

大数据发展对制造业企业 全要素生产率的影响机理研究*

史 丹 孙光林

内容提要:本文以 2014—2018 年中国制造业企业上市公司数据为样本,在构建大数据发展影响制造业企业全要素生产率理论框架的基础上,实证检验了大数据发展对制造业企业全要素生产率的影响机理。研究发现:(1)大数据发展对制造业企业全要素生产率具有显著正向影响,提高大数据发展水平有利于提升制造业企业全要素生产率。(2)内在机理表明,企业创新、要素配置与数据赋能的中介效应显著,大数据发展可以通过促进企业创新、优化资本与劳动要素配置效率,以及数据赋能来提升制造业企业全要素生产率。(3)异质效应表明,大数据发展对民营企业和小企业全要素生产率的作用效果要大于国有企业和大企业。

关键词:大数据发展 制造业企业 要素配置 数据赋能 全要素生产率

作者简介:史 丹,中国社会科学院工业经济研究所所长、研究员、博士生导师,100006;

孙光林(通讯作者),中国社会科学院工业经济研究所博士后,100006,南京财经大学金融学院讲师、博士、硕士生导师,210023。

中图分类号:F062.5 **文献标识码:**A **文章编号:**1002-8102(2022)09-0085-16

一、问题提出

党的十九届四中全会将数据与传统的劳动力、资本、土地、技术等要素并列作为一种新的生产要素,提出通过分配机制的不断健全和完善,数据可作为生产要素按贡献参与分配。事实上,大数据技术发展通过数据要素发挥作用使制造业的发展方式由生产者主导向消费者主导方向转变(史丹,2018),然而,数据要素如何在提升制造业企业全要素生产率的过程中发挥作用,是值得探讨的理论问题。

2011 年,全球知名咨询公司麦肯锡在报告《大数据:创新、竞争和生产力的下一个前沿领域》中

* 基金项目:国家社会科学基金重大项目“推进新型工业化与经济体系优化升级研究”(21ZDA021);中国社会科学院登峰战略优势学科(产业经济)成果。作者感谢大连外国语大学商学院副教授艾永芳、匿名审稿人与编辑部老师提出的宝贵意见,当然,文责自负。孙光林电子邮箱:sunguanglin008@126.com。

首次提出“大数据”的概念,认为大数据是超出常规数据库获取、存储、管理与分析的数据集。国务院 2015 年在《促进大数据发展行动纲要》中指出“大数据是以容量大、类型多、存储速度快、应用价值高为主要特征的数据集合”。许宪春等(2019)认为大数据并不简单的是数据量大,还包括数据的价值高,大数据的核心是从海量数据中获取更准确、更深层次的价值,而不是简单的统计数据。Farboodi 等(2019)认为大数据是经济活动的副产品,其本质是一种信息,利用大数据能提升企业的经营效率。张叶青等(2021)将“大数据应用”定义为“企业收集、处理与利用的海量、高速、多样化的数据要素或资产”。关于大数据的作用,有学者认为,大数据技术能够降低交易成本和扩大市场范围(Lyytinen 等,2016),以及降低连接成本和提高资源配置效率。也有学者认为,数据资源要发挥作用必须与其他要素进行融合,才能实现数据价值增值(Chen 等,2015;肖静华等,2018),企业基于大数据资源分析出商业机会,与消费者形成良好的互动,能够做出更明智的生产决策(谢康等,2020)。事实上,上述对大数据的理解和定义主要是描述大数据本身的特性和作用,而忽视了大数据发挥作用所需要的制度条件 and 市场条件。因此,本文将大数据发挥作用所需要的制度条件 and 市场条件定义为“大数据发展”,即大数据发挥作用依赖的制度条件 and 市场条件构成了大数据发展的基本要件。

有关大数据与制造业之关系的研究主要集中于以下两个方面:一是关于大数据在制造业企业应用的相关研究,着重分析大数据在制造业企业产品设计、生产制造、行为决策和市场营销等方面的应用价值(Pietro,2018;孙新波等,2019);二是关于大数据与制造业产业发展关系的相关研究,如Zhang 和 Yang(2018)认为如果大数据在能源密集型制造业得到有效利用,能有效促进能源密集型制造业的绿色可持续发展;吕明元和麻林宵(2022)发现我国各省域大数据与制造业融合水平在不断提升。然而,对于大数据发展会对制造业企业全要素生产率造成怎样的影响,其内在作用机制是什么,已有研究尚未给出明确的解答。

本文的创新点主要体现在以下两个方面:一是根据大数据发展相关定义,从制度条件与市场条件两个方面构建大数据发展测度体系,在此基础上,从理论与实证双重角度系统探讨了大数据发展对制造业全要素生产率的直接效应与作用路径,有助于厘清大数据发展影响制造业企业全要素生产率的微观基础与作用机制;二是在我国加快建设创新型国家和实现数字发展战略的大背景下,全面刻画大数据发展对制造业企业全要素生产率的影响机理,对于解决我国具有丰富大数据资源的潜在价值和客观利用率偏低的现实矛盾,以及为更好地发挥大数据的作用服务于制造业提供了新的思路,具有重要的政策价值。

二、大数据发展指数构建与测度结果分析

(一)大数据发展指数构建

由于大数据发展是一个比较抽象的概念,很难找到一个单一指标来准确反映大数据发展水平。已有衡量大数据指标的相关方法主要集中在以下三个方面。一是使用文本分析法。该方法是基于上市公司年报的文本信息,抓取与大数据相关的关键词汇,“大数据”相关关键词汇在公司年报中出现的频率越高,表示该公司大数据应用水平越高(Farboodi 等,2019;张叶青等,2021;赵宸宇等,2021)。二是问卷调查法。如Chen 等(2015)从产品需求、设计与测试等方面设计了5个题目进行主观调查;谢康等(2020)基于该方法从大数据资源和大数据分析两个方面各设计若干问题对374家企业大数据情况进行了测度。三是指数构建法。如赵云辉等

(2019) 和王欣亮等 (2022) 从商用指数 (电子商务企业占总企业比重与每百家企业拥有网站数等) 和民用指数 (人均互联网接入量与各省百人计算机数等) 两个方面选择合适指标对大数据发展进行测度。

通过对比可以发现,文本分析法的基本假设是上市公司年报披露的信息是基于公司客观情况的陈述,然而,不少企业在大数据应用方面只是空喊口号,外界无法观察到公司的真实情况,不可避免会存在较大偏差。问卷调查法虽然能够通过设计问题与企业管理者直接对接,再以实地调查的方式揭示企业有关大数据的实际情况,但是调研样本数量相对较少,不易展开较大规模的调查以保证数据结果的可信度。根据本文大数据发展相关定义,本文研究更注重大数据发展的制度条件和市场条件等外部环境的好坏,因此,本文借鉴已有文献的做法,选择指数构建法对我国各省市自治区大数据发展水平进行测度 (赵云辉等,2019)。

