

· 数字经济 ·

互联网使用对个人收入的提升作用及 分化效应研究



□李梦凡

[河海大学 南京 211100]

[摘要] **【目的/意义】** 中国互联网基础设施建设和应用发展迅速，但个人的互联网应用能力还存在发展不平衡、不充分的问题。在日益深化的互联网时代，探讨互联网使用对个人收入的影响，对优化收入分配格局有重要实践价值。**【设计/方法】** 基于CGSS微观社会调查数据，利用基本回归比较、探索性因子分析、倾向得分匹配方法识别因果效应，并辅之以敏感性分析、稳健性检验和异质性分析，实证检验互联网使用对个人收入的提升作用与分化效应。**【结论/发现】** 考虑应用能力，互联网使用的个人收入提升作用显著且稳定，异质性分析表明相较于城市而言农村人口在互联网使用受益更多，相对于男性来说女性收入提高效应更明显，年轻群体的互联网使用收入溢价更高，技能分层后互联网使用红利分化特征显著。在接入可及不均衡状况改善阶段，互联网使用收入提升效应显著。跨越该阶段以后，需重视如何应对互联网使用“红利分化”相关问题。

[关键词] 互联网使用；数字能力；处理效应

[中图分类号] F49

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2022)-4006

Research on the Promotion and Differentiation Effect of Personal Income from Internet Use

LI Meng-fan

(Hohai University Nanjing 211100 China)

Abstract [Purpose/Significance] China's internet infrastructure construction and application have developed rapidly, but there are still unbalanced and insufficient development of individuals' capabilities. In the deepening Internet era, exploring the impact of Internet use on personal income has important practical value for optimizing the pattern of income distribution. [Design/Methodology] Based on CGSS micro-social survey data, the causal effects were identified by basic regression comparison, exploratory factor analysis, and propensity score matching methods, supplemented by sensitivity analysis, robustness test and heterogeneity analysis, to empirically test the enhancement and differentiation effect of Internet use on personal income. [Conclusions/Findings] Considering individuals' application ability, the personal income improvement effect of Internet use is significant and stable, the heterogeneity analysis shows that the rural population benefits more compared with the urban population, females' income enhances more than that of men, young group gets more than old group, and the skill stratification is significant. In the stage of improving the imbalance in access availability, the income increasing effect of Internet use is significant. Crossing this stage, we need to pay attention to how to deal with the problem of "dividend differentiation".

Key words internet use; digital capability; treatment effect

[收稿日期] 2022-07-04

[基金项目] 国家社科基金（18CJL046）。

[作者简介] 李梦凡（1987-）男，博士，河海大学公共管理学院讲师。

引言

从农耕时代到工业时代再到信息时代,技术不断推动人类社会的发展。进入新世纪以来,互联网正以不可逆转的力量,在全球范围掀起一场涉及诸多层面的深刻变革,不断刷新生产、流通、分配和消费的模式。人类正经历所谓的“互联网时代”,互联网技术的大众化应用,日渐成为人们社会生产以及日常生活的重要组成部分。2002年,党的十六大报告中首次出现“互联网”相关内容。2017年,党的十九大报告多处提及互联网,涵盖网络文化、网络安全、网络管理等多个方面。2020年,党的十九届五中全会又特别强调发展数字经济。近年来,中国的网络建设持续加速、应用环境全方位优化、基础资源保有量不断攀升,不仅在成功应对突发公共卫生、安全等事件上提供了关键技术支撑,也逐步成长为推动经济迈向高质量发展的新引擎之一。2021年,中国互联网络信息中心(CNNIC)正式发布的第47次《中国互联网络发展状况统计报告》指出,截至2020年12月,中国网民规模达9.9亿,互联网普及率达70%。其中,农村网民规模达3.1亿(城镇6.8亿),占网民群体的31%;受教育程度在初中及以下的网民群体占比近60%;有收入但月收入在1000元以下的网民群体占比超过15%^①。这些数据表明,中国互联网行业和相关基础设施建设发展迅速,惠及人群愈加广泛。但在互联网普及率与互联网应用能力等方面,还有一定程度的发展不平衡、不充分问题。正像历次技术变革存在的多方面伴生现象一样,随着互联网技术的深化和扩散,“数字鸿沟”“数字贫困”等现象也在频频进入世界范围内的公众视野。已有研究指出^[1]，“数字鸿沟”会经历两个发展阶段：首先是接入可及性差异阶段，这曾是数字鸿沟的基本形态；其次是互联网应用差异阶段，这个阶段会触发互联网红利差异。也就是说，随着数字鸿沟的发展，收入分配格局也将产生重要的动态变化。那么，在当前系列复杂综合因素的作用背景下，互联网使用对居民个人收入的影响如何，其作用有无分化趋势，是本文尝试研究的主要问题。

一、文献回顾

(一) 互联网使用与劳动者收入提升

2019年中国社会科学院发布的《人口与劳动绿

皮书：中国人口与劳动问题报告No.19》中提及，互联网使用行为可使个体年收入增加46.5%，而且这对中低收入、中高技能劳动群体影响显著^②。陈玉宇和吴玉立较早地研究了个人电脑的使用与工资之间的关系。该研究利用2005年全国家庭普查数据，采用虚拟变量和趋势得分模型估计使用个人电脑的工资回报率。在控制性别、年龄、教育水平、职业以及工资黏性等因素的情况下，回报率高达20%^[2]。以农业领域的劳动者为例，许竹青等研究了信息的有效供给对农民跨越“数字鸿沟”“享受信息红利”的影响。这项研究通过对农业信息短信服务与农产品的售价的因果关系以及内在作用机制进行分析，发现信息有效供给能显著提高易腐产品（豆角）售价，从而提高农业从业者的收入^[3]。曾亿武等研究了电子商务与农户农业收入的关系，发现前者显著地促进了后者，而增收差异可以追溯到物质、人力和社会资本在内的资本禀赋差异^[4]。程名望、张家平研究了信息与通讯技术与城乡收入差距之间的关系。该项研究先基于理论分析假设二者间的非线性关系，然后采用2003~2016年中国省级面板数据进行验证。发现互联网普及对城乡收入差距呈现先增后降的“倒U型”影响，并确认时间拐点在2009年，这意味着互联网技术给中国城乡融合发展带来了重要机遇。最后，还利用中国社会状况调查数据（CGSS2015），发现现阶段互联网普及能够缩小城乡居民收入差距的直接原因是其对农村居民收入的提升效应相比城镇居民更大^[5]。赵羚雅、向运华基于CFPS2016数据，考察互联网使用和社会资本对非农就业的影响，发现互联网使用使得农村居民非农就业概率提高了49.3%，社会资本是互联网使用影响农民非农就业的重要途径^[6]。邱子迅和周亚虹通过匹配“中国电子商务发展指数”与CFPS2016和CFPS2018数据，从供需有效对接的角度，研究了电子商务与农村家庭增收的关系以及异质性^[7]。刘生龙等研究了农村居民使用互联网的收入回报。该研究利用CFPS2010、CFPS2014和CFPS2018的个人调查数据，采用广义随机森林模型、工具变量法发现互联网使用对农村人口收入增加有显著因果效应，发现该效应在青年和较高教育程度人群中尤为明显，提出并验证了互联网影响农村居民收入的路径机制，如更多的就业机会、更灵活的就业形式等^[8]。罗楚亮和梁晓慧采用CMDS2016数据研究也发现，借助互联网找工作，可以降低信息不对称，产生工资溢价效应，而且该效应在教育程度高、累计流动时间长的群体中更强^[9]。

（二）提升效应论争与异质性

但是同时还有学者（如刘生龙等）指出，互联网对个人收入的影响，学术界仍有争论^[8]。尽管该研究领域较有影响的克鲁格很早就利用美国人口普查数据实证分析得出，电脑使用对个人收入具有10%~15%的提升作用^[10]。然而后续许多研究者质疑了克鲁格的研究与结论。汉德尔指出，该研究利用截面数据做回归，有可能存在遗漏变量导致的偏误，从而使得估计结果不具备说服力。贝尔的研究支持了这种质疑，他利用固定效应控制不可观测的个体差异之后，发现互联网使用对工资的影响不显著^[11]。类似的，皮施克做了一项很有意思的研究。他考虑办公工具的选择性，利用德国的截面数据，发现白领工人在工作中使用的工具，无论是电脑还是普通铅笔，都会与高收入有很强的相关性。借此他指出，在不控制个体不可观测特征的情况下，简单的回归结果难以保证可靠的因果推断^[12]。赵建国、周德水利用2016年中国流动人口动态监测调查数据，研究了互联网使用与大学毕业生工资的关系。基于信息搜寻理论、运用倾向值匹配法和分位数回归法，发现前者对后者提高有显著作用，但随着分位点的提高，互联网使用的影响程度不断减弱，呈倒U形趋势。此外，影响效应存在区域差异性、户籍差异性^[13]。谭燕芝等考察了信息化与个人收入的影响以及城乡差异。该研究基于CFPS2014数据，发现在控制个人异质性特征的情况下，互联网使用能够对个人收入提升14%，城市样本中约为20%，但是农村样本并不显著。值得关注的是，这个研究还发现城乡互联网重视程度并无显著差异，回报率差异根源可能在互联网应用能力差异上^[14]。蒋琪等基于CFPS2010、CFPS2014的面板数据，运用FE和PSM-DID模型，估计互联网使用对居民个人收入的影响，发现互联网对中年人、农村户籍群体、高等教育群体的影响相对较大，而认知能力和互联网用的学习用途，是可能的作用机制^[15]。迪马吉奥和豪尔吉陶伊基于美国综合社会调查发现教育、认知能力和收入高的群体更多利用互联网“积累资本”，而非单纯娱乐^[16]。这一观点，在邦法德利的研究中得到再次验证，后者利用瑞士数据发现，高教育、高收入者一般利用互联网获取经济利益，社会经济地位低的用户娱乐居多^[17]。郝大海和王磊应用中国的数据（CFPS）获得了与上述研究类似的结论。该研究通过多层次回归模型，区分地区和个体因素，发现地区因素影响互联网接入可及性，而接入网络后地区因素的影响下降，网络应用

