数据劳动者对地区间收入差距的作用：弥合还是扩大？

高 立，张 漪

（中国石油大学(北京) 经济管理学院，北京 100249）

摘要： 数据要素与劳动要素的融合催生了数据劳动者，这一新兴群体的出现创造了改善收入分配格局与区域发展不平衡的新路径。基于2014—2022年我国31个省份的面板数据，本文系统考察了数据劳动者对地区间收入差距的影响效应及作用机制。研究发现：数据劳动者能够弥合地区间的收入差距，且该结果经过一系列内生性检验和稳健性检验后依然成立；在机制分析中，检验了产业间收入差距和技术进步的中介效应，证实了其能作为数据劳动者影响地区间收入差距的中介变量；在东部地区、城镇化水平和市场化程度更高的地区，数据劳动者能释放更大的数字红利，其规模增加对地区间收入差距的缩小效果更为明显。为数字经济时代下新型劳动者如何影响收入差距提供了新的理论视角和政策启示。

关键词： 数据劳动者；数据要素；收入差距；地区间收入差距

中图分类号： F49；F124.7 **文献标志码：** A **文章编号：** 1671－1807(2026)08－0152－08

党的十九届四中全会提出“健全劳动、资本、土地、知识、技术、管理、数据等生产要素由市场评价贡献、按贡献决定报酬的机制”，首次将数据纳入生产要素范畴。2023年国家数据局等17部门联合出台《“数据要素×”三年行动计划(2024—2026年)》，其落脚点之一在于推动数据要素与劳动要素协同，此后一大批新兴职业应运而生。2025年，17个新职业、42个新工种的诞生，肯定了前几年新认证职业的方向，展现了新技术的职业内涵拓展和劳动分工细化的潜力。这些新职业的涌现，反映了近年来“数据”相关职业供需两旺的现象。

与此同时，这些凝结了数据价值的劳动者对经济社会的巨大影响将直接体现在收入分配差距的变化上。国家统计局数据显示，2019—2024年，中国低收入组家庭居民人均可支配收入与高收入组家庭居民人均可支配收入的差距持续扩大，从6.90万元扩大至8.92万元。同期，居民人均可支配收入基尼系数呈现波动态势，从2019年的0.465增加至2020的0.468，又回落至2024年的0.465，出现向好迹象。那么，该类型劳动者在复杂多变的收入差距问题中是否发挥了作用、发挥了何种作用，以及如何因势利导借此良机改善收入差距现状，是学界需要明晰并解答的问题，更是数字中国背景下实现共同富裕目标过程中必须面对的重要议题。

事实上，数据要素作为一种桥梁型生产要素不能单独发挥作用，它需要与其他生产要素结合变为现实生产要素[1]。已有研究证明了数据要素可以优化劳动和资本要素配置[2-3]，可通过与人力资本匹配赋能企业创新和区域创新[4-5]、与资本要素结合驱动中国经济增长[6]和全球价值链分工的分散网络化[7]。广大新型劳动者正是数据要素和劳动要素融合的产物。有学者基于劳动关系将新型劳动者划分为三类[8]，第一类是与数字化平台建立传统雇佣关系的“数据劳动者”；第二类是与平台保持灵活零工关系的“数字劳动者”；第三类是兼具生产者和消费者双重身份的“数字产消者”。前两类劳动者即为本文定义的数据劳动者，他们通过接收数据信息、应用数字技术、反馈数据问题、优化数字化平台等一系列与“数据”直接相关的行为释放数据价值并获取相应的劳动报酬，这类劳动者的工作无法脱离“数据”单独进行，但目前尚未有文献以“数据劳动者”这一整体为研究对象。

总体收入差距包括区域内收入差距和区域间收入差距，而区域间收入差距是形成总体收入差距的主要原因。从收入分配环节来看，税收[9]、养老保险[10]、公积金制度[11]等皆通过调节初次分配、再分配环节影响最终收入分配格局，从而影响收入差距。从外部环境来看，市场垄断[12]和技术进步[13]

收稿日期： 2025-12-01

作者简介： 高立(1987—)，男，广西南宁人，博士，副教授，研究方向为能源经济与政策评估；张漪(2001—)，女，浙江台州人，硕士研究生，研究方向为数字经济。

也是收入差距天平上的砝码。众学者对收入差距的重视归根结底源于其对促进社会公平和在高质量发展中实现共同富裕具有深刻意义,一旦收入差距过大,则可能引发犯罪率上升[14]、社会福利减少[15]等消极影响。

当前学术界尚未建立数据劳动者与收入差距之间的联系,未将数据劳动者这一新兴群体纳入影响收入差距因素的考虑之内,相关研究集中在数字经济对收入差距的影响,关注的重点包括数字经济如何影响行业[16]、城乡[17-18]和地区收入[19]差距。在此基础上,本文聚焦于"数据劳动者"这一新兴群体,试图捕捉和挖掘其对收入差距的影响和作用路径,以弥补现有研究的缺失,为理解数字经济时代的收入差距变化提供崭新视角。

1　理论分析

基于新古典经济增长理论,在传统 Cobb-Douglas 生产函数中引入新核心生产要素——数据劳动者的人力投入,并将技术进步、产业间收入差距等因素纳入模型,构建并推导了数据劳动者影响地区间收入差距的理论分析模型。

