发生。

基于血缘、亲缘、地缘和业缘的中国乡土社会关系呈现一种“差序格局”特征（费孝通, 1998）。源于人们日常交往的知识溢出效应，顺着乡土社会关系的亲疏变化轨迹，同样呈现着差序化的结构特征。在差序格局的各个层次中，最内圈的人与中心人物的关系最紧密、感情最深厚、交往最频繁，知识溢出效应也最大；随着圈子向外推，关系和情感的因素逐渐减弱，交往也越来越少，知识溢出效应逐渐变小。社会认同理论使用“内群体”（in-group）这一概念衡量个人在社会关系中对于自身群体成员身份的认同与归属，个体会不自觉地进行圈内人和圈外人的比较，进而产生对自己所属群体的积极认同，并倾向于给内群体成员较多的资源和互助（Turner et al., 1983）。人们的知觉、态度和行为受到个体对内群体的偏爱以及对外群体的偏见的影响，内群体认同感越强，其行为决策越容易受群体成员身份以及群体内部其他成员的影响（Tajfel and Turner, 1986; Ellemers et al., 1999）。在中国乡村地区，内群体交往是影响农村居民享有知识溢出效应的一个极为关键的因素。对于电商农户，其内群体具体指“从事网络营销的亲戚朋友”，他们之间具有紧密的联系和较高的相似性。电商农户的内群体交往有助于促进与电子商务相关的大量信息快速分享，从而增加电商农户触碰并深入了解大数据分析技术的可能性。正如曾亿武等（2019）所指出的，新技术和新机会不断从外部传

导到农村地区，拥有丰富社会资本的农户，通常会有更高的概率从中受益。此外，在电商农户内群体交往的过程中，率先使用了大数据产品的内群体成员所产生的先驱效应和榜样力量会比外群体成员强大得多。内群体成员之间的相似程度高，使得他们采取相同行为的预期结果也趋于近似，并且，对内群体的认同和偏好，也促进了他们采取较为一致的行为决策。

基于以上分析，本文提出以下研究假说：

H1：培训经历丰富的电商农户具有更高的大数据使用倾向；

H2：内群体交往紧密的电商农户具有更高的大数据使用倾向。

### （二）大数据使用对电商农户收入水平的影响

在前互联网时代，农户的经营决策基于主观感受和经验积累，具有局部性、滞缓性、粗略性等缺陷。从生产者到消费者的整个对接过程中，链条长，环节多，信息收集、处理和传递的效率很低，信息滞后和失真现象严重。市场上的需求信息无法及时有效地反馈给生产主体，导致经营决策的科学性和计划性差。互联网带来的最大变化是信息，包括信息的流动和信息的分析。信息的流动从原本的不方便变成了以纳秒为计量单位的流动速度，从原来的定向流动变成了不定向流动，打破了信息不对称的堡垒（罗珉、李亮宇，2015）。数据井喷式增长和数据分析技术的开发，使得对信息的分析能力突飞猛进地提升。大数据让电商农户的信息获取变得更及时、更全面、更精准、更接近于完全信息，大幅降低主观偏差，突破时空的限制。互联网减少了中间环节，去中介化，去渠道化，点对点，端对端，直通直达，彻底拉近经营者与消费者的距离，让经营者更容易且更有效地做出需求导向型决策，实现用户本位主义（李海舰等，2014）。通过大数据产品的使用，电商农户可以获得一系列有关自身产品的最优参数组合，包括颜色、重量、口味、价格、物流等方面，从而对比发现自身经营存在哪些不利或不足的环节，以进行有针对性的改进。这一点极为关键，因为电子商务市场高度竞争，微弱的决策偏差将会在互联网的乘数效应机制下被不断放大，对经营绩效产生显著的影响。电子商务市场瞬息万变，经营者必须时刻保持警惕，做到与时俱进，甚至引领潮流。电商农户通过大数据产品可实时了解网络市场的营销动态，尤其是同类竞争者的最新进展以及成长最快的产品的有关信息。另外，大数据分析技术不仅能够实时处理并提供精准信息，而且还可以开发智能预测功能（Lokers et al., 2016）。以往人们只能对已发生事件的信息进行搜集、加工和处理，而大数据不仅包括已发生事件和正在发生事件的数据，而且包括通过机器学习挖掘的将来要发生事件的数据（何大安、任晓，2018），使经营者获得更强的前瞻力，积极谋求先动优势。

基于以上分析，本文提出以下研究假说：

H3：大数据使用对电商农户收入水平产生正向作用。

### （三）大数据使用对电商农户内部收入差距的影响

发生在传统社会的市场交易通常是以面对面为基础和在互动的情景中完成的。于是，农户的经营决策能力需要较长时期的经验积累，经验积累越深厚，则其经营决策的准确性越高，经营收入也越多；换言之，经验积累差距较大程度地导致了经营收入差距。卡斯特（2006）指出，在前互联网时代，作为一种与空间形态相匹配的存在形式，时间始终代表着事项的前后次序或脉络。到了大数

据时代，时间不再保持着事项固有的次序，而是一种“非序列化时间”，事项的脉络可以分割，推展的次序可以打乱，不同的事项可以相互交织，也可以齐头并进（张兆曙，2017）。在大数据的支持下，电商农户的经营收入不再严格遵循时间维度上的经验积累机制，低收入农户借助大数据，可以弥补自身在经验积累上的不足，甚至实现经营决策上的快速赶超。有学者研究表明，电子商务加剧了农户内部的收入不平等程度（曾亿武等，2018），但是大数据产品的使用将有助于缩小电商农户内部的经验积累差距和人力资本差距，进而促进缩小收入差距。以“生意参谋”为例，其为商家提供数据披露、分析、诊断、建议、优化、预测等一站式数据产品服务，全面展示店铺经营全链路的各项数据，包括店铺实时数据、商品实时排行、店铺行业排名、店铺经营概况、流量分析、商品分析、交易分析、服务分析、营销分析和市场行情。大数据在经营决策上的这些辅助作用，使低收入农户具有更大的报酬递增空间及获得更高边际效益的可能。

