

# 大数据发展和企业劳动收入份额\*

——来自“国家级大数据综合试验区”的证据

宋华盛 卢历祺

**摘要:**本文通过构建理论模型刻画了大数据影响劳动收入份额的经济逻辑,并利用“国家级大数据综合试验区”作为准自然实验,使用双重差分法来检验大数据发展对中国上市公司劳动收入份额的影响。研究发现,大数据发展显著降低了企业的劳动收入份额。异质性结果表明,“国家级大数据综合试验区”显著降低了劳动密集型企业、融资约束更高的企业以及高新技术行业的劳动收入份额。进一步的影响机制分析发现,大数据发展通过促进企业技术进步对劳动收入份额产生影响;大数据发展促进了企业数字化转型,企业增加数字化资产投入,挤出了劳动投入,进而导致劳动收入份额下降。

**关键词:**大数据 数字化转型 劳动收入份额 技术进步

## 一、引言

随着以人工智能、区块链、云计算和大数据为代表的信息技术迅速发展,数字化的信息和知识在实体经济中发挥着越来越重要的作用。与此同时,信息技术的快速发展也引发了人们对于新技术会挤压劳动力的担忧。习近平总书记指出:“新一轮科技革命和产业变革有力推动了经济发展,也对就业和收入分配带来深刻影响,包括一些负面影响,需要有效应对和解决”。<sup>①</sup>因此,数据要素如何影响企业劳动收入份额,是值得探讨的现实问题。Karabarounis & Neiman(2014)认为,信息化是导致全球范围内劳动收入份额占比下降的一个关键因素。

全球知名咨询公司麦肯锡(Mckinsy)在2011年提出“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,并指出大数据正在成为重要的生产要素。我国高度重视大数据的开发和应用,2014年以来,“大数据”相关内容已连续8年被写入国务院《政府工作报告》。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。信息技术与经济社会的交汇融合引发了数据迅猛增长,数据已成为国家基础性战略资源。现有文献从经济增长(杨俊等,2022;Goldfarb & Tucker,2019)、产业转型(Einav & Levin,2014)绿色发展(许宪春等,2019)、企业绩效(张叶青等,2021)和政府治理(Sarker et al.,2018)等方面探讨了大数据发展的经济后果,但是少有文章聚焦于大数据对劳动收入份额的影响。

从大数据发展的角度探讨劳动收入份额的影响因素是非常重要的议题。我国自改革开放以来就不断推进收入分配制度改革,但目前收入分配不均问题仍然突出(苏桂芳等,2021)。党的二十大

\* 宋华盛,浙江大学民营经济研究中心、浙江大学经济学院,邮政编码:310058,电子邮箱:songzju@zju.edu.cn;卢历祺(通讯作者),浙江大学经济学院,邮政编码:310058,电子邮箱:luliqi@foxmail.com。基金项目:教育部人文社会科学重点研究基地重大项目“产业链创新链融合推动民营经济高质量发展研究”(22JD790074);浙江省自然科学基金项目“贸易政策不确定性、需求冲击与企业出口行为和绩效”(LZ23G020001)。感谢匿名审稿专家的宝贵意见,文责自负。

<sup>①</sup>习近平:《扎实推动共同富裕》,《求是》2021年第20期。

报告中指出“居民收入增长和经济增长基本同步,劳动报酬提高与劳动生产率提高基本同步”。<sup>①</sup>现有研究对劳动收入份额的决定因素的探讨已有很多,刘亚琳等(2022)从结构性因素、经济波动、技术进步偏向性、不完全竞争和经济全球化等五个方面进行了详细的总结。但是少有文章从大数据角度探讨劳动收入份额。与本文较为相似的文章是邵文波和盛丹(2017),他们使用2004—2013年规模以上工业企业年度调查数据探讨了信息化发展对中国企业就业吸纳能力的影响,并认为信息化会带来劳动收入占比的下降,但是他们的文章主要从就业的角度分析劳动力市场需求方面的变化,没有对劳动力收入份额的变动趋势及影响机制进行详细的论述。本文从收入分配的角度,研究大数据对劳动相对于其他投入要素在产出中所占份额的影响,并且时间范围聚焦于2011—2019年,这段时期是我国信息技术蓬勃发展的时期。

本文首先构建理论模型刻画了大数据影响劳动收入份额的经济逻辑。接着,在实证方面,由于存在大量优质劳动力和较高劳动收入份额的企业往往更容易应用大数据技术,可能存在反向因果问题。另外,管理者的偏好等第三方因素可能同时影响企业大数据使用程度和劳动收入份额。因此,普通最小二乘法很难估计出准确的结果。为了解决潜在的内生性问题,本文利用“国家级大数据综合试验区”的外生政策来识别因果关系。“国家级大数据综合试验区”是我国首个数字经济试点政策,主要通过承担七大任务推进地区信息化发展,包括大数据制度创新、公共数据开放共享、大数据创新应用、大数据产业集聚、大数据流通、数据中心整合利用、大数据国际交流合作(邱子迅、周亚虹,2021)。“国家级大数据综合试验区”会通过两种途径影响我国上市企业的劳动收入份额。一方面,大数据发展可以促进企业技术进步。试验区通过大数据的开放和共享提高了信息透明度,降低企业信息不对称;通过完善数据交易制度缓解垄断对企业的不利影响,提高企业生产率;通过建立大数据产业联盟和推动大数据产业集聚来促进技术知识的扩散,推动企业技术进步。企业生产率的提高会进一步影响劳动收入份额。另一方面,大数据发展影响企业要素配置,促进了企业数字化转型和数字化资产投资,从而挤出了劳动投入,降低了平均工资,但并没有显著改变员工总人数和不同岗位的员工人数,最终导致劳动收入份额下降。

本文使用双重差分法(difference-in-differences, DID)进行回归,发现“国家级大数据综合试验区”对中国上市企业的劳动收入份额具有显著的负向影响,这一结果通过了安慰剂检验,并且当本研究使用不同方法衡量劳动收入份额、剔除同期其他政策的影响、使用双重差分倾向得分匹配方法(PSM-DID)进行回归后,结论依然稳健。进一步分析表明,“国家级大数据综合试验区”对不同公司的影响不同,它显著降低了高新技术行业的劳动收入份额,对非高新技术行业的影响不显著。“国家级大数据综合试验区”对劳动密集型企业影响显著,对资本密集型企业影响不显著;试验区建设显著降低了融资约束较高的企业的劳动收入份额,对融资约束较低的企业没有显著影响。

文章的贡献主要体现在以下三个方面:第一,本文丰富了关于大数据发展的经济后果的文献。现有文献大多发现包含大数据在内的信息技术发展会对企业和宏观经济有着积极影响,部分学者探讨了人工智能应用对劳动力就业的替代效应和创造效应,如王军和常红(2021)。本文通过使用“国家级大数据综合试验区”的外生政策冲击,从收入分配角度探讨大数据发展可能产生的不利影响,从而帮助我们更全面地认识信息技术发展的经济后果,制定更准确的信息化政策。

第二,从大数据的研究视角,实证识别了一个劳动收入份额的决定因素。目前少有研究探讨大数据发展对劳动收入份额的影响,如Aghion et al.(2019)通过构建企业异质性模型,认为IT、大数据等技术的发展使加工效率更高的企业加速扩张生产,而低效企业很难进入市场,因此阻碍创新,导致国民收入中劳动收入份额的下降。还有部分学者研究了人工智能对劳动力市场的影响。郭凯明(2019)通过构建多部门一般均衡模型,认为人工智能服务或人工智能拓展型技术提高,都会促使生

<sup>①</sup>习近平:《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》,人民出版社2022年版。

产要素在产业部门间流动,这一结构转型过程会导致劳动收入份额变动,具体方向取决于人工智能的产出弹性和人工智能与传统生产方式的替代弹性。然而上述研究都没有提供经验证据。本文从实证层面,利用外生政策冲击识别了因果关系,检验了大数据发展对劳动收入份额的影响。

第三,文章从理论模型和实证证据两方面揭示了大数据发展影响劳动收入份额的路径。现有关于数字技术对劳动收入份额的分析多侧重于就业促进效应和负向的替代效应孰占主导,而本文从企业技术进步和要素配置这两个方面进行机制分析,发现大数据作为资本增强型技术进步直接影响了劳动收入份额,同时会改变要素配置,通过促进企业数字化转型和增加数字化资产投资,挤出劳动投入,降低了平均工资,但是并没有显著改变员工总人数和不同岗位的员工人数,最终导致劳动收入份额下降。这有助于我们更为全面地认识大数据发展影响企业劳动收入份额的作用机制。

本文接下来的结构安排如下:第二部分为文献综述;第三部分建立理论分析框架;第四部分介绍“国家级大数据综合试验区”的政策背景及实证策略;第五部分为主要回归结果、稳健性检验和异质性分析;第六部分进行机制分析;第七部分总结全文。