近年来,国家信息中心、中国电子信息产业发展研究院以及一些高校科研机构等都曾发布过有关省域层面的大数据发展指数,但是,这些机构发布的大数据发展指数均未逐年连续编制,且不同机构发布的大数据发展指数因指标体系与评价方法等存在显著差异,因此本文不能直接借鉴现有的指数进行研究。同时,在我国大数据发展过程中,贵州省大数据产业发展起步较早、行动快、影响较大,而 2014 年是贵州省发展大数据的起始之年,且构建大数据发展指数的一些指标只更新到 2018 年。因此,本文选择构建我国 2014—2018 年省域大数据发展指数。根据本文大数据发展的相关定义,分别从制度条件与市场条件两个维度出发,本着代表性、系统性、独立性、动态性以及可操作性的原则进行指标选择。其中,制度条件指标包括大数据政策力度、大数据试点创新、数据开放平台与大数据服务通达度;市场条件指标包括电子商务发展指数、科技投入、大数据支持产业规模、消费能力、智能终端普及率和互联网普及率。

上述指标的度量方式及含义见表 1。鉴于熵值法可以客观地确定各个指标权重,因而,本文利用该方法对上述 10 个代表性指标进行加权平均以获得大数据发展指数 *bigdata*。另外,本文还通过计算这些指标的算术平均值 (*bigdata2*) 来度量大数据发展指数,以进行稳健性检验。

表 1 大数据发展指标评价体系		
一级指标	二级指标	指标含义
制度条件	大数据政策力度	大数据政策发表次数,发布次数越多表示政府对大数据的政策支持力度越大
	大数据试点创新	反映省域大数据发展在国家战略中的地位,使用是否存在大数据综合试验区来衡量:是 = 1,否 = 0
	数据开放平台	反映数据开放的范围和多样性,使用是否存在政府数据开放平台来表示:是 = 1,否 = 0
	大数据服务通达度	反映政府利用大数据提高公共服务的能力,该指数越大表示政府利用大数据提高公共服务的能力越强
市场条件	电子商务发展指数	反映大数据商用的成熟程度,该指数越大表示大数据商用的成熟度越高
	科技投入	反映科研投入规模,科研投入规模越大,对大数据相关产业的支持力度越大,使用规模以上工业企业 R&D 经费占 GDP 比重衡量

续表 1

一级指标	二级指标	指标含义
市场条件	大数据支持产业规模	反映该省域大数据相关产业发展情况,使用软件和信息技术服务业收入占 GDP 比重衡量
	消费能力	反映公众对信息消费的承受能力,该指标越大表示消费者消费潜力越大,使用居民交通通信支出占总消费支出的比重衡量
	智能终端普及率	反映大数据信息采集与市场应用的广泛度,智能终端普及率越高表示大数据信息采集和市场应用的范围越广泛
	互联网普及率	反映网络信息基础设施的完备度,互联网普及率越高表示大数据发展的基础设施支持越强

(二)测度结果分析

2014—2018 年各省区市大数据发展水平综合得分测度结果如表 2 所示,可以发现,我国大数据发展得分值在 0. 0042 和 0. 1298 之间,表明我国各省区市大数据发展水平较低,还有较大提升空间。由表 2 可知,大数据发展水平最高的是贵州省,大数据发展得分为 0. 1298,北京市、上海市、广东省与浙江省分别位于第 2~5 名,山西省、吉林省、海南省与青海省等 9 个省份大数据发展水平不足 0. 01,且集中分布在中西部地区,表明我国大数据发展水平具有较强的区域不均衡性。

表 2 2014—2018 年我国各省区市大数据发展得分值

排名	省区市	得分值	排名	省区市	得分值
1	贵州	0. 1298	16	宁夏	0. 0158
2	北京	0. 1142	17	四川	0. 0147
3	上海	0. 1010	18	陕西	0. 0137
4	广东	0. 0735	19	江西	0. 0134
5	浙江	0. 0706	20	新疆	0. 0104
6	湖北	0. 0552	21	安徽	0. 0103
7	山东	0. 0366	22	山西	0. 0098
8	天津	0. 0321	23	吉林	0. 0087
9	河南	0. 0318	24	海南	0. 0085
10	重庆	0. 0311	25	湖南	0. 0077
11	江苏	0. 0283	26	甘肃	0. 0063
12	辽宁	0. 0279	27	青海	0. 0058
13	河北	0. 0263	28	黑龙江	0. 0057
14	福建	0. 0248	29	云南	0. 0049
15	内蒙古	0. 0168	30	广西	0. 0042

我国 30 个省区市大数据发展水平的均值(M)为 0. 05,标准差(SD)为 0. 042,本文借鉴魏敏和李书昊(2018)的做法,将各省区市大数据发展水平按照明星型(得分等于或高于 $M + 0. 5SD$)、平庸型(得分介于 $M - 0. 5SD$ 至 $M + 0. 5SD$ 之间)与落后型(得分低于或等于 $M - 0. 5SD$)三个不同的

等级(见表 3)。明星型大数据发展省区市应该高于 0.07,分别是贵州、北京、上海、广东和浙江 5 个省市,大数据发展得分值依次为 0.1298、0.1142、0.1010、0.0735 和 0.0706,占全部省份的比例为 16.67%,上述省市中贵州得益于得天独厚的地理条件,而其他 4 个省市都是我国经济发展最好的省市,数字技术发展起步较早。平庸型省区市大数据发展得分值介于 0.029 和 0.070 之间,包括山东、天津、河南、湖北和重庆 5 个省市,占全部省份的比例为 16.67%,这些省市虽然重视大数据发展,但关注程度仍然不足。落后型省区市大数据发展得分值低于 0.029,包括辽宁、河北、江西、宁夏、四川和广西等 20 个省份,占全部省份的比重为 66.67%,可能这些省份对大数据发展的关注程度仍然不足,大数据发展较为落后,仍有较大的提升空间。进一步地,通过对比分析不同省区市的区域差异,可以发现,明星型省市除了贵州由于地理优势较早发展大数据产业以外,其他省市均位于东部经济发达地区。大数据发展较为平庸的省市分布较为均衡、东部和中部各有 2 个省市,西部地区有 1 个。对于落后型省区市而言,西部地区占据一半,东部和中部地区分别有 5 个和 6 个,这个结果进一步地表明我国大数据发展虽然具有明显的空间差异性,整体上东部地区要好于中西部地区,但目前差距仍然较小。