基于以上分析，本文提出以下研究假说：

H4：大数据使用能够缩小电商农户内部收入差距。

### 三、实证研究设计

#### （一）数据来源

本文研究所使用的数据来源于课题组2018年4月至7月对山东省菏泽市曹县、江苏省宿迁市沭阳县和浙江省杭州市临安区三个地区50个村庄电商农户的问卷调查。这三个地区的农村电子商务发展迅速，形成了较大规模的“淘宝村集群”，受到广泛宣传和关注，并且在主营产品类型、所在地区的经济发展水平方面存在明显差异，因而具有代表性和层次性。其中，山东省菏泽市曹县属于欠发达地区，其网络营销的主营产品是演出服饰和木制品，属于工业品类型，2017年全县“淘宝村”发展到74个；江苏省宿迁市沭阳县属于中等发达地区，其网络营销的主营产品是花卉苗木，属于鲜活农产品类型，2017年全县“淘宝村”发展到41个；浙江省杭州市临安区属于发达地区，其网络营销的主营产品是坚果炒货，属于干货农产品类型，2017年全区“淘宝村”发展到19个。课题组先后对这三个地区开展了入户调查，采取“一对一”的访谈形式，受访者是家庭网店经营的主要决策者，原则上每个地区随机抽取8个“淘宝村”和8个非“淘宝村”，最终实际完成了50个村的调研，其中有26个村属于“淘宝村”。平均每个村随机抽取10个电商农户进行调查，回收有效问卷405个，剔除存在极端值和缺失值的问卷后，最终用于本文实证研究的有效样本容量是397个。

#### （二）变量说明及描述性统计

本文研究所涉及的被解释变量和核心解释变量包括电商农户的大数据使用行为、产品销售净收入、培训经历和内群体交往。对于电商农户大数据使用行为，本文采用“是否使用大数据产品辅助网店经营决策”进行测度，表1的描述性统计结果显示，约28%的样本农户在其网店经营过程中使用了大数据产品，符合现实情况。大数据应用作为农村电子商务运营领域的新兴技术、高级工具和付费行为，目前尚处于早期发展阶段，用户规模还不是很大。对于电商农户的产品销售净收入，课题组成员询问的是调查（时点）前一年的情况，表1结果显示，2017年样本农户的家庭产品销售净

收入均值是 29.27 万元。对于电商农户的培训经历，询问的是受访者使用大数据产品之前平均每年参加电商技能培训的次数。从表 1 可以看到，受访者在使用大数据产品之前平均每年参加电商技能培训 1.15 次。对于电商农户的内群体交往，本文基于内群体的概念内涵，结合乡土社会的特点，采用“从事网络销售且经常联系的亲友户数”进行测度，调查结果显示，受访者平均拥有 8.32 个从事网络销售且经常联系的亲友家庭。

根据既有的理论和前期经验研究，在分析的过程中，本文控制了被访问者的性别、年龄、受教育年限、是否党员、工作经验、工作经验平方、创业经历、网店经营年限、主营产品类型等有关变量。通常情况下，性别、年龄、受教育年限和是否党员等人口学变量在技术采纳行为方程和收入决定方程中需加以控制。人力资本是影响农民技术采纳行为和收益的重要因素，除了受教育年限以外，工作经验和创业经历尤其重要（曾亿武等，2019）。本文还借鉴 Mincer（1974）的工资决定方程，引入工作经验的平方项作为控制变量。网店经营年限反映“干中学”和知识积累的程度。鉴于农产品的产品特性可能带来的影响，本文还引入主营产品是否属于农产品对其进行控制。从表 1 的描述性统计结果可以看到，网店经营的主要决策者以男性为主，平均年龄 34 岁左右，平均接受约 11 年的正规教育，约 10% 属于党员，开网店前从事与电商相关的工作时间接近 1 年，开网店前创业累计次数是 0.68 次，平均网店经营时间超过 4 年，略超过一半的电商农户主要销售农产品。

表 1 变量说明及描述性统计

变量名称	变量定义说明	全部样本	
		均值	标准差
大数据使用行为	是否使用大数据产品辅助网店经营决策（是=1，否=0）	0.282	0.451
产品销售净收入	2017 年家庭产品销售净收入，单位：万元	29.275	5.907
培训经历	使用大数据产品之前平均每年参加培训的次数，单位：次	1.148	3.174
内群体交往	从事网络销售且经常联系的亲友户数，单位：户	8.315	11.788
性别	男=1，女=0	0.760	1.282
年龄	单位：周岁	33.823	8.951
受教育年限	单位：年	10.683	3.151
是否党员	是=1，否=0	0.105	0.308
工作经验	开网店前从事与电商相关的工作年限，单位：年	0.936	2.107
工作经验平方	工作经验的平方项	5.306	16.721
创业经历	开网店前创业累计次数，单位：次	0.678	1.271
网店经营年限	从第一家网店开始经营至今的累计年限，单位：年	4.218	2.636
主营产品类型	主营产品是否属于农产品（是=1，否=0）	0.554	0.498

根据是否使用大数据产品辅助网店经营决策，本文将电商农户划分为大数据使用型电商农户和普通电商农户。表 2 是两组农户各变量均值 t 检验的统计结果。可以看到，大数据使用型电商农户与普通电商农户在产品销售净收入、培训经历、内群体交往、年龄、网店经营年限和主营产品类型上表现出显著差异；具体地，相比较于普通电商农户，大数据使用型电商农户的产品销售净收入水

平更高，培训经历更丰富，内群体交往更紧密，年龄上更为年轻，网店经营年限更长，主营产品类型更加偏向于非农产品。

表2 大数据使用型电商农户与普通电商农户各变量均值的t检验结果

变量名称	大数据使用型电商农户		普通电商农户		均值差 (t检验)
	均值	标准差	均值	标准差	
产品销售净收入	52.042	8.673	20.328	2.125	31.714***
培训经历	1.944	5.505	0.915	1.630	1.029***
内群体交往	11.600	16.337	7.145	9.350	4.455***
性别	0.869	0.339	0.739	1.555	0.130
年龄	32.380	7.797	33.920	8.996	-1.540*
受教育年限	10.730	3.256	10.630	3.097	0.100
是否党员	0.060	1.925	0.125	0.316	-0.065
工作经验	1.164	2.192	0.753	1.892	0.411
工作经验平方	6.113	17.855	4.134	14.314	1.979
创业经历	0.654	1.001	0.691	1.419	-0.037
网店经营年限	4.804	2.745	3.956	2.463	0.848***
主营产品类型	0.421	0.496	0.590	0.493	-0.169***