## 二、文献综述

### (一)劳动收入份额的现有研究

劳动收入份额在早期并未引起经济学界的充分重视,这是因为著名的“卡尔多”事实认为,各种生产要素的收入在国民收入中所占的比重在长期保持稳定。但是,20世纪80年代以来,西欧和美国的劳动收入份额开始下降,学者发现我国的劳动收入份额也出现下降趋势(Karabarbounis & Nerman, 2014)。因此,劳动收入份额开始成为国内外学者的研究热点。跨国比较显示,我国劳动收入份额明显低于世界平均水平。Karabarbounis & Nerman(2014)测算了1975—2012年全球59个国家的公司部门的劳动收入份额,发现美国、日本和德国等国家的劳动收入份额在0.60~0.65左右,而我国劳动收入份额在0.35~0.45之间。在劳动收入份额的变动趋势方面,虽然不同学者测算的具体数值存在差异,但是无论是我国宏观层面的测算,还是微观企业层面的劳动收入份额,整体都呈现自20世纪90年代中期不断下降、金融危机之后开始有所回升的趋势(施新政等,2019)。

已有研究表明,劳动收入份额的变动会对宏观和微观经济产生重要影响。在宏观层面,劳动收入份额会影响收入差距和经济增长。劳动收入份额和收入不平等虽然分属于功能性与规模性收入分配,其范畴不同,但是两者具有内在联系,现有研究认为资本收入增长过快而劳动收入增长缓慢会导致收入差距增大(Daudey & Garcia-Peñalosa, 2007),因此,不利于社会稳定。同时,劳动收入份额的不断下降会导致国内消费能力减弱,进一步导致对外贸易的依存度过高,不利于拉动内需和经济增长。在微观层面,柏培文和罗永春(2022)通过构建理论模型以及使用1998—2018年我国上市公司数据发现,劳动收入占比与企业业绩呈正相关关系,并存在时间、地区、企业性质等方面的差异。

鉴于劳动收入份额下降和收入差距扩大对经济的不利影响,诸多学者探讨了影响劳动收入份额的内在因素(Autor et al., 2020)。早期研究大多是从制度因素和经济结构等宏观视角进行考察。如李稻葵等(2009)、刘亚琳等(2018)均认为产业结构转型会使劳动收入份额呈现U型演变趋势。一些研究聚焦于要素市场的不完全竞争,认为劳动力市场分割和资本市场摩擦都会对劳动收入份额产生负向影响(施新政等,2019)。近些年来,诸多学者从微观企业的视角研究劳动收入份额的影响因素,其中的一个重要的方向是讨论要素偏向型技术进步。有偏技术进步非对称性地作用于生产要素的边际产出,从而引起不同生产要素所有者报酬以及要素投入不同比例的变化,进而改变要素收入分配情况。黄先海和徐圣(2009)、陈宇峰等(2013)都发现劳动节约型(资本偏向型)技术进步是我国企业劳动收入份额下降的重要原因。

### (二)大数据对劳动收入份额的影响机制研究

本文主要研究大数据发展对劳动收入份额的影响,因此,下文主要介绍大数据通过技术进步和

企业要素配置机制对劳动收入份额产生影响的文献。

1. 技术进步机制。大数据发展可以促进企业技术进步,而有偏技术进步会通过非对称地改变要素生产率进而影响要素收入分配。当劳动和资本替代时,资本偏向型(劳动节约型)技术进步会降低劳动收入份额。反之,如果技术进步偏向于劳动,技术进步对劳动收入份额的作用为正。根据现有文献,大多数学者认为我国目前技术进步为资本偏向型,一项新技术的应用带来的资本边际产出提高大于劳动边际产出提高,并且这是劳动收入份额下降的重要因素(文雁兵、陆雪琴,2018)。同时,Nordhaus(2021)、Graetz & Michaels(2018)等文献均将人工智能(包括工业机器人)、大数据和云计算为代表的信息技术建模为资本增强型技术进步,导致劳动力被替代。

大数据发展可以通过降低信息不对称、完善数据交易制度和知识溢出效应促进企业技术进步,进而影响企业劳动收入份额。首先,企业应用大数据技术可以提高信息透明度,减少因为信息不对称给企业造成的搜寻成本和决策成本。大数据技术可以快速收集和存储海量的信息,通过计算和匹配分析,能够及时、清晰地识别市场需求,面向不同的客户群体输出定制服务(Farboodi et al., 2019; 张叶青等, 2021)。在生产过程中应用大数据技术,可以实时收集、上传原材料和中间产品数据,同时借助智能的优化算法,综合考虑资金约束、人员技能约束和机器设备约束等,动态调整生产计划,辅助企业进行生产决策。

其次,大数据区域共享机制和交易制度的完善,可以缓解垄断对企业的不利影响,提高企业生产率。数据要素类似于公共品,其复制、传播和使用过程不受到时空限制,体现出低成本特征和非竞争性的特征(Goldfarb & Tucker, 2019),这使得数据要素可以在企业生产活动中被广泛应用。但是,具有先发优势的企业凭借积累的海量数据,通过更精准的服务和算法迭代,形成消费者集聚,进而积累更多的数据资源,形成“数据反馈循环”和行业垄断地位(Farboodi et al., 2019)。而通过对数据中心的整合利用,促进大数据流通,打破数据资源壁垒,可以缓解垄断对企业的不利影响,促进企业技术进步。

最后,大数据产业集聚和共享可以促进技术知识的扩散和企业间协同创新,通过知识溢出效应推动企业技术进步。Akcigit & Liu(2016)认为创新本身是一个试错过程,研发过程中的失败尝试所形成的大量数据和信息对企业也具有重要意义。而如果企业之间没有进行大数据信息共享,那么研究的资源和精力就可能浪费在其他企业已经发现毫无意义的项目上,从而导致无效率的均衡(徐翔等, 2021)。大数据开放共享和产业集聚可以促进企业交流、共享研发过程中的各种数据,通过知识溢出效应提高创新活动的效率,进而实现生产效率的提升。

2. 企业要素配置。在数字经济时代,数据被广泛视作一种新的生产要素,它会促进现有生产要素之间形成更密切的交互关系,并引起广泛的资源重组与聚合,并且影响资本和劳动要素配置(杨俊等, 2022)。徐翔等(2021)综述了数据生产要素的经济学概念和主要特征,认为数据在很多时候被视为信息的一种表现形式或者信息化载体,并分析了数据生产要素对微观企业决策的影响。Jones & Tonetti(2020)认为,数据可以被视为信息中不属于创意和知识的部分。作为生产要素的数据,其本身不能被直接应用于生产经济物品,但是却能在生产过程中发挥作用——如创造新的知识或者形成对于未来的预测,进而指导企业要素配置和经济物品的生产。

大数据会影响资本要素配置,促进企业数字化转型。一方面,大数据发展有助于降低资本市场摩擦,提高资本市场配置效率。企业通过应用大数据,提炼资本需求者和供给者的有效信息,提高供需信息的传递效率,债权人能够以更低的成本搜集企业的基本面和资金使用状况等信息,债务人可以更有针对性地寻找投资方进行融资。因此,大数据应用可以提高资本要素市场配置效率,降低企业融资成本。而当资本要素和劳动要素之间是替代关系时,<sup>①</sup>资本市场配置效率提高可以让企业进

<sup>①</sup>现有文献对我国微观企业的资本—劳动替代弹性的测算结果普遍大于1(陈晓玲等, 2019; 陈登科、陈诗一, 2018; Berkowitz et al., 2015),即资本要素和劳动要素呈现替代关系。同时, Karabarbounis & Neiman(2014)的跨国比较也发现,各国的资本—劳动替代弹性平均值为1.28。

行更多资本融资(施新政等,2019),从而购买更多的机器设备,降低劳动收入份额。另一方面,大数据和实体经济的融合为企业数字化转型提供了宝贵的机遇(吴非等,2021;袁淳等,2021),企业进行更多的信息化和数字化资本投资。大数据发展通过改善数字发展环境,促进数字化资源的集聚,为企业数字化转型提供了一定支持。而企业在数字化转型期间需要投入大量资金,用于购买和开发信息系统、计算机软件、云服务等数字化资产,因此会增加资本要素投入,降低劳动收入份额。

大数据会影响劳动要素配置。大数据发展对劳动要素的影响体现在替代效应和创造效应两个方面(Acemoglu & Restrepo,2018;王军、常红,2021)。一方面,随着大数据和实体经济的不断融合,企业利用人工智能和自动化技术将替代部分劳动力,对劳动要素产生替代效应。与此同时,每位工人的产出增加,但是工资并没有相应增加,产出和工资的反向分离导致劳动收入份额的下降。部分学者使用人工智能衡量信息技术水平,来考察其对劳动要素和劳动收入份额的影响(Agrawal et al.,2019)。Gries & Naudé(2018)通过建立包含人工智能资本的内生增长理论模型,结合需求约束条件对模型进行均衡求解。结果表明,在人工智能和劳动力进行相互替代的情况下,人工智能将会降低劳动力工资水平,并减少劳动收入份额。另一方面,大数据发展在一些新的领域创造出新的任务,增加了数字型劳动力要素的需求规模和工资水平,对劳动要素产生创造效应,这会抵消信息技术对劳动收入份额的负面影响。因此,大数据发展对企业劳动要素配置的影响是替代效应和创造效应的综合结果。

