

社会关系、信息网络对新农民工收入的影响

——基于熵均衡法的实证分析

陆文聪 谢昌财

【摘 要】文章采用中国家庭收入调查(CHIP)数据,根据准实验研究设计原理将初次外出的新农民工按其拥有的社会关系与信息网络差异分为7个不同情境的处理组与对照组,运用熵均衡处理方法估算组间收入差异,以分析社会网络对初次外出新农民工收入的影响。研究表明,亲戚关系对新农民工的收入效应强于熟人关系,但老乡信息网络比血缘信息网络更利于新农民工找到收入更高的工作;既有亲戚关系又有老乡信息网络的新农民工的平均收入比其他新农民工高50%以上。新农民工在初次务工城市通过强关系的人情资源与老乡网络的信息渠道实现就业是成本低、效率高、效果好的路径。基于这些结论,文章建议构建以政府为主导、以企业为基石、以社区为纽带的多元社会网络支持体系,打破新农民工传统闭塞的人际交流圈,重构其在务工城市的新型社会网络。

【关键词】新农民工 社会关系 信息网络 收入差异 熵均衡

【作 者】陆文聪 浙江大学中国农村发展研究院,教授;谢昌财 浙江大学管理学院,博士研究生。

一、引 言

随着农业新技术的推广和农民人力资本素质的提升,农业劳动生产率不断提高,中国农村对劳动力的需求量也逐年下降,大量农村剩余劳动力随之向城市第二、第三产业转移就业。根据国家统计局农村司的监测调查报告,2016年中国进城务工的农民工数量达2.82亿人,比上年增加424万人,增长1.5%;预计至2020年,每年将新增约500万新农民工。如何使新农民工在经济社会发展中获得相应利益、更好地融入城市生活是政府及学者关心的热点问题。对新农民工来说,第一份工作是其职业生涯的起点,由于缺乏工作经验,离开熟悉的“乡土社会”进入陌生的城市求职,最理性的选择是依靠亲朋

好友的社会关系与信息渠道(赵延东、王奋宇,2002)。因此,研究社会关系和信息网络对新农民工寻求非农就业机会及其收入的影响,具有重要的现实意义。

劳动者收入差异是多年来学术界一直关注的重要问题。早在20世纪60年代,人力资本理论就将劳动者收入差距归结为个人能力水平的差异。此后,社会资本理论认为嵌入在社会网络结构中的社会资本是引起收入差异的另一个重要因素(Granovetter, 1973)。虽有大量文献考察了社会关系与信息网络对劳动者收入水平的影响,但得出的结论并不一致。例如,一些研究认为,社会关系与信息网络有助于劳动者社会经济地位的获得并影响其收入(王春超、周先波,2013);另一些研究却认为,社会关系不影响工作机会(Lin, 1999),社会网络对收入不产生直接影响(Mouw, 2003)。导致这种分歧的原因可能有以下两个方面。一方面,在理论研究上对社会网络的界定和测量不同。例如,章元和陆铭(2009)、叶静怡和武玲蔚(2014)分别以赠送给亲友的礼金价值占家庭总支出的比重和职业权力与职业声望分数度量农民工的社会网络。另一方面,在实证研究上采用不同的数据与计量方法。例如,吴愈晓(2011)采用社会网络与职业经历问卷调查数据及多元对数回归模型方法分析社会关系对初职获得方式的影响,宋月萍、马腾(2015)采用2013年全国流动人口动态监测社会融合专题调查数据和倾向得分匹配法估算同乡会对农民工劳动收入的影响。

总的来看,关于外出农民工收入问题的研究还需进一步探索。一是从研究对象上看,鲜有学者重视新农民工如何通过社会关系与信息网络获取非农就业岗位,以及由此导致的收入差异。二是从研究方法上看,早期关于农民工收入的研究主要采用均值回归模型,由于存在内生性问题,并且容易受到极端值的影响,其结果往往不可靠。虽然学者运用匹配法来解决这类问题,在一定程度上有效控制了选择偏误,结果的可靠性也得到一定的提升。但这些方法(如倾向得分匹配、基因匹配、马氏距离匹配等)在解决处理组与对照组协变量均衡问题上还存在不足,同时在匹配过程中造成大量样本数据缺失,并最终影响分析结果的可靠性。Hainmueller(2012)提出的熵均衡数据处理方法可以较好地解决上述问题。该方法通过预先设置一组平衡性约束与规范性约束条件,确保处理组和对照组在特定矩下实现数据的精确匹配,并自动计算一组与约束条件相匹配的最优权重。与匹配法相比,熵均衡可以在整体样本中为每一个处理组样本匹配一个与之相似的对照组,进而保留所有样本信息,并最大限度地消除样本中的内生偏误,实现处理组与对照组之间样本的精确匹配,协变量标准化均数差及均值差异性检验更为稳健,结果更为可靠。