假设有一经济体,由存在收入差距的地区 1 和地区 2 组成,其产出 Y_i 可表示为数据劳动者 D_i 的函数,即 $Y_i = f(D_i)$,且满足 $Y_1 > Y_2$,该经济体的数据劳动者总和 $D = D_1 + D_2$。较之欠发达地区,经济发达地区在数字技术和数字基础设施等方面的财政支出强度更大,且相关配套资源投入齐全,而数据劳动者高度依赖于当地数字技术和数字基础设施的支持,会形成典型的数据劳动者"虹吸效应"。但与此同时,后发优势使得欠发达地区存在追赶效应,以更快的速度推动数据劳动者规模的扩大,实现经济的追赶。故在此假设两地的数据劳动者存在以下关系:$D_1 > D_2$,且 $\frac{\partial Y_2}{\partial D_2} > \frac{\partial Y_1}{\partial D_1}$。进一步地,地区间收入差距的定义如(1)式所示。

$$G = \frac{Y_1 - Y_2}{Y_1 + Y_2} \tag{1}$$

根据式(1),有

$$\frac{\partial G}{\partial Y_1} = \frac{2Y_2}{(Y_1 + Y_2)^2}, \frac{\partial G}{\partial Y_2} = -\frac{2Y_1}{(Y_1 + Y_2)^2} \tag{2}$$

进一步求导可以得到

$$\frac{\partial G}{\partial D} = \frac{\partial G}{\partial Y_1} \cdot \frac{\partial Y_1}{\partial D} + \frac{\partial G}{\partial Y_2} \cdot \frac{\partial Y_2}{\partial D} \tag{3}$$

结合式(1),可知地区间收入差距对数据劳动者的偏导小于 0。

$$\frac{\partial G}{\partial D} = \left(Y_2 \frac{\partial Y_1}{\partial D_1} - Y_1 \frac{\partial Y_2}{\partial D_2}\right) \frac{2}{(Y_1 + Y_2)^2} < 0 \tag{4}$$

据此,提出以下假设。

H1:数据劳动者增加会缩小地区间收入差距。

此外,数据劳动者的兴起和壮大,能在技术进步特别是数字技术进步中发挥巨大作用;同时,作为数据要素与劳动要素融合的产物,数据劳动者能够促进产业间资源流动,缩小产业间生产效率。基于此,从以上两个角度出发,分析数据劳动者如何通过促进技术进步、缩小产业间收入差距,进而缩小地区间收入差距。

1.1　技术进步

"第二次机器革命"时代,技术进步被描述为指数级增长、数字化的进步和组合式创新。而持续的技术进步依赖于数据积累和人力资本投入,其创造的数字化劳动力与一般劳动力之间的关系正在转变为替代关系[20],但数据劳动者作为数据要素的创造者、改造者,以及数字技术的开发者、使用者,其在推进技术进步中发挥着越来越稳定的作用。此外,根据本文设定,区域技术进步同时受到本地数据劳动者规模的直接影响和其他地区溢出效应的影响,距离衰减效应的存在使后者影响程度弱于前者[21],即

$$A_i = F(D_i, D_{-i}), \text{且} \frac{\partial A_i}{\partial D_i} > \frac{\partial A_i}{\partial D_{-i}} > 0 \tag{5}$$

式中:A 为技术进步,从而可得

$$\frac{\partial G}{\partial D} = \frac{\partial G}{\partial A_1}\frac{\partial A_1}{\partial D} + \frac{\partial G}{\partial A_2}\frac{\partial A_2}{\partial D} = \frac{\partial G}{\partial A_1}\underbrace{\left(\frac{\partial A_1}{\partial D_1} + \frac{\partial A_1}{\partial D_2}\right)}_{>0} + \frac{\partial G}{\partial A_2}\underbrace{\left(\frac{\partial A_2}{\partial D_1} + \frac{\partial A_2}{\partial D_2}\right)}_{>0} \tag{6}$$

据此,提出以下假设。

H2:数据劳动者能够通过推动技术进步来缩小地区间收入差距。

1.2　产业间收入差距

跨界融合是数字经济时代的特征之一,三次产业之间的清晰界限由于虚拟经济与实体经济的深度融合而模糊[22],产业间资源的合理配置可以缩小第一、二、三产业间的劳动生产率差距,从而达到缩小产业间收入差距的目的[23],而产业间收入差距是总体收入差距的重要组成部分。在此假定三次产业的生产函数为

$$Y_{ij} = A_{ij} K_{ij}^{\alpha_{ij}} L_{ij}^{1-\alpha_{ij}} \tag{7}$$

其中,各产业的劳动投入与数据劳动者关系表达式为

$$L_{ij}^{1-\alpha_{ij}} = \lambda_{ij} D_{ij} \tag{8}$$

式中：$i = 1,2$，表示地区；$j = 1,2,3$，分别表示第一、二、三产业。根据利润最大化原则，三大产业平均工资可表示为

$$W_{i,j} = (1-\alpha_{ij}) A_{ij} K_{ij}^{\alpha_{ij}} L_{ij}^{-\alpha_{ij}} \tag{9}$$