### 三、理论框架

#### (一)模型设定

本文考虑大数据影响劳动收入份额带来的两种渠道。一方面,参考Nordhaus(2021)的机制,大数据作为资本增强型技术进步通过影响生产率,间接对劳动收入份额产生影响。另一方面,参考Acemoglu & Restrepo(2018)的机制,大数据通过直接改变生产任务的分配方式,即资本和劳动要素之间的配置,影响劳动收入份额。建立基于生产任务的静态模型,设定最终品  $Y$  是由生产任务  $y(i)$  生产,将生产任务的总数量标准化为1。生产函数设定如下:

$$Y = \left( \int_{N-1}^N y(i)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} \quad (1)$$

其中,  $\sigma$  表示不同生产任务之间的替代弹性。积分限处于  $N-1$  到  $N$  之间意味着生产最终品过程中始终保证投入一单位任务。新的任务会代替原有的任务,所以  $N$  增大代表质量提升。

企业有两种类型的生产任务。存在技术约束  $I \in [N-1, N]$ , 当  $i \leq I$  时,企业可以选择传统的资本或劳动进行生产,生产任务可以被自动化。当  $i > I$  时,生产任务不能被自动化。生产任务  $i$  的生产函数为:

$$y(i) = \begin{cases} A_k(i)D^\eta(i)K(i) + A_l(i)L(i) & \text{if } i \leq I \\ A_l(i)L(i) & \text{if } i > I \end{cases} \quad (2)$$

其中,  $A_k(i)$ 、 $A_l(i)$  分别为劳动增强型技术进步和资本增强型技术进步,  $K(i)$ 、 $L(i)$  分别是任务  $i$  中的资本和劳动要素投入。参考Abis & Veldkamp(2022)的假设,数据可以作为资本要素进行积累和投入生产,因此,将数据与资本相结合,  $D(i)$  为在任务  $i$  中投入的原始数据,  $\eta$  表示数据的重要程度或应用效率,且  $D(i)$  关于  $i$  严格递增,这表示大数据的规模报酬递增特征。模型根据生产任务的质量来区分高技能劳动力和低技能劳动力。假设  $A_l(i)$  关于  $i$  严格递增,即质量更高的生产任务需要投入技能更高的劳动力。当生产任务的质量低于技术约束的门槛值( $I$ )时,任务可以被自动化,也就是低技能劳动力可以被资本完全替代。

本文考虑大数据影响劳动收入份额的两种渠道。一方面,大数据作为资本增强型技术进步通过影响生产率间接对劳动收入份额产生影响。有偏技术进步会通过非对称地改变要素生产率进而影响要素收入分配,但影响方向取决于资本—劳动的替代弹性。在本文的模型中表现在大数据的应用程度 $\eta$ 的增大。另一方面,大数据通过直接改变生产任务的分配方式,即资本和劳动要素之间的配置,影响劳动收入份额。包括大数据在内的新一代信息技术的应用可以促使企业进行数字化转型,更多生产和业务流程被自动化,即模型中 $I$ 的增大。同时,技术变革的一个重要特征是会创造出新任务,比如应用大数据后需要数据分析师、系统架构师等新型数字化人才,这种新任务的创造在模型中表现为 $N$ 的增大。

### (二)模型均衡

在静态模型中,资本存量 $K$ 是给定的。假设产品市场和生产任务市场是完全竞争市场,则生产任务 $i$ 的价格为:

$$p(i) = \begin{cases} \min \left\{ \frac{R}{A_k(i)D^\eta(i)}, \frac{W}{A_l(i)} \right\} & \text{if } i \leq I \\ \frac{W}{A_l(i)} & \text{if } i > I \end{cases} \quad (3)$$

其中, $R$ 表示资本价格, $W$ 表示劳动工资。当生产任务 $i \leq I$ 时,企业决定是否进行自动化取决于资本的效率工资和劳动的效率工资。根据企业利润最大化条件,可以得出任务 $i$ 的需求 $y(i) = Yp(i)^{-\sigma}$ 。进一步可以得到生产任务 $i$ 的要素需求:

$$k(i) = \begin{cases} Y \frac{1}{A_k(i)D^\eta(i)} \left( \frac{R}{A_k(i)D^\eta(i)} \right)^{-\sigma} & \text{if } i \leq I \\ 0 & \text{if } i > I \end{cases} \quad (4a)$$

$$l(i) = \begin{cases} 0 & \text{if } i \leq I \\ Y \frac{1}{A_l(i)} \left( \frac{W}{A_l(i)} \right)^{-\sigma} & \text{if } i > I \end{cases} \quad (4b)$$

根据式(4)可知,现有任务中 $i \leq I$ 的部分由资本完成, $i > I$ 的部分由劳动完成。进一步得到要素市场出清条件:

$$K = Y \int_{N-1}^I \frac{1}{A_k(i)D^\eta(i)} \left( \frac{R}{A_k(i)D^\eta(i)} \right)^{-\sigma} di \quad (5a)$$

$$L = Y \int_I^N \frac{1}{A_l(i)} \left( \frac{W}{A_l(i)} \right)^{-\sigma} di \quad (5b)$$

在完全竞争市场条件下,可以得到总产出:

$$Y = \left[ K^{\frac{\sigma-1}{\sigma}} \left( \int_{N-1}^I (A_k(i)D^\eta(i))^{\sigma-1} di \right)^{\frac{1}{\sigma}} + L^{\frac{\sigma-1}{\sigma}} \left( \int_I^N A_l(i)^{\sigma-1} di \right)^{\frac{1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \quad (6)$$

### (三)大数据和劳动收入份额

对式(6)求劳动的偏导数可以得到劳动收入份额。令 $\omega = \frac{WL}{RK}$ ,则 $\omega$ 上升意味着劳动收入份额

上升。令  $L^s(\omega)$  为劳动供给函数,  $\frac{\partial L^s}{\partial \omega} > 0$ , 劳动供给函数向上倾斜。那么根据式(6)可以得到:

$$\ln(\omega) = \frac{\sigma-1}{\sigma} \ln(L^s(\omega)) + \frac{1}{\sigma} \ln\left(\int_I^N A_i(i)^{\sigma-1} di\right) - \frac{\sigma-1}{\sigma} \ln(K) - \frac{1}{\sigma} \ln\left(\int_{N-1}^I (A_k(i) D^\eta(i))^{\sigma-1} di\right) \quad (7)$$

通过对式(7)分别求  $\eta$ 、 $N$ 、 $I$  的偏导数, 可以得到:

$$\frac{\partial \ln(\omega)}{\partial \eta} = -\frac{\sigma-1}{\sigma} \frac{1}{\int_{N-1}^I (A_k(i) D^\eta(i))^{\sigma-1} di} \int_{N-1}^I (A_k(i) D^\eta(i))^{\sigma-1} \ln(D(i)) di \quad (8)$$

$$\frac{\partial \ln(\omega)}{\partial N} = \frac{1}{\sigma} \frac{1}{\int_I^N A_i(i)^{\sigma-1} di} A_i(N)^{\sigma-1} + \frac{1}{\sigma} \frac{1}{\int_{N-1}^I (A_k(i) D^\eta(i))^{\sigma-1} di} (A_k(N-1) D^\eta(N-1))^{\sigma-1} > 0 \quad (9)$$

$$\frac{\partial \ln(\omega)}{\partial I} = -\frac{1}{\sigma} \frac{1}{\int_I^N A_i(i)^{\sigma-1} di} A_i(I)^{\sigma-1} - \frac{1}{\sigma} \frac{1}{\int_{N-1}^I (A_k(i) D^\eta(i))^{\sigma-1} di} (A_k(I) D^\eta(I))^{\sigma-1} < 0 \quad (10)$$

以上三个算式分别显示了大数据影响劳动收入份额的两种渠道。其中, 式(8)显示了大数据作为资本增强型技术进步通过影响生产率, 间接对劳动收入份额产生影响, 影响方向取决于资本—劳动替代弹性。当  $\sigma > 1$ , 即资本和劳动是替代关系时, 大数据导致的资本增强型技术进步进一步使劳动收入份额下降; 反之, 当  $\sigma < 1$ , 劳动收入份额上升。式(9)和式(10)显示了大数据通过直接改变生产任务的分配方式, 即资本和劳动要素之间的配置, 影响劳动收入份额, 这种渠道与资本—劳动替代弹性的大小无关。其中, 根据式(9), 大数据应用会创造新任务, 这些新任务由高技能劳动力完成, 因此劳动收入份额上升。由式(10)可知, 大数据发展导致  $I$  上升, 更多任务可以进行自动化, 劳动收入份额下降。图1显示了模型得到的大数据影响劳动收入份额的路径。在理论上, 大数据对劳动收入份额的影响方向不确定, 需要进行实证探讨。