目的差异更多地与个人社会结构性因素有关^[18]。申广军、刘超使用企业层面微观数据（中国工业企业数据库）发现，使用信息技术的企业其劳动收入份额更高，也就是说信息技术存在分配效应，且具有异质性^[19]。这说明就业机会、形式和类别很可能会影响互联网使用的收入提高效果。王元超基于互联网对个人收入提高具有的技术效应和资本效应，探讨两种效应的阶层异质性，通过CFPS2014数据实证发现“倒U型”的效应-阶层关系^[20]。叶明睿和蒋文茜基于信息鸿沟和相对贫困视角，通过田野调查，对信息增收行为路径做出感知、筹划、行动和评估的阶段式划分，研究发现，拟态环境构建、结构系统性认知、社会资本现实转化等方面的信息能力困境是制约贫困群体信息增收的重要表现^[21]。冯喜良等利用CLDS2014和CLDS2016数据定量研究互联网使用的性别工资收入差距，发现农民工整体的工资溢价效应明显，且互联网使用可以缩小性别工资收入差距。互联网使用技能进行精准化培训，尤其是提高女性农民工互联网的使用能力，是缩小农民工的性别工资收入差距的关键步骤^[22]。

（三）文献启示与本文思路

梳理文献发现，尽管大部分研究支持互联网使用对个人收入的正向影响，但仍有存疑之处，如影响的不确定性、影响的群体差异性。对于识别互联网使用的收入处理效应，多数研究在“直接利用OLS回归结果很可能存在偏误”上能达成共识，但是究竟应使用什么识别方法，看法及做法并不统一。相当一部分研究采用倾向得分匹配（PSM）方法。另有研究尽管使用PSM方法，也意识到该方法并不能很好解决内生性问题，因为不可观测因素的干扰。但是只进行了平衡性检验、共同支撑区间检验和多种匹配方法的稳健性检验，这些检验措施仍然不能或不能很好地解决不可观测因素干扰问题。除了直接测算的经验研究，一些文献还探究了更深层次的因素，如互联网应用能力、互联网资本等。如邱泽奇等基于互联网资本视角指出，因接入机会差异导致的数字鸿沟，这是信息富有和信息贫穷之间的差异，可以由基础设施的发展建设来填平。这种“改善”同时触发了互联网红利差异，关乎人们互联网运用上的差别而产生的数字不平等^[1]。

这些深层因素可否作为分析不可观测干扰的切入点呢？正是在上述文献的启发下，本文使用倾向得分匹配思考设计互联网使用的收入处理效应估计，考察互联网使用能力（数字能力）因素，并通过敏感性检验来确定模型是否恰当，直到找到对不

可观测干扰不敏感的模型为止,从而较为准确地识别出互联网使用的收入提升效应,然后在此基础上展开稳健性检验、异质性分析与相关讨论。本文的边际工作主要体现在:首先,考虑并测算了不可直接观测的潜变量——数字能力;其次,在因果效应识别方法倾向得分匹配分析中加入了敏感性分析与检验;最后,在因果效应识别的基础上的异质性分析,突出强调新阶段互联网使用的红利分化效应。希冀对相关领域的治理政策设计提供一些经验依据。

二、互联网使用的收入提升与分化效应

(一) 理论分析

一般认为,影响微观个人收入主要有其掌握的物质资本、人力资本和社会资本等因素。随着新工具、新基础设施、新科学知识、新经济空间秩序等产生和发展,物质资本、人力资本和社会资本的外延也得到拓展和丰富。然而在大多数情况下,这些发展并没有从根本上改变上述三个维度资本的根本内涵。因此,长期以来传统的三维资本理论,是解释个人收入、收入变化及其差异的基础。如果计算机和互联网的应用对个人收入带来了影响,那么我们不禁会问:在解释个人收入时,它应当归为上述哪个类别呢?

看似都有关联。克鲁格考察计算机革命给工资结构带来的影响时,提出“互联网工资溢价”现象,并通过经验数据,在同等条件下工作中是否使用计算机导致的个体工资差异^[10]。为什么互联网使用会带来工资溢价或收入增加呢?此后一些研究者,将技术效应和资本效应作为重要阐释视角。技术效应采取典型的新古典分析视角强调互联网技术应用带来的直接收益,通过新技术(计算机与互联网)的应用,一方面将传统劳动领域中的劳动者解放出来,这部分人从事更具创造性、更高效的工作,工资收入会随着工作效率的提高而增加。另一方面,随着互联网技术的扩散普及、基础设施的投资建设,会对掌握互联网使用技术的劳动者产生新增需求,这在劳动力市场上会增加互联网使用者的工资议价能力。与技术效应相对,资本效应强调互联网技术应用带来的间接收益。计算机和互联网技术发展,与以往技术进步存在明显差异,它在减少信息获取的成本、扩展人们获取信息的能力等方面具有显著优势。从劳动力市场上看,无论对于更好工作的求职机会,还是增加市场急需的特定类型的人力资本积累等方面,互联网的使用者掌握着更多

机会,而这会间接实现工资增长。由上述讨论就可以看出,这些解释并没有跳出三维资本的范围。新技术的发生天然就与物质资本积累相关,其作用的发挥往往是通过一定的物质基础设施来实现。计算机和互联网的应用也不例外,大到公共网络基础设施,小到每个人手中的通讯设备。此外,由巴赫等提出过“数字人力资本”(Digital Human Capital)的概念^[23],也印证了互联网越来越成为把人的能力转化为生产、教育以及参与社会的工具,又与人力资本相关。互联互通,还可能使得互联网使用者的社会关系网络得到强化与拓展,降低信息获取的成本,提高获取信息的能力,又对应着社会资本理论的部分要义。

但以互联网为代表的新技术作用又有其独特性。这里本文借鉴了索托的概念界定与思想观点,他将“资本”界定为“凝聚以往投入而成、具有市场进入机会、能够经由市场获益的资产”,它首先是需要并且可以被积累的生产要素,同时它的作用过程又体现出一定的社会经济机制^[24]。综合来说,可以将资本看作内含特定社会机制的发展要素。人们不断积累它,通过它获取收入、培养社会能力和实现个人发展。本文认为,从发展要素的意义上说,物质资本、人力资本和社会资本等,也可以视为种种具体的内含特定社会机制的发展要素或资产。人们掌握有各式各类、形态各异的资产,从劳动力到物质资产,如房屋、土地、自然资源等。如果这些资产无法进入市场,就是“僵化”的资本^[24]。这些资产获得进入市场机会,进而得到获取收益的可能,称作资产转化为资本,即资本化。经历资本化过程,人们就不仅是各类资产占有者,而且是与市场形成有机联系的参与主体,资产就经历实际意义上的生产要素“投入”环节、其价值在市场运行中得到评价、完成向“产出”的转化,进而成为资本。邱泽奇等将索托的逻辑引入互联网应用中,提出“互联网资本”的概念:任何因既往投入形成的、具有互联网市场进入机会并可以通过互联网市场获益的资产^[1]。如此一来,无论是三维资本中的哪一维度,人们占有的资产都存在一定程度上的“闲置”“僵化”,并未时时刻刻“开足马力”给其所有者增加收入。失业或工作机会的缺乏,可以视为人力资本的闲置。局部供需失衡、停工停产,可看作物质资本的闲置。地理交通不便、通讯障碍等导致本可共享分享的社会网络信息滞后或缺失,则体现了社会资本的闲置。不仅如此,产品声誉、技术能力,甚至零碎时间等都属于索托所提到的

“僵化”资本，就明显带着互联网时代的特征，即发现新发展要素的赋能效应。所以说，互联网技术的应用，就像市场经济中的“扩音器”“放大镜”“探测仪”“传呼机”等，有利于发现、激活和强化这些“僵化”资产，为其打通投入通道和实现产出过程拓展了机会空间，让曾经难以转化为资本的资产转化为互联网资本。

互联网使用需要两方面的条件：一个互联网使用的可及性，这需要通过各类互联网基础设施的不断建成、完善、接入来达成的，是一个客观的条件；另一个是互联网使用的能力，这是劳动者通过学习或培训提高互联网工具的使用能力来达到的，是一个近似主观能动的因素。本文在此需要加以明确的是，这两个方面之间并不存在简单的互补关系，尤其是在基础设施普及率较高的阶段。当互联网工具作为一种新技术简单应用到生产生活的推广阶段，往往方便快捷，对使用技术要求不高，这时是否互联网使用对于劳动者收入差异的影响，表现为接入机会差异导致的“数字鸿沟”。但是，随着基础设施的逐步建设和完善，对应用能力的要求不断提高，即使拥有同等接入机会，互联网使用能力上的差异开始在个体收入分布上影响力凸显。而且更为重要的是，随着生产、生活中互联网接入场景的增多、深化，数字能力较低的群体甚至会被排斥、隔绝在互联网赋能效应圈外部，不仅未获得便利，反而转变为发展劣势。例如，一些老年人面临的“数字贫困”现象就属此类。当出现这种情况的时候，就不能单靠进一步加大对互联网基础设施的投入来解决，该措施有时甚至就是导致阶段性“数字贫困”现象的直接原因。所以说，在考察互联网使用的工资溢价效应的时候，除了考虑传统三维资本，还需考虑互联网使用在主、客观等方面的多重赋能作用。