产业间收入差距可表示为第二、三产业平均工资之平均值与第一产业平均工资之比，即

$$S_i = \frac{W_{i3} + W_{i2}}{2W_{i1}} \tag{10}$$

具体地，产业间收入差距如(10)式所示。

$$S_i = \frac{(1-\alpha_{i3}) A_{i3} K_{i3}^{\alpha_{i3}} L_{i3}^{-\alpha_{i3}} + (1-\alpha_{i2}) A_{i2} K_{i2}^{\alpha_{i2}} L_{i2}^{-\alpha_{i2}}}{2(1-\alpha_{i1}) A_{i1} K_{i1}^{\alpha_{i1}} L_{i1}^{-\alpha_{i1}}} \tag{11}$$

进一步地，以地区 1 为例，求解产业间收入差距对数据劳动者的偏导：

$$\frac{\partial S_1}{\partial D_1} = \frac{\partial S_1}{\partial D_{11}} \frac{\partial D_{11}}{\partial D_1} + \frac{\partial S_1}{\partial D_{12}} \frac{\partial D_{12}}{\partial D_1} + \frac{\partial S_1}{\partial D_{13}} \frac{\partial D_{13}}{\partial D_1} \tag{12}$$

在第一产业中，以农业为例，数字技术带动了智慧农业发展，互联网、物联网、云计算、大数据等数字技术与农业的深度融合显著提升了农业生产效率[24-25]。但具体而言，传统农业生产如种植、养殖、采收等与数据处理、分析等工作大相径庭，数据劳动者生存空间狭窄。现代农业逐步引入区块链、物联网、大数据等数字技术[26]，对于参与了诸如农产品供应链智能化、农业生态环境实时监测等环节和过程的数据劳动者而言，将其简单归类于第一产业范畴，将造成第一产业数据劳动者数量的高估与第二、三产业数据劳动者数量的低估，或是重复统计的问题；换言之，能精准定位为第一产业数据劳动者的规模很小。综上，就第一产业全局而言，其尚未实现与数字技术的紧密“挂钩”，故为便于分析，将假定第一产业的数据劳动者数量为 0，即 $D_{11} = 0$，$D_1 = D_{12} + D_{13}$，故有

$$\begin{aligned}\frac{\partial S_1}{\partial D_1} &= \frac{\partial S_1}{\partial D_{12}} \frac{\partial D_{12}}{\partial D_1} + \frac{\partial S_1}{\partial D_{13}} \frac{\partial D_{13}}{\partial D_1} = \\ &\frac{\partial S_1}{\partial D_{12}} + \frac{\partial S_1}{\partial D_{13}} = \\ &-\frac{1}{W_{11}} \left(\frac{\alpha_{12} W_{12}}{D_{12}} + \frac{\alpha_{13} W_{13}}{D_{13}} \right) < 0\end{aligned} \tag{13}$$

同理，

$$\frac{\partial S_2}{\partial D_2} < 0 \tag{14}$$

为突出分析重点，再假设某地区的产业间收入差距不会受到另一地区的数据劳动者影响，结合上述推导，有

$$\frac{\partial G}{\partial D} = \frac{\partial G}{\partial S_1} \underbrace{\frac{\partial S_1}{\partial D_1}}_{<0} + \frac{\partial G}{\partial S_2} \underbrace{\frac{\partial S_2}{\partial D_2}}_{<0} < 0 \tag{15}$$

基于上述分析，提出以下假设。

H3：数据劳动者能够通过缩小产业间收入差距来缩小地区间收入差距。

2　研究设计

2.1　模型构建

为验证上述研究假设，构建如下面板数据双向固定效应模型，以检验数据劳动者规模的增加对地区间收入差距的影响。

$$\mathrm{Gap}_{it} = \alpha_0 + \alpha_1 \mathrm{dl}_{it} + \gamma X_{it} + \lambda_i + \mu_t + \varepsilon_{it} \tag{16}$$

式中：i 为省份；t 为时间；Gap 为地区间收入差距；dl 为数据劳动者规模；X 为一系列控制变量的集合；λ_i、μ_t 分别为地区固定效应、时间固定效应；ε_{it} 为随机扰动项。

2.2　数据来源

基于数据可得性前提，选取的研究样本为 2014—2022 年 31 个省份(因数据缺失，未包含港澳台地区)的面板数据。所使用的原始数据来源于《中国统计年鉴》《中国区域经济统计年鉴》，各省份统计局及各市、州、区国民经济和社会发展统计公报。除此之外，数据劳动者规模估计部分数据来源于中华人民共和国交通运输部、美团公开数据和天眼查，对以上数据存在的部分缺失值，采用插值法进行补充。

2.3　变量说明

2.3.1　被解释变量：地区间收入差距(Gap)

地区间收入差距即各省份间可支配收入的差距，由于农村居民人均可支配收入数据存在大量缺失，使用各市、区城镇居民可支配收入数据进行计算。特殊地，上海市城镇及农村人均可支配收入数据皆存在大量缺漏，故其采用各区全体居民人均可支配收入计算。利用 Dagum 基尼系数分解方法[27]对各省份可支配收入加以分解计算，将其中的组间基尼系数进行提取并定义其为地区间收入差距，组间基尼系数的计虎公式为

$$G_{jh} = \frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}{n_j n_h (\bar{y}_j - \bar{y}_h)} \tag{17}$$