注：\*、\*\*和\*\*\*分别表示在10%、5%和1%水平上显著。

### (三) 研究方法

本文将首先检验培训经历、内群体交往对电商农户大数据使用的影响效应，由于大数据使用行为属于二元选择变量，故采用二元Probit模型进行分析。具体模型如下：

$$P(D_i=1|X_i)=\Phi(X_i)=\alpha+\beta_1T_i+\gamma_1I_i+\lambda_1Z_i+\mu_i \quad (1)$$

(1)式中， $i$ 表示不同的农户， $P(D_i=1|X_i)$ 表示农户 $i$ 选择使用大数据产品的概率， $\Phi(X_i)$ 为标准正态分布的累积分布函数， $T$ 是培训经历变量， $I$ 是内群体交往变量， $Z$ 是控制变量， $\alpha$ 是常数项， $\mu$ 是随机干扰项。考虑进一步验证模型结果的稳健性，本文将在进行二元Probit模型估计的同时，开展二元Logit模型以及线性概率模型（Linear Probability Model，简称LPM）估计作为对照。

其次，为了测度大数据使用对电商农户收入水平的影响效应，构建如下半对数模型：

$$\ln Y_i=\alpha+\delta D_i+\beta X_i+\varepsilon_i \quad (2)$$

(2)式中， $i$ 表示不同的农户， $Y$ 表示电商农户的产品销售净收入， $D$ 为农户是否使用大数据产品辅助网店经营决策的虚拟变量， $X$ 是控制变量， $\alpha$ 是常数项， $\varepsilon$ 是随机干扰项。若电商农户的大数据使用行为是随机的，则半弹性系数 $\delta$ 能够精确地反映大数据使用的增收效应。然而，现实中，电商农户的大数据使用行为是具有主观能动性的决策结果，并非随机事件，如果没有考虑这个潜在的选择过程而直接使用OLS估计上述模型，参数估计结果将是有偏的，因为潜在选择模型和上述模

型的非观测因素是相互关联的 (Shaver, 1998), 导致  $\varepsilon$  与  $D$  相关, 换言之, 电商农户的大数据使用行为是内生的。

为克服大数据使用行为的内生性问题, 本文引入熵均衡 (Entropy Balancing) 方法对数据进行预处理, 以在最大限度上控制自选择性所导致的估计偏误。熵均衡法由 Hainmueller (2012) 提出, 它预先设置一组平衡性约束与规范性约束条件, 并通过计算一组与约束条件相匹配的最优权重, 使得处理组和对照组在特定矩阵下实现样本的精确匹配。熵均衡法可以在整体样本中为每一个处理组样本匹配一个与之极为相似的对照组样本, 做到保留所有样本的有用信息, 匹配后的协变量标准化均值差及均值差异性检验也更为稳健, 结果更为可靠 (Hainmueller and Xu, 2013; 邓睿, 2019)。考虑进一步验证模型结果的稳健性, 本文还将同时使用 OLS 回归和倾向得分匹配方法进行对照。

最后, 为了验证大数据使用对电商农户内部收入差距的影响, 本文采用分位数回归 (Quantile Regression) 的方法进行实证分析。在 (2) 式基础上, 构建分位数计量模型如下:

$$\ln Y_{i,q} = \alpha + \delta_q D_{i,q} + \beta_q X_{i,q} + \varepsilon_{i,q} \quad (3)$$

(3) 式中,  $Y_{i,q}$ 、 $D_{i,q}$  和  $X_{i,q}$  分别表示第  $i$  个农户在  $q$  分位数上的产品销售净收入、大数据使用行为和控制变量。分位数回归模型能够分别估计大数据使用行为对于各收入组电商农户收入的边际贡献, 如果大数据使用对于低收入电商农户收入的边际贡献大于中等收入电商农户和高收入电商农户, 说明大数据使用具有缩小电商农户内部收入差距的作用, 反之则是扩大收入差距。

## 四、实证结果分析

### (一) 培训经历、内群体交往对电商农户大数据使用的影响

表 3 报告了电商农户大数据使用行为影响因素的回归结果。其中, 方程 (1) 和方程 (2) 对应的是二元 Probit 模型, 方程 (3) 对应的是二元 Logit 模型, 方程 (4) 对应的是 LPM; 方程 (1) 仅引入培训经历和内群体交往两个核心解释变量, 没有考虑其他控制变量, 方程 (2)、(3) 和 (4) 在方程 (1) 的基础上增加性别等控制变量。从表 3 可以看到, 培训经历和内群体交往均对电商农户的大数据使用行为产生显著的正向影响, 方程 (2) 估计结果显示, 电商农户参与电商技能培训的次数每增加 1 次, 其使用大数据产品的可能性将提升 2% 左右; 电商农户拥有的从事网络销售且经常联系的亲友家庭每增加 1 户, 其使用大数据产品的可能性将提升 0.5%。实证结果具有良好的稳健性, 表明以培训为主要形式的知识转移和以内群体交往为核心的知识溢出对于驱动电商农户使用大数据产品辅助其网店经营决策具有重要的积极作用, 与理论分析预期一致, 假说 1 和假说 2 得到验证。

控制变量方面, 工作经验、工作经验平方、网店经营年限和主营产品类型同样显著影响电商农户的大数据使用行为, 而性别、年龄、受教育年限、是否党员和创业经历没有产生显著影响。工作经验的回归系数为正, 与此同时, 工作经验平方的回归系数为负, 说明工作经验对电商农户大数据使用行为产生先促进后抑制的作用, 两者呈现“倒 U 型”关系。网店经营年限的回归系数为正, 说明从事网店经营的时间越久, 则电商农户使用大数据产品的可能性越大。主营产品类型的回归系数

