

数字赋能下高学历为什么带来了更高的工资溢价

——基于 CFPS 数据的实证研究

张国胜 吴 晶*

内容提要 本文使用 2018 年中国家庭追踪调查 (CFPS) 数据, 分析数字赋能是否能够促使高学历劳动者获得更高的工资溢价, 并尝试揭示其作用机制。研究发现, 数字赋能与劳动者的工资溢价具有显著的正向关系, 且受教育程度越高, 数字赋能的工资溢价效应越明显, 劳动者的受教育程度每增加 1 年, 数字赋能带来的工资回报随之提高 2.62%。互联网使用行为上的结构差异, 是致使不同受教育程度劳动者工资溢价幅度产生差异的主要原因。一方面, 劳动者的受教育程度越高, 在工作中使用互联网的频率越高, 数字赋能所产生的工资溢价效应越明显; 另一方面, 劳动者的受教育程度越高, 在学习与社交中使用互联网的频率越高, 通过增强人力资本积累与社会资本积累, 也能够产生更高的工资溢价效应。

关键词 数字赋能 工资溢价 受教育程度

一 引言

伴随互联网的全方位商业化, 数字资源开始成为一种重要的资源, 赋能经济社会发展日趋普遍。根据中国互联网络信息中心发布的第 47 次《中国互联网络发展状况统计报告》, 中国网民规模达 9.89 亿, 互联网普及率达 70.4%。在上述现实背景下, 数

* 张国胜, 云南大学经济学院, 电子邮箱: zhangguosheng@ynu.edu.cn; 吴晶, 云南大学经济学院, 电子邮箱: 503898233@qq.com。本研究得到研究阐释党的十九届五中全会精神国家社会科学基金重大项目“推进以农业转移人口市民化为首要任务的新型城镇化研究”(项目编号: 21ZDA068)、国家社会科学基金重点项目“户籍制度城乡双向改革与新型城镇化研究”(项目编号: 20AJL012)的资助。

字赋能是指互联网等数字技术的广泛使用而产生的对经济社会发展的驱动效应^①。从劳动力市场来看，数字赋能不但创造了新的就业机会，而且能够提升劳动者的生产效率与工资报酬（世界银行，2017）。通过控制个体异质性、样本选择偏误等问题，国内外众多学者均认为劳动者使用互联网等数字技术能够带来明显的工资溢价（Krueger, 1993; DiMaggio & Bonikowski, 2008; 陈玉宇、吴玉立，2008），其原因是数字赋能可以降低信息获取成本、增强人力资本（Anderson et al., 1997）、促进工作与市场的更好衔接等（世界银行，2017）。

现有研究也注意到了数字赋能的工资溢价效应在城乡（程名望、张家平，2019）、性别（庄家焯等，2016）、阶层（王元超，2019）、受教育水平（李雅楠、谢倩芸，2017）等群体之间存在明显差异。就数字赋能在群体之间的工资溢价差异而言，现有文献从数字设备接入、数字技能（digital skills）水平等方面进行了探索性研究，并认为数字设备接入更便捷、数字技能水平更高能够导致更高的工资溢价（Spitz - Oener, 2008; 陈玉宇、吴玉立，2008）。然而，随着智能手机、笔记本电脑等数字设备的广泛应用，数字赋能对互联网接入的物质要求和互联网使用的技能水平的门槛正在逐步下降（李雅楠、谢倩芸，2017）。也就是说，数字设备接入与数字技能水平固然重要，但已经不是数字赋能在群体之间产生异质性工资溢价的主要原因。那么是什么因素导致了数字赋能在群体之间的工资溢价效应差异？这显然是一个值得深入研究的问题。据此，本文立足于中国互联网的广泛使用，围绕不同受教育程度劳动者在工作、学习、社交等方面的互联网使用行为差异，深入研究数字赋能是否会对不同受教育程度劳动者产生差异化的工资溢价效应，并揭示其作用机制。

本文研究发现，数字赋能与劳动者的工资溢价具有明显的正向关系，受教育程度越高，数字赋能的工资溢价效应越明显；定量分析表明，劳动者的受教育程度每增加 1 年，数字赋能带来的工资回报随之提高 2.62%，高学历的确带来了更高的工资溢价效应。就其作用机理而言，互联网使用行为上的结构差异，是致使不同受教育程度劳动者工资溢价幅度产生差异的主要原因。一方面，劳动者的受教育程度越高，在工作中使用互联网的频率越高，数字赋能所产生的工资溢价效应越明显；另一方面，劳动者

^① 现有文献大多将数字赋能定义为通过大数据、移动互联和人工智能等数字技术对特定的人群进行赋能，使他们获得特定的能力。但现阶段的数字赋能更多是利用数字技术赋能农业、商业、金融、医疗等各行各业（夏显力等，2019），不仅局限于对微观个体的赋能，因此本文参照世界银行对数字红利的定义，认为数字赋能是指互联网等数字技术的广泛使用而产生的对经济社会发展的驱动效应。

的受教育程度越高，在学习与社交中使用互联网的频率越高，通过增强人力资本积累与社会资本积累，也能够产生更高的工资溢价效应。

与现有文献相比，本文可能的创新之处有：第一，在研究方法上，本文综合了工具变量法（Instrumental Variable，简称IV）、处理效应模型（Treatment Effect Model，简称TEM）以及倾向值匹配（Propensity Score Matching，简称PSM）等多种方法，控制了计量模型可能存在的多种内生性问题，获得了相对稳健的经验结论，得到了数字赋能的工资溢价效应在不同受教育程度劳动者中间存在显著差异的稳健结论。第二，在研究内容上，本文立足于个体互联网行为，从工作中使用互联网、学习中使用互联网和社交中使用互联网三个角度，分析了为何数字赋能对不同受教育程度劳动者产生差异化的工资溢价效应。

二 文献综述与研究假设

（一）数字赋能与工资溢价

数字赋能对劳动者的工资溢价效应体现在直接和间接两个方面。首先，数字技术不仅能够替代劳动者完成常规性、重复性的工作，把劳动者从传统劳动中解放出来，使其关注于更具创造性、价值更高的生产活动（Bresnahan, 1999），而且可以增强个人完成大量工作任务的能力，因此在工作中数字技术的使用可大幅提高劳动者的生产率（Krueger, 1993），直接提升劳动者工资收益。其次，互联网等数字技术的广泛应用减少了信息获取成本，使个体更便捷地获取海量资源，通过在线学习、求职、医疗、交友等方式提升个人能力、积累人力资本、扩充社会资本，从而增强自身在劳动力市场中的竞争力（庄家焯等，2016），进而获得更多工资回报，间接实现工资溢价。