式中：G_{jh} 为省份 j 和省份 h 的组间基尼系数；y_{ji} 为第 j 个省份第 i 个市(区)的城镇居民人均可支配收入；$\bar{y}_j$ 为第 j 个省份的城镇居民人均可支配收入；n_j 为第 j 个省份的市(区、县)个数。参照洪俊杰等[19]的做法，对得到的地区间收入差距数据矩阵进

行均值处理,以其均值作为衡量各省间收入差距的依据。

2.3.2　核心解释变量:数据劳动者(dl)

由于数据可获得性,选取以下四类典型劳动者代表数据劳动者。

(1)邮政业就业人员数。

(2)信息传输、软件和信息技术服务业城镇单位就业人员。

(3)网约配送员。为得到其大致数据,聚焦于美团公开数据,并做了如下处理:根据公开数据获得其 2017—2022 年全国规模,由于美团外卖于 2013 年末上线,默认 2013 年骑手规模为 0,利用插值法获得 2014—2016 年全国美团骑手规模;参考高子茗[28]的做法,利用天眼查获取各省历年外卖配送企业累计数量,计算得到各省份历年外卖配送企业数量占全国数量的比例,再乘以全国美团骑手规模。

(4)网约车司机。以 2019 年为界,将网约车司机规模分为两个阶段进行粗略估计:①2014—2018 年,出租车数量大于网络预约出租车驾驶员证数量。在网约车发展初期,网约车司机几乎都是从出租车司机转化而来的。基于一人一车的假设,以各省份出租车数量代指网约车司机数量。②2019—2022 年,出租车数量小于网约证数量。网络预约出租车驾驶员证相关条例于 2016 年 7 月发布,基于一人一证的情况,以全国网络预约出租车驾驶员证发放数量代指网约车司机规模,再乘以各省出租车数量占全国比例分配各省司机数量。

2.3.3　控制变量

模型中还纳入一系列控制变量,具体包括:为刻画政策环境,选取政府干预程度(gov),以地方财政一般预算支出占地区生产总值比例衡量;对外开放程度(ie),以按当年美元兑人民币汇率折算的货物进出口总额与地区生产总值占比衡量;经济发展水平(gnp0),以人均地区生产总值(万元)衡量;为表征经济结构,采用第三产业增加值与第二产业增加值之比(ts)和第一产业增加值与地区生产总值之比(fic)共同衡量产业升级;引入人力资本水平(hc),以普通高校在校生人数与常住人口之比衡量;此外,以人均城市道路面积(pcr) 与对数化人均公路里程(lmt) 作为基础交通设施的代理变量。通过控制上述涵盖政策环境、发展基础、产业结构和基础设施等可能影响到收入差距的变量,旨在尽可能剥离传统因素的作用,从而更准确地识别数据劳动者对地区间收入差距的净效应。

3　实证分析

3.1　基准回归

为探究数据劳动者规模增加对地区间收入差距的影响,使用双向固定效应模型进行回归,表 1 为基准回归结果。列(1)和列(2)皆加入了年份和省份固定效应,前者未加入控制变量而后者加入了控制变量。结果显示,数据劳动者规模增加会缩小地区间收入差距,H1 得以验证。

表 1　基准回归结果(因变量:Gap)

变量	(1)	(2)
数据劳动者	−0.002 6 (0.007 2)	−0.025 8*** (−0.009 3)
截距项	0.126 5*** (0.002 4)	0.371 5*** (−0.109 2)
控制变量	否	是
年份固定效应	是	是
省份固定效应	是	是
调整后 R^2	0.838 2	0.851 4
观测值	279	279

注:***、**、* 分别表示 1%、5%、10% 的显著性水平;括号内为省份层面聚类标准误。

3.2　内生性处理

为缓解数据劳动者和地区间收入差距之间存在的内生性问题,考虑以全国互联网宽带接入用户数(ibu)作为数据劳动者的一个工具变量。为避免弱工具变量问题,参考李健旋[29],纳入各省份移动电话用户数(mpu),以其与全国互联网宽带接入用户数的交乘项(ibu×mpu)作为数据劳动者的工具变量。表 2 列(1)展示了第一阶段的估计结果,表明所选的工具变量与数据劳动者呈现显著的正相关关系。使用两阶段最小二乘法的第二阶段估计结果如列(2)所示,显著为负,与基准回归一致,Cragg-Donald Wald F 值为 237.999,且 Kleibergen-Paaprk Wald F 统计量为 17.926,大于 10% 的 Stock-Yogo 临界值,通过了弱工具变量检验。

3.3　稳健性检验

(1)子样本回归。考虑到 2019 年底暴发的新冠肺炎疫情,剔除受疫情影响最严重的 2020 年和 2021 年的观测样本,回归结果如表 3 列(1)所示。其次,由于上海市的收入差距是基于全体居民人均可支配收入得到的,与其他样本的原始数据内容存在一定的差异,故剔除上海的观测样本,回归结果如表 3 列(2)所示。结果均显示数据劳动者的增加会缩小地区间收入差距,佐证了基准回归结果的稳健性。