表 3 三类大数据发展水平的区域分布

类型	东部地区	中部地区	西部地区
明星型	北京、上海、广东、浙江	—	贵州
平庸型	山东、天津	河南、湖北	重庆
落后型	江苏、辽宁、河北、福建、海南	江西、安徽、山西、吉林、湖南、黑龙江	内蒙古、宁夏、四川、陕西、新疆、甘肃、青海、云南、广西

三、理论框架与研究假说

(一)大数据发展对制造业企业全要素生产率影响的直接效应

大数据发展能够使大数据资源更合理地流动与共享,从而有助于提高制造业企业全要素生产率。具体来看:第一,大数据发展水平越高,往往意味着该区域政府部门和商业部门的电子化水平越高,由此积累的数据资源越丰富,从而能够为制造业企业发展提供更多的数据支持。大数据发展不仅能够使丰富的数据资源成为制造业企业的生产要素,还能够增加信息的有效性,为制造业产品市场交易提供更优的匹配路径,减少因信息不对称、机会主义与有限理性等给制造业企业造成的搜寻成本与决策成本。第二,大数据发展水平越高的区域数据共享机制和交易机制更完善,有利于缓解大数据垄断对制造业企业全要素生产率的不利影响。数据要素的显著特点是低成本性、大规模性与易得性,这使得数据要素可以在制造业企业生产活动中被广泛使用。但是,大数据垄断会阻碍数据要素的合理流动,使数据要素的非竞争性功能无法在制造业企业中发挥作用,抑制了数据价值的创造能力,从而降低微观制造业企业的全要素生产率。第三,大数据发展水平越高的地域能够在制造业企业中形成良好的学习和使用大数据的氛围,制造业企业管理层更倾向于在实际生产决策中利用大数据资源,从而有助于提升制造业企业全要素生产率。一方面,制造业企业管理层利用大数据能够提高市场透明度,有助于制造业企业发现客户需求、细分市场、辅助决策支持和产品创新等,从而可以提高制造业企业的全要素生产率。另一方面,制造业企业有效率

地管理数据不仅可以提高劳动生产率与资本使用效率,还有利于提高制造业企业的创新能力,促进新产品开发以及服务创新等。基于以上理论分析,本文提出基本研究假说 H1。

H1:大数据发展能够提升制造业企业全要素生产率。

(二)大数据发展影响制造业企业全要素生产率的理论机理分析

1. 企业创新的中介理论机理

大数据发展能够通过促进制造业企业创新来提升全要素生产率,具体来看:第一,大数据记录了消费者选择制造业商品的行为,利用大数据技术可以挖掘消费者的隐藏需求,使消费者积累的各类数据流动起来,从而依靠数据信息互通将制造业产品生产者和消费者连接起来,这一过程将大幅度提升数据要素的边际产出,从而引发企业创新的正反馈效应。相对而言,大数据发展水平越高的区域越可能积累更多的数据资源,数据交易机制更为完善,制造业企业也更容易获取大数据,有利于制造业企业开展创新,进而提升制造业企业全要素生产率。第二,大数据发展有助于优化制造业企业创新模式,通过促进制造业企业创新来提高企业全要素生产率。事实上,技术的突破性创新会引起生产方式的革命性变化,催生制造业新模式与新业态。海量的数据信息能被制造业企业获取和利用是企业提高创新能力的必要条件,大数据发展为制造业企业利用大数据创造了契机,即大数据发展能赋予制造业企业更高的创新能力,实现对制造业企业创新过程的动态跟踪,从而通过促进制造业企业创新来提升企业全要素生产率。第三,大数据发展可以优化制造业企业的生产流程、降低生产成本,实现对生产流程的动态与模块化管理,提升制造业企业整合内外部产业链的协同能力,从而有利于制造业企业更好地开展创新活动来提升全要素生产率。由此,本文提出基本研究假说 H2。

H2:大数据发展可以通过企业创新来提高制造业企业全要素生产率。

2. 要素配置的中介理论机理

(1)资本要素配置的中介理论机理。大数据发展有助于降低资本市场交易成本,提高资本市场的交易效率,以优化资本要素配置,从而降低资本扭曲造成的制造业企业全要素生产率损失。具体来看:第一,大数据发展水平越高的区域更容易形成公开透明的市场环境,增加了资本供需双方在匹配过程中的信息有效性。大数据时代资本市场产生了更加丰富的信息,信息冗余也变得更加复杂,资本供需双方匹配也会产生新的问题。但是,大数据发展越好意味着拥有更高的大数据技术水平,可以有效地将烦琐的数据转化为可用的信息,从而更便利地对制造业行业资本供需双方进行匹配,使制造业企业资本配置效率达到更优,以降低资本错配造成的制造业企业全要素生产率损失(孙光林等,2021)。第二,大数据发展越好的区域制造业企业拥有更好的数据平台来充分挖掘数据的价值,从而产生新的价格机制优化资本要素。在经济市场中,资本需求者和供给者是相互独立分开的,资本要素市场配置效率的高低需要价格机制予以调节。但是,价格机制的功能发挥依赖于供需信息的传递效率,基于大量数据中提炼出的有效信息,有助于减少制造业行业的资本错配,从而使资本更好地为制造业企业服务来提升全要素生产率(史丹、孙光林,2022)。第三,从交易成本角度来看,大数据发展为降低资本市场交易成本提供了契机,有利于降低资本交易过程中的搜寻成本、议价成本与决策成本等,使资本更便利地支持优质制造业企业发展,防止因资本错配而导致的制造业企业的“逆淘汰现象”,从而通过优化资本要素配置来提升制造业企业全要素生产率(高培勇等,2019)。基于以上理论分析,本文提出基本研究假说 H3。

H3:大数据发展可以通过提高资本要素配置效率来提升制造业企业全要素生产率。

(2)劳动要素配置的中介理论机理。大数据发展对制造业部门劳动要素的影响表现在替代效

应与创新效应两个方面。一方面,随着大数据发展与制造业的不断融合,制造业通过推动自动化与智能化等新生产模式的建设,将替代部分劳动力,表现为大数据发展的要素替代效应降低了制造业对劳动力的需求,从而降低生产劳动部门的劳动收入水平;另一方面,大数据发展通过创新效应鼓励高技能劳动者从生产部门转移到研发部门,提高了制造业创新部门的工资水平。然而,低技能劳动力只能接受更低的工资从事制造业生产劳动,或者选择失业丧失工资收入。换言之,大数据发展增加了数字型劳动力要素的需求规模,即大数据技术在扩大数字技能型岗位就业市场的同时降低了传统劳动力市场的就业范围,导致制造业行业中不同部门的劳动力市场工资出现分化,有助于优化劳动要素配置。