受社会结构因素影响，劳动力市场中的数字赋能并不是一个同质化、均等化的过程，数字赋能的工资溢价在不同劳动者群体之间存在明显差异。技术偏向理论认为，互联网等数字技术的使用对于受教育程度高的个体来说是互补的，而对于受教育程度低的个体是替代的（Autor et al., 2002），受教育程度高的个体更能利用数字技术提升生产率，从而提升个人工资收入。Spitz - Oener（2008）基于德国的劳动力调查数据发现，受过高等教育的个体使用数字技术时能获得更高的工资收益。Liu & Mithas（2016）的研究进一步指出，劳动者教育程度每增加1年，数字技术使用的工资溢价将提升1.90%。陈玉宇和吴玉立（2008）、蒋琪等（2018）也得出高学历群体使用数字技术可以带来额外的工资收入的一致结论。

本文据此提出假说 1：数字赋能对工资溢价呈正向影响，且受教育程度越高，数字赋能的工资溢价效应越明显。

（二）受教育程度差异与数字赋能的差异化工资溢价效应

首先，劳动者的受教育程度越高，数字赋能工作的效应越明显，产生的工资溢价效应越高。受教育程度作为劳动力市场中的一个重要信号（杨铭、王任远，2019），反映了劳动者的能力差异，学历越高的劳动者越有可能从事技术含量较高的工作。在数字赋能时代，这些工作岗位可能是数字技术催生下的新兴岗位，也有可能是数字技术与传统产业融合发展之后的岗位，但都需要密集使用互联网等数字技术。这就意味着高学历劳动者更需要在工作中使用互联网。同时，在工作中使用互联网具有技能或知识门槛，教育程度较高的劳动者通常具备一定的数字素养（经济合作与发展组织，2013），已经掌握或更容易通过培训或自学掌握高级的数字技能（如创建新的知识对象、信息单元、媒体产品或其他数字输出），而不仅仅只是停留在简单的网络接入和文字处理的初级阶段（蒋琪等，2018），因此高学历人群更有能力在工作中使用互联网。从这些逻辑出发，劳动者受教育程度越高，在工作中使用互联网的频率就越高，数字赋能工作的效应越明显，即工资溢价效应就越明显。

其次，劳动者的受教育程度越高，数字赋能增强人力资本积累的效应越明显，进而带来的工资溢价效应越高。数字技术既是一个学习对象，更是一个辅助学习的工具，它可以转变学习的方式、地点和时间（经济合作与发展组织，2009）。基于个体职业发展目的和“终生学习”理念，受教育程度更高的劳动者往往在使用数字技术时更具主动性，其使用也偏向信息导向型（Bonfadelli，2002）。同时，受教育程度更高的劳动者更偏向于互联网中的学习而非休闲娱乐（经济合作与发展组织，2013），借助互联网这一种便捷可及的社会资源，通过在线培训等可将之转换为自身发展所需要的人力资本。同时，在数字赋能的条件下，数字技术不仅可以帮助受教育程度更高的劳动者重新整合碎片化的时间并加以利用（庄家焱等，2016），而且通过互联网提供的海量多样资源，受教育程度更高的劳动者可以有选择、有针对性地提高个人素质和提升在劳动力市场中的竞争力（王元超，2019），从而强化高学历劳动者的人力资本积累。从这些逻辑出发，劳动者的受教育程度越高，数字赋能增强人力资本积累的效应越明显，这会进一步带来生产率的提高（DiMaggio & Bonikowski，2008），最终使得劳动者获得更高的工资收入或更好的雇佣机会。

最后，劳动者的受教育程度越高，数字赋能强化社会资本积累的效应越明显，进而带来的工资溢价效应越高。劳动者的受教育水平可以看作是社会经济地位的替代指标（Gaziano，1997）。一般来说，个人社会地位越高，因工作关系和职业关系等所能接触到

的社会关系网络越多（边燕杰，2004），为了拓展和维持自己的社会关系网络以调动社会资源，其更倾向于在社交中使用互联网。与传统社交媒体相比，以QQ和微信为代表的数字社交媒体具有移动便捷性、传递即时性、无时空边界线性、交流多向互动性、图文音频多样性和低成本廉价性等特点，使用数字社交媒体可以打破以血缘、地缘为纽带的社会网络的“内卷化”趋势（刘传江等，2018），使人与人之间的沟通更为便捷，信息传递更为迅速。这不仅可以整合和延展原有社会网络，而且能够通过“跨越”不同社交网络发展新型社会资本并扩大社交网络规模。从这些逻辑出发，劳动者的受教育程度越高，数字赋能强化社会资本积累的效应越明显，而社会资本的完善和加强又可以向劳动者提供更多的机会（王春超、张承莎，2019），进而获得更高的工资收入。

据此，本文提出以下三个假说：假说2，受教育程度越高，劳动者在工作中使用互联网的频率越高，数字赋能带来的工资溢价更高；假说3，受教育程度越高，劳动者在学习中使用互联网的频率越高，数字赋能带来的工资溢价更高；假说4，受教育程度越高，劳动者在社交中使用互联网的频率越高，数字赋能带来的工资溢价更高。

三 研究设计

（一）数据来源和变量说明

本文数据来源于北京大学中国社会科学调查中心设计和实施的2018年中国家庭追踪调查（Chinese Family Panel Studies，简称CFPS）。该调查采用多阶段、内隐分层和与人口规模成比例的系统概率抽样设计，调查对象涵盖31个省（市、自治区）的家庭户及家庭成员，总样本规模为14241户家庭和32669个成员，具有良好的代表性。根据研究需要，本文将所选样本限制在调查时有工作且获得工资性收入的劳动力，在剔除了离退休人员、失业人员、在校学生、学龄前儿童以及样本缺失值后，最终得到7040个劳动力样本。此外为确保基本结论的可靠性，本文也引入2016年CFPS数据进行分析。结合研究主题，本文的变量选择如下：