鉴于此,本文以新农民工为研究对象,采用中国家庭收入调查(CHIP)数据,基于社会网络视角对样本进行分组,在控制个人资本、人力资本和政治资本变量的同时,采用熵均衡处理方法,分别实证检验社会关系和信息网络对其收入的影响。

二、文献综述与理论假设

社会网络有3个明显特征:一是达高性,即个体通过社会关系在等级制结构中可触及的最顶端位置的资源;二是异质性,即个体通过社会网络可触及的信息资源幅度;三是广泛性,即个体通过社会网络可触及的信息资源数量(Lin, 1999)。本文选择社会关系和信息网络两个维度作为社会网络的代理变量,研究其对新农民工收入的影响。其中,社会关系是指新农民工初次外出务工时在务工城市拥有的人际关系,在社会网络结构中发挥的是人情作用。根据格兰诺维特的定义,强关系是指交往频繁、关系紧密,有很强情感因素维系着的人际关系,如亲戚;弱关系和强关系相对应,其特征是交往面广,交往对象可能来自各行各业的人际关系,如朋友、熟人等。信息网络是指新农民工在初次外出务工时获取就业信息的渠道。

本文将通过亲戚提供的就业信息定义为亲戚网,通过老乡提供的就业信息定义为老乡网。社会关系与信息网络分别关注社会网络的网络关系强弱与结构,他们在为新农民工初次进入城市务工时提供社会支持、信誉保证和就业信息。新农民工求职过程中,社会关系的本质在于信任关系,在面对雇主时依靠亲朋好友的关系进行推荐,以亲朋好友的信誉作为自身求职提供担保,进而获取更好的工作岗位;信息网络的本质在于信号传递,在尚不完善的劳动力市场上进行职业信息搜寻时,新农民工通过其拥有的信息网络可以快速筛选职业信息,缩短搜寻时间和减少搜寻成本。

(一) 社会关系与农民工收入

费孝通(1997)指出,中国社会结构是一种“差序格局”,社会关系以个人为中心逐渐向外推进,表明了自己与他人的亲疏远近,以亲戚为纽带的宗族关系是中国农村最重要和最稳定的社会网络之一。格兰诺维特(Granovetter, 1973)提出了“弱关系假说”,而边燕杰(Bian, 1997)则认为,在中国文化背景下,通过强关系能够获得更好的工作。张顺、郭小弦(2011)认为,强关系假设更适用于中国劳动力市场,尤其在低端劳动力市场中作用更加明显。强关系在权力、财富、声望等方面的优势明显,关系越强,触及的资源位势就越高,社会关系就越可能正向影响情感和社交支持,从而发挥其作为人情资源作用的优势。基于上述分析,本文提出假设1:社会关系作为一种人情资源,强关系对新农民工收入的影响大于弱关系。

(二) 信息网络与农民工收入

初次外出农民工由于和城市接触较少,在搜寻就业信息时寻求亲戚和老乡是较好的选择。在向新农民工传递劳动力市场信息的过程中,信息网络扮演了重要角色。章元等(2008)的研究发现,农民工利用血缘信息网络找到的工作不能直接提高工资水平,而基于同村关系的老乡信息网络有助于提高初次外出农民工的收入,其原因是血缘信息

网络提供的劳动力市场信息是有限和重复的,即这些信息具有“同质性”;而通过老乡信息网络,新农民工可以认识更多的非重复关系人,掌握更多的异质性资源,因此,与血缘信息网络相比,老乡信息网络传递的劳动力市场信息价值更高。张春泥、谢宇(2013)的研究表明,同乡聚集有助于提高农民工的工资收入,倾向于同乡聚集的农民工从同乡聚集中获得的收入回报更高。李根强等(2016)指出,同乡聚集的正向选择作用说明同乡聚集作为农民工信息网络的一种形式有助于提高农民工的劳动力市场收益。基于上述分析,本文提出假设 2:信息网络作为一种信息资源,老乡信息网络对新农民工的收入效应强于血缘信息网络。