为负，说明相比于农产品，主营非农产品的电商农户具有更高的大数据使用倾向。

表3 电商农户大数据使用行为影响因素的回归结果

变量	(1)	(2)	(3)	(4)
	Probit	Probit	Logit	LPM
培训经历	0.026** (0.012)	0.021* (0.012)	1.107* (0.062)	0.020* (0.011)
内群体交往	0.006*** (0.002)	0.005*** (0.002)	1.027*** (0.012)	0.005*** (0.002)
性别	—	0.004 (0.019)	1.107 (0.093)	0.004 (0.019)
年龄	—	-0.004 (0.004)	0.980 (0.018)	-0.003 (0.003)
受教育年限	—	0.006 (0.009)	1.032 (0.045)	0.005 (0.008)
是否党员	—	0.019 (0.042)	1.090 (0.208)	-0.002 (0.018)
工作经验	—	0.099*** (0.037)	1.627*** (0.296)	0.093*** (0.034)
工作经验平方	—	-0.015*** (0.005)	0.930*** (0.026)	-0.013*** (0.004)
创业经历	—	-0.016 (0.253)	0.925 (0.115)	-0.011 (0.019)
网店经营年限	—	0.027*** (0.010)	1.144*** (0.059)	0.026*** (0.010)
主营产品类型	—	-0.119** (0.052)	0.560** (0.144)	-0.105** (0.050)
常数项	—	—	0.286 (0.247)	0.214 (0.165)
Pseudo-R <sup>2</sup> 或 R <sup>2</sup>	0.035	0.088	0.087	0.102
LR 或 F 统计量	16.060	38.120	37.850	3.540
样本容量	397	397	397	397

注：\*、\*\*和\*\*\*分别表示在10%、5%和1%水平上显著，括号内为稳健标准误；方程（1）和（2）汇报的是自变量的边际效应，方程（3）汇报的是自变量的优势比（odds ratios）。

### （二）大数据使用对电商农户收入水平的影响

1.简单 OLS 回归结果。在表 4 中，方程（5）暂不考虑其他控制变量，仅估计大数据使用行为对电商农户产品销售净收入的影响效应，结果显示，大数据使用显著促进了电商农户增收；方程（6）引入其他控制变量，大数据使用行为的增收效应依然成立，大数据使用型电商农户比普通电商农户的收入高 97.0% ( $e^{0.678}-1=0.970$ )。然而，简单的 OLS 回归由于没有考虑到自选择的内生性问题，其

估计结果需谨慎对待。此外，培训经历和内群体交往也对电商农户的收入产生了显著的正向作用，综合方程（4）、方程（6）和方程（7）可以发现，大数据使用行为在电商农户“培训经历—收入水平”和“内群体交往—收入水平”的正向关系中发挥了部分中介作用，换言之，培训经历和内群体交往的部分增收效应是通过激发电商农户的大数据使用行为来实现的。其他控制变量包括性别、工作经验、工作经验平方和网店经营年限同样显著影响了电商农户的收入水平。

表 4 大数据使用对电商农户收入影响的 OLS 回归结果

变量	(5)	(6)	(7)
大数据使用行为	1.230*** (0.185)	0.678*** (0.174)	— —
培训经历	—	0.067** (0.030)	0.081*** (0.028)
内群体交往	—	0.020*** (0.008)	0.024*** (0.007)
性别	—	0.158*** (0.056)	0.161** (0.066)
年龄	—	-0.017 (0.012)	-0.019 (0.012)
受教育年限	—	0.020 (0.027)	0.024 (0.028)
是否党员	—	-0.076 (0.048)	-0.077 (0.047)
工作经验	—	0.222** (0.110)	0.286** (0.115)
工作经验平方	—	-0.036** (0.014)	-0.045*** (0.015)
创业经历	—	0.055 (0.042)	0.047 (0.041)
网店经营年限	—	0.219*** (0.036)	0.237*** (0.036)
主营产品类型	—	-0.197 (0.174)	-0.272 (0.173)
常数项	11.118*** (0.103)	10.445*** (0.556)	10.582*** (0.563)
R <sup>2</sup>	0.102	0.258	0.227
F 统计量	52.940	13.450	12.290
样本容量	397	397	397

注：\*、\*\*和\*\*\*分别表示在 10%、5%和 1%水平上显著，括号内为稳健标准误。

2.熵均衡法估计结果。由于 OLS 回归无法有效控制处理组（即大数据使用型电商农户样本）和控制组（即普通电商农户样本）的协变量差异，本文进一步采用熵均衡方法对协变量设定一阶矩（均

值)、二阶矩(方差)和三阶矩(偏度)等约束条件并采用其自动计算出来的最优权重作为平衡权重进行调整,使两组样本在满足约束条件下实现精确匹配,最大限度上控制样本的选择性偏误。表5报告了协变量在熵均衡处理前后的均值、方差及匹配性检验结果。可以看到,匹配前,处理组和控制组的协变量的均值和方差均有较大差异,经熵均衡处理后,协变量的均值和方差的差异明显缩小。为进一步检验熵均衡结果的可靠性,可以进一步计算处理组和控制组之间的标准化均数差(SMD)并进行均值差异t检验,结果显示,处理后两组样本之间的标准化均数差全部为0,协变量的均值差异t检验的P值全部为1,表明处理组和控制组的各个协变量数据已实现精确匹配。