首先看被解释变量。本文使用小时工资的自然对数来衡量工资溢价。为排除劳动者因工作时长不同而造成的偏误，借鉴程虹和李唐（2017）的研究方法，本文采用个体过去一年内每小时平均工资的自然对数作为工资水平的测度指标。

其次看核心解释变量。本文使用是否采用互联网来衡量数字赋能。借鉴张景娜和张雪凯（2020）的做法，根据2018年CFPS问卷中“是否电脑上网”和“是否移动上网”来定义劳动者互联网使用状态，若劳动者未使用电脑或移动设备上网则取值为0，否则

为 1。

最后看控制变量。本文的控制变量包括年龄、受教育程度、性别等经典工资式中的控制变量 (Mincer, 1974), 以及户籍、婚姻、健康、政治面貌和是否有行政管理职务等个体特征变量。考虑到企业规模越大, 其员工的工资越高, 以及企业的所有制特征也会显著影响劳动者的工资收入 (程虹、李唐, 2017), 本文的控制变量还包含企业规模、企业所有制类型等企业特征变量。本文也将控制劳动者权益保护方面的一系列指标, 涵盖劳动者是否参加工会、是否签订合同以及雇主是否提供养老、医疗保险等相关指标。

表 1 主要变量名称、定义及基本统计特征

变量名称	定义	观测值个数	均值	最小值	最大值
小时工资对数	过去一年每小时平均工资的自然对数	7040	2.557	0	9.028
数字赋能	是否使用互联网 (是=1, 否=0)	7040	0.762	0	1
年龄	实际年龄	7040	38.463	16	83
受教育程度	实际受教育程度 (年)	7040	10.272	0	22
性别	是否女性 (是=1, 否=0)	7040	0.427	0	1
户籍	户籍情况 (城市=1, 农村=0)	7040	0.334	0	1
婚姻状况	实际婚姻状况 (已婚=1, 未婚=0)	7040	0.747	0	1
健康状况	实际健康状况 (不健康=1, 一般=2, 比较健康=3, 很健康=4, 非常健康=5)	7040	3.244	1	5
政治面貌	是否党员 (是=1, 否=0)	7040	0.016	0	1
行政管理职务	是否有行政管理职务 (是=1, 否=0)	7040	0.104	0	1
企业规模	企业总人数的自然对数	7040	3.919	0	12.429
国有企业	是否国有企业 (是=1, 否=0)	7040	0.124	0	1
外资企业	是否外资企业 (是=1, 否=0)	7040	0.031	0	1
工会	是否参加工会 (是=1, 否=0)	7040	0.135	0	1
签订合同	是否签订合同 (是=1, 否=0)	7040	0.507	0	1
养老保险	雇主是否提供养老保险 (是=1, 否=0)	7040	0.367	0	1
医疗保险	雇主是否提供医疗保险 (是=1, 否=0)	7040	0.366	0	1

资料来源: 根据 2018 年中国家庭追踪调查 (CFPS) 数据计算得到。

(二) 模型构建

本文在扩展的工资决定式的基础上 (Mincer, 1974), 以数字赋能、劳动者受教育程度及其交互项作为解释变量, 构建半对数模型:

$$\ln wage_i = \beta_0 + \beta_1 (Internet_i \times edu_i) + \beta_2 Internet_i + \beta_3 edu_i + \beta_4 CV_i + \varepsilon_i \quad (1)$$

其中 i 表示劳动者个体， $\ln wage_i$ 为劳动者小时工资的自然对数， $Internet_i \times edu_i$ 为互联网使用与劳动者受教育水平的交互项， $Internet_i$ 表示劳动者互联网使用情况， edu_i 为劳动者受教育水平， CV_i 为控制变量， ε_i 表示随机扰动项。此外，由于在计量模型中引入了交互项 $Internet_i \times edu_i$ ，为了避免单变量 edu_i 和 $Internet_i$ 与交互项 $Internet_i \times edu_i$ 之间产生比较严重的共线性问题，因此参照 Algina & Moulder (2001) 的研究方法，将单变量 edu_i 和 $Internet_i$ 做中心化处理以减少两个变量乘积项和单变量之间的共线性^①。中心化实际上是一种线性变换，不会改变各变量与被解释变量之间的相关性（陶锋，2011）。

四 数字赋能的工资溢价效应

（一）基准回归结果

本文使用最小二乘法对式（1）进行基准回归，以分析数字赋能对不同受教育程度劳动者能否产生差异化的工资溢价效应，回归结果如表 2 所示。表 2 中的第（1）列未加入数字赋能和劳动者受教育程度的交互项，回归结果显示数字赋能的系数显著为正，表明数字赋能提高劳动者个体的工资回报。表 2 的第（2）列加入数字赋能和劳动者受教育程度的交互项，而第（3）列和第（4）列在第（2）列的基础上继续加入省份和行业的虚拟变量。表 2 的第（2）列至第（4）列回归结果均显示数字赋能和受教育程度的交互项与劳动者工资水平显著正相关，这表明数字赋能对不同受教育程度劳动者的工资溢价效应存在显著差异，劳动者的受教育程度越高，数字赋能的工资溢价效应越明显，这验证了假说 1。由表 2 的第（4）列回归结果可知，在控制了行业和省份固定效应后，劳动者受教育程度每增加 1 年，数字赋能所带来的工资回报增加 2.62%。这一影响高于 Liu & Mithas (2016) 得出的结果，原因可能在于 Liu & Mithas (2016) 估计的是不同受教育程度劳动者使用计算机这一数字设备带来的工资溢价效应差异，而相较于传统的计算机，互联网的普及更广泛、赋能更明显，因此数字赋能的工资溢价效应在不同受教育程度劳动者间的差异会更显著。

^① 中心化处理过程具体步骤是：首先将 $Internet$ 和 edu 分别取其均值 $\overline{Internet}$ 和 \overline{edu} ，即“取中”（centering），然后用 $Internet$ 和 edu 的原始值减去各自的均值，其结果就是原始值的离差，即 $Internet_m$ 、 edu_m ，最后，将取中后的 $Internet_m$ 、 edu_m 相乘即可。

表 2 数字赋能的工资溢价效应 (OLS 基准回归)