讨论互联网使用对个人收入影响，本文主要是通过多种路径来进行分析的：一是互联网使用直接影响个人收入，这里称之为“技术效应”（或接入效应）；二是互联网通过对经典三维资本（物质、人力和社会资本）“赋能”，从而间接影响个人收入，这即是“资本效应”（或接通效应）；三是考虑互联网使用能力或数字能力，会影响个人对互联网使用的概率和效果，这里统称为“强化效应”。复杂性一方面体现在数字能力难以观测，可能存在“内生性问题”；另一方面在微观个体之间数字能力差异会引起互联网使用红利的“分化效应”。作为不可观测因素的数字能力，对于本文的研究主题

非常重要，因此被纳入到理论分析以及后续的经验研究中。

（二）实证策略

作为参照标准，本文先估计线性回归模型。借鉴克鲁格“互联网-工资”方程^[10]，构建如下基准模型：

$$y_i = \alpha + \alpha_1 \text{treat}_i + \beta_i X_i + u_i \quad (1)$$

其中， i 表示不同个体，被解释变量 y_i 为小时工资对数，核心解释变量 treat_i 为互联网使用情况，控制变量 X_i 为一系列由人口、经济、社会、家庭、省份等相关特征变量组成的向量。待估参数 α_1 表示互联网使用对于个人收入的影响，这是本文关注的估计系数。 u_i 是本模型中的干扰项。简单线性回归模型，可能存在由样本选择偏误引起的估计偏差，也即 $\text{cov}(\text{treat}_i, u_i) \neq 0$ 。

为了避免模型设定引起的偏差，本文还使用匹配方法来估计因果效应。倾向得分匹配方法借鉴Rosenbaum和Rubin的分析思路，为解决维数诅咒导致无法匹配难题，从由系列观测变量 X_i 进行匹配转向一维的倾向指数进行匹配。倾向指数是具有特征 X_i 的个体工作使用互联网（ $\text{treat}=1$ ）的可能性，也即个体干预分配概率，一般采用限制因变量（如logit、probit）模型估计得到。在此基础上，分别使用一对一近邻匹配、卡尺匹配、核匹配、局部线性匹配等方式，估计得到互联网使用对于个人收入的平均处理效应，并进行相应的检验。这是一种非参数方法，但不能规避不可观测因素的干扰。前面的理论分析指出，在互联网使用与收入之间，的确存在着不可忽视的不可观测因素（互联网使用能力）的影响。

当存在不可观测的选择（隐性偏差），为了评估倾向得分匹配估计的可靠性，需要进行相应检验。Rosenbaum提出倾向得分匹配相应的敏感性分析思路^[25]：当两个个体接受处理的倾向值不同时，会存在隐藏性偏差，敏感性分析主要是研究这种偏差大到何种地步会改变之前做出的因果效应推断结论。敏感性越低，表明之前的结论越可靠。否则，应该对倾向得分匹配估计结果保持谨慎，此时就要考虑重新设计匹配模型。本文后面将结合具体实证过程详细展开分析。

在处理效应对不可观测因素的干扰敏感性很高时，就要考虑如何应对这种影响。这里本文将采用探索性因子分析的方法，在模型（1）的基础上，将原本不可观测的互联网应用能力（ability）纳入模型，进行回归和匹配分析：

$$y_i = \gamma + \gamma_1 \text{treat}_i + \gamma_2 \text{ability}_i + \beta_i X_i + \varepsilon_i \quad (2)$$

其中, ε_i 是本模型中的干扰项, 与模型(1)有所差异, 因为控制了(1)中的不可观测因素, 这里假设干扰项与处理变量无关, 即 $\text{cov}(\text{treat}_i, \varepsilon_i) = 0$ 成立。当然, 这种假设是否站得住脚, 或者说在多大程度上可被接受, 还需后面的敏感性分析来检验。我们关注的估计系数是 γ_1 , 除上述对应用能力变量的控制和对干扰项的假定之外, 其他方面与模型(1)一样。通过倾向得分匹配方法估计出处理效应后, 考察所得效应是否通过检验和敏感性分析。确定可以通过敏感性分析后, 将进一步分性别、年龄层、城乡户籍、技能高低等进行处理效应异质性分析。

三、互联网使用对收入影响的实证分析

(一) 数据来源、主要变量选取与统计分析

本文使用的是中国综合社会调查(Chinese General Social Survey, CGSS) 2017年的数据。CGSS是中国最早的全国性、综合性、连续性学术调查项目, 由中国人民大学中国调查与数据中心负责执行, 旨在系统、全面地收集社会、社区、家庭、个人多个层次的数据, 总结社会变迁的趋势, 探讨具有重大科学和现实意义的议题, 推动国内科学研究的开放与共享, 为国际比较研究提供数据资料, 充当多学科的经济与社会数据采集平台。CGSS2017共完成有效样本12582份, 问卷由3大模块构成, 分别是A核心模块、C社会网络和网络社会(含ISSP2017)模块和D家庭问卷(含EASS2016)模块, 公布数据包含783个变量。尤其是A和C模块中包含居民使用互联网情况数据, 是目前国内罕见的、具有全国代表性的、个体互联网使用数据, 特别是有数字能力(互联网使用能力)的测量指标, 较为符合本文的研究需要。

本文首先关注工作中是否使用互联网对个人收入的影响。互联网使用采取“工作中使用互联网所占的比例”来表示, 将其处理为一个二值变量, 大于0赋值为1, 表示工作中使用互联网, 否则赋值为0, 表示工作中没有使用互联网^⑥。如表1所示, 个人收入采取对数形式, 用个人全年总收入的对数(tinc)来表示^⑦。对于被解释变量和主要解释变量, 在删除缺失值、无效或异常数据之后, 共得到9990个样本。为了尽量降低混淆和干扰, 研究还控制了可能影响个人收入的其他变量, 主要包括人口统计特征变量、社会特征变量、经济特征变量、家

庭特征变量等。首先, 个体人口统计学特征包括性别、年龄、受教育年限、健康状况等。性别(male)采取“是否是男性”来表示^⑧; 年龄(age)用调查年份(2017年)减去出生年份得到; 受教育程度(edu)采用受教育年限来测量^⑨; 健康状况(health)是受访个人的主观健康水平^⑩, 分为五个等级, 数字越大代表健康水平越高。其次, 经济特征包括拥有房产数量、是不是工会会员、有无金融资产投资、工作类型等。房产(house)记录包括与他人共有在内的所拥有的房产数量^⑪; 工会会员(union)不一定与福利挂钩, 但在信息获取、权益保障等方面具有优势, 会影响到个人收入^⑫; 工作类型根据问卷问答来分类, 每种类型的工作生成0-1二值变量^⑬。再次, 社会特征包括民族、有无宗教信仰、政治面貌、户籍状态、社会福利状况。民族采用“是否是汉族(han)”来表示^⑭; 宗教信仰(religion)是一个二值变量^⑮; 政治面貌(party)采用“是否为中共党员”来表示^⑯; 户籍状态(hukou)也是一个二值变量, 非农户口赋值为1, 否则为0。

描述性统计显示, 工作中有互联网使用的群体占总样本约15%, 相比未使用互联网群体而言, 使用群体个人年收入对数值平均要高出73%左右。他们的受教育程度更高, 其母亲的受教育程度也明显更高。这部分人更加年轻, 未婚者更多, 拥有投资和房产的概率更大, 参与工会的比例较高(详见表1)。我们从现有文献和研究结论可知, 除收入外的这些变量都是直接或间接影响个人收入的重要因素。这个统计显示出, 工作中互联网使用和互联网使用者本身之间存在较为明显的“选择性”特征。同时也说明, 在进行简单回归分析时, 这些变量都有被纳入模型进行控制的必要性。

(二) 回归分析

以便于比较研究, 本文先对工作中互联网使用与个人年收入之间的关系, 参照模型(1), 进行基于OLS回归模型的简单估计。为了更好的分析, 还加入协变量、控制区域变量等。如表2所示, 列(1)未添加任何控制, 列(2)控制了区域互联网平均应用水平, 列(3)控制了省份变量, 列(4)控制了个人(如受教育程度、性别、年龄、健康等)特征, 列(5)增添社会特征控制变量(如民族、户籍、政治身份等), 列(6)增加控制了经济与家庭特征(如是否加入工会、投资、住房、婚姻、子女数量等)。需要说明的是, 这里本文没有控制工作类型, 主要是考虑到工作类型与互联网使

表 1 主要变量描述性统计表

变量名	变量含义	控制组 (treat=0)	处理组 (treat=1)	全样本
收入 tinc	个人年收入 (对数)	9.833 [1.328]	10.569 [1.178]	9.944 [1.333]
教育 edu	受教育年限0-19	9.489 [4.472]	11.96 [4.105]	9.898 [4.508]
性别 male	男性=1, 否则=0	0.457 [0.498]	0.511 [0.500]	0.466 [0.499]
年龄 age	连续整数	45.49 [12.43]	40.21 [11.63]	44.62 [12.46]
婚否 marriage	已婚=1, 否则=0	0.809 [0.393]	0.775 [0.418]	0.803 [0.398]
健康 health	健康等级1-5	3.571 [1.078]	3.844 [0.919]	3.616 [1.058]
汉族 han	汉族=1, 否则=0	0.916 [0.277]	0.945 [0.227]	0.921 [0.270]
户籍 hukou	非农户口=1, 否则=0	0.415 [0.493]	0.612 [0.487]	0.448 [0.497]
宗教 religion	有宗教信仰=1, 否则=0	0.099 1 [0.299]	0.086 1 [0.281]	0.096 9 [0.296]
党员 party	党员=1, 否则=0	0.088 1 [0.284]	0.148 [0.355]	0.098 [0.297]
投资 invest	有投资=1, 否则=0	0.107 [0.309]	0.208 [0.406]	0.124 [0.329]
房产 house	连续整数	1.104 [0.652]	1.183 [0.754]	1.117 [0.671]
工会 union	工会成员=1, 否则=0	0.169 [0.375]	0.256 [0.437]	0.183 [0.387]
子女 nchild	连续整数	1.455 [0.950]	1.138 [0.855]	1.402 [0.942]
母亲教育 medu	受教育年限0-19	3.517 [4.370]	5.704 [4.566]	3.879 [4.477]