进一步来看,大数据发展通过优化劳动要素配置对制造业全要素生产率的影响是替代效应和创新效应的综合结果。余东华和韦丹琳(2021)认为大数据等互联网技术表现出的要素替代性会导致劳动力和劳动力中间产品出现冗余,降低了制造业生产部门的要素配置效率,不利于数字技术在制造业部门的普及,从而抑制制造业生产效率的提升。然而,大数据发展在劳动市场引起的创新效应会促使更多高技能劳动力进入制造业研发部门,有利于促使制造业企业开展创新活动,从而提高制造业企业全要素生产率。事实上,如果大数据发展引致的要素替代效应大于创新效应,可能会使制造业发展陷入“索洛悖论”,但是,现阶段我国制造业部门劳动力技能溢价产生的创新效应要大于替代效应,从而有效缓解了劳动变化对制造业全要素生产率产生的非适应性冲击(余东华、韦丹琳,2021)。基于以上理论分析,本文提出基本研究假说 H4。

H4:大数据发展会通过优化劳动要素配置来提升制造业企业全要素生产率。

3. 数据赋能的中介理论机理

数据要素的显著特点是低成本性、大规模性与易得性,大数据发展使数据要素可以在制造业企业生产活动中被广泛使用。然而,数据要素在制造业企业生产过程中能够在多大程度上被利用却取决于企业的数字化水平。如果制造业企业数字化水平较高,便能够利用大数据形成有用的分析结果为制造业企业生产决策提供参考,数据赋能效应就较强,有助于降低消费者需求与产品供给等不确定性因素对制造业企业生产造成的冲击,以提升制造业企业全要素生产率。一方面,大数据发展为制造业企业数字化转型提供了契机,而企业数字化转型又能够促进企业优化商业模式,对改善企业管理决策具有积极影响。事实上,大数据发展的数据赋能效应可以使制造业企业更充分地利用市场,有助于企业发现客户需求、细分市场、辅助决策支持和产品创新等,从而提升制造业企业全要素生产率。另一方面,大数据发展为所有制造业企业提供了均等的受益机会,但是,制造业企业要想获取大数据发展产生的数据红利,就必须进行数字化转型,从而有效发挥数据赋能效应。数据赋能可以强化制造业企业处理信息的能力,助力制造业企业基于大数据挖掘数据要素的潜在价值,以提升制造业企业管理层对市场的洞察力,使企业能够更好地进行试错式学习来提高全要素生产率。基于以上理论分析,本文提出基本研究假说 H5。

H5:大数据发展可以通过数据赋能来提升制造业企业全要素生产率。

四、研究设计

(一)数据来源

本文以 2014—2018 年我国 A 股上市公司中的制造业企业为研究对象。企业层面的数据均来自国泰安数据库;构造大数据发展指数的基础数据分别通过查阅相关统计年鉴、收集地方政府官

方网站数据、整理相关研究报告所得;区域经济相关数据来源于国家统计年鉴。遵循已有研究的惯例,本文剔除了被标记为 ST、相关财务数据有缺失值的样本。此外,为消除异常值对回归结果的影响,本文对相应变量进行了 Winsorize 处理和对数化处理。最终共得到公司-年度观测值 8276 个。值得注意的是,本文在研究过程中,采用主流文献的做法(唐松等,2020),使用微观上市公司注册地与所在省域进行匹配,构建 2014—2018 年的面板数据集。^①

(二) 变量选取

1. 被解释变量:全要素生产率

本文借鉴宋敏等(2021)的做法,分别采用 LP 法和 OP 法对企业全要素生产率进行测度。

2. 核心变量:大数据发展

对于大数据发展指数的详细测度方法与结果可参考本文第二部分。

3. 中介变量

(1) 企业创新。借鉴王玉泽等(2019)的做法,本文以企业研发资金投入的对数值作为企业创新水平的度量变量(rd)。

(2) 要素配置。由于我国要素市场不完善,要素错配是一种很常见的现象,主要表现为资本和劳动力要素价格扭曲,即要素配置不合理。事实上,要素配置不合理指标应该包含价格和数量两个方面,即价格和数量配置不合理导致的扭曲,但是,根据市场均衡调整的一般原理,数量导致的要素配置不合理最终会表现在价格上。

基于此,我们参照吕承超和王志阁(2019)的做法,采用同一行业、同一产权性质和同一年度组企业使用要素成本的偏离度来度量企业的要素配置情况,既能够反映企业要素配置不合理的价格含义,又考虑到企业可能遭受的要素市场外部差异。之所以按照该方式测度,是因为虽然不同行业、不同所有制企业或不同产权性质,以及不同年度所属企业会面临不同的竞争优势,比如国企会比民企更受金融市场和劳动力市场的青睐,投资回报周期相对较短的行业会比较长的行业更受投资者青睐,企业在不同年度由于受金融市场和政策因素的影响也会面临不同的融资成本。但是,同一行业、同一产权性质和同一年度组的企业在要素市场会面临较为类似的竞争条件,差异相对较小。

具体地,首先按照不同行业、不同产权性质和不同年度对企业进行分组,然后计算不同分组企业资本成本的均值,再根据样本企业资本成本与其所在组的均值作差,采用企业资金使用成本与其所在组平均资金使用成本的差值来度量企业资本要素配置(kms),该指标越大,代表企业资本要素的扭曲情况越严重,表明资本要素的配置效率越低,其中,资金使用成本为利息支出与扣除应付账款后负债总额的比值。针对劳动要素配置变量,首先,对不同行业、不同产权性质和不同年度的企业进行分组,计算不同分组劳动成本的均值、最大值和最小值;其次,分别计算企业劳动成本与上述均值的差值和上述最值之间的差值;最后,计算上述两个差值之间的比值,便得到企业劳动价格扭曲程度指标。本文使用企业职工平均薪酬与其所在组平均薪酬的偏离度来测度企业劳动要素配置(lms),该指标越大代表企业劳动收入报酬差距越大,说明数字技能型高素质员工收入越