变量	(1)	(2)	(3)	(4)
数字赋能	0.099*** (0.026)	0.173*** (0.029)	0.166*** (0.028)	0.155*** (0.028)
数字赋能 × 受教育程度		0.034*** (0.006)	0.031*** (0.006)	0.026*** (0.006)
受教育程度	0.031*** (0.003)	0.034*** (0.003)	0.037*** (0.003)	0.033*** (0.003)
年龄	0.057*** (0.006)	0.061*** (0.006)	0.065*** (0.006)	0.063*** (0.006)
年龄的平方	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
性别	-0.254*** (0.019)	-0.264*** (0.019)	-0.267*** (0.019)	-0.254*** (0.020)
户籍	0.089*** (0.022)	0.082*** (0.022)	0.055** (0.023)	0.051** (0.023)
婚姻状态	0.125*** (0.027)	0.132*** (0.027)	0.138*** (0.027)	0.140*** (0.027)
健康状态	0.028*** (0.009)	0.028*** (0.009)	0.035*** (0.009)	0.033*** (0.009)
党员	-0.061 (0.082)	-0.073 (0.082)	-0.052 (0.078)	-0.055 (0.079)
行政管理职务	0.317*** (0.029)	0.309*** (0.029)	0.287*** (0.028)	0.297*** (0.029)
企业规模	0.016*** (0.006)	0.015*** (0.006)	0.020*** (0.006)	0.022*** (0.006)
国有企业	-0.115*** (0.030)	-0.111*** (0.029)	-0.078*** (0.029)	-0.103*** (0.030)
外资企业	0.003 (0.051)	0.012 (0.051)	-0.091* (0.049)	-0.069 (0.049)
工会	0.145*** (0.027)	0.142*** (0.027)	0.117*** (0.026)	0.128*** (0.027)
签订合同	0.129*** (0.025)	0.128*** (0.025)	0.090*** (0.025)	0.086*** (0.025)
养老保险	0.054 (0.054)	0.048 (0.054)	0.023 (0.053)	0.039 (0.052)
医疗保险	0.331*** (0.054)	0.318*** (0.054)	0.284*** (0.055)	0.268*** (0.054)
常数项	0.636*** (0.129)	0.463*** (0.133)	0.613*** (0.142)	0.581*** (0.156)
地区虚拟变量	否	否	是	是
行业虚拟变量	否	否	否	是
R 平方	0.251	0.255	0.290	0.296
观测数	7040	7040	7040	7040

注：括号内为稳健标准误；*、**、*** 分别表示在 10%、5%、1% 的水平上显著。
资料来源：根据 2018 年中国家庭追踪调查 (CFPS) 数据计算得到。

(二) 控制模型内生性的回归结果

1. 工具变量回归结果

由于收入高的人群更可能购买数字设备、更容易接入互联网 (经济合作与发展组

织, 2013), 工资水平高的人群更倾向于使用互联网等数字技术, 因此工资溢价也会反过来影响数字赋能, 这会导致式 (1) 存在内生性问题。为克服由于核心变量间的逆向因果关系导致的估计偏误, 本文参考已有文献, 选取“个体对互联网作为信息获取渠道的重要程度”(互联网偏好) 作为互联网使用的工具变量(杨碧云等, 2019), 这一方面是因为个体对于互联网作为信息获取渠道的重要程度越高, 则很有可能使用互联网, 满足工具变量所需的相关性要求, 另一方面是因为个体对于互联网的重要程度属于个人层次上的一种偏好, 与劳动者的工资回报并没有直接关系, 满足工具变量所需的外生性要求。工具变量的回归结果如表 3 所示。在第一阶段回归中, F 统计量均超过了经验标准值 10, 说明工具变量和内生变量具有较强的相关性; 在第二阶段回归中, Durbin-Wu-Hausman 内生性检验在 1% 的显著性水平上拒绝了解释变量没有内生性的零假设, 表明式 (1) 确实存在内生性问题, 采用工具变量处理内生性问题的做法是合理的。

表 3 数字赋能的工资溢价效应 (IV 回归)

变量	第一阶段		第二阶段	第一阶段		第二阶段
	数字赋能	数字赋能 × 受教育程度	小时工资对数	数字赋能	数字赋能 × 受教育程度	小时工资对数
数字赋能 × 受教育程度			0.075 *** (0.011)			0.060 *** (0.010)
数字赋能			0.601 *** (0.088)			0.566 *** (0.086)
受教育程度	0.034 *** (0.003)	-0.693 *** (0.021)	0.031 *** (0.004)	0.032 *** (0.003)	-0.688 *** (0.021)	0.030 *** (0.004)
互联网偏好	0.093 *** (0.004)	0.018 (0.017)		0.092 *** (0.004)	0.014 (0.017)	
互联网偏好 × 受教育程度	-0.007 *** (0.001)	0.180 *** (0.005)		-0.006 *** (0.001)	0.177 *** (0.005)	
其他控制变量	是	是	是	是	是	是
地区虚拟变量	否	否	否	是	是	是
行业虚拟变量	否	否	否	是	是	是
观测数	7035	7035	7035	7035	7035	7035
R 平方	0.441	0.488	0.232	0.449	0.499	0.277
第一阶段的 F 值	363.508	142.309		123.181	92.383	
内生性检验: Durbin-Wu-Hausman F 检验量 [p 值]			17.093 [0.000]			14.183 [0.000]
弱工具变量检验: Cragg-Donald Wald F 统计量			455.628			445.082

注: 括号内为稳健标准误; *、**、*** 分别表示在 10%、5%、1% 的水平上显著; 限于篇幅, 未报告其他控制变量和常数项的回归结果。

资料来源: 根据 2018 年中国家庭追踪调查 (CFPS) 数据计算得到。

弱工具变量检验 Cragg-Donald Wald F 统计量远超过 10% 偏误下的临界值 7.03, 即选定的工具变量不存在弱工具变量问题。最终检验结论显示, 数字赋能和受教育程度的交互项系数仍高度显著为正 (1% 的水平上显著), 这说明在控制核心变量因逆向因果关系而产生的内生性问题之后, 仍然可以证实学历越高, 数字赋能的工资溢价效应越强这一结论。