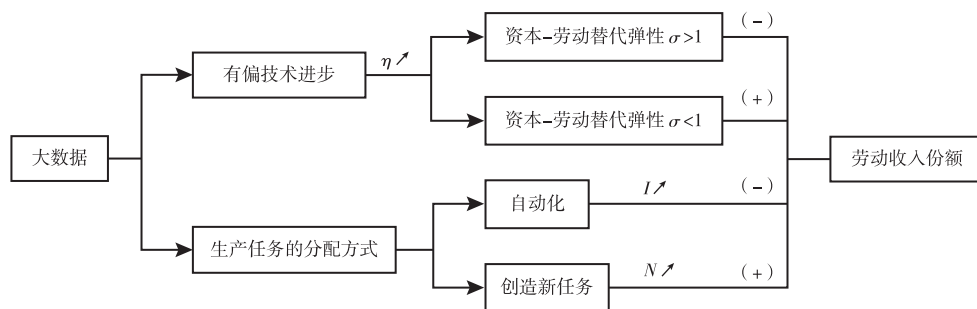


图1 大数据影响劳动收入份额的路径

#### 四、政策背景和实证策略

##### (一)“国家级大数据综合试验区”实施的政策背景

2015年8月31日, 国务院印发《促进大数据发展行动纲要》(以下简称《纲要》), 定义“大数据是以

容量大、类型多、存取速度快、应用价值高为主要特征的数据集合”。为了全面推进我国大数据发展和应用,加快建设数据强国,《纲要》明确提出“开展区域试点,推进贵州等大数据综合试验区建设,促进区域性大数据基础设施的整合和数据资源的汇聚应用”。2016年,为加快实施国家大数据战略,贯彻落实国务院《促进大数据发展行动纲要》,国家发展和改革委员会同工业和信息化部、国家互联网与信息化办公室先后批复同意贵州省、京津冀、珠江三角洲、上海市、河南省、重庆市、沈阳市、内蒙古等八大区域开展“国家级大数据综合试验区”建设。

大数据发展可以通过降低信息不对称、完善数据交易制度和知识溢出效应影响企业技术进步,进而影响企业劳动收入份额。一方面,“国家级大数据综合试验区”提高了数据信息透明度,降低了信息不对称的可能性。具体而言,试验区上线公共数据资源开放系统,免费向公众提供多领域公共数据;建立宏观经济数据库和大数据监测分析系统,开展宏观经济、价格、能源、投资等信息的监测和开放,大大降低了企业需要搜寻相关信息的成本。同时促进推动政府数据共享开放,包括开放政府税务等数据给试点银行,帮助解决小微企业因为信息不对称出现的融资问题。另一方面,在完善数据交易制度方面,当地建立公共数据标准化技术委员会,对数据进行分类管理;试验区的大数据交易所制定《数据确权暂行管理办法》《数据交易结算制度》等一系列交易规则,开展大数据制度的探索和实践。最后,大数据产业集聚和共享促进了技术知识的扩散。具体而言,试验区建立大数据产业联盟,通过核心区引领、大数据产业园区等节点带动的大数据发展格局,实现集聚发展;将大数据产品纳入当地创新产品推荐目录的重点支持领域,享受政采优惠政策,吸引大数据相关企业落户基地。

根据各地发展报告,“国家级大数据综合试验区”建设确实提高了当地的大数据发展水平。第一,“国家级大数据综合试验区”推动了当地的数据中心建设。贵州省在建立大数据试验区后,打造国家“东数西算”重要算力高地,贵阳贵安新区成为全球集聚超大型数据中心最多的地区之一。第二,作为国家级的标杆平台,大数据综合试验区有力促进了大数据产业集聚。贵州省在被选定为试验区后,积极推进大数据重点企业、科研院所和行业协会开展技术交流合作,全力打造“三个千亿级”产业集群。河南省也基于“国家级大数据综合试验区”积极招商引资,截至2018年,已有华为、甲骨文、浪潮、超图、中诚信、用友等102家大数据企业落地河南。第三,“国家级大数据综合试验区”推动了数据要素流通,实现数据跨区域共享。京津冀立足各自特色和比较优势,共同建设以“中关村研发—张北、承德、廊坊数据存储—天津数据装备制造”为主线的“京津冀大数据走廊”,其中,已有286家关联企业在天津武清区落地。

## (二) 计量模型

本文将“国家级大数据综合试验区”建设视为一项准自然实验,使用双重差分模型(difference-in-differences, DID)来考察信息化建设对企业劳动收入份额的影响,基准模型的形式如下:

$$LaborShare_{it} = \alpha + \beta treat_i \times post_t + \gamma X_{it} + \delta_i + \lambda_t + \varepsilon_{it} \quad (11)$$

其中,被解释变量  $LaborShare_{it}$  表示企业  $i$  在第  $t$  年的劳动收入份额。 $treat_i$  为处理组虚拟变量,当企业  $i$  为“国家级大数据综合试验区”试点企业时,  $treat_i = 1$ , 否则为0。 $post_t$  为政策时间虚拟变量,政策实施前  $post_t = 0$ , 政策实施后  $post_t = 1$ 。下文使用  $treat_i$  和  $post_t$  交乘项的系数  $\beta$  来衡量“国家大数据综合试验区”建设对中国上市公司劳动收入份额的影响。 $\delta_i$  为企业固定效应,用以剔除企业层面不随时间变化的不可观测因素的影响。 $\lambda_t$  是年份固定效应,用来排除特定年份同时影响所有企业的外部冲击的干扰。 $X_{it}$  是控制变量集,包括企业层面的企业年龄、规模、杠杆率、资产收益率、所有权性质、管理层持股比例、留存收益资产比及出口贸易,以及城市层面的经济发展水平、产业结构、金融水平、对外投资水平。

## (三) 数据来源和变量构建

下文实证研究所使用的基础数据样本为2011—2019年所有A股上市公司,数据来源于国泰安数据库(CSMAR),并对数据进行如下处理:第一,考虑到金融业在财务数据和资产结构方面的特殊性,



本文剔除金融类企业;第二,删除样本期内被ST和\*ST处理的经营异常样本;第三,只保留在政策实施前后均有观测值的样本;第四,为减少异常值的影响,对所有微观层面的连续变量进行1%和99%的缩尾处理。删去关键变量缺失样本后,文章共得到2165家企业的16977个企业一年份观测值。其中,有918家企业受到了“国家级大数据综合试验区”政策的影响,占全部样本企业数目的42.4%。

文章从微观企业视角研究大数据发展对劳动收入份额的影响,一个重要的基础就是较为准确地计算企业层面的劳动收入份额。借鉴方军雄(2011)以及罗明津和铁瑛(2021)等已有文献,本文使用成本增加值法计算企业劳动收入份额要素,劳动收入份额定义为企业员工支付占工业增加值的比重( $LS$ )。其中,员工支付使用现金流量表中“付给职工以及为职工支付的现金”一项,它包含了给职工支付的工资、奖金、津贴补贴、养老保险、待业保险、补充养老保险、公积金和住房困难补助等。工业增加值的计算方法为“营业收入-营业成本+员工支付+固定资产折旧”。同时,因为劳动收入份额是在0~1之间波动,为使其更符合正态分布,参考施新政等(2019)的处理方式,对其进行logistic的转换,然后取自然对数,即得到 $\ln LS = \ln(LS/(1-LS))$ 。

参考现有文献对企业层面劳动收入份额的影响研究,本文控制了企业层面的其他变量,具体包括:(1)企业年龄( $age$ ),使用样本观测时间与企业成立时间的差值,反映企业的生产经验;(2)企业规模( $size$ ),使用总资产对数值来衡量;(3)资产负债率( $lev$ ),即负债合计与资产合计的比值;(4)资产收益率( $roa$ ),公司净利润除以总资产;(5)所有权性质( $soe$ ),国有企业取值为1,非国有企业取值为0;(6)两职合一( $ceo\_duality$ )公司中董事长和总理由同一人担任即取值为1,否则为两职分离,取值为0;(7)管理层持股比例( $ceo\_share$ ),管理层持股占总股数比例;(8)留存收益资产比( $re$ ),使用盈余公积和未分配利润之和再除以资产总额来衡量;(9)是否进行出口贸易( $export$ )。

地区层面的控制变量包括:(1)经济发展水平( $\ln gdp$ ),即人均地区生产总值的对数值;(2)产业结构( $second\_share$ ),使用第二产业占地区生产总值的比重来衡量;(3)金融水平( $finance$ ),使用年末金融机构各项存贷款余额的对数值来衡量;(4)外商直接投资水平( $open$ ),即当年实际使用的外商直接投资金额的对数值。城市层面的数据来源于历年《中国城市统计年鉴》以及各省和各地级市统计年鉴。<sup>①</sup>

## 五、实证结果

### (一)基准估计结果

表1报告了“国家级大数据综合试验区”建设对中国上市公司劳动收入份额影响的回归结果。在列(1)中,当因变量为要素成本增加值法测算得到的劳动收入份额( $LS$ )时,交互项系数为负数并且在1%的水平上显著,其经济含义是,“国家级大数据综合试验区”使得企业劳动收入份额降低了1.69%。<sup>②</sup>列(2)使用经过logistic的转换,然后取自然对数得到的劳动收入份额( $\ln LS$ ),结果显示交互项的估计系数在5%的水平上显著为负,表明“国家级大数据综合试验区”降低了企业劳动收入份额。