(三) 社会关系+信息网络与农民工收入

在目前市场经济条件下,劳动就业市场信息在一定程度上还处于非对称状态,新农民工很难获得有价值的信息,与雇主的互信关系也无从建立。在这种情况下,社会关系与信息网络将是信息的桥梁、信任的基础和作为人际关系约束的保证(边燕杰、张文宏,2001)。张顺、郭小弦(2011)认为,社会网络中人情资源对入职收入的影响大于信息资源,社会网络资源与收入存在正相关关系。基于上述分析本文提出假设 3:社会关系的人情资源作用对新农民工的收入效应强于信息网络。

本文以初次外出的新农民工为研究对象,从社会关系与信息网络 2 个维度和 7 个情景考察社会网络对新农民工的收入效应及作用形式,同时考察 Lin(1999)指出的社会网络的 3 个特征。

三、方法、变量与数据

为了避免均值回归模型可能出现的不满足因变量对称分布的条件假设及模型可能产生的内生性问题,本文在匹配法的基础上,引进熵均衡方法对数据进行预处理,在最大程度上控制样本的选择性偏误。

(一) 方法与模型

近年来,经济学研究中解决二值选择问题通常采用平均处理效应的计算框架。首先构造一个二元虚拟变量 $D_i = \{0, 1\}$,其中 $D_i = 1$ 表示处理组。引入 J 个协变量,构建矩阵 $X_i = \{X_{i1}, X_{i2}, \dots, X_{iJ}\}, j \in \{0, 1\}$ 。则样本的平均处理效应为:

$$ATE = E[Y(1)|D=1] - E[Y(0)|D=1] \quad (1)$$

其中, $E[Y(0)|D=1]$ 是一个反事实结果,一般通过构建 Logit 或 Probit 模型,运用倾向得分法进行估算。Diamond 等(2013)指出,倾向得分法的缺点是处理高维度数据时协变量平衡效果不佳,且估算的倾向得分不可靠。传统倾向得分匹配方法需要找到一个特定待估模型,然后用 Logit 回归估计单元数据的权重,最后通过平衡性检验检查待估权

重是否与协变量分布一致。与倾向得分法不同的是,熵均衡方法试图为对照组寻找一组最优权值,使其与处理组在各个矩条件下相匹配。估算公式为:

$$E[Y(0)|D=1] = \frac{\sum_{\{i|D=0\}} Y_i w_i}{\sum_{\{i|D=1\}} w_i} \quad (2)$$

其中, w_i 是对照组样本的权值。权值选择遵循: $\min D(w) = \sum_{\{i|D=0\}} d(w_i)$, 服从以下平衡性约束和规范性约束条件:

$$\sum_{\{i|D=0\}} w_i CR_i(X_i) = m_r, \quad r = 1, 2, \dots, R \quad (3)$$

$$\sum_{\{i|D=0\}} w_i = 1, \quad w_i \geq 0, \quad \{i|D=0\} \quad (4)$$

其中, $d(w_i)$ 为距离函数, 由最小偏差估计 CR 发散度 (Read 等, 1988)。根据 Kullback (1959), 可计算 $d(w_i) = w_i \log(w_i/b_i)$, 其中 w_i 是待估权重, b_i 是基准权重。 $CR_i(X_i) = m_r$ 描述了 R 组关于赋值后对照组的平衡性约束条件。熵均衡需要在式(3)和式(4)约束下求损耗函数 $D(w)$ 的最小值。引入拉格朗日函数, 可得到以下优化问题:

$$\min_{w, \lambda_0, G} L = \sum_{\{i|D=0\}} w_i \log \frac{w_i}{b_i} + \sum_{r=1}^R \lambda_r (\sum_{\{i|D=0\}} w_i CR_i(X_i) - m_r) + (\lambda_0 - 1) (\sum_{\{i|D=0\}} w_i - 1) \quad (5)$$