表5 熵均衡法处理后的协变量匹配性检验

协变量		均值		方差		标准化均数差 SMD	均值差异 t检验P值
		处理组	控制组	处理组	控制组		
培训经历	熵均衡前	1.944	0.915	30.300	2.657	0.063	0.004
	熵均衡后	1.944	1.944	30.300	10.010	0	1
内群体交往	熵均衡前	11.600	7.145	266.900	87.430	0.001	0.001
	熵均衡后	11.600	11.600	266.900	236.000	0	1
性别	熵均衡前	0.869	0.739	0.115	2.419	0.063	0.274
	熵均衡后	0.869	0.869	0.115	3.171	0	1
年龄	熵均衡前	32.380	33.920	60.790	80.930	-0.020	0.068
	熵均衡后	32.380	32.380	60.790	54.780	0	1
受教育年限	熵均衡前	10.730	10.630	10.600	9.589	0.010	0.938
	熵均衡后	10.730	10.730	10.600	13.160	0	1
是否党员	熵均衡前	0.060	0.125	3.705	0.100	-0.033	0.108
	熵均衡后	0.060	0.060	0.057	0.057	0	1
工作经验	熵均衡前	1.164	0.753	4.804	3.581	0.104	0.275
	熵均衡后	1.164	1.164	4.804	4.779	0	1
工作经验平方	熵均衡前	6.113	4.134	318.800	204.900	0.008	0.663
	熵均衡后	6.113	6.113	318.800	251.600	0	1
创业经历	熵均衡前	0.654	0.691	1.002	2.013	-0.021	0.771
	熵均衡后	0.654	0.654	1.002	1.009	0	1
网店经营年限	熵均衡前	4.804	3.956	7.537	6.067	0.130	0.007
	熵均衡后	4.804	4.804	7.537	6.423	0	1
主营产品类型	熵均衡前	0.421	0.590	0.246	0.243	-0.693	0.002
	熵均衡后	0.421	0.421	0.246	0.245	0	1

注:  $SMD = (\bar{X}_t - \bar{X}_c) / \sqrt{[S_t^2(n_t - 1) + S_c^2(n_c - 1) / (n_t + n_c - 2)]}$ , 其中  $\bar{X}_t$ 、 $\bar{X}_c$  分别为处理组和控制组中各变量的均值,  $S_t^2$ 、 $S_c^2$  分别为处理组和控制组中各变量的方差,  $n_t$ 、 $n_c$  分别为处理组和控制组的样本容量。

表6是采用经过熵均衡方法处理后的数据重新估计的电商农户大数据使用的增收效应。从表6可以看到,剔除样本选择偏误以后,大数据使用行为在10%水平上显著正向影响电商农户的收入水平。具体地,大数据使用型电商农户比普通电商农户的收入高出53.4% ( $e^{0.428} - 1 = 0.534$ ),增收效果依然相当明显。调查显示,大数据使用型电商农户在大数据产品上平均花费6272.2元,可见电商农

户每投入 1000 元的大数据使用成本，大约平均增收 8.5%，单位成本的收益增长率较高。

与表 4 方程（6）比较可知，由于自选择性问题的存在，直接的 OLS 回归会导致严重高估大数据使用的增收效应，产生夸大的判断。从理论上讲，大数据产品的使用通过为电商农户带来覆盖面广、时效性强、精准度高、前瞻力大的信息服务和决策指导，能够显著提升电商农户的经营能力，促进收入增长。但在实践过程中，大数据使用的增收效果具体达到何种程度，跟平台企业的大数据产品开发能力以及电商农户的使用深度密切相关。一方面，在目前的中国，基于电子商务平台的大数据产品尚处于早期开发的发展阶段，大数据技术在信息收集、处理和分析上达到比较成熟的高度还需要经历一个过程；另一方面，电商农户对大数据产品的使用也在初探时期，并且他们根据大数据产品的分析结果对生产经营中的薄弱环节有针对性地改进本身也需要一定的时间。关于大数据使用对电商农户收入的提升作用，熵均衡匹配方法的测算结果要比 OLS 回归结果更加合理，更加契合客观的实际情况。

表 6 熵均衡处理后电商农户大数据使用的增收效应

变量	回归系数	稳健标准误	t 统计量	P 值	95%水平下置信区间
大数据使用行为	0.428*	0.243	1.83	0.069	[-0.033, 0.888]
常数项	11.869***	0.190	62.42	0.000	[11.495, 12.243]
R <sup>2</sup>			0.018		
F 统计量			3.330		
样本容量			397		

注：\*、\*\*和\*\*\*分别表示在 10%、5%和 1%水平上显著。

3.稳健性检验：倾向得分匹配。除了熵均衡法以外，另一种处理自选择性问题的经典方法是倾向得分匹配（Propensity Score Matching，简称 PSM）。PSM 以倾向得分为依据，从控制组的样本中为每个处理组样本寻找并构造一个除了处理行为以外其他可观测特征都极为接近的参照对象，于是这个参照对象的结果可视为处理组样本对象在接受处理之前的结果，因此处理组样本对象的结果减去这个参照对象的结果所得的变量差值即为处理行为所带来的净效应。表 7 报告了电商农户大数据使用增收效应的 PSM 测算结果<sup>①</sup>，可以看到，大数据使用对电商农户的收入水平产生了显著正向作用的论断同样得到支持。PSM 的测算结果与熵均衡法相比稍有分歧，存在一定程度的高估，但与 OLS 回归结果相比，它已经大幅降低了选择性偏差。综合 OLS 回归、熵均衡法和 PSM 的估计结果，可以认定前文的假说 3 得到验证。

表 7 电商农户大数据使用增收效应的倾向得分匹配估计结果

匹配方法	处理组	对照组	ATT 值	标准误
最近邻匹配（1—1 匹配）	12.238	11.523	0.715***	0.283
最近邻匹配（1—5 匹配）	12.238	11.573	0.666***	0.203

<sup>①</sup>匹配后全部协变量的标准化偏差都在 10%以内，两组农户的可观测特征不再存在显著差异。限于篇幅，平衡性检验结果在此不做详细汇报，感兴趣的读者可向作者索要。

半径匹配 (r=0.005)	12.189	11.531	0.660**	0.215
半径匹配 (r=0.010)	12.222	11.624	0.597**	0.200
核匹配 (窗宽=0.06)	12.238	11.598	0.641**	0.188
核匹配 (窗宽=0.10)	12.239	11.573	0.655**	0.185
马氏匹配	12.296	11.468	0.828***	0.183