① 公司注册地与省域匹配可能存在漏洞,为此,我们做了如下尝试。一是收集样本企业子公司的相关数据,并提取每个企业对应的资产规模最大的子公司,分别计算子公司规模占母公司规模的比例,子公司规模超过 50% 的样本约占总样本的 8.5%,经过对比最大子公司注册地和母公司注册地,发现在上述 8.5% 中仅有 23.74% 存在母公司与最大子公司注册地不一致的情况。总体来看,只有 2.018% 的匹配偏差,不会对实证结果造成明显的影响。二是企业注册地与实际办公地不一致的偏差。经测算,该类样本仅占 3.79%,同时,我们以企业办公地为标准,重新进行了数据匹配,实证结果与基本结论一致。

高,越能激励高技能劳动者从制造业企业生产部门转移出来,而制造业企业低技能劳动者只能接受更低的工资或失业,劳动要素配置程度越高,其中,职工平均薪酬为支付给职工以及为职工支付的现金与员工数量的比值。

(3)数据赋能。借鉴吴非等(2021)的做法,本文采用文本挖掘法对制造业企业数据赋能(*dcg*)进行测算。具体来看,首先,确定企业年报中有关数据要素的语言或关键词的表述,包括互联网+、区块链、机器人、机器学习、数字技术、数据要素、智能制造与工业互联网等;其次,基于Python 软件抓取关键词,计算关键词出现的频数,某个关键词出现的频率越高表明该制造业企业对发挥数据要素作用越重视;最后,对所有关键词出现的频数进行累加,值越大表明该企业的数据赋能效应越高。

4. 控制变量

为了避免其他因素可能对企业全要素生产率产生影响,本文选取了如下控制变量。企业规模(*size*),使用企业年末总资产的对数值表示;企业盈利情况(*roa*),使用企业年末总资产报酬率表示;财务杠杆水平(*lev*),使用资产负债率表示;企业收入增长率(*grow*),用企业营业收入增长率表示;企业年龄(*age*),即企业成立年限;股权集中度(*first*),使用第一大股东持股比例表示;企业产权性质(*soe*),国有企业取1,民营企业取0;经济发展水平(*lgdp*),使用省域人均GDP的对数值来衡量。所有变量的描述性统计见表4。

表 4 主要变量的描述性统计

变量	变量描述	均值	标准差	最小值	最大值
<i>TFP_LP</i>	LP 法企业全要素生产率	14.752	0.931	12.456	17.662
<i>TFP_OP</i>	OP 法企业全要素生产率	14.908	0.929	12.599	17.813
<i>bigdata</i>	基于熵值法计算的省域大数据发展指数	0.050	0.042	0.002	0.409
<i>bigdata2</i>	基于算数平均法计算的省域大数据发展指数	0.051	0.026	0.006	0.118
<i>kms</i>	企业资本要素配置,本文测度所得	0.000	0.017	-0.041	0.112
<i>lms</i>	劳动要素配置,本文测度所得	0.000	0.143	-0.634	0.990
<i>rd</i>	企业研发资金投入的对数值	7.253	10.309	0	69.73
<i>dcg</i>	数据赋能,本文测度所得	1.412	1.332	0	6.176
<i>lev</i>	财务杠杆水平,资产负债率	0.399	0.194	0.038	1.685
<i>grow</i>	企业收入增长率	0.327	1.385	-0.957	26.129
<i>roa</i>	企业盈利情况	0.039	0.068	-0.475	0.258
<i>soe</i>	企业产权性质,国有企业取1,民营企业取0	0.386	0.487	0	1
<i>first</i>	股权集中度,即第一大股东持股比例	34.418	14.615	7.31	80.34
<i>size</i>	企业规模,企业年末总资产的对数值	22.097	1.255	19.033	26.553
<i>age</i>	企业成立年限,单位:年	17.185	5.352	3	39
<i>lgdp</i>	经济发展水平,省域人均GDP的对数值,单位:元	11.395	0.873	9.272	12.803

(三)计量模型构建

首先,为了检验大数据发展能否对制造业企业全要素生产率产生直接影响,本文构建如下面板数据回归模型:

$$TFP_{i,j,t} = \beta_0 + \beta_1 bigdata_{j,t} + \sum \beta_k controls_{i,j,t} + \varepsilon_{i,j,t}$$
 (1)

其中,下标 i,j,t 分别代表企业、省区市和年份; TFP 代表制造业企业全要素生产率,本文分别采用 TFP_LP (LP 法全要素生产率)和 TFP_OP (OP 法全要素生产率)来衡量; $bigdata$ 代表省域层面的大数据发展水平; $controls$ 为控制变量向量集。

其次,为了检验大数据发展影响制造业企业全要素生产率的机理,构建如下中介效应模型:

$$media_{i,j,t} = \beta_0 + \beta_2 bigdata_{j,t} + \sum \beta_k controls_{i,j,t} + \varepsilon_{i,j,t}$$
 (2)

$$TFP_{i,j,t} = \beta_0 + \beta_3 bigdata_{j,t} + \beta_4 media_{j,t} + \sum \beta_k controls_{i,j,t} + \varepsilon_{i,j,t}$$
 (3)

其中,方程(2)检验了大数据发展对中介变量的影响,方程(3)是将核心变量与中介变量放入同一方程的回归结果,如果 β_2 、 β_3 和 β_4 的系数值在置信水平上均显著,说明中介效应显著,即大数据发展能通过中介变量对制造业企业全要素生产率产生影响。

五、实证结果分析

(一)大数据发展对制造业企业全要素生产率的直接影响

表 5 给出了本文的基准回归结果,其中,第(1)列是使用 OP 法测度企业全要素生产率的回归结果,第(2)列是使用 LP 法测度企业全要素生产率的回归结果。根据第(1)、(2)列的回归结果可知,大数据发展的系数值分别为 0.538 和 0.522,均在 1% 的置信水平上显著,表明大数据发展对制造业企业全要素生产率具有显著正向影响。为了检验回归结果的稳健性,在第(3)、(4)列回归中更换大数据发展的度量指标,使用算术平均值的方法来度量各省区市的大数据发展水平。由第(3)、(4)列的回归结果可知,大数据发展的系数值分别为 1.691 和 1.656,在 1% 的置信水平仍然显著为正,再次说明大数据发展对制造业企业全要素生产率的影响是正向显著的。

表 5 基准回归结果分析

变量	(1)	(2)	(3)	(4)
	TFP_OP	TFP_LP	TFP_OP	TFP_LP
$bigdata$	0.538 *** (2.94)	0.522 *** (2.89)		
$bigdata2$			1.691 *** (3.90)	1.656 *** (3.85)
常数项	-1.757 *** (-4.40)	-1.618 *** (-4.08)	-0.839 * (-1.68)	-0.717 (-1.45)
行业、省份、年份固定效应	控制	控制	控制	控制
观测值	8276	8276	8276	8276
R^2	0.725	0.728	0.726	0.729