(2)增加控制变量。金融发展水平是影响收入

表 2　内生性处理:工具变量回归结果

变量	(1)dl	(2)Gap
	第一阶段	第二阶段
Ibu×mpu	0.001 4***	
	(0.000 3)	
数据劳动者		−0.033 3**
		(0.015 2)
Kleibergen-Paap rk LM 统计量		3.923 0
		[0.047 6]
Kleibergen-Paap rk Wald F 统计量		17.926 0
		{16.380 0}
控制变量	是	是
年份固定效应	是	是
省份固定效应	是	是
观测值	279	279

注:***、**、* 分别表示 1%、5%、10% 的显著性水平;括号内为省份层面聚类标准误;[ ]中的数值为 LM 检验的 P 值;{ }中的数值为 Stock-Yogo 10%的临界值。

表 3　稳健性检验

变量	(1) Gap	(2) Gap	(3) Gap	(4) Gap
数据劳动者	−0.027 1***	−0.024 2**	−0.027 7***	−0.026 0***
	(0.008 5)	(0.0094)	(0.008 4)	(0.007 7)
截距项	0.384 5***	0.372 1***	0.422 9***	0.367 7***
	(0.107 0)	(0.113 4)	(0.108 2)	(0.090 2)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
省份固定效应	是	是	是	是
调整后 R^2	0.818 5	0.821 8	0.822 9	0.890 1
观测值	217	270	279	279

注:***、**、* 分别表示 1%、5%、10% 的显著性水平;括号内为省份层面聚类标准误。

分配的重要因素,故在基准回归的基础上加入了金融发展水平(loan),以金融机构贷款余额与 GDP 的比例衡量。人口老龄化可能会通过人口结构、劳动收入份额和代际人力资本投资等对收入分配产生重要影响[30],还需加入可度量老龄化程度的老年抚养比(old)。此外,失业人数直接影响失业者及其家庭的收入水平[31],而农村最低生活保障制度的建立可部分缓解农村总体贫困状况[32]。分别选用城镇登记失业人数(ue)和农村居民最低生活保障与农村人口之比(rr)衡量上述两个因素。控制了 4 个新变量后,回归结果如表 3 列(3)所示,仍支持基准回归结果。

(3)更换被解释变量。参考洪俊杰等[22]的做法,以历年某省份与其他省份之间的基尼系数同该省份与其他省份之间的基尼系数总和之比为权重,利用基尼系数的加权平均值代替原地区间收入差距进行再次回归,回归结果如表 3 列(4)所示。结果显示,数据劳动者规模的增加对地区间收入差距的缩小作用再次得以验证。

4　机制检验

根据前文理论分析部分,数据劳动者能够缩小产业间从业人员的工资差距、推动技术进步,进而促进地区间收入差距的减小。参考江艇[33]的做法,通过以下模型检验上述两条作用机制的存在性。

$$M_{it} = \beta_0 + \beta_1 \mathrm{dl}_{it} + \gamma X_{it} + \lambda_i + \mu_t + \varepsilon_{it} \quad (18)$$

式中:M_{it} 为机制变量。

4.1　技术进步

以本年度各省份专利申请授权量占全国比例(pat)为各省份技术进步水平的表征变量,根据表 4 列(1)结果,数据劳动者规模增加能够显著促进各省技术进步。不同类型的专利由于技术含量和应用价值不尽相同,对收入差距的影响也存在差异[34],为进一步研究各类型专利对收入差距发挥的异质中介作用,再分别对发明专利占比(ipat)、实用型专利占比(ppat)和外观设计专利占比(apat)进行回归,结果如表 4 列(2)~列(4)所示。结果显示,相较于实用型专利和外观设计专利,数据劳动者对发明专利的正向影响更为显著,进一步支持了数据劳动者能够通过推动技术进步来缩小地区间收入差距的结果,H2 得证。

4.2　产业间收入差距

考虑到各省份优势产业的差异,以第一产业与第二、三产业城镇单位就业人员平均工资差值的绝对值与第二、三产业城镇单位就业人员平均工资之比($absw_2$)衡量产业间劳动收入份额差距;为进一步探究各产业城镇单位就业人员平均工资较于三大产业总体工资平均水平的相对变动在其中发挥的作用大小,分别构建第一产业与城镇单位就业人员平均工资差值的绝对值与城镇单位就业人员平均工资之比(absf)、第二、三产业与城镇单位就业人员平均工资差值的绝对值与城镇单位就业人员平均工资之比(absst),计算公式为

表 4　机制检验:技术进步

变量	(1) pat	(2) ipat	(3) ppat	(4) apat
数据劳动者	0.014 5***	0.023 1***	0.010 0**	0.064 8**
	(0.004 5)	(0.005 2)	(0.004 9)	(0.030 6)
截距项	0.030 6	0.005 8	0.042 5	−0.040 8
	(0.062 7)	(0.032 9)	(0.075 7)	(0.089 5)
控制变量	是	是	是	是
年份固定效应	是	是	是	是
省份固定效应	是	是	是	是
调整后 R^2	0.9796	0.988 3	0.968 2	0.940 2
观测值	279	279	279	279