2. 处理效应模型的回归结果

劳动者是否使用互联网并非是完全随机分配的, 其家庭收入状况、父母受教育程度以及所在企业的行业职业特征、工作任务要求等都会对劳动者是否使用互联网产生影响, 从而使劳动者是否使用互联网并不是随机的, 而是与个体特征、企业特征之间存在相关性, 因此很可能导致式 (1) 存在样本自选择并诱致内生性。为解决此类问题, 借鉴程虹和李唐 (2017) 的研究方法, 采用处理效应模型 (TEM) 对数字赋能在不同受教育程度劳动者间的异质性工资溢价效应进行测度。表 4 报告了采用两步法 (2SLS) 和极大似然估计法 (MLE) 进行回归的结果。回归结果显示逆米尔斯比率 λ 均表现为高度显著 (1% 的水平上显著), MLE 估计方法中方程独立性检验在 1% 的水平上显著, 这说明式 (1) 确实存在样本自选择问题, 采用处理效用模型进行回归是合理的。检验结果再次说明, 在排除样本自选择偏误之后, 仍然可以证实劳动者的受教育程度越高, 其数字赋能的工资溢价效应越强。

表 4 数字赋能的工资溢价效应 (处理效应模型)

变量	2SLS		MLE	
	(1)	(2)	(3)	(4)
数字赋能 × 受教育程度	0.050 *** (0.010)	0.039 *** (0.010)	0.042 *** (0.009)	0.032 *** (0.009)
数字赋能	1.147 *** (0.166)	0.747 *** (0.143)	0.956 *** (0.068)	0.584 *** (0.094)
受教育程度	0.009 (0.006)	0.006 (0.006)	0.014 *** (0.004)	-0.012 (0.009)
其他控制变量	是	是	是	是
行业虚拟变量	否	是	否	是
省份虚拟变量	否	是	否	是
λ (逆米尔斯比率)	-0.577 *** (0.088)	-0.590 *** (0.087)	-0.477 *** (0.032)	-0.462 *** (0.031)
方程独立性检验 [p 值]			73.140 *** [0.000]	158.010 *** [0.000]
观测值	4260	4260	4260	4260

注: 括号内为稳健标准误; *, **, *** 分别表示在 10%、5%、1% 的水平上显著; 限于篇幅, 未报告其他控制变量和常数项的回归结果。

资料来源: 根据 2018 年中国家庭追踪调查 (CFPS) 数据计算得到。

3. 增加控制变量的回归结果

个人能力是影响工资溢价的重要因素（王春超、张承莎，2019），但是现实中“能力”不可观测，其遗漏会导致式（1）存在内生性问题。因此本文首先增加 2018 年 CFPS 访员对受访者的智力水平（IQ）评价作为个人能力的代理变量进行回归，回归结果如表 5 的第（1）列和第（2）列所示。其次，由新人力资本理论可知，能力分为认知能力和非认知能力，因此本文借鉴陈雨露和秦雪征（2018）的研究思路，一方面利用 2018 年 CFPS 提供的劳动者字词识记能力得分与数学计算能力得分，将其标准化后相加作为认知能力的代理变量；另一方面，本文将受访者自评的人缘关系和自信程度得分标准化，加总后作为非认知能力的代理变量，回归结果如表 5 的第（3）列、第（4）列所示。整体来看，无论是利用 IQ 作为能力的代理变量还是加入认知能力和非认知能力的代理变量，数字赋能和受教育程度的交互项系数依然高度显著，符号没有变化，表明在控制一些不可观测的个体遗漏特征之后，数字赋能下高学历劳动者能够获得更高的工资溢价这一核心结论仍然不会改变。

表 5 数字赋能的工资溢价效应（增加控制变量）

变量	(1)	(2)	(3)	(4)
数字赋能 × 受教育程度	0.034 *** (0.006)	0.026 *** (0.006)	0.036 *** (0.007)	0.028 *** (0.007)
数字赋能	0.170 *** (0.029)	0.150 *** (0.028)	0.206 *** (0.035)	0.187 *** (0.034)
教育程度	0.033 *** (0.003)	0.032 *** (0.003)	0.037 *** (0.005)	0.034 *** (0.005)
其他控制变量	是	是	是	是
行业虚拟变量	否	是	否	是
省份虚拟变量	否	是	否	是
R 平方	0.255	0.297	0.264	0.323
观测数	7040	7040	4376	4376

注：括号内为稳健标准误；*、**、*** 分别表示在 10%、5%、1% 的水平上显著；限于篇幅，未报告其他控制变量和常数项的回归结果。

资料来源：根据 2018 年中国家庭追踪调查（CFPS）数据计算得到。

（三）稳健性分析

为确保基本结论的可信度，本文首先采用倾向值匹配（PSM）方法进行稳健性检

验。具体做法：第一步，按照劳动者是否使用互联网，将样本分为两组以构造匹配样本，将使用互联网的劳动者定义为实验组，从未使用互联网的劳动者中寻找控制组；第二步，以是否使用互联网作为被解释变量，以年龄、性别、受教育程度、户籍等作为解释变量，利用 Logit 模型估计劳动者使用互联网的概率，并把回归预测值作为倾向得分；第三步，采用一对四近邻匹配和核匹配的方法对样本进行匹配，匹配后的样本分别包含 3364 个和 6880 个样本；最后运用匹配样本再次进行回归。表 6 的结果显示数字赋能在不同受教育程度劳动者间的工资溢价效应确实存在差异性，与上文的研究结论一致。

表 6 数字赋能的工资溢价效应 (PSM 筛选样本)

变量	近邻匹配(1:4)		核匹配	
	(1)	(2)	(3)	(4)
数字赋能 × 受教育程度	0.017 *** (0.008)	0.012 ** (0.008)	0.032 *** (0.006)	0.026 *** (0.006)
数字赋能	0.155 *** (0.032)	0.128 *** (0.031)	0.170 *** (0.029)	0.150 *** (0.028)
受教育程度	0.027 *** (0.005)	0.027 *** (0.005)	0.034 *** (0.003)	0.033 *** (0.003)
其他控制变量	是	是	是	是
行业虚拟变量	否	是	否	是
省份虚拟变量	否	是	否	是
R 平方	0.183	0.238	0.249	0.291
观测数	3364	3364	6880	6880