注：括号内为标准差。全样本数N=9990，其中控制组N=8489，处理组N=1501。

表 2 OLS回归结果表

	(1)	(2)	(3)	(4)	(5)	(6)
			被解释变量个人总收入对数 (tinc)			
treat	0.727*** (20.216)	0.545*** (16.996)	0.531*** (16.379)	0.206*** (7.241)	0.172*** (6.130)	0.152*** (5.355)
edu				0.095*** (32.719)	0.067*** (20.215)	0.057*** (16.361)
male				0.295*** (13.853)	0.322*** (15.302)	0.326*** (15.331)
age				0.107*** (15.865)	0.100*** (15.076)	0.084*** (11.459)
age2				-0.001*** (-16.785)	-0.001*** (-16.817)	-0.001*** (-13.423)
health				0.192*** (17.454)	0.184*** (17.045)	0.176*** (15.990)
han					0.235*** (5.945)	0.231*** (5.781)
hukou					0.398*** (15.421)	0.328*** (11.911)
religion					0.021 (0.578)	0.041 (1.111)
party					0.173*** (4.881)	0.129*** (3.589)
union						0.078** (2.562)
invest						0.267*** (7.772)
house						0.077*** (4.888)
marriage						0.235*** (7.474)
nchild						-0.073*** (-4.921)
medu						0.005 (1.527)
区域互联网	NO	YES	NO	YES	YES	YES
省份变量	NO	NO	YES	NO	NO	NO
常数项	9.826*** (704.709)	10.805*** (285.142)	6.911*** (116.292)	4.399*** (26.756)	4.788*** (28.592)	5.218*** (29.412)
样本数量	9990	9990	9990	7486	7486	7093

注：括号内为标准差。

用之间存在较强的关联,且互联网使用会影响工作类型,若控制住则会低估互联网使用的作用。本部分的估计结果如表2所示。

回归结果显示,在没有任何协变量的情况下,平均而言,工作中的互联网使用使得个人年收入增加107% (计算方法为 $e^{0.727}-1$),且非常显著。但这个影响太大了,由于可能存在选择性偏差,所以引入协变量进行回归。在此之前,考虑到工作种类极可能影响个人收入,而且在不同工作或行业互联网使用的频率本身有差异,加上区域互联网应用程度的差异会影响互联网应用,本文在不产生冲突的情况下又分别控制了区域互联网使用频率、省份类别变量(参见表2中第(2)(3)列结果)。控制区域互联网应用程度和控制省份类别变量的区别不大(系数分别为54.5%与53.1%),所以在随后的估算分析中,我们统一控制区域互联网应用程度,不再控制省份类别变量,避免一定程度的共线或过度控制问题出现。本文引入个人特征协变量(教育、性别、年龄、年龄平方项、健康等,参见表2第(4)列结果),现有研究表明,这些变量大多对于个人收入带着很强的正向影响,本文的结果也证实了这一点,可以看到互联网使用的收入影响依然非常显著,但系数大大降低(20.6%)。接着又分别引入与个人相关的社会、经济、家庭等特征协变量(参见表2第(5)(6)列结果)进行回归,互联网使用的收入影响仍然具有很高的显著性,但影响在进一步降低(系数分别为17.2%、15.2%)。也就是说,系列协变量稀释了处理变量对收入的影响。本部分通过OLS回归分析,我们得到初步结论:在尽可能控制影响收入的其他变量的情况下,工作中互联网使用会使得个人收入提高16.4% ($e^{0.152}-1$)。下面依据倾向得分匹配方法估计和分析互联网使用的收入处理效应。

(三) 匹配估计、平衡性检验与敏感性分析

考虑使用倾向得分匹配方法构造匹配样本。倾向得分值最接近的控制组个体即为实验组的配对样本,通过这种方法可以最大程度减少实验组与控制组个体存在的系统性差异,从而减少估计偏误。采用probit模型估计倾向得分指数,从一个最基本的模型出发,将主要协变量均引入模型,利用倾向得分指数的平衡指数特征,检验倾向指数模型是否充分。经过多次调整、重设模型和筛选协变量之后,本部分最终得到一个通过平衡指数特征检验的模型,协变量包括年龄(age)、母亲教育(medu)、工会成员(union)、教育(edu)、房产(house)、

户籍(hukou)、婚姻(marriage)、子女数量(nchild)、性别(male)、汉族(han)、投资(invest)、健康(health)、党员(party)^④。

表3上半部分显示了在估计倾向得分值时的probit回归结果。表3下半部分显示,处理效应估计值ATT为18%,即表明工作中互联网使用会使得个人收入平均增加19.7% ($e^{0.180}-1$),大于前述添加系列协变量的OLS回归模型得到的结果16.4%。对应的t值为3.58,远大于1.96的临界值,很显著。这说明倾向得分匹配分析处理效应是有必要的(这里仅针对匹配相对于OLS回归的必要性而言,该做法对于估计处理效应是否充分,后续还有更进一步的分析),简单的OLS回归分析低估了互联网使用对收入增加的作用,大概低估1/5 ((19.7-16.4)/16.4)。

接着检验协变量的平衡性。无论是从简单统计(见表1)还是从平衡性检验(见表4)来看,匹配之前控制组和处理组之间的差异都很显著,这也说明了直接利用线性回归得到结果不可信,不足以表达互联网使用对个人收入的处理效应。

再由表4还可以看出,经过匹配两组协变量标准化平均值的差异明显下降,平均降低了84.4%。匹配之后协变量的平衡性变好,除婚姻状态变量外,t检验显示绝大多数变量基本上没有显著差异。表4最后一列显示的是两组协变量的方差比,结果显示方差比基本接近于1,满足匹配条件,可以进行倾向得分匹配(还可参考图1(a)所示的结果)。

从干预组和控制组的倾向指数分布情况来看,满足共同区间要求(图1(b)),大多数观测值均在共同取值范围内,在进行倾向得分匹配时损失的样本数量极少,匹配效果较好。需要指出的是,以往使用倾向得分匹配方法研究互联网使用与个人收入关系的文献,大多在此止步。但如本文前面指出,倾向得分匹配方法仅仅对于可观测变量存在样本选择偏误时有效,但当存在不可观测的选择,就无法只依据平衡性检验和倾向得分共同取值范围来得知,为了评估倾向得分匹配估计的可靠性,这里借鉴Rosenbaum提出的敏感性分析,进行相应检验,考察隐藏性偏差对因果效应推断结论的影响程度^[25]。

观察表5中符号秩检验结果,当无法观测到的因素对于两种处理($treat=0$ 和 $treat=1$)发生比(Γ)为1时,也即不存在隐藏性偏差时,Wilcoxon符号秩检验的上界(sig+)和下界(sig-)都远小于0.0001(参考标准是0.05),此时前述估

表 3 倾向得分值与处理效应估计

treat	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
age	-0.01956	0.002114	-9.25	0	-0.0237	-0.01541
medu	0.012651	0.005067	2.5	0.013	0.002719	0.022583
union	0.126545	0.049187	2.57	0.01	0.03014	0.22295
edu	0.033642	0.006377	5.28	0	0.021144	0.046141
house	0.015806	0.025974	0.61	0.543	-0.0351	0.066714
hukou	0.244903	0.046671	5.25	0	0.153429	0.336377
marriage	0.093767	0.052366	1.79	0.073	-0.00887	0.196403
nchild	0.03837	0.027321	1.4	0.16	-0.01518	0.091917
male	0.033347	0.036732	0.91	0.364	-0.03865	0.10534
han	0.188088	0.074586	2.52	0.012	0.041902	0.334273
invest	0.105499	0.053103	1.99	0.047	0.001419	0.209579
health	0.001434	0.019898	0.07	0.943	-0.03757	0.040434
party	0.014098	0.058197	0.24	0.809	-0.09997	0.128162
_cons	-1.01199	0.163098	-6.2	0	-1.33166	-0.69233
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
tinc	Unmatched	10.64972	9.987975	0.661744	0.03673	18.02
	ATT	10.64972	10.4695	0.180221	0.050314	3.58

表 4 协变量平衡性检验

变量	未匹配U/匹配M	处理组mean	控制组mean	%bias	%reduct bias	t	p> t	V(T)/V(C)
age	U	40.395	45.6	-43.1		-15.21	0	0.88*
	M	40.213	39.645	4.7	89.1	1.29	0.199	0.90*
medu	U	5.2842	2.9168	53.4		21.18	0	1.25*
	M	5.7036	5.7951	-2.1	96.1	-0.53	0.598	0.89*
union	U	0.27062	0.20376	15.8		6.3	0	.
	M	0.25615	0.23839	4.2	73.4	1.11	0.266	.
edu	U	11.459	8.6937	60.2		22.87	0	0.89*
	M	11.958	11.81	3.2	94.6	0.99	0.32	1.08
house	U	1.1703	1.0932	11.1		4.51	0	1.30*
	M	1.1831	1.1749	1.2	89.4	0.3	0.767	1.03
hukou	U	0.61178	0.43801	35.3		13.62	0	.
	M	0.61202	0.59973	2.5	92.9	0.68	0.496	.
marriage	U	0.75643	0.75288	0.8		0.32	0.749	.
	M	0.77459	0.77391	0.2	80.7	0.04	0.965	.
nchild	U	1.2618	1.7293	-42.4		-15.43	0	0.69*
	M	1.138	1.1189	1.7	95.9	0.61	0.541	1.04
male	U	0.50372	0.46654	7.4		2.89	0.004	.
	M	0.51093	0.50751	0.7	90.8	0.18	0.853	.
han	U	0.94111	0.92218	7.5		2.79	0.005	.
	M	0.94536	0.95765	-4.9	35	-1.55	0.122	.
invest	U	0.20011	0.0973	29.2		12.75	0	.
	M	0.20833	0.19467	3.9	86.7	0.92	0.357	.
health	U	3.7679	3.4124	34		12.58	0	0.76*
	M	3.8443	3.8443	0	100	0	1	0.90*
party	U	0.15666	0.1044	15.6		6.45	0	.
	M	0.14754	0.1332	4.3	72.6	1.12	0.264	.