注:括号内为基于稳健标准误计算的 t 值;***、** 和 * 分别表示 1%、5% 和 10% 的显著性水平。下同。

(二)大数据发展的机制分析

1. 企业创新的中介效应实证结果

表 6 给出了企业创新的中介效应回归结果。由表 6 第(1)列可知,在 5% 的置信水平上,大数据发展对制造业企业创新研发投入具有显著正向影响,大数据发展有利于企业创新。同时,由表 6 第(2)、(3)列可知,在 1% 的置信水平上,大数据发展和企业创新对企业全要素生产率均具有显著正向影响,说明企业创新的中介效应显著。此外,将企业创新的衡量指标由研发投入规模改为研发人员占比(*lrd*),回归结果如表 6 第(4)~(6)列所示,根据企业创新中研发人员占比和大数据发展的系数值的显著性和作用方向可知,企业创新的中介效应仍然显著,表明大数据发展可以通过提高企业创新水平对企业全要素生产率产生促进作用。因此,基本研究假说 H2 得到验证。

表 6 企业创新的中介效应回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)
	<i>rd</i>	<i>TFP_LP</i>	<i>TFP_OP</i>	<i>lrd</i>	<i>TFP_LP</i>	<i>TFP_OP</i>
<i>bigdata</i>	5. 703 ** (2. 48)	0. 544 *** (2. 96)	0. 528 *** (2. 91)	2. 396 ** (2. 15)	0. 529 *** (2. 92)	0. 513 *** (2. 87)
<i>rd</i>		0. 002 *** (2. 81)	0. 002 *** (2. 87)			
<i>lrd</i>					0. 012 *** (6. 48)	0. 012 *** (6. 39)
其他变量	控制	控制	控制	控制	控制	控制
观测值	8276	8276	8276	8276	8276	8276
R ²	0. 405	0. 724	0. 728	0. 374	0. 726	0. 730

2. 要素配置的中介效应实证结果

表 7 给出了要素配置的中介效应回归结果。由表 7 第(1)列可知,在 5% 的置信水平上,大数据发展对资本要素配置具有显著影响,说明提高大数据发展水平有利于优化资本要素配置。由表 7 第(2)(3)列可知,无论 LP 法还是 OP 法测度的全要素生产率,大数据发展对企业全要素生产率的影响均显著为正。同时,在 1% 的置信水平上,资本要素配置对全要素生产率具有负向影响,提高资本要素配置效率能提升全要素生产率。由此,根据第(1)~(3)列中大数据发展和资本要素配置系数值的显著性,可知资本要素配置的中介效应显著。因此,基本研究假说 H3 得到验证。

表 7 第(4)~(6)列给出了劳动要素配置的中介效应回归结果。由第(4)列可知,在 1% 的置信水平上,大数据发展对劳动要素配置具有显著促进作用,表明大数据发展有助于优化劳动要素配置。与此同时,由第(5)、(6)列可知,在 1% 的置信水平上,大数据发展对企业全要素生产率具有显著正向影响,劳动要素配置对企业全要素生产率的影响也在 1% 的置信水平上显著,表明劳动要素配置在大数据发展影响企业全要素生产率的过程中具有显著的中介效应。因此,基本研究假说 H4 得到验证。

表 7 要素配置的中介效应回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)
	<i>kms</i>	<i>TFP_LP</i>	<i>TFP_OP</i>	<i>lms</i>	<i>TFP_LP</i>	<i>TFP_OP</i>
<i>bigdata</i>	- 0. 112 ** (- 2. 02)	0. 537 *** (2. 68)	0. 521 *** (2. 64)	0. 198 *** (3. 99)	0. 575 *** (3. 10)	0. 562 *** (3. 06)
<i>kms</i>		- 1. 409 *** (- 3. 00)	- 1. 282 *** (- 2. 75)			
<i>lms</i>					0. 583 *** (12. 95)	0. 597 *** (13. 40)
其他变量	控制	控制	控制	控制	控制	控制
观测值	8276	8276	8276	8276	8276	8276
R ²	0. 199	0. 724	0. 723	0. 116	0. 732	0. 736

3. 数据赋能的中介效应实证结果

表 8 给出了制造业企业数据赋能的中介效应回归结果。由表 8 第(1)列可知,大数据发展对企业数据赋能的影响为正,且在 5% 的置信水平下显著。与此同时,表 8 第(2)、(3)列的回归结果表明,在 1% 的置信水平上,数据赋能与大数据发展对企业全要素生产率的影响均显著为正,说明数据赋能的中介效应显著,即大数据发展可以通过提升企业数据赋能效应来提高全要素生产率。因此,基本研究假说 H5 得到验证。

表 8 数据赋能的中介效应回归结果

变量	(1)	(2)	(3)
	<i>dcg</i>	<i>TFP_LP</i>	<i>TFP_OP</i>
<i>bigdata</i>	0. 868 ** (2. 56)	0. 513 *** (2. 80)	0. 499 *** (2. 75)
<i>dcg</i>		0. 054 *** (9. 76)	0. 052 *** (9. 52)
其他变量	控制	控制	控制
观测值	8276	8276	8276
R ²	0. 241	0. 727	0. 731

(三)稳健性检验与内生性讨论

1. 稳健性检验

为了保证基准回归结果的稳健性,本文采用以下方式进行稳健性检验。首先,在样本中剔除包括计算机、通信和其他电子设备制造业,电信、广播电视和卫星传输服务,互联网和相关服务在内的三个行业的观测值,利用剩余样本进行回归。回归结果如表 9 第(1)、(2)列所示,结果表明,在 1% 的置信水平上,大数据发展对企业全要素生产率的作用仍然显著为正。其次,使用联合固定效应方法进行稳健性检验,回归结果如表 9 第(3)、(4)列所示,不难发现大数据发展对企业全要素

生产率的影响展现出显著的正向作用效果,且在 1% 的置信水平上显著。最后,采用更换核心变量的方式进行稳健性检验,本文借鉴张叶青等(2021)的做法,使用文本分析法构建微观企业的大数据指标,根据该制造业上市公司年报大数据相关关键词^①出现的次数来衡量,大数据相关词汇越多,表明该企业大数据发展水平越高。回归结果如表 9 第(5)、(6)列所示,大数据发展对制造业企业全要素生产率的影响仍然显著为正。因而,本文关于大数据发展有助于提高企业全要素生产率的核心结论是稳健的。