注:***、**、* 分别表示 1%、5%、10% 的显著性水平;括号内为省份层面聚类标准误。

$$absw_2=\frac{|第一产业城镇单位就业人员平均工资-第二三产业城镇单位就业人员平均工资|}{第二、三产业城镇单位就业人员平均工资} \tag{19}$$

$$absf=\frac{|第一产业城镇单位就业人员平均工资-城镇单位就业人员平均工资|}{城镇单位就业人员平均工资} \tag{20}$$

$$absst=\frac{|第二三产业城镇单位就业人员平均工资-城镇单位就业人员平均工资|}{城镇单位就业人员平均工资} \tag{21}$$

产业间收入差距的机制检验结果如表5所示。列(1)结果显示,数据劳动者规模增加能够显著缩小产业间劳动收入份额差距。列(2)和列(3)结果显示,相较于数据劳动者规模增加对第二、三产业就业人员与城镇人员平均工资相对差距的扩大作用,其对第一产业与城镇人员平均工资相对差距的缩小作用更为显著。上述结果进一步补充了数据劳动者对产业间劳动收入份额差距的缩小作用,H3得以验证,表明数据劳动者能够通过缩小产业间就业人员收入差距来弥合地区间收入差距。

5　异质性

为进一步验证数据劳动者规模对地区间收入差距的弥合作用,基于地区、城镇化水平以及市场化程度三个角度对该弥合作用展开异质性分析。

5.1　东部和中西部地区

表6列(1)和列(2)为不同区域数据劳动者对地区间收入差距的异质性作用结果。进一步利用邹氏检验对二者系数进行组间差异检验,发现东部地区与中西部地区的系数存在显著差异。中西部地区基础设施和数字技术发展滞后,数字劳动者发展队伍受限,进一步限制了地区间收入差距作用的发挥。而东部地区蕴含的巨大数字红利吸引了众多数据劳动者,其完善的数字基础设施使得数字资源能更好地为数据劳动者所用,且辐射效应使得东部各省份发展状况更为均衡,从而一定程度上降低地区间收入差距[35-36]。

5.2　城镇化水平

城镇化实为农村人口向城镇转移的过程,其本质即为劳动力流动。劳动人口在城乡间、地区间的流动,可以缩小地区间收入差距[37-38]。鉴于人口城镇化是城镇化的核心要素,以城镇人口与常住人口比值衡量城镇化,按照各省份城镇化率分年度计算中位数,并以中位数为界将各省份分为城镇化高水平组和低水平组。表6列(3)和列(4)分别为低、高城镇化水平下劳动者对地区间收入差距的异质性实证结果。究其原因,城镇化水平较高地区,如长三角、京津冀、珠三角,出现了高中心性城市集群[39],此现象有助于区域经济一体化、促进政策协同强化强技术创新等的良性溢出效应,对弥合地区间收入差距意义非凡。

表5　机制检验:产业间收入差距

变量	(1) absw2	(2) absf	(3) absst
数据劳动者	−0.145 5* (0.071 9)	−0.153 4** (0.0710)	0.0180* (0.0092)
截距项	2.0592** (0.7883)	2.116 9*** (0.768 9)	−0.062 6 (0.122 1)
控制变量	是	是	是
年份固定效应	是	是	是
省份固定效应	是	是	是
调整后 R^2	0.776 2	0.768 3	0.629 7
观测值	279	279	279

注:***、**、*分别表示1%、5%、10%的显著性水平;括号内为省份层面聚类标准误。

5.3　市场化程度

中国自改革开放以来获得的经济成就部分得益于以市场化改革为方向的经济转型,市场化改革带来了经济、社会、法律体制等一系列的变革[40]。在规范、完善的制度支撑下,市场化意味着资源的公平竞争和充分利用,提高市场化程度能有效缩小收入差距[41]。选用樊纲指数衡量各省份的市场化程度,以2019年樊纲指数中位数为划分依据,将省份划分为市场化程度较低地区和市场化程度较高地区,结果如表6列(5)和列(6)所示。市场化程度较高的地区通常具备更加自由和开放的经济环境,信息的自由流动使得知识溢出更为活跃[42],能实现地区间资源的高效配置与协同利用,从而有效降低地区间收入差距。

6　结论与启示

新就业形态劳动者创造改善收入分配格局新路径,挖掘数据劳动者这一新兴群体的价值和潜能,把握数据劳动者作用于收入差距的深层逻辑,是数字中国背景下走向共同富裕道路的重要举措。结果表明:数据劳动者规模的扩大能够弥合地区间收入差距,经过工具变量检验和一系列内生性检验后,此结论依然成立;数据劳动者通过缩小产业间收入差距和推动技术进步实现地区间收入差距的弥合;异质性分析表明,在东部地区、城镇化水平和市场化程度更高的地区,数据劳动者规模增加对地区间收入差距的缩小效果更为明显。

表6　异质性分析(因变量:Gap)