计的处理效应是有效的。但是，当发生比大于1.5时，即考虑隐藏性偏差时，Wilcoxon符号秩检验的上界已不满足0.05的参考标准，也就是说即使存在足够小的隐藏性偏差，就可能改变之前估计的因果处理效应值。换句话说，我们通过匹配得到的处理效应对于隐藏性偏差的扰动非常敏感。

再看表5中Hodges-Lehmann点估计（t-hat）和置信区间检验（CI），当Gamma=1时，工作中互联网使用相对于其他人的收入高出15.5%（ $e^{0.144}-1$ ），与我们之前估计出的结果差距不大，且系数在95%的置信区间[9.1%，21.5%]包含了我们之间的估计结果18%，且不包含0。但是，当Gamma=

1.5时，点估计的上下界分别是-6.5%和35.9%，95%的置信区间[-13.0%，43.8%]包含了0，表示这个点估计是统计不显著的。这样一来，我们按照Gamma值至少是2的判断标准，上述处理效应对于隐藏性偏差很敏感，此前估计结果的可信度不足。

综合上述检验过程，没有充分理由相信原有估计结果，那么我们为了估计出工作中互联网使用对个人收入的影响大小，就不得不思考如何处理隐藏性偏差的问题。接下来，就按照本文前面尤其是理论分析部分的思路，尝试以互联网应用能力（数字能力）为视角，对不可观测因素的偏差扰动进行探索。

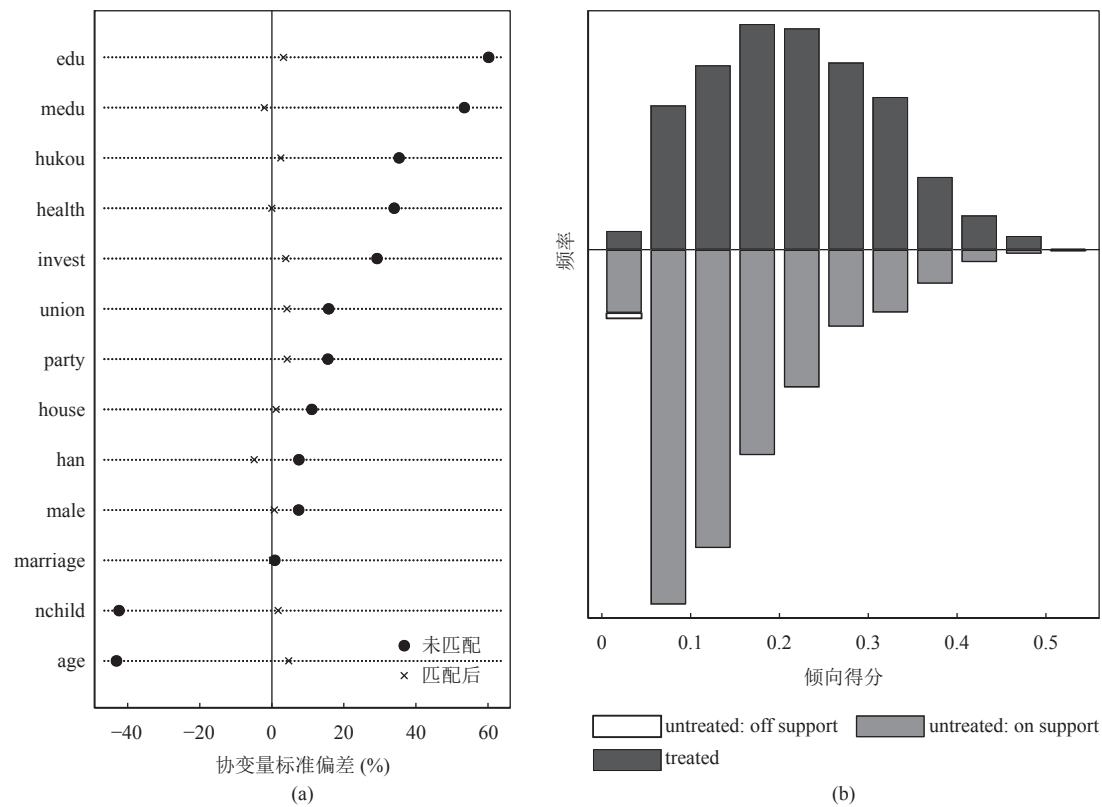


图 1 平衡性检验（左）与倾向得分共同取值范围（右）

表 5 敏感性分析

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	<0.000 1	<0.000 1	0.143 841	0.143 841	0.091 16	0.214 59
1.5	0.966 886	0	-0.065 03	0.359 232	-0.129 64	0.437 735
2	1	0	-0.202 73	0.517 037	-0.275 82	0.601 986
2.5	1	0	-0.326 16	0.640 467	-0.394 23	0.713 558
3	1	0	-0.417 32	0.733 169	-0.490 42	0.804 719
3.5	1	0	-0.490 42	0.804 719	-0.569 72	0.895 88
4	1	0	-0.549 31	0.895 88	-0.640 47	0.972 955

四、基于互联网使用能力的实证分析

（一）互联网应用能力测度

由于互联网应用能力不可以直接观测，我们使用探索性因子分析（EFA）模型来测度互联网应用能力。根据CGSS2017问卷中如“我会使用电脑打开网站”“我会使用智能手机下载安装APP”“在网上查找自己想要的信息并不难”“网上（如微信、微博）看到周围人转发的重要消息，我会先验证再相信”“当我想在网上表达自己的想法时，我

知道怎么操作”和“在网上进行支付或交易时，我会观察使用环境来确定是否使用”等6个问题的回答（“非常不符合”“不符合”“无所谓符合不符合”“符合”和“非常符合”，分别赋值1、2、3、4和5）来进行探索性因子分析。

在因子分析之前，必须对指标变量进行Bartlett检验、KMO检验、Cronbach信度检验、多元正态检验等先行检验，考察因子分析可行性，如表6所示。

Bartlett检验p值为0.000，非常显著的拒绝了互

表 6 因子分析先行检验结果

检验类型	检验结果		
Bartlett检验	$\chi^2(15)=8\,265.767$; $p=0.000$		
KMO检验	KMO=0.901		
Cronbach's α	0.8998		
多元正态检验	Mardia mSkewness = 8.226 168	chi2(56) = 3 183.333	Prob>chi2 = 0.0000
	Mardia mKurtosis = 80.655 71	chi2(1) = 6 437.250	Prob>chi2 = 0.0000
	Henze-Zirkler = 169.940 5	chi2(1) = 34 728.662	Prob>chi2 = 0.0000
	Doornik-Hansen	chi2(12) = 3 017.433	Prob>chi2 = 0.0000

联网应用能力各指标之间不相关的原假设。KMO 检验值为0.901，属于非常好的检验值，说明本文选择的6个指标之间具有较高的共同性，能够用来测量公因子。Cronbach信度检验系数为0.899 8，远超过可接受的临界值0.7。上述三项检验都说明可以使用这些指标变量来进行因子分析。这里的多

元正态检验都显著的拒绝了原假设，说明这些指标不满足联合正态分布假设。但是在大量样本条件下，按照中心极限定理，对样本均值统计量的推断并无影响。从本文使用的样本量来看，可以进行相关分析。表7是主成分因子分析结果与保留因子的载荷系数。

表 7 主成分因子分析结果与保留因子的载荷系数

Factor	Eigenvalue	Difference	Proportion	Cumulative	Variable	Factor1	Uniqueness
Factor1	4.04825	3.41923	0.6747	0.6747	able1	0.8124	0.3401
Factor2	0.62902	0.20114	0.1048	0.7795	able2	0.8611	0.2584
Factor3	0.42787	0.07847	0.0713	0.8509	able3	0.8764	0.2319
Factor4	0.34940	0.05312	0.0582	0.9091	able4	0.7237	0.4763
Factor5	0.29628	0.04711	0.0494	0.9585	able5	0.8478	0.2813
Factor6	0.24917	.	0.0415	1.0000	able6	0.7977	0.3637
LR test: independent vs. saturated: chi2(15) = 8269.34 Prob>chi2 = 0.0000							

第一个因子有一个很强的特征值4.05，能解释这些指标67.5%的方差。除此之外的因子的特征值都是小于1的，按照取舍规则可以舍去。同时，因子1各指标的因子载荷全部大于0.4（最低是0.723 7），符合要求。总体而言，各指标的因子载荷权重系数差异不大，另一个为了不改变各项指标的原始量纲，本文使用简单均值来得到互联网应用能力的测量值。

从图2（a）可以看到，互联网应用能力与个人收入之间呈现明显的正相关关系。这也可以说明，在此前将互联网应用能力作为一种不可观测因素加以忽略，确实可能对本文要识别的因果效应带来干