注：括号内为稳健标准误；*、**、*** 分别表示在 10%、5%、1% 的水平上显著；限于篇幅，未报告其他控制变量和常数项的回归结果。

资料来源：根据 2018 年中国家庭追踪调查 (CFPS) 数据计算得到。

其次，由于 2016 年 CFPS 数据同样包含有互联网使用、工资、个人特征等基本信息，本文进一步采用 2016 年 CFPS 数据进行稳健性检验。基于式 (1) 主要控制变量的选取与测度思路，表 7 为使用 2016 年 CFPS 数据的回归结果。表 7 的第 (1) 列、第 (2) 列以互联网使用作为数字赋能的衡量指标，第 (3) 列、第 (4) 列以手机上网作为数字赋能的衡量指标，结果显示互联网使用与受教育水平的交互项均显著为正，说明数字赋能下高学历劳动者的工资溢价效应更明显这一结论是稳健可靠的。

表 7 数字赋能的工资溢价效应（更换 2016 年 CFPS 数据）

变量	(1)	(2)	(3)	(4)
数字赋能×受教育程度	0.030 *** (0.009)	0.023 ** (0.009)		
手机上网×受教育程度			0.026 *** (0.009)	0.019 ** (0.009)
数字赋能	0.070 (0.045)	0.054 (0.044)		
手机上网			-0.134 (0.089)	0.045 (0.043)
受教育程度	0.023 *** (0.005)	0.023 *** (0.005)	0.013 * (0.007)	0.024 *** (0.005)
其他控制变量	是	是	是	是
行业虚拟变量	否	是	否	是
省份虚拟变量	否	是	否	是
R 平方	0.216	0.259	0.217	0.261
观测数	3948	3928	3948	3928

注：括号内为稳健标准误；*、**、*** 分别表示在 10%、5%、1% 的水平上显著；限于篇幅，未报告其他控制变量和常数项的回归结果。

资料来源：根据 2016 年中国家庭追踪调查（CFPS）数据计算得到。

五 作用机制分析

本部分将进一步检验不同受教育程度的劳动者在工作、学习、社交等方面的互联网使用行为，验证数字赋能通过何种渠道，让高学历劳动者产生更强的工资溢价效应。分析将从三方面进行。

（一）受教育程度与数字工作赋能

如前文所述，高学历劳动者由于工作上的需要和技能上的优势在工作中使用互联网的频率更高，数字赋能工作带来的工资溢价效应随之也更高。为检验这一作用机制，本文分两步来验证：第一步是使用受教育程度作为关键解释变量，以使用互联网进行工作的频率作为被解释变量，加入年龄、性别、户籍等控制变量并控制了行业和省份固定效应，考察劳动者的教育程度与互联网工作行为（使用互联网进行工作的频率）之间的关系；第二步是将 2018 年 CFPS 的“使用互联网进行工作的频率”作为数字赋

能的替代变量，验证互联网工作行为对劳动者工资的影响。鉴于劳动者互联网工作行为有 7 种不同程度的选择（分别为“从不”、“几个月一次”、“一月一次”、“一月 2 至 3 次”、“一周 1 至 2 次”、“一周 3 至 4 次”和“几乎每天”），且不同选项之间具有一定顺序，在此我们第一步主要选用多元有序 Probit（Ordered Probit）模型和多元有序 Logit（Ordered Logit）模型来分析受教育程度对劳动者互联网工作行为的影响；为了消除模型误设问题，本文也兼顾使用 OLS 回归方法。

表 8 的第（1）列、第（2）列和第（3）列展示了劳动者的教育程度和互联网工作行为的基本回归结果，其中第（1）列、第（2）列分别采用有序 Probit 方法和有序 Logit 方法进行回归，第（3）列采用 OLS 方法进行回归。多元有序 Probit 模型和多元有序 Logit 模型的检验结果均显示，劳动者的受教育程度和互联网工作行为之间保持着显著的正相关关系，表明劳动者的受教育程度越高，使用互联网进行工作的频率越高。基于 OLS 的回归结果同样显示，劳动者的受教育程度和互联网工作行为之间存在显著的正相关关系，OLS 同多元有序 Probit（Ordered Probit）模型和多元有序 Logit（Ordered Logit）模型的回归结果在显著性和系数符号方面均一致，进一步凸显了劳动者受教育程度与互联网工作行为正相关这一客观事实。表 8 的第（4）列回归结果显示，在工作中使用互联网的频率越高，劳动者获得的工资溢价越高，这说明受教育程度越高，劳动者在工作中使用互联网的频率越高，数字赋能带来的工资溢价更高，验证了本文的假说 2。

表 8 受教育程度与数字工作赋能

变量	互联网工作行为			小时工资对数
	(1)	(2)	(3)	(4)
	Ordered Probit	Ordered Logit	OLS	OLS
受教育程度	0.161 *** (0.007)	0.273 *** (0.011)	0.303 *** (0.011)	0.034 *** (0.004)
互联网工作行为				0.026 *** (0.005)
其他控制变量	是	是	是	是
省份虚拟变量	是	是	是	是
行业虚拟变量	是	是	是	是
Pseudo R 平方/ R 平方	0.137	0.137	0.313	0.301
观测数	5016	5016	5016	5016

注：括号内为稳健标准误；*、**、*** 分别表示在 10%、5%、1% 的水平上显著；限于篇幅，未报告其他控制变量和常数项的回归结果。

资料来源：根据 2018 年中国家庭追踪调查（CFPS）数据计算得到。

（二）受教育程度与数字学习赋能

根据前文分析，高学历劳动者基于个体职业发展目的和“终生学习”理念在学习中使用互联网的频率更高，数字赋能增强人力资本积累的效应越明显，因而带来的工资溢价效应越高。为检验这一作用机制，本文以 2018 年 CFPS 的“使用互联网进行学习的频率”作为衡量互联网学习行为的代理变量，同样分两步来验证：第一步是使用受教育程度作为关键解释变量，以使用互联网进行学习的频率作为被解释变量，年龄、性别、户籍等作为控制变量并加入行业和省份虚拟变量，考察劳动者的教育程度与互联网学习行为之间的关系；第二步是将互联网学习行为作为数字赋能的替代变量，验证互联网学习行为对劳动者工资的影响。因互联网学习行为也具有一定顺序的不同选择，故第一步主要选用多元有序 Probit 模型和多元有序 Logit 模型进行回归。具体回归结果如表 9 所示。