其中, $G = [\lambda_0, \dots, \lambda_R]^T$ 是平衡性约束条件的拉格朗日乘数。令一阶条件 $\partial L / \partial w_i = 0$, 可得到每个权重的解, 问题转化为一个无约束条件的最小化求值:

$$\min_G L = \log \left\{ \sum_{\{i|D=0\}} b_i \exp \left[- \sum_{r=1}^R \lambda_r CR_i(X_i) \right] \right\} + \sum_{r=1}^R \lambda_r m_r \quad (6)$$

借用列文伯格—马夸尔特法 (Levenberg-Marquardt, LM) 求解最优值 G^* , 定义 $(R \times n_0)$ 维约束矩阵 $C = [C_1(X_i), \dots, C_R(X_i)]^T$ 和矩向量 $M = [m_1, \dots, m_R]^T$, 则 $\min D(w)$ 可由 $CW = M$ 给出, 其中 C^T 必须是满秩的, 因此熵均衡法对处理组与对照组数据有一定要求, 否则将无法找到可行解, 这也是该方法的缺陷。式(6)可改写为: $\min_G L = \log [B^G \exp(-C^T G)] + M^T G$ 。

根据 Hainmueller (2012) 的熵均衡法原理可求得每个待估权重的最优解为:

$$w_i^* = B \exp(-C^T G) \div B^T \exp(-C^T G) \quad (7)$$

为避免遗漏变量所带来的遗漏偏误问题, 引入 Rosenbaum 等 (1983) 提出的条件独立假定, 即给定 X_i , $E(Y_{0i} | X_i, D_i) = E(Y_{0i} | X_i)$, 且 $E(Y_{1i} | X_i, D_i) = E(Y_{1i} | X_i)$, 意味着在给定 X_i 的情况下, Y_{0i} 与 Y_{1i} 都独立于 D_i 。如果该假定成立, 可以将协变量引入以下回归方程, 进而解决遗漏变量带来的遗漏偏差问题。即:

$$y_i = \alpha + x_i' \beta + \gamma D_i + \varepsilon_i \quad (8)$$

其中, x'_i 为控制变量向量, D_i 为核心解释变量, γ 为核心解释变量的回归系数。

(二) 变量选择

1. 被解释变量。中国家庭收入调查问卷中仅提供新农民工获得第一份工作时的月收入 and 离开第一份工作时的月收入, 但考虑到用人单位通常采用劳动试用期制度, 且在试用期内可能实行低工资水平, 为了避免这种试用期制度对新农民工工资收入的影响, 本文选取新农民工离开第一份工作时的月收入作为被解释变量。

2. 核心解释变量。本文的核心解释变量为社会关系和信息网络。度量新农民工社会关系的题目是“第一次来城市时是否在城里有认识的人”, 选项为: (1) 有亲戚; (2) 有朋友或熟人; (3) 有其他认识的人; (4) 没有认识的人。度量新农民工信息网络的题目是“谁向您提供了有关第一次外出找工作的信息”, 选项为: (1) 亲戚; (2) 本村其他外出务工经商者; (3) 熟人; (4) 没有人。具体分为 7 个解释变量: $treat_1$ 为初次务工城市有亲戚关系; $treat_2$ 为初次务工城市有熟人关系; $treat_3$ 为初次务工城市有血缘信息网络; $treat_4$ 为初次务工城市有老乡信息网络; $treat_5$ 为有亲戚关系且有老乡信息网络; $treat_6$ 为有亲戚关系但无老乡信息网络; $treat_7$ 为有老乡信息网络但无亲戚关系。

3. 协变量。协变量由个人特征 (IC)、人力资本 (HC)、社会资本 (SC) 和政治资本 (PC) 构成。个人特征变量由性别、婚姻状况、冒险精神和信任水平 4 个协变量组成, 其中, 冒险精神衡量农民工对初次外出务工的态度, 并按由低到高划分 10 个不同等级; “对身边大多数人的信任情况” 衡量新农民工日常的信任水平。人力资本变量由年龄、教育和健康状况 3 个协变量组成。社会资本变量主要由 7 个模型中的核心解释变量 (社会关系与信息网络) 来反映, 加入兄弟姐妹数量和第一次进城时是否与人结伴 2 个协变量。政治资本变量由外出务工前是否有当兵、村干部或其他非农工作经历来描述^①。