表1 “国家级大数据综合试验区”对劳动收入份额的影响

	(1)	(2)
	$LS$	$\ln LS$
$treat \times post$	-0.00476*** (0.00184)	-0.0244** (0.00955)
控制变量	是	是
Year FE	是	是
Firm FE	是	是
观测值	16977	16977
$R^2$	0.824	0.835

注:(1)\*、\*\*、\*\*\*分别表示在10%、5%、1%水平上显著;(2)括号中为聚类稳健标准误;(3)限于篇幅,控制变量回归结果省略,下同。

<sup>①</sup>限于篇幅,正文未汇报主要变量的描述性统计。感兴趣的读者可联系作者索取。

<sup>②</sup> $0.00476/0.2815 \times 100\% \approx 1.69\%$ ,其中0.2815是要素成本增加值法测算得到的劳动收入份额( $LS$ )均值。

(二) 稳健性检验

1. 动态效应分析。双重差分模型估计结果一致性的前提是样本企业服从平行趋势假设,即在无外生政策冲击时,被解释变量在处理组和控制组的趋势趋于一致。同时,考虑到基准模型只汇报了平均处理效应,没有考虑综合试验区建设后效果的动态变化,本文设立双重差分的动态形式,如式(12)所示。其中,  $time_t$  表示 2011—2020 年的一系列年份虚拟变量,以政策实施前一年(2015 年)作为基准组,控制变量集  $X_{it}$  与基准模型相同,模型同时控制了企业固定效应  $\delta_i$  和年份固定效应  $\lambda_t$ 。

$$LaborShare_{it} = \alpha + \sum_{t=2011}^{2020} \beta_t treat_i \times time_t + \gamma X_{it} + \delta_i + \lambda_t + \epsilon_{it} \quad (12)$$

动态效应检验结果如图 2 所示。在政策实施前,即 2016 年之前,无论是以要素成本增加值法计算的劳动收入份额  $LS$ (左图)还是对数化之后的  $\ln LS$ (右图)为被解释变量,交乘项的系数均不显著异于 0,这表明处理组和对照组在事前满足共同趋势。在“国家大数据综合试验区”建设之后,交乘项系数均显著为负值,说明信息化建设降低了企业劳动收入份额。进一步地,从政策实施后的效果来看,信息化建设对企业劳动收入份额的影响效应从第二年开始显著,说明存在一定的政策时滞;样本期内后续几年的系数均显著,说明“国家大数据综合试验区”建设产生了持续影响。

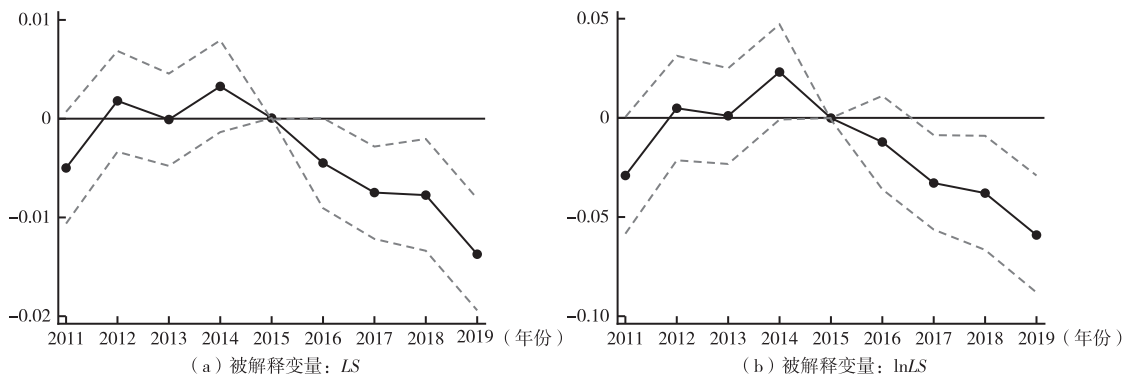


图 2 政策的动态效应

2. 改变劳动收入份额的衡量方法。为了避免度量偏差问题,本文参考施新政等(2019)的指标选择方法,使用劳动分配率法来计算被解释变量,用员工薪酬总额占营业总收入比例来衡量劳动收入份额( $LS2$ )。根据劳动与社会保障部 2004 年发布的《行业人工成本调查和分析方法》,劳动分配率反映了企业人工投入和总产出之间的关系,也能说明在企业总收入中职工收入的比例。因此企业劳动分配率可以用来衡量劳动收入份额。同样地,也对使用劳动分配率法计算得到的劳动收入份额进行 logistic 转换,并取自然对数( $\ln LS2$ )。回归结果显示,在更换劳动收入份额的衡量方法之后,交乘项系数依然显著为负,说明“国家级大数据综合试验区”降低了上市企业劳动收入份额,支持了本文的结论。<sup>①</sup>

3. 排除其他政策干扰。“国家级大数据综合试验区”建设同期也有其他政策出台,有些政策可能会对企业的劳动收入份额产生影响,如国家发展和改革委员会在 2014 年 6 月 23 日颁布的《关于同意深圳市等 80 个城市建设信息惠民国家试点城市的通知》(下称“信息惠民”)。“信息惠民”试点以推动跨层级、跨部门信息共享和业务协同为抓手,有效整合孤立、分散的公共服务资源,促进公共服务的多方协同合作、资源共享、制度对接。为了排除此政策对回归结果的影响,下文加入“信息惠民”试点城市虚拟变量和政策实施时间虚拟变量的交互项,衡量该企业是否受到该政策的影响。回归结果显示,控制了“信息惠民”试点后,“国家级大数据综合试验区”仍然显著降低了企业劳动收入份额。文

<sup>①</sup>限于篇幅,正文未汇报稳健性检验的回归结果。感兴趣的读者可联系作者索取。下同。

章进一步考虑了最低工资标准对回归结果的影响。由于《最低工资规定(2004)》要求各城市逐年提高最低工资标准,劳动力成本的上升会直接影响企业的劳动要素需求,从而影响劳动收入份额。为了排除该政策的干扰,本文加入企业所在地当年的最低工资变量。由回归结果可知,在控制了最低工资水平后交互项系数仍然显著为负,支持了本文的主要结论。

4. PSM-DID。尽管上文对 DID 模型的假设条件和估计结果进行了相关检验,但“国家级大数据综合试验区”建设并非严格意义上的自然实验,因此结果仍然可能存在处理组样本的选择性偏差问题。为保证控制组的选择和处理组样本更高地匹配,以检验基准回归结果的可靠性,本文进一步使用倾向得分匹配双重差分法(PSM-DID)进行稳健检验。采用半径匹配的方法,选取企业年龄、规模、杠杆率、资产收益率、所有权性质、两职合一、管理层持股比例、留存收益资产比及出口贸易作为匹配协变量,并且分别采用截面匹配与逐年匹配两种方法获得匹配样本。以匹配后的样本企业作为基础,再使用双重差分法估计“国家大数据综合试验区”建设对劳动收入份额的影响。回归结果显示,无论是进行截面匹配还是逐年匹配,使用 PSM-DID 方法的回归系数的大小、方向、显著性均与基准回归结果一致,因此,本文估计的“国家级大数据综合试验区”建设降低了上市企业劳动收入份额的结果是稳健的。

5. 安慰剂检验。为了检验本文的研究结论不是由于某些偶然因素造成的,本文借鉴 Cantoni et al.(2017)的方法进行安慰剂检验,以判断“国家级大数据综合试验区”建设对上市公司劳动收入份额的作用是否是由其他随机性因素引起的。分别使处理组随机化和  $treat \times post$  交互项随机化,以要素成本增加值法得到的劳动收入份额( $\ln LS$ )为被解释变量,重新估计式(11),并将随机抽样的过程重复 500 次,保留每次回归得到的  $treat \times post$  项系数。图 3 绘制了 500 次随机抽样后估计系数的概率密度分布。可以看到,无论是随机选择处理组(左图),还是随机选择交互项(右图),所有  $treat \times post$  项的估计系数集中分布在 0 点附近,且显著异于基准回归的系数 -0.0244。这些结果表明,本文的估计结果比较稳健,不太可能由其他不可观察因素驱动。<sup>①</sup>