表 9 的第（1）列、第（2）列和第（3）列展示了劳动者的受教育程度和互联网学习行为的基本回归结果，多元有序 Probit 模型、多元有序 Logit 模型和 OLS 的回归结果显示劳动者的受教育程度和互联网学习行为显著正相关，表明随着受教育程度的增加，劳动者在学习中使用互联网的频率会增加，第（4）列汇报了互联网学习行为与劳动者工资回报显著正相关的 OLS 回归结果，说明互联网学习行为频率越高，劳动者的工资溢价效应越显著。这意味着，劳动者受教育程度越高，在学习中使用互联网的频率越高，数字赋能带来的工资溢价更高，验证了本文的假说 3。

表 9 受教育程度与数字学习赋能

变量	互联网学习行为			小时工资对数
	(1)	(2)	(3)	(4)
	Ordered Probit	Ordered Logit	OLS	OLS
受教育程度	0.126 *** (0.006)	0.215 *** (0.010)	0.236 *** (0.010)	0.036 *** (0.004)
互联网学习行为				0.015 *** (0.005)
其他控制变量	是	是	是	是
省份虚拟变量	是	是	是	是
行业虚拟变量	是	是	是	是
Pseudo R 平方/ R 平方	0.074	0.076	0.222	0.299
观测数	5367	5367	5367	5367

注：括号内为稳健标准误；*、**、*** 分别表示在 10%、5%、1% 的水平上显著；限于篇幅，未报告其他控制变量和常数项的回归结果。

资料来源：根据 2018 年中国家庭追踪调查（CFPS）数据计算得到。

(三) 受教育程度与数字社交赋能

根据前文分析,高学历劳动者为了拓展和维持自己的社会关系网络,其在社交中使用互联网的频率更高,数字赋能强化社会资本积累的效应越明显,因而带来的工资溢价效应越高。为检验这一作用机制,以 2018 年 CFPS 的“使用互联网进行社交的频率”来衡量互联网社交行为,即人们利用互联网来与他人实现社交互动的行为,仍然分两步进行考察:第一步是使用受教育程度作为关键解释变量,以使用互联网进行社交的频率作为被解释变量,加入影响互联网使用行为的个体特征变量并控制了行业虚拟变量和省份虚拟变量,考察劳动者受教育程度与互联网社交行为之间的关系;第二步是将使用互联网进行社交的频率作为数字赋能的替代变量,验证互联网社交行为对劳动者工资的影响。基本回归结果如表 10 所示。

多元有序 Probit 模型、多元有序 Logit 模型和 OLS 的回归结果均显示,劳动者的受教育程度和互联网社交行为之间保持着显著的正相关关系,说明劳动者受教育水平越高,在社交中使用互联网的频率越高。表 10 的第(4)列回归结果显示,互联网社交行为与劳动者工资回报显著正相关,表明在社交中使用互联网的频率越高,劳动者的工资回报越高。综合多元有序 Logit 模型、多元有序 Probit 模型的回归结果可知,劳动者受教育程度越高,劳动者在社交中使用互联网的频率越高,数字赋能带来的工资溢价更高,验证了本文的假说 4。

表 10 受教育程度与数字社交赋能

变量	互联网社交行为			小时工资对数
	(1)	(2)	(3)	(4)
	Ordered Probit	Ordered Logit	OLS	OLS
受教育程度	0.037 *** (0.006)	0.061 *** (0.011)	0.035 *** (0.007)	0.039 *** (0.004)
互联网社交行为				0.018 *** (0.006)
其他控制变量	是	是	是	是
省份虚拟变量	是	是	是	是
行业虚拟变量	是	是	是	是
Pseudo R 平方/ R 平方	0.056	0.056	0.091	0.299
观测数	5365	5365	5365	5365

注:括号内为稳健标准误;*、**、*** 分别表示在 10%、5%、1% 的水平上显著;限于篇幅,未报告其他控制变量和常数项的回归结果。

资料来源:根据 2018 年中国家庭追踪调查(CFPS)数据计算得到。

六 结论与政策建议

数字赋能不但创造了新的就业机会，而且能够提升劳动者的生产效率与工资报酬，存在明显的工资溢价效应。但劳动力市场中的数字赋能并不是一个同质化、均等化的过程，数字赋能产生的工资溢价效应，在不同受教育程度的劳动者群体之间存在明显差异，其原因主要是劳动者在互联网使用行为上的结构差异。本文利用北京大学中国社会科学调查中心发布的2018年中国家庭追踪调查（CFPS）数据，从理论和实证两个层面考察了数字赋能是否能对不同受教育程度劳动者产生差异化的工资溢价效应，并揭示其作用机制。

本文研究发现，其一，数字赋能存在显著的工资溢价效应，且随着劳动者受教育水平的提高，数字赋能对工资溢价的正向影响增强；定量分析表明，受教育程度每增加1年，数字赋能所带来的工资回报随之提高2.62%，高学历确实能使数字赋能产生更高的工资溢价效应。其二，作用机制分析表明，互联网使用行为上的结构差异，是引发高学历劳动者更大程度享受到了数字赋能的工资溢价效应的主要原因，劳动者的受教育程度越高，在工作、学习与社交中使用互联网的频率越高，通过增强人力资本积累与社会资本积累，能够产生更高的工资溢价效应。

本文的政策含义如下：一是政府应继续推动互联网等数字技术基础设施的建设，进一步加快宽带高速网络建设，提高宽带普及率，降低宽带使用费用，降低因供给不足带来的数字鸿沟，发挥数字赋能的工资溢价效应。二是应增加针对低教育程度劳动者的数字技能培训，使其拥有一些更高级的数字技术使用和处理能力，提高其数字技术使用技能和数字素养，有效弥合不同受教育程度劳动者之间的数字技能鸿沟。三是应增强低教育程度劳动者的数字技术使用意识，帮助其建立终身学习的理念，支持、引导和鼓励其利用数字技术更频繁、更高效地进行互联网发展行为，增加其在工作、学习和社交活动中使用数字技术的频率。四是政府应努力普及九年义务教育之外的高等教育，数字技术飞速进步意味着对劳动力的能力要求越来越高，应增加教育投资，通过教育大众化来提高全民终身教育水平和平均受教育程度。