为检验本文的理论假设, 对被解释变量取对数, 式 (8) 的回归模型可拓展为:

$$\ln y_i = \alpha + \beta_1 IC_i + \beta_2 HC_i + \beta_3 SC_i + \beta_4 PC_i + \gamma D_i + \varepsilon_i \quad (9)$$

(三) 数据来源

本文采用中国收入分配研究院 CHIP 数据库中的“中国城市、农村和流动人口居民收入调查 (CHIP2008)”数据^②, 调查样本涉及广东、江苏、浙江、上海、湖北、重庆、

① 需说明的是, 问卷中的“非农工作经历”并不能完全衡量政治资本变量。但本文研究对象是初次外出务工者, 受访对象在回答该问题时主要集中在“外出务工前是否有当兵或村干部经历”, 具有“非农工作经历”的比例很少。因此, 即使无法将这部分人从统计中分开, 对结果也不存在较大影响。

② 中国家庭收入调查 (CHIP) 中只有 2008 年调查涉及本文所研究的初次外出农民工数据, 并且没有后续跟踪调查。除 CHIP 数据库之外, 目前还没有其他专项数据涉及新农民工的收入问题。需要指出的是, 本文探讨的重点是社会网络对新农民工收入的相对影响, 而不是收入水平的绝对增长, 因此, 数据的时间问题对本文的结论不会构成较大的影响。

表 1 变量描述性统计(n=1870)

变 量	均值	标准差	最小值	最大值
收入	883.6727	485.7792	50	4000
性别	0.6460	0.4783	0	1
婚姻状况	0.5802	0.4936	0	1
冒险精神	4.2299	2.6445	0	10
信任水平	0.5716	0.4950	0	1
年龄	30.3492	9.5330	16	65
受教育年限	9.3455	2.4847	1	16
健康状况	0.9786	0.1447	0	1
兄弟姐妹数量	2.2117	1.6055	0	11
结伴出行	0.7091	0.4543	0	1
党政经历	0.0973	0.2964	0	1

注:收入变量最小值为 50 元,显然存在调查错误。但为了尊重原始数据,本文未对其删除处理。

劳动收入,运用 OLS 回归来估计收入效应可能出现较大的误差,因为这两类群体在年龄、受教育程度、健康状况、对他人的信任水平、过往经历等诸多因素上存在差别,同时,这些差异会在不同程度上影响其自身社会网络的构建。为了与熵均衡处理的结果进行比较,本文首先运用 OLS 回归方法,实证估计 7 个模型参数,结果如表 2 所示。

(一) OLS 回归结果

模型 1 和模型 2 考察社会关系的收入效应。回归结果显示,新农民工在其务工城市“有亲戚关系”比“没有亲戚关系”的收入高 18.48% ($1-e^{0.1696}=0.1848$),“有熟人关系”比“没有熟人关系”的收入低 4.78%。性别、冒险精神、信任水平、受教育年限、健康状况、结伴出行及党政经历等变量的回归系数显著为正。年龄变量的回归系数显著为负,表明新农民工年龄越大,收入越低;兄弟姐妹数量的回归系数为负且非常显著,表明农民工兄弟姐妹人数的增加会稀释“强关系”的作用。模型 2 中,婚姻状况变量的回归系数为负,表明未婚者受自身家庭因素影响较小,可以有更多时间与精力投入劳动。

模型 3 和模型 4 考察信息网络的收入效应。回归结果显示,新农民工在其务工城市“有血缘信息网络”和“有老乡信息网络”分别比对照组的收入高 11.35% 和 14.05%。模型 5 至模型 7 进一步考察“亲戚关系+老乡信息网络”的联合效应。回归结果显示,新农民工在其务工城市“有亲戚关系且有老乡信息网络”、“有亲戚关系但无老乡信息网络”和“有老乡信息网络但无亲戚关系”的月收入分别比对照组高 53.85%、13.79% 和 7.05%,收入效应均为正,协变量系数与其他模型相比未出现明显变化。

(二) 熵均衡处理结果与分析

由于处理组与对照组样本的协变量之间存在差异,OLS 回归尽管对协变量进行了

四川、河南、安徽等省份共计 15 个大中城市的外来务工者。剔除无效数据和“自我雇佣”样本后,本文采用的数据样本总量为 1 870 个,样本描述性统计如表 1 所示。