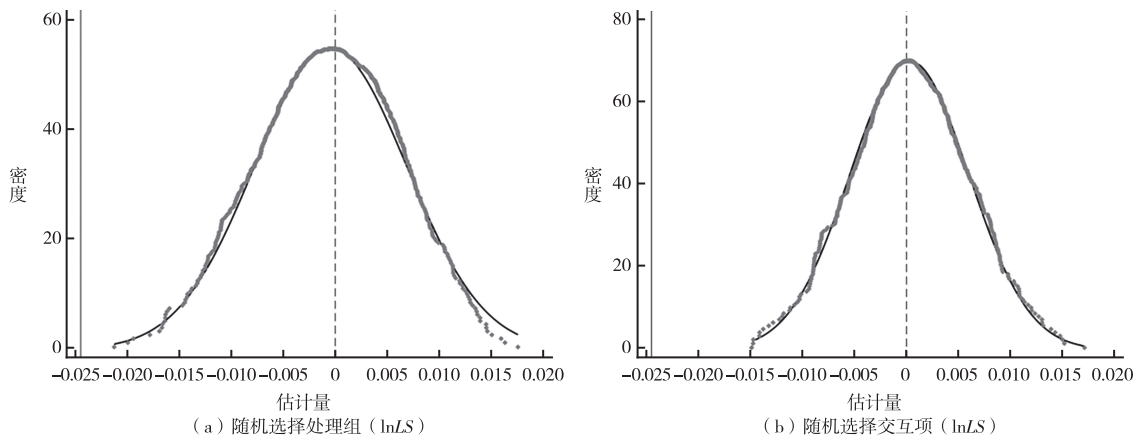


图3 安慰剂检验概率密度分布

### (三) 异质性分析

本文在异质性分析部分,讨论了“国家级大数据综合试验区”对不同要素密集度、不同融资约束程度和不同行业技术特征的企业劳动收入份额的影响。

1. 要素密集度异质性。本文首先考虑大数据发展对不同要素密集度的企业劳动收入份额的异质性影响。根据机制分析,“国家级大数据综合试验区”对企业劳动要素配置存在替代效应和创造效

<sup>①</sup>限于篇幅,正文未汇报安慰剂检验的P值分布情况,感兴趣的读者可联系作者索取。感谢匿名审稿人的建议。

应两方面的影响。由于劳动密集型企业的员工技能水平平均更低,总体议价水平更弱(施新政等,2019),所以当大数据等信息技术对企业劳动要素配置产生冲击时,这部分劳动力更容易被替代,其收入份额更容易被压缩。

本文按照企业资本劳动比区分了劳动密集型企业 and 资本密集型企业,分别考察“国家级大数据综合试验区”对不同要素密集度的企业劳动收入份额的影响。根据表2列(1)(2)的回归结果,“国家级大数据综合试验区”显著降低了劳动密集型企业的劳动收入份额,而对资本密集型企业的影 响不显著。组间差异的经验P值在10%的水平上显著。

2. 融资约束异质性。根据机制分析,“国家级大数据综合试验区”可能提高资本市场配置效率,进而影响企业要素配置和劳动收入份额。因此,大数据发展与劳动收入份额的关系在不同融资约束水平下可能存在差异。根据Neumeyer & Perri(2005)的研究结论,当企业面临的债务融资约束程度较高时,资本要素除了在提供基本的边际产出价值之外,还能凭借较高的抵押价值为企业带来融资收入,所以企业会更加偏好固定资产投资,此时,“国家级大数据综合试验区”提高资本市场配置效率可以帮助企业更及时地筹措充足地资金,支持企业发展。

本文用KZ指数衡量企业的融资约束情况。在计算得到KZ指数之后,按照样本的年度平均值进行分组,如果公司的KZ指数大于平均值,则视为融资约束较高的组,否则为融资约束较低的组。分组回归结果如表2列(3)(4)所示。对于融资约束较高的企业,“国家级大数据综合试验区”降低了其劳动收入份额,并且在1%的水平上成立。但是对于融资约束较低的企业,大数据发展对劳动收入份额的影响虽然为负,但是不显著。组间差异的经验P值在5%的水平上显著。

3. 行业技术异质性。如果大数据发展促进企业技术进步和数字化转型是“国家级大数据综合试验区”影响企业劳动收入份额的重要机制,那么可以预期,大数据发展与劳动收入份额的负向关系应当在高新技术产业中更加显著。这是因为高新技术企业的生产经营注重科技创新,数字技术的创新与转型作为新时代下的前沿阵地,自然也是高新技术企业关注并投入的重点领域。同时高新技术企业具有更高的技术条件,能够将数字化转型深度嵌入到自身的组织框架、决策体系和生产流程中(吴非等,2021)。而大数据发展通过改善数字发展环境,促进数字化资源的集聚,为企业数字化转型提供了宝贵的机遇,更容易促进高新技术企业投入更多资金购买数字化和信息化资产,增加资本要素投入,助力其数字化转型,进而影响劳动收入份额。

本文将生物医药制造业、通用设备制造业、专用设备制造业、交通运输设备制造业、电气机械及器材制造业、计算机及其他电子设备制造业、通信设备仪器仪表制造业,以及信息传输、软件和信息技术服务业界定为高新技术产业,其他行业界定为非高新技术产业。分样本回归结果见表2列(5)(6)。结果显示“国家级大数据综合试验区”只显著降低了高新技术产业的劳动收入份额,组间差异的经验P值在10%的水平上显著,与本文的逻辑相符。

表2 要素密集度、融资约束和行业技术的异质性分析

	(1)	(2)	(3)	(4)	(5)	(6)
	lnLS	lnLS	lnLS	lnLS	lnLS	lnLS
	资本密集	劳动密集	融资约束高	融资约束低	高新技术	非高新
<i>treat</i> × <i>post</i>	-0.00242 (0.0180)	-0.0379*** (0.0106)	-0.0461*** (0.0134)	-0.00438 (0.0123)	-0.0225* (0.0126)	-0.00698 (0.0125)
控制变量	是	是	是	是	是	是
Year FE	是	是	是	是	是	是
Firm FE	是	是	是	是	是	是
观测值	4620	12386	9033	7944	6597	10409
R <sup>2</sup>	0.843	0.828	0.829	0.846	0.859	0.834
经验P值	0.073*		0.023**		0.092*	

注:“经验P值”用于检验 *treat* × *post* 交互项的组间系数差异,通过自抽样(bootstrap)1000次得到。

## 六、影响机制分析

## (一)促进企业技术进步

“国家级大数据综合试验区”能够使大数据资源更合理地流动和共享,通过大数据产业集聚促进技术知识扩散,从而有助于提高企业技术水平,降低企业劳动收入份额。表3显示了“国家级大数据综合试验区”对上市公司全要素生产率的影响。其中,列(1)是使用LP方法测度的企业全要素生产率的回归结果,列(2)是使用OP方法测度的企业全要素生产率的回归结果。根据回归结果可知,交互项的系数值分别为0.0310和0.0250,均在5%的水平上显著,说明“国家级大数据综合试验区”显著提高了企业全要素生产率。大数据发展促进了企业技术进步,从而使劳动收入份额下降。同时考虑到政策的时滞和持续性问题,表3的列(3)(4)分别对企业全要素生产率的滞后期进行回归,结果显示试验区建设对后续年份的TFP也有显著影响。

表3 “国家级大数据综合试验区”和全要素生产率

	(1)	(2)	(3)	(4)
	$TFP(LP)$	$TFP(OP)$	$TFP(LP)_{t+1}$	$TFP(OP)_{t+1}$
$treat \cdot post$	0.0310** (0.0142)	0.0250** (0.0114)	0.0380** (0.0175)	0.0215* (0.0127)
控制变量	是	是	是	是
Year FE	是	是	是	是
Firm FE	是	是	是	是
观测值	16601	16801	13681	13826
R <sup>2</sup>	0.904	0.899	0.895	0.899

进一步,考虑到“国家级大数据综合试验区”带来的技术进步的偏向性问题,文章通过欧拉方程分解得到企业的劳动生产率,回归结果显示,“国家级大数据综合试验区”对劳动生产率没有显著的影响,可以说明当前我国试验区的大数据发展所带来的技术进步不是劳动偏向型。<sup>①</sup>同时,如果大数据发展是资本偏向型技术进步,那么劳动密集型企业的初始资本投入更低,技术进步带来的生产率提高程度更大,所以大数据发展应该对劳动密集型企业的生产率正向影响更大。根据表4中“国家级大数据综合试验区”对全要素生产率的异质性影响的回归结果,对劳动密集型企业而言,无论使用OP法还是LP法衡量,试验区均对其全要素生产率有显著提升作用;对资本密集型企业而言,试验区对其使用LP法衡量的全要素生产率有轻微的负向影响,对使用OP法衡量的全要素生产率没有显著影响。这符合大数据发展是资本偏向型技术进步的假设。

表4 “国家级大数据综合试验区”对全要素生产率的异质性影响

	(1)	(2)	(3)	(4)
	$TFP(LP)$	$TFP(OP)$	$TFP(LP)$	$TFP(OP)$
	劳动密集型		资本密集型	
$treat \times post$	0.0540*** (0.0175)	0.0381*** (0.0129)	-0.0372* (0.0219)	-0.00809 (0.0230)
控制变量	是	是	是	是
Year FE	是	是	是	是
Firm FE	是	是	是	是
观测值	12255	12255	4546	4546
R <sup>2</sup>	0.892	0.897	0.937	0.905