表 9 稳健性检验						
变量	(1)	(2)	(3)	(4)	(5)	(6)
	TFP_OP	TFP_LP	TFP_OP	TFP_LP	TFP_OP	TFP_LP
<i>bigdata</i>	0.432 *** (2.71)	0.427 *** (2.66)	0.441 *** (2.96)	0.462 *** (3.06)	0.018 *** (2.76)	0.017 *** (4.70)
控制变量	控制	控制	控制	控制	控制	控制
行业、省份、年份固定效应	控制	控制	未控制	未控制	控制	控制
时间 × 行业	未控制	未控制	控制	控制	未控制	未控制
R ²	0.720	0.721	0.718	0.722	0.642	0.743

2. 内生性讨论

测量误差、遗漏变量与双向因果关系等均有可能导致内生性问题。为此,本文采用工具变量法进行回归,借鉴施炳展和李建桐(2020)的做法,使用新中国成立初期(1953—1957 年)人均函件数量作为大数据发展的工具变量。工具变量的合理性在于:一方面,人均函件数量表明该省份居民对传统通信的偏爱程度,这种偏爱会延伸到对互联网通信媒介的热爱,有利于大数据发展,与核心解释变量密切相关;另一方面,人均函件数量更多的是用于普通居民的日常通信,不会对制造业企业全要素生产率产生直接影响。与此同时,新中国成立初期的人均函件数量对近期企业全要素生产率的影响是相对外生的。工具变量的回归结果如表 10 第(1)(2)列所示,工具变量 *t* 值在 1% 的置信水平上显著,一阶段 *F* 值大于 10,工具变量回归结果是有效的。结果表明,在 1% 的置信水平上,大数据发展对制造业企业全要素生产率的影响仍然显著为正,再次说明回归结果的稳健性。

表 10 内生性讨论		
变量	(1)	(2)
	TFP_OP	TFP_LP
<i>bigdata</i>	0.474 *** (2.88)	0.468 *** (2.73)
工具变量 <i>t</i> 值	4.90	
一阶段 <i>F</i> 值	221.68	
其他变量	控制	控制
R ²	0.783	0.776

① 大数据涉及相关词汇及变量构建方法可参考张叶青等(2021),为节约版面,不再赘述。

(四)异质效应分析

为了深入分析大数据发展对企业全要素生产率影响的多元化特征,实证检验大数据发展对企业全要素生产率的异质效应,我们分别将样本按照国有企业和民营企业、大规模企业和小规模企业进行分类并分别进行回归,估计结果如表 11 所示。由表 11 第(1)(2)列可知,大数据发展对民营企业全要素生产率的影响在 1% 的置信水平上显著,系数值分别为 0.613 和 0.601;与此同时,由表 11 第(3)(4)列可知,大数据发展对国有企业的影响在 10% 的置信水平上显著,系数值分别为 0.314 和 0.299,表明大数据发展对民营企业全要素生产率的影响要大于国有企业,说明大数据发展更有利于提升民营企业的全要素生产率。此外,我们进一步将样本按照企业规模均值分为大企业和小企业两类,进行分组回归,估计结果如表 11 第(5)~(8)所示。结果表明,在 1% 的置信水平上,大数据发展对小企业全要素生产率的影响显著为正,系数值分别为 0.928 和 0.908;大数据发展对大企业全要素生产率的影响在 5% 的置信水平上显著为正,系数值分别为 0.118 和 0.127,表明大数据发展对小企业全要素生产率的作用效果要大于大企业。

表 11 异质效应回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>TFP_LP</i>	<i>TFP_OP</i>	<i>TFP_LP</i>	<i>TFP_OP</i>	<i>TFP_LP</i>	<i>TFP_OP</i>	<i>TFP_LP</i>	<i>TFP_OP</i>
	民营企业	民营企业	国有企业	国有企业	小企业	小企业	大企业	大企业
<i>bigdata</i>	0.613*** (2.14)	0.601*** (2.38)	0.314* (1.76)	0.299* (1.78)	0.928*** (4.01)	0.908*** (3.96)	0.118** (2.32)	0.127** (2.28)
其他变量	控制	控制	控制	控制	控制	控制	控制	控制
<i>N</i>	4981	4981	3295	3295	4675	4675	3601	3601

六、结论及建议

本文从直接作用效应、中介传导机理与异质效应三个方面分析了大数据发展对制造业企业全要素生产率的作用效果。主要结论如下:(1)大数据发展对制造业企业全要素生产率具有显著正向影响,大数据发展有利于提高制造业企业全要素生产率;(2)资本要素配置、企业创新与数据赋能的中介效应显著,大数据发展可以通过缓解资本要素配置的扭曲程度、促进制造业企业创新与提高企业数据赋能效应来提高制造业企业全要素生产率;(3)劳动要素配置的中介效应显著,大数据发展可以通过增大劳动要素配置的优化程度来提升制造业企业全要素生产率;(4)异质效应结论表明,大数据发展更有利于提升民营企业的全要素生产率,对小企业全要素生产率的作用效果要大于大企业。

根据研究结论,本文提出以下政策建议。(1)加强数据资源管理与大数据标准体系建设,有序建设大数据的开放与共享平台。随着大数据在经济社会各领域的广泛应用,大数据资源出现爆炸式增长,如何有效利用大数据资源是当前有待解决的现实问题,只有加强数据管理,才能使数据要素得到更深入的应用。在企业层面,应当及时出台数据要素管理的规章制度,使数据要素在生产过程中得到应用,以优化传统要素配置效率,从而提升企业全要素生产率。在政府层面,应当建立大数据标准化体系以及数据要素流通的体制机制,以打破企业之间存在的数据壁垒,通过数据要

素规范化传输协议与格式等,将数据要素由互联网企业巨头主导变为各个领域协同参与,以最大化开发大数据的内在价值。与此同时,在保证政府数据安全的前提下,应该构建政府数据与企业间的共享机制,使大数据更好地服务企业,以释放政府大数据的经济价值。(2) 加快培育大数据技能人才。未来大数据在企业生产过程中会得到更广泛的应用,企业对大数据技能人才的需求量也会越来越大。大数据技能人才不只是掌握编程的数字技术类人才,还包括能够根据实际生产场景挖掘大数据潜在价值的人才。大数据技能人才的培养是一个系统性的工程,政府机构应当倡导建立大数据相关学科,为培养大数据人才奠定基础;学历教育和职业教育机构应该探索跨学院、跨专业与跨学科交叉培养人才的数据专业技能;企业与社会培训机构应当为在职人员提供接受大数据知识和技能培训的机会,从而提高一线生产人员应用大数据的能力。(3) 努力提升劳动者数字技能,构建数字要素收入分配机制。大数据发展会对传统就业市场造成冲击效应,数据要素对传统劳动份额的替代效应也会扩大收入差距。为此,从劳动者角度来看,应当紧随时代潮流,根据当前劳动市场对劳动者数据技能的新要求,提升自身的劳动技能以更好地适应社会。与此同时,从企业角度来看,应当为在职员工提供数据技能培训的机会。此外,应当进一步构建数据要素参与分配的社会分配格局,以激励企业进行创新,推动大数据技术和企业生产过程的有机结合,特别针对拥有数据优势的企业与个体,还可以提升居民收入水平,从而规范收入分配秩序。(4) 构建大数据技术企业和其他产业交流的平台,充分发挥数据要素在企业内外部优化资源配置的作用,提升企业创新能力。大数据平台的建设可以利用大数据企业的技术优势消除企业间内外部数据资源流通的障碍,特别需要构建大数据企业和制造业企业交流的平台。此外,基于大数据技术对制造业生产过程进行有效分析,可以重塑分工和创新生产模式,从而提升制造业企业的智能化和数字化水平。