四、实证结果与分析

本文试图探究新农民工是否拥有社会网络对其劳动收入的影响,但现实中只能观测到每位农民工“有社会网络”或者“没有社会网络”的劳

表 2 OLS 参数估计

变 量	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7
treat _{1~7}	0.1696*** (0.0301)	-0.0489* (0.0290)	0.1075*** (0.0300)	0.1315*** (0.0375)	0.4308*** (0.0867)	0.1292*** (0.0308)	0.0681* (0.0390)
性别	0.1018*** (0.0301)	0.0915*** (0.0304)	0.0943*** (0.0302)	0.0813*** (0.0307)	0.0880*** (0.0303)	0.0994*** (0.0303)	0.0856*** (0.0306)
婚姻状况	-0.0580 (0.0374)	-0.0669* (0.0376)	-0.0654* (0.0376)	-0.0723* (0.0374)	-0.0753** (0.0374)	-0.0577 (0.0375)	-0.0685* (0.0366)
冒险精神	0.0116** (0.0058)	0.0099* (0.0058)	0.0110* (0.0058)	0.0103* (0.0058)	0.0099* (0.0058)	0.0112* (0.0058)	0.0102* (0.0058)
信任水平	0.0523* (0.0289)	0.0523* (0.0291)	0.0516* (0.0290)	0.0579** (0.0292)	0.0622** (0.0290)	0.0493* (0.0291)	0.0535* (0.0292)
年龄	-0.0112*** (0.0025)	-0.0117** (0.0025)	-0.0112*** (0.0025)	-0.0118*** (0.0025)	-0.0111*** (0.0025)	-0.0115*** (0.0025)	-0.0118*** (0.0025)
受教育年限	0.0572*** (0.0060)	0.0555*** (0.0060)	0.0568*** (0.0060)	0.0557*** (0.0060)	0.0552*** (0.0060)	0.0567*** (0.0060)	0.0554*** (0.0060)
健康状况	0.4680*** (0.1332)	0.4646*** (0.1328)	0.4668*** (0.1318)	0.4489*** (0.1324)	0.4646*** (0.1330)	0.4667*** (0.1328)	0.4559*** (0.1323)
兄弟姐妹数量	-0.0236** (0.0106)	-0.0217** (0.0107)	-0.0226** (0.0107)	-0.0213** (0.0107)	-0.0219*** (0.0107)	-0.0232** (0.0107)	-0.0216** (0.0107)
结伴出行	0.0653** (0.0316)	0.0595* (0.0319)	0.0589* (0.0318)	0.0456 (0.0319)	0.0527* (0.0318)	0.0648** (0.0317)	0.0521* (0.0320)
党政经历	0.1337*** (0.0491)	0.1270*** (0.0490)	0.1289*** (0.0493)	0.1195*** (0.0486)	0.1161** (0.0485)	0.1351*** (0.0494)	0.1246** (0.0488)
常数项	5.7775*** (0.1582)	5.9047*** (0.1570)	5.8162*** (0.1580)	5.8944*** (0.1571)	5.8687*** (0.1576)	5.8128*** (0.1578)	5.8982*** (0.1570)
F 检验	29.65	27.10	27.97	28.44	29.93	28.34	27.36
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Adj-R ²	0.1695	0.1569	0.1614	0.1611	0.1652	0.1634	0.1570
处理组 n ₁	624	706	663	332	45	579	287
对照组 n ₀	1246	1164	1207	1538	1825	1291	1583

注：括号内数据为稳健标准误；*、**、*** 分别表示在 10%、5%和 1%的水平上显著。

控制,但仍然很难得到无偏的估计结果。本文采用熵均衡方法,将处理组与对照组样本的协变量一阶矩(均值)和二阶矩(方差)两个维度进行调整,使两组样本协变量在满足约束条件下实现精确匹配。结果如表 3 所示。

表 3 列出的是模型 8 协变量在熵均衡处理前后的均值、方差及匹配性检验(限于篇幅,模型 9 至模型 14 检验结果文中未列出)。匹配前协变量在一阶矩和二阶矩上均有较大差异,经熵均衡法匹配后,处理组与对照组协变量的均值、方差基本一致。为进一步检验熵均衡结果的可靠性,可以计算熵均衡处理前后处理组与控制组的标准化均数