<sup>①</sup>限于篇幅,正文未汇报具体回归结果,感兴趣的读者可联系作者索取。

(二)影响企业要素配置

“国家级大数据综合试验区”提升了当地的信息基础设施水平,并且加速完善数字经济的相关规章制度,可以有效改善当地数字发展环境(邱子迅、周亚虹,2021),进而有助于企业数字化转型。“国家大数据综合试验区”通过促进企业进行数字化转型,企业更多投入前沿科技数字技术与硬件系统中,推动企业生产资料与生产过程的数字化,投入更多资本要素,进一步降低了劳动收入份额。

首先考虑到“国家级大数据综合试验区”对企业数字化转型的影响。本文参考吴非等(2021),使用上市公司年报中有关人工智能、区块链、云计算、大数据等的关键词频数来衡量企业数字化转型,并剔除了关键词前存在“没有”“无”“不”等否定性词语的表述。本文对数字化的关键词词频进行加总,得到数字化转型的代理变量 *digTrans*。由于该类数据存在明显的右偏特征,本文进一步对其进行对数化处理。由于大数据产业本身属于信息技术类行业,因此与大数据直接相关的行业可能与其他行业受到的影响有所不同(张叶青等,2021),这样可以进一步剔除信息运输、软件和信息技术服务业的上市公司。此外,考虑到年报文本长度的差异,参考袁淳等(2021)的做法,将数字化词频除以年报中管理层经营讨论与分析内容(MD&A)的文字数目。

“国家大数据综合试验区”对企业数字化转型的回归结果如表 5 所示。列(1)的被解释变量为数字化转型相关词频数,列(2)取对数,在此基础上列(3)删除互联网及相关服务业,列(4)是将数字化词频除以年报中管理层经营讨论与分析内容的文字数目。结果现实,无论使用何种方法衡量被解释变量,“国家大数据综合试验区”均显著促进了企业的数字化转型。

表 5 “国家级大数据综合试验区”和企业数字化转型

	(1)	(2)	(3)	(4)
	<i>digTrans</i>	$\ln(\textit{digTrans})$	$\ln(\textit{digTrans})$	$\ln(\textit{digTrans}/\textit{MD\&A})$
<i>treat</i> × <i>post</i>	3.167*** (0.418)	0.0776*** (0.0221)	0.0843*** (0.0222)	0.0101*** (0.00367)
控制变量	是	是	是	是
Year FE	是	是	是	是
Firm FE	是	是	是	是
观测值	16977	16977	16801	16919
R <sup>2</sup>	0.739	0.795	0.793	0.718

接着考虑“国家级大数据综合试验区”对企业资产投资的影响。“国家级大数据综合试验区”出台了相关产业政策如税收优惠或者补贴,降低企业数字化相关成本,促进企业投入更多数字化资产,从而降低了劳动收入份额。参考祁怀锦等(2020)的指标选择方法,本文使用上市公司财务报告附注披露的年末无形资产明细项中与数字化转型相关的部分占无形资产总额的比重来度量企业的数字化资产。<sup>①</sup>回归结果如表 6 所示,“国家级大数据综合试验区”显著提高了企业数字化无形资产占比。更多的数字化资产投资挤出了企业对于劳动力的投入,进而降低了劳动收入份额。列(2)使用人均资本作为被解释变量,也证明了“国家大数据综合试验区”提高了企业固定资产投资积极性。同时考虑到政策的时滞和持续性问题,表 6 列(3)(4)分别对数字化无形资产和人均资本的滞后期进行回归,结果显示试验区建设对后续年份的企业人均资本也有显著影响。

此外,考虑到大数据发展可能通过在一些新的领域创造新的任务和职位,如数据分析师、系统架

<sup>①</sup>具体地,当无形资产明细项包含“软件”“网络”“客户端”“管理系统”“智能平台”等与数字化转型技术相关的关键词以及与此相关的专利时,将该明细项目界定为“数字化技术无形资产”,再对同一公司同年度多项数字化技术无形资产进行加总,计算其占本年度无形资产的比例。

构师等,因此对劳动收入份额产生正向的影响,所以,本文实证分析了“国家级大数据综合试验区”对企业雇佣员工数目的对数的影响。回归结果显示,试验区建设并没有显著影响企业总的雇佣人数和普通员工数目,虽然高管数目在10%的水平上略有下降,但是系数很小。因此,“国家级大数据综合试验区”通过影响企业要素配置对劳动收入份额的影响为负。同时考虑到不同岗位的雇佣人数变化,本研究区分了工作岗位的回归,结果显示“国家级大数据综合试验区”对于生产人员、财务人员、销售人员和技术人员人数均没有显著影响。<sup>①</sup>

表6 “国家级大数据综合试验区”和企业数字资产

	(1)	(2)	(3)	(4)
	数字化无形资产占比	人均资本	数字化无形资产占比(t+1)	人均资本(t+1)
<i>treat</i> × <i>post</i>	0.0144*** (0.00546)	0.0249* (0.0145)	0.00846 (0.00589)	0.0327* (0.0182)
控制变量	是	是	是	是
Year FE	是	是	是	是
Firm FE	是	是	是	是
观测值	15612	16977	12945	13931
R <sup>2</sup>	0.673	0.944	0.688	0.934

最后,考虑“国家级大数据综合试验区”对企业平均工资的影响。通过“支付给职工以及为职工支付的现金”与“员工人数”之比得到平均工资,并对其取自然对数,回归结果如表7所示。表7列(1)显示,“国家级大数据综合试验区”显著降低了平均工资率,并且列(2)高管的平均工资和列(3)普通雇员的平均工资都显著下降。这和现有学者使用人工智能(包括工业机器人)来衡量信息技术水平,发现其对劳动报酬的存在负向影响的结论一致。例如 Acemoglu & Restrepo(2020)通过对1990—2007年间机器人与美国劳动力市场关系的研究发现,大数据、人工智能等信息技术增加了分配给资本的任务,均衡工资率下降。因此,“国家级大数据综合试验区”通过促进企业数字化转型从而增加资本投入、降低平均工资对劳动收入份额产生负向影响,并没有显著改变员工总人数和不同岗位的员工人数。

表7 “国家级大数据综合试验区”和工资率

	(1)	(2)	(3)
	平均工资	高管工资	普通雇员工资
<i>treat</i> × <i>post</i>	-0.0301*** (0.00771)	-0.0214** (0.00984)	-0.0295*** (0.00789)
控制变量	是	是	是
Year FE	是	是	是
Firm FE	是	是	是
观测值	16977	16954	16952
R <sup>2</sup>	0.827	0.835	0.825

## 七、结论和启示

在经济全球化背景下,以人工智能、区块链、云计算和大数据为代表的信息技术成为经济活动的

<sup>①</sup>限于篇幅,正文未汇报具体回归结果。感兴趣的读者可联系作者索取。

新动能,但是,信息化发展可能会存在一些潜在的弊端,尤其是加剧了人们对就业和收入分配的担忧。本文通过构建理论模型,并且使用“国家级大数据综合试验区”这一外生冲击考察了大数据发展对劳动收入份额的影响。文章发现,“国家级大数据综合试验区”对中国上市企业的劳动收入份额产生显著的负向影响。进一步的研究表明,“国家级大数据综合试验区”对劳动收入份额的抑制作用只在高新技术行业、劳动密集型企业 and 融资约束程度较低的企业中显著。机制分析表明,大数据发展通过影响企业技术进步和要素配置进而对劳动收入份额产生影响。文章还进行了一系列稳健性检验和安慰剂检验,研究结论不变。

本文的研究结论具有重要的理论和现实意义。在理论上,丰富了信息化的经济后果和劳动收入份额的影响因素的相关文献,同时揭示了“国家级大数据综合试验区”影响劳动收入份额的作用机理。在政策含义上,第一,本文研究发现“国家级大数据综合试验区”对劳动收入份额有负面影响,这表明我们应该客观、全面地认识包括大数据在内的信息技术发展带来的利弊。在进一步推动数字经济与实体经济融合的同时,应关注企业数字化转型和产业结构转型带来的包含劳动挤出在内的负面问题。第二,本文的研究发现劳动密集型企业、融资约束更高的企业以及高新技术行业受到的负面冲击更大,因此在推动信息技术发展时,应该有针对性的对重点企业进行扶持,包括补贴、税收优惠和信贷机制等,帮助企业平稳度过转型期。第三,为了缓解大数据发展和新一代信息技术变革可能带来的负面影响,首先应该通过再分配来调节初次分配中“强资本、弱劳动”的情况,坚持就业优先和就业稳定的战略,构建合理的收入分配体系,从而维持劳动收入份额的相对稳定,扎实推动共同富裕。同时,应进一步完善劳动保护制度和失业保障制度,保护劳动者权益,缓解信息技术发展和结构转型升级可能会带来的结构性失业的影响。最后,应该大力支持劳动力就业和再就业培训,提高劳动力技能水平。大数据和其他信息技术的发展以及推动产业转型升级和经济发展需要大规模的高技能劳动力作为支撑,需要完善劳动力教育和培训体系,使劳动力技能可以匹配现有的信息技术水平,满足市场对于数字技能人才的迫切需求。