扰偏差。这种偏差仅仅依靠由可观测变量进行的匹配方法是不可以得到解决的，这一观点也在前面的敏感性分析得到反映。这里还需要加以交待的是，因为调查问卷中包含“互联网应用能力”指标变量的样本只有2429个，那么后续分析中所涉及的样本与前面实证所用样本已经发生改变，所以对前后参数估计值的绝对比较没有太大意义。然而，这并不影响这里的分析重点——如何消除不可观测因素的干扰。稳妥起见，本文还比较了测算前后样本收入分布的特征，图2（b）显示的是全部个体和有互联网应用能力数据的个体的收入分布情况，大致相同。此外，图2（a）显示了互联网应用能力与个人

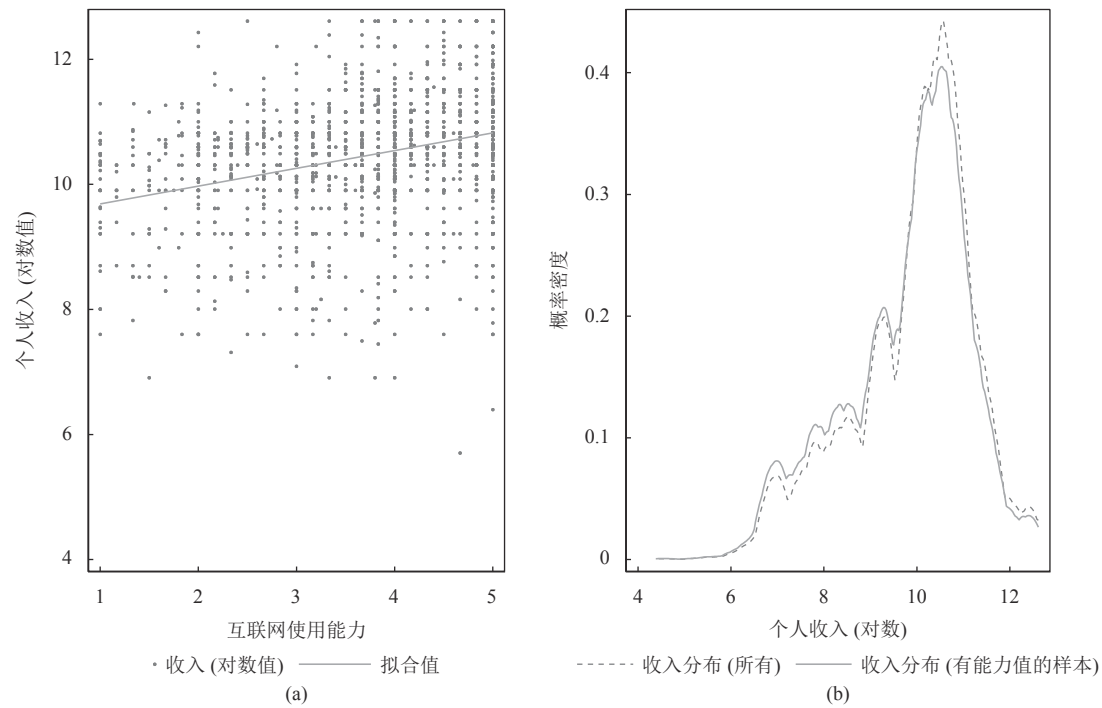


图 2 互联网应用能力、收入及其分布

收入之间的统计关系，存在很明显的正相关关系。

（二）回归分析

参照模型（2），加入个人的互联网应用能力的回归结果（详见表8）显示，不考虑其他协变量的情况下，工作中的互联网使用平均会使得个人年收入显著增加70%（ $e^{0.532}-1$ ，第（1）列）。控制区域互联网应用程度和控制省份类别变量的区别不大（系数分别为53.3%与50.1%，第（2）列和第（3）列）。引入个人特征协变量，可以看到互联网使用的收入影响依然非常显著，但系数大大降低

（39.3%，第（4）列）。接着引入社会特征变量、经济家庭等特征变量（第（5）、（6）列结果）发现，互联网使用的收入影响在进一步降低（37.0%、32.9%），但是依然是显著的。综合来看，通过OLS回归分析得到：在尽可能控制影响收入的其他变量的情况下，工作中互联网使用会使得个人收入提高39%（ $e^{0.329}-1$ ）。这个回归结果很显著，对个人收入提高效应是不考虑个人互联网使用能力时的2倍多。

表 8 OLS回归结果（考虑互联网应用能力）

	(1)	(2)	(3)	(4)	(5)	(6)
	tinc	tinc	tinc	tinc	tinc	tinc
treat	0.532*** (12.401)	0.533*** (13.474)	0.501*** (12.503)	0.393*** (9.451)	0.370*** (8.879)	0.329*** (7.808)
ablilty	0.189*** (9.153)	0.146*** (7.662)	0.157*** (8.137)	0.119*** (4.984)	0.098*** (4.089)	0.071*** (2.930)
省份	—	控制	—	—	—	—
区域互联网	—	—	控制	控制	控制	控制
个人特征变量	—	—	—	控制	控制	控制
社会特征变量	—	—	—	—	控制	控制
家庭经济变量	—	—	—	—	—	控制
常数项	9.519*** (124.324)	10.263*** (116.018)	7.932*** (67.597)	5.580*** (18.705)	5.673*** (18.330)	6.133*** (18.730)
N	1951	1951	1951	1764	1764	1672

（三）PSM估计、敏感性分析与稳健性检验

表9的上半部分显示了在考虑了个人互联网使用能力时，倾向得分值的probit回归结果。表9下半部分显示，处理效应估计值ATT为41.5%，即表明工作中互联网使用会平均使得个人收入增加51%

（ $e^{0.415}-1$ ），大于前述添加系列协变量的OLS回归模型得到的结果（39%）。对应的t值为4.85，远大于1.96的临界值，很显著。这说明简单的OLS回归分析低估了互联网使用对收入增加的作用，大概低估31%（ $(51-39)/39$ ），不容小觑。

表 9 倾向得分值与处理效应估计（考虑互联网应用能力）

treat	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
ability	0.17646	0.045924	3.84	0	0.086451	0.266469
age	0.122695	0.025422	4.83	0	0.072869	0.172522
age2	-0.00173	0.000296	-5.85	0	-0.00231	-0.00115
medu	0.03051	0.009149	3.33	0.001	0.012578	0.048442
edu	-0.0564	0.046113	-1.22	0.221	-0.14678	0.033982
edu2	0.006236	0.002134	2.92	0.003	0.002052	0.010419
union	0.008867	0.09471	0.09	0.925	-0.17676	0.194496
house	0.016152	0.04969	0.33	0.745	-0.08124	0.113542
hukou	0.292304	0.088003	3.32	0.001	0.119821	0.464787
marriage	0.114029	0.10872	1.05	0.294	-0.09906	0.327115
invest	0.255725	0.101894	2.51	0.012	0.056017	0.455434
male	-0.0044	0.069107	-0.06	0.949	-0.13985	0.13105
health	0.514442	0.231126	2.23	0.026	0.061443	0.96744
health2	-0.06271	0.031657	-1.98	0.048	-0.12476	-0.00067
Party	0.100386	0.11636	0.86	0.388	-0.12767	0.328446
nchild	-0.05006	0.056655	-0.88	0.377	-0.1611	0.060978
religion	0.186259	0.118559	1.57	0.116	-0.04611	0.41863
_cons	-4.29913	0.701014	-6.13	0	-5.67309	-2.92517
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Tinc	Unmatched	10.85218	10.14107	0.711116	0.043534	16.33
	ATT	10.85218	10.43716	0.415021	0.085496	4.85

接着检验协变量的平衡性。从平衡性检验（见图3（a））来看，匹配之前控制组和处理组之间的差异都很显著，匹配之前控制组和处理组之间的差

异变小（几乎都在10%以内），可以视为两组无差异。

从干预组和控制组的倾向指数分布情况来看，满足共同区间要求（图3（b）），大多数观测值均

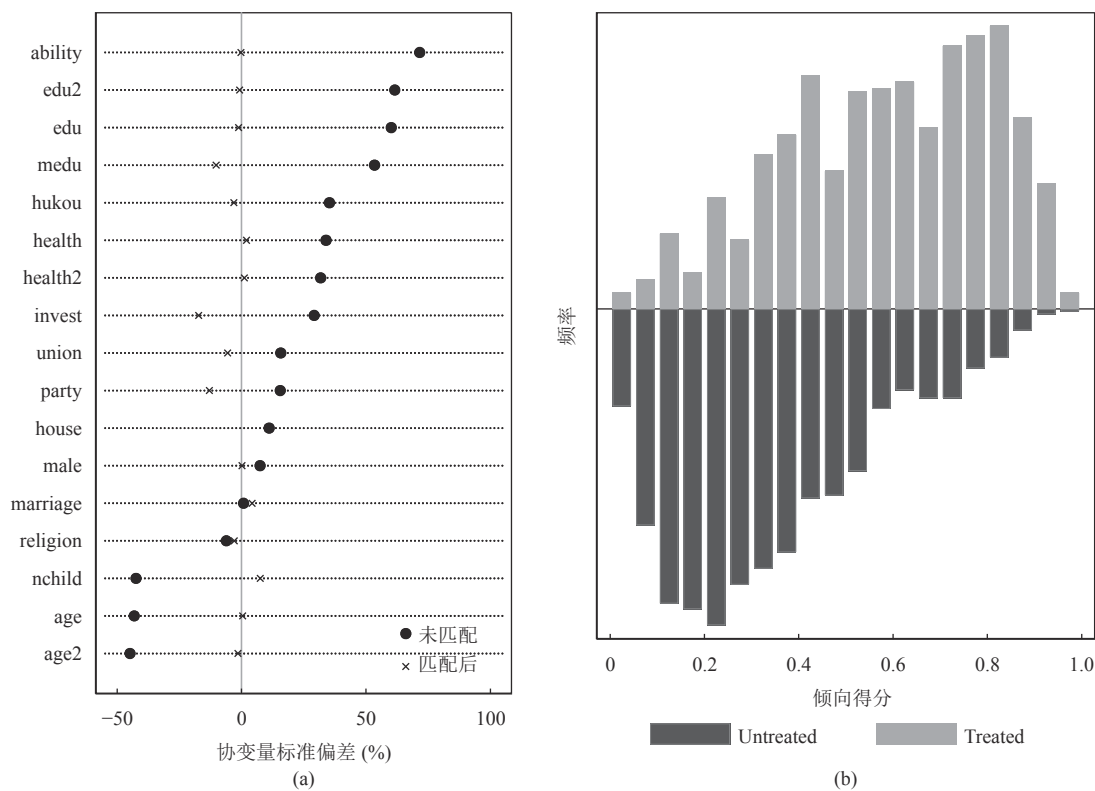


图 3 平衡性检验（左）与倾向得分共同取值范围（右）

在共同取值范围内，在进行倾向得分匹配时损失的样本数量极少，匹配效果较好。下面进行敏感性分析，考察隐藏性偏差对因果效应推断结论的影响程度。

表10中的符号秩检验结果显示，当仍可能存在的无法观测因素对于两种处理（ $treat=0$ 和 $treat=1$ ）发生比（Gamma）为2.5时，Wilcoxon符号秩检验