表3 协变量匹配性检验

变 量	均值		方差		SMD	T-test	
	处理组	对照组	处理组	对照组		t-score	p> t
性别							
处理前	0.5897	0.6742	0.2423	0.2198	-0.1772	-3.5535	0.0004
处理后	0.5897	0.5898	0.2423	0.2421	-0.0002	-0.0041	0.9968
婚姻状况							
处理前	0.5449	0.5979	0.2484	0.2406	-0.1074	-2.1783	0.0296
处理后	0.5449	0.5449	0.2484	0.2425	0	0	1
冒险精神							
处理前	4.0370	4.3270	7.3970	6.770	-0.1098	-0.0139	0.0277
处理后	4.0370	4.0370	7.3970	6.854	0	0	1
信任水平							
处理前	0.5705	0.4972	0.2454	0.2450	-0.0034	-0.0700	0.9450
处理后	0.5705	0.5705	0.2454	0.2452	0	0	1
年龄							
处理前	29.63	31.55	92.47	89.77	-0.1134	-2.2998	0.0217
处理后	29.63	29.63	92.47	84.82	0	0	1
受教育年限							
处理前	9.163	9.1850	5.44	6.521	-0.1104	-2.3182	0.0206
处理后	9.163	9.1640	5.44	6.551	-0.0004	-0.0085	0.9936
健康状况							
处理前	0.9760	0.9799	0.0235	0.0197	-0.0269	-0.5531	0.5941
处理后	0.9760	0.9760	0.0235	0.0235	0	0	1
兄弟姐妹数量							
处理前	2.234	2.304	2.372	2.683	-0.0205	-0.4274	0.6692
处理后	2.234	2.234	2.372	2.760	0	0	1
结伴出行							
处理前	0.6795	0.8605	0.2181	0.2000	-0.0978	-1.9647	0.0498
处理后	0.6795	0.6795	0.2181	0.2180	0	0	1
党政经历							
处理前	0.0785	0.1067	0.0725	0.0954	-0.0951	-2.0278	0.0429
处理后	0.0785	0.0786	0.0725	0.0724	-0.0001	-0.0015	0.9992

注：(1) $SMD=(\bar{X}_T-\bar{X}_C)/\sqrt{[S_T^2(n_T-1)+S_C^2(n_C-1)]/(n_T+n_C-2)}$ ；(2) $t=(\bar{X}_T-\bar{X}_C)/\sqrt{[S_T^2/n_T+S_C^2/n_C]}$ ，其中， \bar{X}_T 、 \bar{X}_C 分别为处理组与对照组各变量的均值， S_T^2 、 S_C^2 分别为处理组与对照组各变量的方差。

模型8和模型9描述了经熵均衡处理后“社会关系”对新农民工的收入效应。结果显示，新农民工在其务工城市有亲戚关系比对照组的收入高17.99%；有熟人关系比对

差(SMD)，并对均值差异进行t检验。经熵均衡调整后所有协变量p值为1(或无限逼近于1)，表明处理组与对照组各个协变量数据已经实现100%匹配。对模型1匹配后的数据进行重新回归，有效样本量为1248个，即从对照组中找到624个与处理组可以完全匹配的样本。模型8至模型14处理后样本回归系数如表4所示，限于篇幅，协变量回归系数未列出。

表 4 熵均衡处理前后收入效应对比

模型	有效样本量		社会网络变量系数		95%水平下置信区间	
	处理前	处理后	处理前	处理后	处理前	处理后
模型 8	1870	1248	0.1696***	0.1655***	[0.111,0.229]	[0.102,0.229]
模型 9	1870	1412	-0.0489*	-0.0485*	[-0.106,0.008]	[-0.111,0.013]
模型 10	1870	1326	0.1075***	0.1044***	[0.049,0.166]	[0.042,0.167]
模型 11	1870	664	0.1315***	0.1321***	[0.058,0.205]	[0.050,0.213]
模型 12	1870	90	0.4308***	0.4275***	[0.261,0.601]	[0.239,0.616]
模型 13	1870	1158	0.1292***	0.1278***	[0.069,0.190]	[0.063,0.193]
模型 14	1870	574	0.0681*	0.0677	[-0.008,0.145]	[-0.017,0.152]