变量	地区		城镇化水平		市场化程度	
	(1)	(2)	(3)	(4)	(5)	(6)
	中西部	东部	较低水平	较高水平	较低程度	较高程度
数据劳动者	−0.013 8	−0.020 3**	−0.074 3	−0.018 0***	0.042 7	−0.030 4***
	(0.048 0)	(−0.008 9)	(0.055 6)	(−0.005 6)	(0.042 8)	(0.010 2)
截距项	0.485 0***	−0.160 2	0.681 8***	0.194 7	0.362 3***	0.313 3*
	(0.167 9)	(0.207 8)	(0.207 5)	(0.112 6)	(0.105 5)	(0.148 7)
控制变量	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
省份固定效应	是	是	是	是	是	是
调整后 R^2	0.498 4	0.928 7	0.644 8	0.946 2	0.673 4	0.915 9
观测值	180	99	139	140	144	135
系数差异 P	1.61*		1.86**		1.70**	

注:***、**、* 分别表示1%、5%、10% 的显著性水平;括号内为省份层面聚类标准误。

本文提出了如下几点政策:①建立数据劳动者规模数据库。数据劳动者是数字经济时代迅速发展起来的一支队伍庞大的新型劳动者群体,但目前仍缺乏官方的统计标准和统计数据,导致其规模、结构以及其对于经济社会的贡献难以估量。部分灵活就业群体本身的估计具有很大的挑战性,可考虑与头部数字平台展开合作,利用平台资料与数据补充政府统计盲区。②统筹优化政策引导,调动一切积极因素。完善推动产业在国内有序梯度转移的协作机制,引导数字产业在地区间的梯度转移;积极培育农、林、牧、渔业等领域数字人才,提升第一产业的数字化水平,缩小产业间数字化水平差距。给予创新研发活动足够的资金投入和政策扶持,同时营造“允许试错、宽容失败”的创新环境。此外,充分发挥技术进步的红利,实现创新成果向现实生产力转化和前沿技术的跨区域应用,以技术进步成果各地区共享引领地区间收入差距问题改善。③实施区域差异化策略,推动数字发展因地制宜。对于数字优势先发地区,如东部地区、高城镇化率和高市场化程度地区,侧重“质”的提升,推动大数据、人工智能等高端数字产业集群发展,打造具有国际竞争力的数字创新高地,在优化数据劳动者从业环境的同时推进数据要素市场化配置改革;对于落后地区,优先实现“量”的突破,破除数据要素流动障碍,加快数字基础设施建设,扩大数字劳动力队伍和数字产业规模,以撬动地区间收入差距沉疴。

参考文献

[1] 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[J]. 中国工业经济,2020(5):42-60.

[2] 曾晶,余泳泽,缪言. 数据要素对劳动和资本的配置效应:机理分析与实证检验[J]. 中南财经政法大学学报,2024(1):148-160.

[3] 王欣亮,张家豪,刘飞. 大数据是经济高质量发展的新引擎吗?——基于数据基础设施与技术应用的双重效应解释[J]. 统计研究,2023,40(5):103-119.

[4] 范德成,肖文雪. 数字经济何以成为区域创新发展新动能——基于要素融合的实证分析[J]. 科技进步与对策,2025,42(5):69-81.

[5] 陶长琪,丁煜. 数据要素何以成为创新红利?——源于人力资本匹配的证据[J]. 中国软科学,2022(5):45-56.

[6] 刘涛雄,戎珂,张亚迪. 数据资本估算及对中国经济增长的贡献——基于数据价值链的视角[J]. 中国社会科学,2023(10):44-64.

[7] 何爱平,李清华. 数字资本价值运动全球化扩张逻辑的空间批判——基于马克思世界市场理论的分析[J]. 经济学家,2024(7):24-34.

[8] 李弦. 数据要素赋能新质生产力的理论逻辑与实践进路——基于马克思劳动过程理论的分析[J]. 上海经济研究,2024(5):25-36.

[9] 汪昊. 中国劳动和资本税收负担及分配效应[J]. 经济研究,2023,58(4):95-113.

[10] 汪伟,靳文惠. 养老保险缴费基数限额、收入不平等与社会福利[J]. 经济学(季刊),2023,23(6):2084-2103.

[11] 卢云鹤,万海远. 住房公积金制度的收入分配效应[J]. 经济学(季刊),2020,20(5):87-106.

[12] 贾俊雪,孙传辉. 公平与效率权衡:垄断、居民收入分配与最优财政货币政策[J]. 管理世界,2019,35(3):48-63.

[13] 郭凯明,罗敏. 有偏技术进步、产业结构转型与工资收入差距[J]. 中国工业经济,2021(3):24-41.

[14] WANG M,LI B. Urban-rural income gap and urban crimerate[J]. Finance Research Letters,2024,63:105285.

[15] 周兴,王芳. 中国城乡居民的收入流动、收入差距与社会福利[J]. 管理世界,2010(5):65-74.