的上界仍然满足0.05的参考标准，同时Hodges-Lehmann 95%的置信区间[9.1%，128.7%]包含了我们之前的估计系数41.5%，且不包含0。这些结果都表明，可以认为上述处理效应对于隐藏性偏差不敏感，此前估计结果可信。也就是说，将互联网使用能力纳入考察模型后，其他不可观测因素的扰动大概率不会对结果造成颠覆性干扰。

表 10 敏感性分析

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	0.675 078	0.675 078	0.583 802	0.759 912
1.5	0	0	0.458 145	0.895 88	0.359 233	0.992 065
2	3.90E-10	0	0.293 893	1.060 13	0.202 733	1.151 29
2.5	0.000 077	0	0.182 321	1.183 56	0.091 161	1.287 26
3	0.029 534	0	0.091 161	1.295 13	-3.4e-07	1.395 72
3.5	0.382 987	0	3.40E-07	1.374 44	-0.091 16	1.497 87
4	0.859 471	0	-0.052 68	1.445 19	-0.150 05	1.565 93

表11显示的是除前述近邻匹配之外，这里采用不同的匹配方法如卡尺匹配、卡尺内近邻匹配、核匹配、局部线性回归匹配、自助法样条匹配、马氏匹配等得到的处理效应估计结果以及T值。无论是从平均处理效应的估计值还是其显著性上看，都说明了本部分估计结果的稳健性。考虑客观赋能效应

和主观赋能等多重效应，平均而言互联网使用约提高了劳动者45.7%（ $e^{0.376}-1$ ）~51%（ $e^{0.415}-1$ ）的个人收入。从已有文献的经验研究来看，本文的结论与中国社会科学院发布的《人口与劳动绿皮书：中国人口与劳动问题报告No.19》中的测算结果（46.5%）较为接近。

表 11 基于多种匹配方式的处理效应稳健性检验

	近邻匹配(n=1)	卡尺匹配	卡尺内近邻匹配	核匹配	局部线性回归匹配	样条匹配 (bs)	马氏匹配
平均处理效应ATT	0.4150	0.3687	0.398 6	0.3848	0.3761	0.3777	0.3787
标准差SE	0.085 5	0.066 0	0.080 2	0.064 9	0.085 5	0.073 2	0.052 5
T值	4.85	5.58	4.97	5.93	4.40	4.36	7.22

(四) 异质性分析

前面基于倾向值方法,在考虑互联网使用能力这个不可观测变量的基础上,较为稳健一致的估计了互联网使用对个人收入的处理效应。由于倾向值实质上是个体进入处理组的概率,进一步思考处理效应随着被处理概率不同而出现差异时,就是在分析处理效应异质性问题。

本部分借助细分-多层次法(stratification-multilevel Method)进行分析。主要考察四类异质性,高技能和低技能、城市和农村户籍、年轻和大龄以及男性和女性之间的处理效应异质性。其中,城乡与性别的数据本身就是分类数据,适用于分层。技能分层采用互联网应用能力值4为临界点,大于等于4分到高技能层、小于4分到低技能层^⑤;年龄分层采取40岁为临界点,大于等于40岁分到大龄层、小于40岁分至年轻层。

为了节省篇幅,这里省略汇报计算结果,直接以图形来直观展示互联网使用对个人收入的处理效应的异质性。图4显示的是总括性结果,在不限特定群体的情况下,随着工作中互联网使用概率的提高(即前面估计出的倾向值得分的提高),处理效应呈现出越来越强的趋势(基本上在95%的显著

水平上)。一方面,这与本文前面的估计结果相互支撑,即互联网使用对劳动者收入具有较稳定的提高效应;另一方面,也不难理解,互联网应用技术像历史上出现的新技术一样,出现之初肯定会存在“使用红利”,其应用场景才会逐步广泛、应用程度才得以深化。

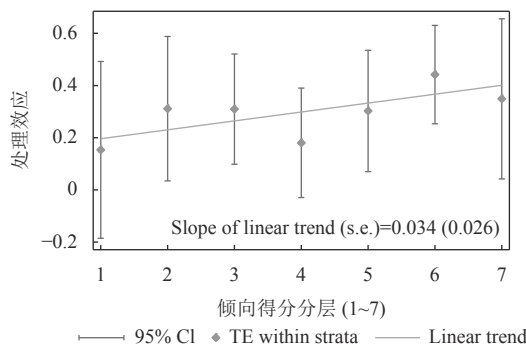


图4 互联网使用概率与处理效应异质性

图5汇报的是从户籍(城市=1;农村=0)、性别(男性=1;女性=0)、年龄层(非较年轻=1;较年轻=0)和数字能力(能力较高=1;能力较低=0)等四个维度考察处理效应在不同群体中的异质性。

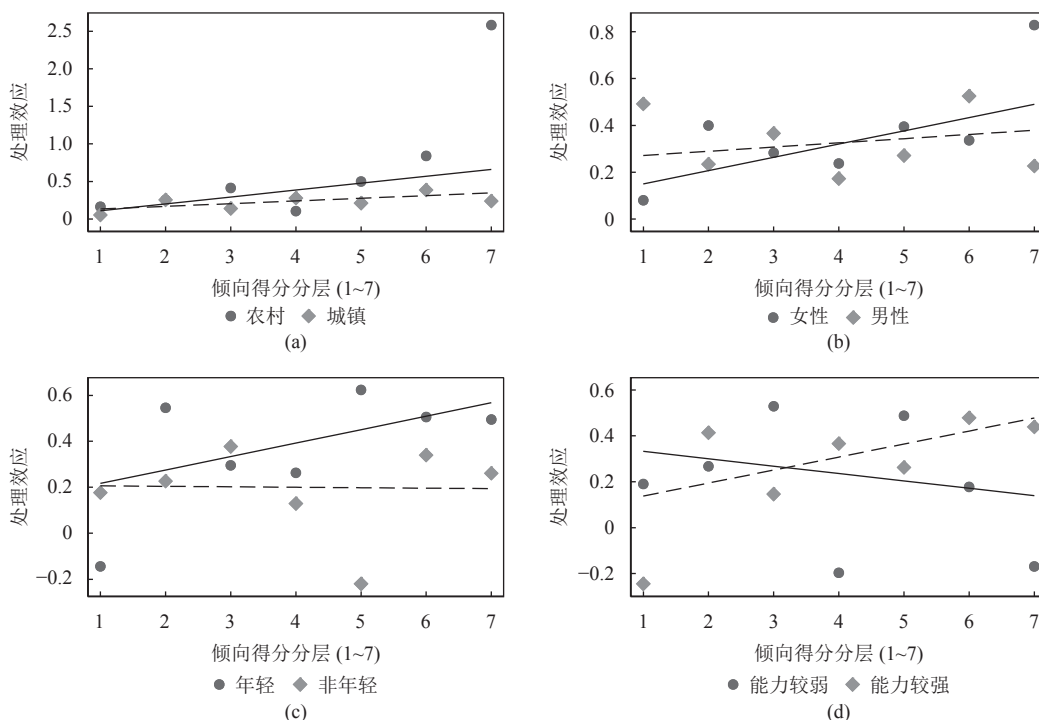


图5 户籍、性别、年龄层、能力维度的处理效应异质性

从图5(a)上看,尽管不同户籍的人都在互联网使用中提高了收入,但相较于城市户籍人口,农村户籍人群在互联网使用中的受益更多,而且随着互

联网使用概率的增加,这种优势愈发明显。这与程名望等的研究结论是一致的^[5]。但同时要注意,这很可能是因为当前阶段农村居民享受到的互联网使

用红利，主要源自“接入可及性”的改善上，随着基础设施建设普及率提高，依照本文前面的理论分析，这种效应可能消失甚至为负。图5（b）上显示，男性和女性在互联网使用获益上的差异不大，有一个趋势性的差异是，随着互联网使用概率的提高，女性收入提高效应更明显。图5（c）对应的是年龄层处理效应异质性。两个年龄层的人群都在互联网使用中提高了收入，但相较于年轻群体，非较年轻人群在互联网使用中的收入提高效应较低，而且变化不明显，随着互联网使用概率的增加，年轻群体表现出的优势特别明显。图5（d）是针对技能分层的结果，关于技能分层处理效应异质性，有两点值得关注：一是较低技能劳动者互联网使用的收入提高效应，随着互联网使用概率的增加而逐步减弱（尽管数据显示一直为正）；二是在使用概率较大的一侧，高低技能分层群体的互联网使用红利也表现出了明显的分化。也就是说，从技能分层的视角来看，较低技能的互联网使用者很可能随着互联网发展而表现出数字相对贫困。