注：\*、\*\*和\*\*\*分别表示在 10%、5%和 1%水平上显著；最近邻匹配、半径匹配和核匹配使用自助法获得标准误，马氏匹配使用 AI 异方差稳健标准误。

### (三) 大数据使用对电商农户内部收入差距的影响

上述实证结果仅反映了大数据使用对电商农户收入水平的平均影响，属于群体平均效应，事实上电商农户在收入水平上往往存在分布异质性。为了进一步观察大数据使用的增收效应是否存在群体内部差异，本文采用分位数回归方法（使用自助法计算协方差矩阵）进行实证检验。

表 8 结果显示，在不同分位数上，大数据使用均对电商农户收入产生了显著正向作用，但从回归系数大小的变动趋势可以发现，随着产品销售净收入分位数的上升，大数据使用行为对电商农户收入的提升效应总体上是逐步减小的，这反映出大数据使用对低收入电商农户的增收作用要更为明显，而这种增收作用在高收入电商农户中相对较小，因此，大数据产品的使用能够发挥出缩小电商农户内部收入差距的作用，前文提出的假说 4 得到验证。这是一个很有意思的研究结果。曾亿武等（2018）以江苏沭阳为例的实证研究发现，电子商务使得农户内部收入差距明显扩大，这是“二级数字鸿沟”的一个体现，即数字化工具使用能力上的差异所导致的发展不平等程度加剧。而本文的研究表明，基于电子商务平台的大数据产品的使用却能够缩小电商农户内部收入差距，这背后隐含的是一种“用更高级的数字技术去缩小二级数字鸿沟”的新思维。一方面，部分信息技术会不可避免地造成社会群体内部出现数字鸿沟；另一方面，通过开发一些契合低收入群体的新技术帮助他们快速且更加充分地获取数字红利，这些技术要有助于切实有效弥补低收入群体的资源禀赋劣势，尤其是信息上和人力资本上的不足，从而实现其收入显著增长。

表 8 大数据使用对电商农户收入影响的分位数回归结果

变量	分位数				
	$\theta=0.10$	$\theta=0.25$	$\theta=0.50$	$\theta=0.75$	$\theta=0.90$
大数据使用行为	1.202** (0.067)	0.795*** (0.237)	0.546*** (0.197)	0.645*** (0.186)	0.399* (0.250)
其他变量	控制	控制	控制	控制	控制
R <sup>2</sup>	0.147	0.160	0.169	0.185	0.163
样本容量	397	397	397	397	397

注：\*、\*\*和\*\*\*分别表示在 10%、5%和 1%水平上显著，括号内为稳健标准误。

## 五、总结与讨论

基于电子商务平台的大数据产品是电子商务与大数据分析技术相结合的产物。伴随着中国农村

电子商务的快速发展，越来越多的农村家庭成为电商专业户。电子商务平台企业开发的大数据产品能够辅助电商农户制定更加科学准确的网店经营决策，基于山东曹县、江苏沭阳和浙江临安三个地区的问卷调查显示，在农村电子商务发展较好的地区，接近三成的电商农户在其网店经营过程中使用了大数据产品。电商农户对大数据的使用本质上是一种新技术采纳行为，受到个体对于新技术的主观感知的影响。以培训为主要形式的知识转移和以内群体交往为核心的知识溢出构成电商农户知识获取的两大路径，其通过改变电商农户对大数据分析技术的触碰概率和主观感知，提升他们的技术警觉性，以及产生先驱效应，最终促进部分电商农户在现阶段已经率先使用了大数据产品辅助网店经营决策。实证研究结果表明，培训经历越丰富、内群体交往越紧密的电商农户具有越高的大数据使用倾向。与普通电商农户相比，大数据使用型电商农户在信息获取方面发生了质的改变，在掌握了更加及时、全面和精准的信息以后，电商农户可以更有针对性地改进自身薄弱环节，更有效地做到需求导向和用户本位主义，以及更有前瞻性地积极谋求先动优势，进而不断提升产品销售绩效。此外，在大数据时代，电商农户的经营收入不再严格遵循以往那种与时间同步演进的经验积累机制，低收入农户在大数据技术的辅助下，可以弥补自身在经验积累和人力资本上的差距，获得更高的边际收益。实证研究证实，大数据使用对电商农户的收入水平产生了显著的正向作用，并有助于缩小电商农户内部收入差距。

本文研究结论对于政府相关政策制定和电商农户经营实践具有重要启示。促进农村电子商务高质量发展以及推动大数据广泛应用和深度服务于农村地区，是适应时代发展潮流的必然选择。随着农村电子商务的持续发展以及大数据分析技术的不断完善，基于电子商务平台的大数据产品应用将逐步进入加速发展阶段，大数据使用的积极效应也将进一步显现出来。为了适应并积极满足客观的发展需要，政府部门应主动增进对大数据的了解和重视，加大对农村电商技能培训的支持力度，不断提升培训的质量和档次，尤其是强化大数据分析方面的培训内容，着力建设促进电商农户之间交流与合作的综合性平台，鼓励电商农户成立相关民间团体，加强信息传播与经验分享，对低收入农户给予大数据应用方面的培训、产品使用上的积极引导和适当补贴。对于电商农户，要提高对大数据产品的学习和使用意识，积极参与电商技能培训，加强与其他网商之间的交流，不断提升自身的认知水平和专业知识。

本文研究存在一定的局限性，尽管本文通过使用比倾向得分匹配方法更具优势的熵均衡法较好地降低了样本选择偏差，但是该方法同倾向得分匹配方法一样，无法避免不可观测因素对估计偏误的潜在影响。受制于数据采集与变量设计，本文无法进一步采用工具变量模型、处理效应模型、内生转换回归模型等更为有效的计量方法。寻找科学合理的工具变量，更加彻底解决内生性问题，提升实证结果的精确度，是后续研究有待拓展的重要方向。另外，本文对电商农户内部大数据使用行为的差异性仅提供了知识获取渠道视角的解释，未来研究还可以寻求其他视角的解释，例如产品属性偏好的视角，可通过采用选择实验法挖掘影响电商农户对大数据产品使用偏好的关键产品属性和支付意愿，并对电子商务平台的大数据产品提出完善建议。