[16] 张云,柏培文. 数智化如何影响双循环参与度与收入差距——基于省级-行业层面数据[J]. 管理世界,2023,

39(10):58-83.
[17] XIAH, YU H, WANG S, et al. Digital economy and the urban-rural income gap: impact, mechanisms and spatial heterogeneity [J]. Journal of Innovation & Knowledge, 2024, 9(3): 100505.
[18] HE B, XU D, NAN G, et al. Does the cross-border ecommerce comprehensive pilot zones policy affect the urban-rural income gap in China? [J]. American Journal of Economics and Sociology, 2024, 83(4): 773-792.
[19] 洪俊杰,李研,杨曦. 数字经济与收入差距:数字经济核心产业的视角[J]. 经济研究,2024,59(5):116-131.
[20] 埃里克·布莱恩约弗森,安德鲁·麦卡菲. 第二次机器革命[J]. 理论与当代,2014(9):53-60.
[21] FIGUEIREDO O, GUIMARÃES P, WOODWARD D. Industry localization, distance decay and knowledge spillovers: following the patent paper trail[J]. Journal of Urban Economics, 2015, 89: 21-31.
[22] 武宵旭,任保平. 数字经济背景下要素资源配置机制重塑的路径与政策调整[J]. 经济体制改革,2022(2):5-10.
[23] 佟伟铭,郭加新,徐维祥,等. 县域视角下乡村发展要素转型对共同富裕的影响研究——以浙江省为例[J]. 地理研究,2023,42(6):1577-1597.
[24] 何可,吴昊,曾杨梅. "双碳"目标下的智慧农业发展[J]. 华中农业大学学报,2023,42(3):10-17.
[25] 易法敏,古飞婷,罗必良. 乡村数字经济包容性发展的政策组合及其治理——基于"电子商务进农村综合示范"的案例研究[J]. 管理世界,2025,41(5):101-116.
[26] 浦徐进,陈柏含,李秀峰. 农产品供应链数智化转型研究的历史演变和未来展望[J]. 中国管理科学,2025,33(4):235-250.
[27] DAGUM C. A new approach to the decomposition of the Gini income inequality ratio[J]. Empirical Economics, 1997, 22(4): 515-531.
[28] 高子茗. 外卖配送新业态发展与工资性别差异变化[J]. 财经研究,2024,50(10):49-63.
[29] 李健旋. 数字化转型降低企业劳动力成本粘性机理及异质性分析[J/OL]. 科研管理,1-20[2025-06-11]. http://kns.cnki.net/kcms/detail/11.1567.G3.20241130.1100.002.html.
[30] 刘李华,孙早. 人口老龄化、居民健康与收入不平等[J]. 经济科学,2022(5):135-149.
[31] 朱奎. 经济危机的收入不平等效应:基于2008年美国金融危机[J]. 马克思主义研究,2016(10):66-74.
[32] 杨穗,高琴,李实. 中国社会福利和收入再分配:1988—2007年[J]. 经济理论与经济管理,2013(3):29-38.
[33] 江艇. 因果推断经验研究中的中介效应与调节效应[J]. 中国工业经济,2022(5):100-120.
[34] 刘清春,张莹莹,LIN L C. 创新与收入不平等[J]. 南方经济,2017,36(2):24-39.
[35] 赵伟,彭玉婷. 数字经济发展是否会影响收入不平等?——基于空间面板模型的实证检验[J]. 经济问题探索,2022(12):35-51.
[36] 詹筱媛,张丽君,刘晴,等. 数字基础设施建设对居民收入机会不平等的影响——基于中国地市级面板数据[J]. 经济问题探索,2024(7):172-190.
[37] 陆铭,向宽虎,李鹏飞,等. 分工与协调:区域发展的新格局、新理论与新路径[J]. 中国工业经济,2023(8):5-22.
[38] 姚枝仲,周素芳. 劳动力流动与地区差距[J]. 世界经济,2003(4):35-44.
[39] 邵云通,吴晓. 中国城际人口流动网络的多尺度空间差异——基于腾讯迁徙大数据的分析[J]. 地理科学,2025,45(1):189-201.
[40] 樊纲,王小鲁,马光荣. 中国市场化进程对经济增长的贡献[J]. 经济研究,2011,46(9):4-16.
[41] 王小鲁,樊纲. 中国收入差距的走势和影响因素分析[J]. 经济研究,2005(10):24-36.
[42] 刘华珂,李旭超,聂禾,等. AI时代:城市数智化转型与企业创新[J]. 中国软科学,2024(2):38-54.

The Role of Data Workers in Interregional Income Gaps: Bridging or Widening?

GAO Li, ZHANG Yi

(School of Economics and Management, China University of Petroleum (Beijing), Beijing 100249, China)

Abstract: The fusion of data and labor factors has given rise to data workers, and the emergence of this new group has created a new path to improve the pattern of income distribution and regional development imbalance. Based on the panel data of 31 provinces in China from 2014 to 2022, the effect of data laborers on the inter-regional income gap and the mechanism of its action were systematically examined. It is found that data workers can bridge the inter-regional income gap, and this result still holds after a series of endogeneity tests and robustness tests. In the mechanism analysis, the mediating effects of inter-industry income gap and technological progress are tested, which confirms that they can serve as the mediating variables of data workers' influence on the inter-regional income gap. In the eastern region, the urbanization level and the marketization of the more advanced regions, data laborers can release greater digital dividends, and the effect of their increased size on the reduction of inter-regional income gap is more obvious. New theoretical perspectives and policy insights on how new types of laborers affect income disparity in the era of digital economy are provided.

Keywords: data laborers; data elements; income gap; interregional income disparities