参考文献:

1. 高培勇、杜创、刘霞辉、袁富华、汤铎铎:《高质量发展背景下的现代化经济体系建设:一个逻辑框架》,《经济研究》2019 年第 4 期。
2. 吕承超、王志阁:《要素资源错配对企业创新的作用机制及实证检验——基于制造业上市公司的经验分析》,《系统工程理论与实践》2019 年第 5 期。
3. 吕明元、麻林宵:《中国省域大数据与制造业融合测度分析——基于 2013—2018 年中国省级面板数据》,《技术经济》2022 年第 1 期。
4. 施炳展、李建桐:《互联网是否促进了分工:来自中国制造业企业的证据》,《管理世界》2020 年第 4 期。
5. 宋敏、周鹏、司海涛:《金融科技与企业全要素生产率——“赋能”和信贷配给的视角》,《中国工业经济》2021 年第 4 期。
6. 孙光林、艾永芳、李森:《资本错配与中国经济增长质量——基于金融效率与产能利用率中介效应实证研究》,《管理科学》2021 年第 5 期。
7. 孙新波、钱雨、张明超、李金柱:《大数据驱动企业供应链敏捷性的实现机理研究》,《管理世界》2019 年第 9 期。
8. 史丹、孙光林:《数字经济、金融效率与我国经济高质量发展》,《企业经济》2022 年第 1 期。
9. 史丹:《绿色发展与全球工业化的新阶段:中国的进展与比较》,《中国工业经济》2018 年第 10 期。
10. 唐松、伍旭川、祝佳:《数字金融与企业技术创新——结构特征、机制识别与金融监管下的效应差异》,《管理世界》2020 年第 5 期。
11. 王欣亮、杜壮壮、刘飞:《大数据发展、营商环境与区域创新绩效》,《科研管理》2022 年第 4 期。
12. 王玉泽、罗能生、刘文彬:《什么样的杠杆率有利于企业创新》,《中国工业经济》2019 年第 3 期。
13. 魏敏、李书昊:《新时代中国经济高质量发展水平的测度研究》,《数量经济技术经济研究》2018 年第 11 期。
14. 吴非、胡慧芷、林慧妍、任晓怡:《企业数字化转型与资本市场表现——来自股票流动性的经验证据》,《管理世界》2021 年第 7 期。
15. 谢康、夏正豪、肖静华:《大数据成为现实生产要素的企业实现机制:产品创新视角》,《中国工业经济》2020 年第 5 期。

16. 肖静华、吴瑶、刘意、谢康：《消费者数据化参与的研发创新——企业与消费者协同演化视角的双案例研究》，《管理世界》2018年第8期。
17. 许宪春、任雪、常子豪：《大数据与绿色发展》，《中国工业经济》2019年第4期。
18. 余东华、韦丹琳：《互联网应用、技能溢价与制造业全要素生产率——兼论如何有效化解“索洛悖论”》，《财经问题研究》2021年第10期。
19. 张叶青、陆瑶、李乐芸：《大数据应用对中国企业市场价值的影响——来自中国上市公司年报文本分析的证据》，《经济研究》2021年第12期。
20. 赵云辉、张哲、冯泰文、陶克涛：《大数据发展、制度环境与政府治理效率》，《管理世界》2019年第11期。
21. 赵宸宇、王文春、李雪松：《数字化转型如何影响企业全要素生产率》，《财贸经济》2021年第7期。
22. Chen, D. Q., Preston, D. S., & Swink, M., How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems*, Vol. 32, No. 2, 2015, pp. 4 – 39.
23. Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L., Big Data and Firm Dynamics. *AEA Papers and Proceedings*, Vol. 109, No. 5, 2019, pp. 38 – 42.
24. Lyytinen, K., Yoo, Y., & Boland, R. J., Digital Product Innovation within Four Classes of Innovation Networks. *Information Systems Journal*, Vol. 26, No. 1, 2016, pp. 47 – 75.
25. Pietro, R. P., A Glimpse on Big Data Analytics in the Framework of Marketing Strategies. *Soft Computing*, Vol. 22, No. 1, 2018, pp. 325 – 342.
26. Zhang, Y., & Yang, M. H., A Big Data Driven Analytical Framework for Energy Intensive Manufacturing Industries. *Journal of Cleaner Production*, Vol. 197, No. 1, 2018, pp. 57 – 72.

Influence Mechanism of Big Data Development on the Total Factor Productivity of Manufacturing Enterprises

SHI Dan (Institute of Industrial Economics, CASS, 100006)

SUN Guanglin (Institute of Industrial Economics, CASS, 100006; Nanjing University of Finance & Economics, 210023)

Abstract: Based on the data of listed companies in China's manufacturing enterprises from 2014 to 2018, this paper constructs a theoretical framework of the influence of big data development on total factor productivity of manufacturing enterprises, and empirically tests the mechanism of the influence of big data development on total factor productivity of manufacturing enterprises. It is found that: (1) Big data development has a significant positive impact on total factor productivity of manufacturing enterprises, and improving the level of big data development is conducive to enhancing total factor productivity of manufacturing enterprises. (2) The internal mechanism shows that the mediating effects of enterprise innovation, factor allocation and data empowerment are significant, and big data development can enhance total factor productivity of manufacturing enterprises by promoting enterprise innovation, optimizing the efficiency of capital and labor factor allocation, and leveraging data empowerment. (3) Heterogeneity effect shows that the effect of big data development on total factor productivity of private enterprises and small enterprises is greater than that on state-owned enterprises and large enterprises.

Keywords: Big Data Development, Manufacturing Enterprises, Factor Allocation, Data Empowerment, Total Factor Productivity

JEL: J24, O14, O32