大数据分析技术的发展无疑为中国“三农”问题之破解带来新的机遇和希望。本文涉及的基于

电子商务平台的大数据产品也仅是大数据分析技术的一种应用类型，它的开发初衷并非专门针对“三农”。但本文关于大数据使用增收效应的研究发现，同样可以为政府加快推进农业大数据建设提供依据和借鉴。2019年中央“一号文件”明确指出：“推进重要农产品全产业链大数据建设，加强国家数字农业农村系统建设。”目前中国正处于推进农产品全产业链大数据建设的起步阶段，主要开展了生猪、大蒜、花生、柑橘等若干种重要农产品的大数据建设试点，探索建立分品种分产业的全产业链大数据。由于农业的产业特性以及中国农业的弱质性尚未从根本上改变，农业大数据建设与交通、医疗、工业等领域的大数据建设相比，面临的困难和问题更为复杂，任重而道远。但不管怎样，加快实现大数据与“三农”的深度融合发展是大势所趋，中国要积极创造条件，抓住机遇，抢占新经济的制高点。

#### 参考文献

- 1.邓睿, 2019:《健康权益可及性与农民工城市劳动供给——来自流动人口动态监测的证据》,《中国农村经济》第4期。
- 2.费孝通, 1998:《乡土中国》,北京:北京大学出版社。
- 3.郭红东、周惠珺, 2013:《先前经验、创业警觉与农民创业机会识别——一个中介效应模型及其启示》,《浙江大学学报(人文社会科学版)》第4期。
- 4.何大安、任晓, 2018:《互联网时代资源配置机制演变及展望》,《经济学家》第10期。
- 5.卡斯特, 2006:《网络社会的崛起》,夏铸九、王志弘等译,北京:社会科学文献出版社。
- 6.李海舰、田跃新、李文杰, 2014:《互联网思维与传统企业再造》,《中国工业经济》第10期。
- 7.李琪、唐跃桓、任小静, 2019:《电子商务发展、空间溢出与农民收入增长》,《农业技术经济》第4期。
- 8.鲁钊阳、廖杉杉, 2016:《农产品电商发展的增收效应研究》,《经济体制改革》第5期。
- 9.路征、张益辉、王坤、董冠琦, 2015:《我国“农民网商”的微观特征及问题分析——基于对福建省某“淘宝镇”的调查》,《情报杂志》第12期。
- 10.罗珉、李亮宇, 2015:《互联网时代的商业模式创新:价值创造视角》,《中国工业经济》第1期。
- 11.屈小博, 2013:《培训对农民工人力资本收益贡献的净效应——基于平均处理效应的估计》,《中国农村经济》第8期。
- 12.芮正云、方聪龙, 2018:《互联网嵌入与农村创业者节俭式创新:二元机会开发的协同与平衡》,《中国农村经济》第7期。
- 13.邵占鹏, 2019:《农民网商网络空间表意能力的局限与反思》,《西北农林科技大学学报(社会科学版)》第2期。
- 14.王金杰、牟韶红、盛玉雪, 2019:《电子商务有益于农村居民创业吗?——基于社会资本的视角》,《经济与管理研究》第2期。
- 15.王昕宇、黄海峰, 2016:《我国农民网商的演进路径及发展对策》,《中州学刊》第8期。
- 16.魏延安, 2017:《农村电商:互联网+三农案例与模式》,北京:电子工业出版社。

- 17.曾亿武、陈永富、郭红东, 2019:《先前经验、社会资本与农户电商采纳行为》,《农业技术经济》第3期。
- 18.曾亿武、郭红东, 2018:《农产品淘宝村集群的形成及对农户收入的影响研究——以江苏沭阳为例》,北京:中国农业出版社。
- 19.曾亿武、郭红东、金松青, 2018:《电子商务有益于农民增收吗?——来自江苏沭阳的证据》,《中国农村经济》第2期。
- 20.展进涛、黄宏伟, 2016:《农村劳动力外出务工及其工资水平的决定:正规教育还是技能培训——基于江苏金湖农户微观数据的实证分析》,《中国农村观察》第2期。
- 21.张庆民、孙树垒、吴士亮、李大芳, 2019:《淘宝村农户网商群体持续成长演化研究》,《农业技术经济》第1期。
- 22.张银、李燕萍, 2010:《农民人力资本、农民学习及其绩效实证研究》,《管理世界》第2期。
- 23.张兆曙, 2017:《“互联网+”的技术红利与非预期后果》,《天津社会科学》第5期。
- 24.Ajzen, I., 1991, “The Theory of Planned Behavior”, *Organizational Behavior and Human Decision Processes*, 50(2):179-211.
- 25.Atasoy, H., 2013, “The Effects of Broadband Internet Expansion on Labor Market Outcomes”, *Industrial & Labor Relations Review*, 66 (2): 315-345.
- 26.Becker, G. S., 1964, *Human Capital: A Theoretical and Empirical Analysis with special reference to education (2nd edition)*, New York: Columbia University Press.
- 27.Carrer, M. J., H. M. de Souza Filho, and M. Batalha, 2017, “Factors Influencing the Adoption of Farm Management Information Systems (FMIS) by Brazilian Citrus Farmers”, *Computers and Electronics in Agriculture*, 138: 11-19.
- 28.Chi, M., A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, 2016, “Big Data for Remote Sensing: Challenges and Opportunities”, *Proceedings of the IEEE*, 104 (11): 2207-2219.
- 29.Cho, K. M., and D. J. Tobias, 2010, “Impact Assessment of Direct Marketing of Small- and Mid-sized Producers through Food Industry Electronic Infrastructure”, International Conference on World Food System.
- 30.Coble, K. H., A. K. Mishra, S. Ferrell, and T. Griffin, 2018, “Big Data in Agriculture: A Challenge for the Future”, *Applied Economic Perspectives and Policy*, 40(1): 79-96.
- 31.Cooper, J., M. Noon, C. Jones, E. Kahn, and P. Arbuckle, 2013, “Big Data in Life Cycle Assessment”, *Journal of Industrial Ecology*, 17(6): 796-799.
- 32.Davis, F. D., 1989, “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology”, *MIS Quarterly*, 13(3): 319-340.
- 33.Ellemers, N., R. Spear, and B. Doosje, 1999, *Social Identity: Context, Commitment and Content*, Oxford: Blackwell Science.
- 34.Fallah, M. H., and S. Ibrahim, 2004, “Knowledge Spillover and Innovation in Technological Clusters”, IAMOT 2004 Conference, Washington D. C.