参考文献：

边燕杰（2004），《城市居民社会资本的来源及作用：网络观点与调查发现》，《中国社

- 会科学》第 3 期，第 136 - 146 页。
- 陈雨露、秦雪征 (2018)，《相貌对个人认知能力与非认知能力的影响——基于中国家庭追踪调查 (CFPS) 数据的研究》，《劳动经济研究》第 4 期，第 71 - 96 页。
- 陈玉宇、吴玉立 (2008)，《信息化对劳动力市场的影响：个人电脑使用回报率的估计》，《经济学 (季刊)》第 4 期，第 1149 - 1166 页。
- 程虹、李唐 (2017)，《人格特征对于劳动力工资的影响效应——基于中国企业—员工匹配调查 (CEES) 的实证研究》，《经济研究》第 2 期，第 171 - 186 页。
- 程名望、张家平 (2019)，《新时代背景下互联网发展与城乡居民消费差距》，《数量经济技术经济研究》第 7 期，第 22 - 41 页。
- 蒋琪、王标悦、张辉、岳爱 (2018)，《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》，《劳动经济研究》第 5 期，第 121 - 143 页。
- 经济合作与发展组织 (2009)，《学会跨越数字鸿沟》，任仲伟、曲囡囡译，北京：教育科学出版社。
- 经济合作与发展组织 (2013)，《互联网经济展望 2012》，张晓译，上海：上海远东出版社。
- 李雅楠、谢倩芸 (2017)，《互联网使用与工资收入差距——基于 CHNS 数据的经验分析》，《经济理论与经济管理》第 7 期，第 87 - 100 页。
- 刘斌、李磊 (2012)，《贸易开放与性别工资差距》，《经济学 (季刊)》第 2 期，第 429 - 460 页。
- 刘传江、覃艳丽、李雪 (2018)，《网络社交媒体使用、社会资本积累与新时代农业转移人口的城市融合——基于六市 1409 个样本的调查》，《杭州师范大学学报 (社会科学版)》第 6 期，第 98 - 108 页。
- 世界银行 (2017)，《2016 世界发展报告：数字红利》，北京：清华大学出版社。
- 陶锋 (2011)，《吸收能力、价值链类型与创新绩效——基于国际代工联盟知识溢出的视角》，《中国工业经济》第 1 期，第 140 - 150 页。
- 王春超、张承莎 (2019)，《非认知能力与工资性收入》，《世界经济》第 3 期，第 143 - 167 页。
- 王元超 (2019)，《互联网工资溢价效应的阶层差异》，《社会学评论》第 2 期，第 27 - 41 页。
- 夏显力、陈哲、张慧利、赵敏娟 (2019)，《农业高质量发展：数字赋能与实现路径》，《中国农村经济》第 12 期，第 2 - 15 页。

- 杨碧云、吴熙、易行健 (2019), 《互联网使用与家庭商业保险购买——来自 CFPS 数据的证据》, 《保险研究》第 12 期, 第 30 - 47 页。
- 杨铭、王任远 (2019), 《受教育程度与农民工性别收入差异——基于动态博弈模型和流动人口监测数据的研究》, 《经济问题》第 9 期, 第 62 - 72 页。
- 张景娜、张雪凯 (2020), 《互联网使用对农地转出决策的影响及机制研究——来自 CFPS 的微观证据》, 《中国农村经济》第 3 期, 第 57 - 77 页。
- 庄家焯、刘爱玉、孙超 (2016), 《网络空间性别不平等的再生产：互联网工资溢价效应的性别差异——以第三期妇女地位调查为例》, 《社会》第 5 期, 第 88 - 106 页。
- Algina, James & Bradley Moulder (2001). A Note on Estimating the Jöreskog-Yang Model for Latent Variable Interaction Using LISREL 8.3. *Structural Equation Modeling*, 8 (1), 40 - 52.
- Anderson, Robert, Tora Bikson, Sally Law & Bridger Mitchell (1997). Universal Access to E-mail: Feasibility and Societal Implications. *Educational Media International*, 34 (2), 86 - 87.
- Autor, David, Frank Levy & Richard Murnane (2002). Upstairs, Downstairs: Computers and Skills on Two Floors of a Large Bank. *ILR Review*, 55 (3), 432 - 447.
- Bonfadelli, Heinz (2002). The Internet and Knowledge Gaps: A Theoretical and Empirical Investigation. *European Journal of Communication*, 17 (1), 65 - 84.
- Bresnahan, Timothy (1999). Computerisation and Wage Dispersion: An Analytical Reinterpretation. *The Economic Journal*, 109 (456), 390 - 415.
- DiMaggio, Paul & Bart Bonikowski (2008). Make Money Surfing the Web? The Impact of Internet Use on the Earnings of U. S. Workers. *American Sociological Review*, 73 (2), 227 - 250.
- Gaziano, Cecilie (1997). Forecast 2000: Widening Knowledge Gaps. *Journalism & Mass Communication Quarterly*, 74 (2), 237 - 264.
- Krueger, Alan (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984 - 1989. *The Quarterly Journal of Economics*, 108 (1), 33 - 60.
- Liu, Che-Wei & Sunil Mithas (2016). The Returns to Digital Skills: Evidence from India, 2005 - 2011. Paper presented at the proceedings of the 35th International Conference on Information Systems, Dublin, Ireland, December 11 - 14.
- Mincer, Jacob (1974). *Schooling, Experience, and Earnings*. New York: Columbia

University Press.

Spitz-Oener, Alexandra (2008). The Returns to Pencil Use Revisited. *ILR Review*, 61 (4), 502 – 517.

Why Does a Higher Education Degree Bring a Higher Wage Premium under the Digital Empowerment: An Empirical Study with CFPS Data

Zhang Guosheng & Wu Jing

(School of Economics, Yunnan University)

Abstract: This study uses data from the China Family Panel Studies in 2018 to explore whether digital empowerment can make higher educated workers to obtain higher wage and try to reveal its mechanism. The study finds that digital empowerment has a significantly positive relationship with the wage premium of workers. Overall, the higher the education level, the stronger the wage premium effect of digital empowerment will be. Every one year increase in a worker's education is associated with a 2.62 percent increase in wage brought by digital empowerment. The structural difference in internet usage behavior is the main factor for the heterogeneous wage premium among people with different education levels. On the one hand, workers with higher education are using internet more frequently at work, then the more significant the wage premium effect caused by digital empowerment will be. On the other hand, workers with higher education are using internet more frequently in learning and social interactions, then a higher wage premium effect can be produced by enhancing the accumulation of human capital and social capital.

Keywords: digital empowerment, wage premium, education level

JEL Classification: I23, J21, J31

(责任编辑: 封永刚)