#### 参考文献:

- 柏培文 罗永春,2022:《提高劳动收入占比是否会降低企业绩效》,《经济学动态》第6期。
- 陈宇峰 贵斌威 陈启清,2013:《技术偏向与中国劳动收入份额的再考察》,《经济研究》第6期。
- 方军雄,2011:《劳动收入比重,真的一致下降吗?——来自中国上市公司的发现》,《管理世界》第7期。
- 郭凯明,2019:《人工智能发展、产业结构转型升级与劳动收入份额变动》,《管理世界》第7期。
- 黄先海 徐圣,2009:《中国劳动收入比重下降成因分析——基于劳动节约型技术进步的视角》,《经济研究》第7期。
- 李稻葵 刘霖林 王红领,2009:《GDP中劳动份额演变的U型规律》,《经济研究》第1期。
- 刘亚琳 申广军 姚洋,2022:《我国劳动收入份额:新变化与再考察》,《经济学(季刊)》第5期。
- 刘亚琳 茅锐 姚洋,2018:《结构转型、金融危机与中国劳动收入份额的变化》,《经济学(季刊)》第2期。
- 罗明津 铁瑛,2021:《企业金融化与劳动收入份额变动》,《金融研究》第8期。
- 祁怀锦 曹修琴 刘艳霞,2020:《数字经济对公司治理的影响——基于信息不对称和管理者非理性行为视角》,《改革》第4期。
- 邱子迅 周亚虹,2021:《数字经济发展与地区全要素生产率——基于国家级大数据综合试验区的分析》,《财经研究》第7期。
- 邵文波 盛丹,2017:《信息化与中国企业就业吸纳下降之谜》,《经济研究》第6期。
- 施新政 高文静 陆瑶 李蒙蒙,2019:《资本市场配置效率与劳动收入份额——来自股权分置改革的证据》,《经济研究》第12期。
- 苏桔芳 陈昌楠 蓝嘉俊,2021:《“营改增”与劳动收入份额:来自中国上市公司的证据》,《财贸经济》第1期。
- 王军 常红,2021:《人工智能对劳动力市场影响研究进展》,《经济学动态》第8期。
- 文雁兵 陆雪琴,2018:《中国劳动收入份额变动的决定机制分析——市场竞争和制度质量的双重视角》,《经济研究》第9期。
- 吴非 胡慧芷 林慧妍 任晓怡,2021:《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》第7期。



- 徐翔 厉克奥博 田晓轩,2021:《数据生产要素研究进展》,《经济学动态》第4期。
- 许宪春 任雪 常子豪,2019:《大数据与绿色发展》,《中国工业经济》第4期。
- 杨俊 李小明 黄守军,2022:《大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论》,《经济研究》第4期。
- 袁淳 肖土盛 耿春晓 盛誉,2021:《数字化转型与企业分工:专业化还是纵向一体化》,《中国工业经济》第9期。
- 张叶青 陆瑶 李乐芸,2021:《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》第12期。
- Abis, S. & L.Veldkamp(2022), “The changing economics of knowledge production”, SSRN Scholarly Paper, No.3570130.
- Acemoglu, D. & P.Restrepo(2018), “The race between man and machine: Implications of technology for growth, factor shares, and employment”, *American Economic Review*, 108(6): 1488—1542.
- Acemoglu, D. & P.Restrepo(2020), “Robots and jobs: evidence from US labor markets”, *Journal of Political Economy*, 128(6): 2188—2244.
- Aghion, P. et al.(2019), “A theory of falling growth and rising rents”, NBER Working Paper, No.26448.
- Agrawal, A. et al.(2019), “Artificial intelligence: The ambiguous labor market impact of automating prediction”, *Journal of Economic Perspectives*, 33(2): 31—50.
- Akcigit, U. & Q.Liu(2016), “The role of information in innovation and competition”, *Journal of the European Economic Association*, 14(4): 828—870.
- Autor, D. et al.(2020), “The fall of the labor share and the rise of superstar firms”, *Quarterly Journal of Economics*, 135(2): 645—709.
- Cantoni, D. et al.(2017), “Curriculum and ideology”, *Journal of Political Economy*, 125(2): 338—392.
- Daudey, E. & C.García-Peñalosa(2007), “The personal and the factor distributions of income in a cross-section of countries”, *Journal of Development Studies*, 43(5): 812—829.
- Einav, L. & J.Levin(2014), “Economics in the age of big data”, *Science*, 346(6210), No.1243089.
- Farboodi, M. et al.(2019), “Big data and firm dynamics”, *AEA Papers and Proceedings*, 109(5): 38—42.
- Goldfarb, A. & C.Tucker(2019), “Digital economics”, *Journal of Economic Literature*, 57(1): 3—43.
- Gries, T. & W.Naudé(2018), “Artificial intelligence, jobs, inequality and productivity: Does aggregate demand matter?”, SSRN Scholarly Paper, No.3301777.
- Graetz, G. & G.Michaels(2018), “Robots at work”, *Review of Economics and Statistics*, 100(5): 753—768.
- Jones, C.I. & C.Tonetti(2020), “Nonrivalry and the economics of data”, *American Economic Review*, 110(9): 2819—2858.
- Karabarbounis, L. & B.Neiman(2014), “The global decline of the labor share”, *Quarterly Journal of Economics*, 129(1): 61—103.
- Neumeyer, P.A. & F.Perri(2005), “Business cycles in emerging economies: The role of interest rates”, *Journal of Monetary Economics*, 52(2): 345—380.
- Nordhaus, W.D.(2021), “Are we approaching an economic singularity? Information technology and the future of economic growth”, *American Economic Journal: Macroeconomics*, 13(1): 299—332.
- Sarker, M.N.I. et al.(2018), “Smart governance through big data: Digital transformation of public agencies”, 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD).

## Big Data and Labor Income Share: Evidence from National Big Data Comprehensive Pilot Zone

SONG Huasheng and LU Liqi  
(Zhejiang University, Hangzhou, China)

**Summary:** As information technologies, notably AI, Blockchain, Cloud Computing, and Big Data, rapidly evolve, digital information and knowledge play an increasingly pivotal role in the real economy. However, the swift progress of information technology has raised concerns about its potential to displace labor. Concurrently, China has been consistently advancing income distribution system reforms since the initiation of reform and opening up. Despite these efforts, the issue of income inequality remains prominent. Therefore, it is worthwhile to explore the research question of how data factor influences the labor income share.

This paper investigates the direction of the impact of big data on the labor income share in enterprises and its pathways. Firstly, this paper constructs a static model based on production tasks, delineating the economic logic of how big data influences the labor income share. Subsequently, it utilizes the policy of the National Big Data Comprehensive Pilot Zone as a quasi-natural experiment to identify causality. The National Big Data Comprehensive Pilot Zone stands as China's first policy experiment in the digital economy, aiming to propel innovation and elevate economic transformation through experimental explorations in areas such as public data openness, the aggregation of the big data industry, and international collaboration. Employing micro-data from Chinese A-share listed companies and city statistical yearbook data from 2011 to 2019, this paper uses difference-in-differences (DID) method and finds that the development of big data significantly reduces firms' labor income share. This result is validated through placebo tests. When we employ different methods to measure labor income share, eliminate the influence of concurrent policies, and utilize the propensity score matching with difference-in-differences (PSM-DID) method, the conclusion remains robust. Furthermore, heterogeneous analysis indicates that the National Big Data Comprehensive Experimental Zone has different impacts on different companies, significantly reducing the labor income share in highly labor-intensive enterprises, enterprises facing higher financial constraints, and those in the high-tech industry.

Lastly, this paper explores the two pathways through which the National Big Data Comprehensive Pilot Zone affects the labor income share of listed companies in China. On one hand, the development of big data stimulates technological progress, and enhances total factor productivity, particularly for labor-intensive enterprises, consequently influencing the labor income share. On the other hand, the development of big data affects factor allocation in enterprises, promoting digital transformation and investments in digital assets, thereby displacing labor input and reducing average wages. In the meantime, the total number of employees and the number of employees in different positions do not exhibit significant changes, ultimately resulting in a decline in the labor income share.

This paper holds significant theoretical and practical implications. Theoretically, it contributes to the literature on the economic consequences of informatization. While existing literature generally finds positive impacts of information technology on enterprises and the macroeconomy, this paper explores the potential adverse effects from the perspective of income distribution. Simultaneously, the paper reveals the pathways through which the development of big data influences labor income share from perspectives of both theoretical model and empirical evidence. Unlike existing literature, which primarily focuses on the employment-promoting and substituting effects of digital technology, this paper conducts the mechanism analysis from the perspectives of technological progress and factor allocation. In terms of policy implications, the paper finds that the National Big Data Comprehensive Pilot Zone has a negative impact on the labor income share. This underscores the need for an objective and comprehensive understanding of the pros and cons brought about by the development of information technologies. As the government continues to promote the integration of the digital economy and the real economy, attention should be paid to the negative effects, including labor displacement, arising from enterprise digital transformation and industrial structural transformation.

**Keywords:** Big Data; Digital Transformation; Labor Income Share; Technological Progress

**JEL Classification:** D33, R58, O33

(责任编辑:金禾)  
(校对:木丰)