- 35.Fan, Q., and V. B. S. Garcia, 2018, "Information Access and Smallholder Farmers' Market Participation in Peru", *Journal of Agricultural Economics*, 69(2): 476-494.
- 36.Galloway, L., J. Sanders, and D. Deakins, 2011, "Rural Small Firms' Use of the Internet: From Global to Local", *Journal of Rural Studies*, 27(3): 254-262.
- 37.Grossman, M., 1972, "On the Concept of Health Capital and the Demand for Health", *Journal of Political Economy*, 80(2): 223-255.
- 38.Hainmueller, J., 2012, "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies", *Political Analysis*, 20 (1): 25-46.
- 39.Hainmueller, J., and Y. Xu, 2013, "ebalance: A Stata Package for Entropy Balancing", *Journal of Statistical Software*, 54(7): 1-18.
- 40.Kim, G. H., S. Trimi, and J. H. Chung, 2014, "Big-data Applications in the Government Sector", *Communications of the ACM*, 57(3): 78-85.
- 41.Leong, C., S. Pan, S. Newell, and L. Cui, 2016, "The Emergence of Self-Organizing E-commerce Ecosystems in Remote Villages of China: A Tale of Digital Empowerment for Rural Development", *MIS Quarterly*, 40(2): 475-484.
- 42.Lokers, R., R. Knapen, S. Janssen, Y. van Randen, and J. Jansen, 2016, "Analysis of Big Data Technologies for Use in Agro-Environmental Science", *Environmental Modelling & Software*, 84: 494-504.
- 43.Lucas, R. E., 1988, "On the Mechanics of Economic Development", *Journal of Monetary Economics*, 22 (1): 3-42.
- 44.Mincer, J., 1974, *Schooling, Experience and Earnings*, New York: Columbia University Press for the National Bureau of Economic Research.
- 45.Pivoto, D., B. Barham, P. D. Waquil, C. R. Foguesatto, V. F. D. Corte, D. Zhang, and E. Talamini, 2018, "Factors Influencing the Adoption of Smart Farming by Brazilian Grain Farmers", *International Food and Agribusiness Management Review*, 22(4): 571-588.
- 46.Rodriguez, D., P. de Voil, M. C. Rufino, M. Odoendo, and M. T. van Wijk., 2017, "To Mulch or to Munch? Big Modelling of big data", *Agricultural Systems*, 153: 32-42.
- 47.Schultz, T. W., 1961, "Investment in Human Capital", *The American Economic Review*, 51(1): 1-17.
- 48.Shaver, J. M., 1998, "Accounting for Endogeneity when Assessing Strategy Performance: Does Entry Mode Choice Affect FDI Survival?", *Management Science*, 44(4): 571-585.
- 49.Tajfel, H., and J. C. Turner, 1986, "The Social Identity Theory of Inter-group Behavior", *Psychology of Inter-group Relations*, 13(3): 7-24.
- 50.Toscano, P., A. Castrignanò, S. F. Di Gennaro, A. V. Vonella, D. Ventrella, and A. Matese, 2019, "A Precision Agriculture Approach for Durum Wheat Yield Assessment Using Remote Sensing Data and Yield Mapping", *Agronomy*, 437(9): 1-18.

51. Turland, M., and P. Slade, 2019, "Farmers' Willingness to Participate in a Big Data Platform", *Agribusiness*, DOI: 10.1002/agr.21627.

52. Turner, J. C., I. Sachdev, and M. A. Hogg, 1983, "Social Categorization, Interpersonal Attraction and Group Formation", *British Journal of Social Psychology*, 22(3): 227-239.

53. Waga, D., and K. Rabah, 2014, "Environmental Conditions' Big Data Management and Cloud Computing Analytics for Sustainable Agriculture", *World Journal of Computer Application and Technology*, 2(3): 73-81.

54. Zapata, S. D., C. E. Carpio, O. Isengildina-Massa, and R. D. Lamie, 2013, "The Economic Impact of Services Provided by an Electronic Trade Platform: The Case of MarketMaker", *Journal of Agricultural and Resource Economics*, 38(3): 359-378.

(作者单位：<sup>1</sup>杭州师范大学经济与管理学院；

<sup>2</sup>浙江大学中国农村发展研究院)

(责任编辑：陈静怡)

## The Driving Forces and Income-added Effects of Big Data Usage by Rural E-commerce Households

Zeng Yiwu Zhang Zenghui Fang Huli Guo Hongdong

**Abstract:** With the advancement of the new technological revolution, big data has been gradually integrated with "agriculture, rural areas and farmers". This article focuses on the phenomenon of rural e-commerce households using big data products to assist them in making online business decisions. After presenting a theoretical analysis, it further empirically analyzes the driving forces and income-added effects of big data usage by rural e-commerce households through adopting the binary probit model, entropy balancing method and quantile regression with the data collected from rural e-commerce households in Caoxian County of Shandong, Shuyang County of Jiangsu and Lin'an District of Zhejiang. It finds that in rural areas with a high level of e-commerce development, nearly 30% of rural e-commerce households use big data products in their operation of online shops, and knowledge transfer in the form of training and knowledge spillover in the form of in-group communication both play an important and positive role in driving rural e-commerce households to use big data. The empirical results also show that the use of big data significantly improves the income level of rural e-commerce households and helps to narrow the income gap within rural e-commerce households.

**Key Words:** Rural E-commerce Household; Big Data; Training Experience; In-group Communication; Income Effect