五、主要结论与相关讨论

互联网使用对个人收入存在“多重赋能”如图6所示：一是接入效应或技术效应；二是接通效应或资本效应。不可观测因素如数字能力，一方面影响互联网使用，另一方面也与个人收入相关，给经验研究带来内生性难题。相比于已有研究，对不可观测因素的处理，是本文想突出表达的重点。

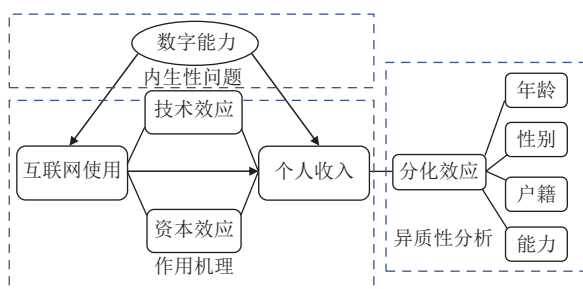


图6 分析框架图

实证研究表明：忽略不可观测的数字能力，互联网使用的个人收入提升效果显著（OLS方法测算结果为16.4%；PSM方法测算结果为19.7%），但敏感性检验发现存在不可观测干扰。通过探索性因子分析（EFA）控制数字能力后，处理效应依然显著（OLS测算结果是39%；多种匹配方法测算结果为45.7%~51%），并通过敏感性检验。处理效应异质性分析发现，依户籍分农村人群在互联网使用中

的受益更多，从性别看女性收入提高效应更显著，从年龄看年轻群体随互联网使用频率增多而表现出的优势明显，按照技能分层考察低技能劳动者互联网使用的收入提高效应有弱化趋势，互联网使用红利分化特征显著。

当前，互联网日渐融入人们的社会生产生活之中。在互联网对各类型、各层次市场参与主体赋能的同时，对个人及其所掌握的各类资本也在赋能，互联网接入可及性红利带着显著的普惠性质。然而，当人人可及互联网的时候，这类红利不仅会逐渐弱化，而且存在以“数字能力”差异为特征的互联网使用红利的分化。而弱化效应与分化效应，必然会给优化收入分配格局带来不利扰动，如数字相对贫困或成为实施乡村振兴战略、推进共同富裕改革过程中的新题和难题。本文认为，从目前的科技水平看，互联网本身尚不具备主观能动性，它的发展与社会经济发展之间不存在固定的线性关系，甚至长期看也不应存在既定的非线性关系。其发展过程中所体现出的动态异质性更需引起关注。所以，并不能仅仅依靠单方面科技建设来实现收入分配格局、城乡发展差距、性别收入差距等社会发展指标的系统性弥合。这还取决于人们对互联网的总体认知和应用水平、个人认识和应用能力的差异以及领域内关键资源在不同微观主体之间的分布情况等。加大对互联网基础设施的投资建设是一种必要手段，但是还必须更加注重对“传统弱势群体”和互联网时代的“新型弱势群体”的互联网使用能力的培训、分享和帮助。同样重要的是，防止互联网平台、关键资源垄断等造成的实际接入可及和有效接通上的非均衡、非充分发展局面，切实将共享作为互联网发展的主题之一。

注释

① 第47次《中国互联网络发展状况统计报告》，<https://zndsssp.dangbei.net/2021/20210203.pdf>。

② https://www.sohu.com/a/287155035_186085。

③ 对应问卷问题c56，“在您一周的工作中，需要使用互联网的工作大概占多大比重”。

④ 分别对应问卷问题a8a，“您个人去年全年的总收入是多少”。

⑤ 对应问卷问题a2，男性为1，女性为0。

⑥ 对应问卷问题a7a，参照李涛等（2021）处理方式：“没有受过任何教育”赋值为0，“私塾、扫盲班”赋值为3，“小学”赋值为6，“初中”赋值为9，“职业高中、普通高中、中专、技校”赋值为12，“大学专科（含成人）”赋值为15，“大学本科（含成人）”赋值为16，研

究生及以上赋值为19。

⑦ 对应问卷问题a15, “您觉得您目前的身体健康状况是”, 相应回答为: 1.很不健康, 2.比较不健康, 3.一般, 4.比较健康, 5.很健康。

⑧ 对应问卷问题a12b, “目前您总共拥有几处房产(包括与他人共同拥有)”。

⑨ 对应问卷问题a45, “请问您是不是工会会员”, 从来都不是赋值为0, 否则为1。

⑩ 对应问卷问题a59a, “以下各种情形, 哪一种更符合你目前的工作的状况?”, 对于每一类型的工作, 符合情况时赋值为1, 否则为0。

⑪ 对应问卷问题a4, 汉族为1, 其他民族为0。

⑫ 对应问卷问题a51, 有宗教信仰为1, 否则赋值为0。

⑬ 对应问卷问题a10, 中共党员赋值为1, 其他为0。

⑭ 本文在这里参考了stata软件中的psestimate命令给出的模型识别策略。

⑮ 这里参考了互联网使用能力在劳动者群体中的分布情况, 取值为4的地方与之前相比, 出现了比较明显的“跳跃”。

参考文献

- [1] 邱泽奇, 张树沁, 刘世定, 等. 从数字鸿沟到红利差异——互联网资本的视角[J]. 中国社会科学, 2016(10): 93-115+203-204.
- [2] 陈玉宇, 吴玉立. 信息化对劳动力市场的影响: 个人电脑使用回报率的估计[J]. 经济学, 2008(4): 1149-1166.
- [3] 许竹青, 郑风田, 陈洁. “数字鸿沟”还是“信息红利”? 信息的有效供给与农民的销售价格——一个微观角度的实证研究[J]. 经济学, 2013, 12(4): 1513-1536.
- [4] 曾亿武, 郭红东, 金松青. 电子商务有益于农民增收吗?——来自江苏沐阳的证据[J]. 中国农村经济, 2018(2): 49-64.
- [5] 程名望, 张家平. 互联网普及与城乡收入差距: 理论与实证[J]. 中国农村经济, 2019(2): 19-41.
- [6] 赵羚雅, 向运华. 互联网使用、社会资本与非农就业[J]. 软科学, 2019, 33(6): 49-53.
- [7] 邱子迅, 周亚虹. 电子商务对农村家庭增收作用的机制分析——基于需求与供给有效对接的微观检验[J]. 中国农村经济, 2021(4): 36-52.
- [8] 刘生龙, 张晓明, 杨竺松. 互联网使用对农村居民收入的影响[J]. 数量经济技术经济研究, 2021, 38(4): 103-119.
- [9] 罗楚亮, 梁晓慧. 互联网就业搜寻对流动人口就业与工资的影响[J]. 学术研究, 2021(3): 72-79.
- [10] KRUEGER A B. How computers have changed the wage structure: evidence from microdata, 1984-1989[J]. Quarterly Journal of Economics, 1993, 108(1): 33-60.
- [11] BELL B D. Skill-biased technical change and wages: evidence from a longitudinal data set[D]. Oxford: Oxford University, IPP6.
- [12] PISCHKE D. The returns to computer use revisited: have pencils changed the wage structure too?[J]. Quarterly Journal of Economics, 1997, 112(1): 291-303.
- [13] 赵建国, 周德水. 互联网使用对大学毕业生就业工资的影响[J]. 中国人口科学, 2019(1): 47-60+127.
- [14] 谭燕芝, 李云仲, 胡万俊. 数字鸿沟还是信息红利: 信息化对城乡收入回报率的差异研究[J]. 现代经济探讨, 2017(10): 88-95.
- [15] 蒋琪, 王标悦, 张辉, 等. 互联网使用对中国居民个人收入的影响——基于CFPS面板数据的经验研究[J]. 劳动经济研究, 2018, 6(5): 121-143.
- [16] DIMAGGIO P, HARGITTAI E. From the “digital divide” to “digital inequality”: studying internet use as penetration increases[EB/OL]. [2022-02-10]. <https://ideas.repec.org/p/pri/cpanda/15.html>.
- [17] BONFADELLI H. The Internet and knowledge gaps—A theoretical and empirical investigation[J]. European Journal of Communication, 2002, 17(1): 65-84.
- [18] 郝大海, 王磊. 地区差异还是社会结构性差异?——我国居民数字鸿沟现象的多层次模型分析[J]. 学术论坛, 2014, 37(12): 88-95.
- [19] 申广军, 刘超. 信息技术的分配效应——论“互联网+”对劳动收入份额的影响[J]. 经济理论与经济管理, 2018(1): 33-45.
- [20] 王元超. 互联网工资溢价效应的阶层差异[J]. 社会学评论, 2019, 7(2): 27-41.
- [21] 叶明睿, 蒋文茜. 信息增收的行为阶梯——一项基于城市“相对贫困”人群信息使用的研究[J]. 南京社会科学, 2021(5): 53-62.
- [22] 冯喜良, 高盼盼, 罗荣波. 互联网使用对农民工性别工资收入差距的影响[J]. 人口与经济, 2021(5): 111-124.
- [23] BACH A, SHAFFER G, WOLFSON T. Digital human capital: developing a framework for understanding the economic impact of digital exclusion in low-income communities[J]. Journal of Information Policy, 2013, 3(1): 247-266.
- [24] 索托. 资本的秘密[M]. 于海生, 译. 北京: 华夏出版社, 2017.
- [25] ROSENBAUM P R. Heterogeneity and causality: unit heterogeneity and design sensitivity in observational studies[J]. The American Statistician, 2005: 59.

编辑 邓婧