注：*、**、*** 分别表示在 10%、5%和 1%水平上显著。

照组的收入低 4.73%，假说 1 成立。这表明对于在初次务工城市拥有一定社会关系的新农民工，如果这种关系是“强关系”，其收入效应为正；相反，依靠“弱关系”的人情资源并不能为新农民工增加收入。强关系的收入效应更高，是因为强关系中包含了义务与责任，能够产生互惠与信任，同时也更容易传播影响力。

模型 10 和模型 11 描述了经熵均衡处理后信息网络对新农民工的收入效应。结果显示，新农民工在其务工城市有亲戚提供务工信息比对照组的收入高 11.00%；有老乡提供务工信息比对照组的收入高 14.12%，假设 2 成立。这表明，信息网络对新农民工的收入效应为正，但老乡信息网络的收入效应相比血缘信息网络更高。新农民工通过老乡信息网络可以接触更多的非重复关系人，他们可以为新农民工提供更多的非重复信息，并且由他们构建起的网络规模和所提供的就业信息明显优于以提供同质性资源为主的血缘信息网络。这一结果表明，信息网络对新农民工发挥作用的机制是通过老乡信息网络接触到的“广泛性”和“异质性”资源。老乡信息网络实际上是一种非正式群体，新农民工在陌生城市与老乡交往，不仅会产生强烈的心理认同，还能获得实质性的帮助。老乡信息网络作为群体沟通的桥梁，能联络不同职业、不同地位的群体，新农民工能借此获得丰富多样的就业信息和求职渠道，提高其社会经济地位，并将新农民工的信息网络触角延伸至不同社会阶层和职业群体中。

模型 12 至模型 14 描述了经熵均衡处理后社会关系和信息网络的联合效应。结果显示，新农民工在务工城市有亲戚关系且有老乡信息网络、有亲戚关系但无老乡信息网络和有老乡信息网络但无亲戚关系的月收入分别比对照组高 53.34%、13.63%和 7.00%。这说明对于在初次务工城市仅有亲戚关系而没有老乡信息网络和仅有老乡信息网络而没有亲戚关系的外出务工者，其收入效应虽然为正，但作用远不及两者的联合效应。进一步对比发现，与模型 11 老乡信息网络的正效应相比，模型 8 的亲戚关系效应更强，这

说明社会网络的人情作用大于信息作用,社会网络对新农民工收入影响的形式是以提供人情为主,以传递信息为辅,本文假设3成立。此外,模型13的收入效应高于模型14,再次证实了社会关系的人情资源作用大于信息网络的信息资源作用。

五、结论与启示

本文基于中国家庭收入调查数据,采用熵均衡方法,实证分析了社会关系与信息网络对新农民工收入的影响。研究结果显示:(1)熵均衡法在估算社会网络效应问题上具有独特优势。运用熵均衡法估算不同情境下新农民工的收入差异可以解决内生性偏误问题,实证估计结果也更为稳健和可靠。(2)亲戚关系对新农民工的收入效应强于熟人关系,老乡信息网络比血缘信息网络更有利于其找到收入较高的工作岗位。新农民工在初次务工城市通过强关系的人情资源与老乡网络的信息资源实现就业是一条“成本低、效率高、效果好”的路径,新农民工在寻找第一份工作时科学有效的方式是通过老乡网络搜寻就业信息,通过亲戚关系实现在收入更高的岗位上就业。(3)社会关系与信息网络对新农民工收入影响的作用形式不同。社会关系依靠他人所占据社会位置上的重要资源,强调的是人情关系的亲密程度和达高性;信息网络依靠他人所占据社交网络上的“信息桥”,强调的是信息资源的广泛程度和异质性。从社会关系与信息网络的联合作用来看,其对于新农民工收入发挥作用的形式是以提供人情为主,以传递信息为辅。

从社会网络对新农民工收入效应的实证分析结论可以看出:(1)提升新农民工的人际交往能力、拓展信息交流网络是新农民工初次外出务工时提高收入的关键。因此,新农民工流出地政府应积极构建更为丰富和开放的农村社会生态网络,从认知和情感上打破传统农村社会保守和闭塞的人际交流圈,让新农民工在外出前有更广泛的人际接触机会,调整 and 改变他们拓展社会网络的态度与能力。(2)对于新农民工流入地,一方面新农民工应主动实现自身社会网络的转变与重构,进而寻求更好的发展机会。另一方面,政府应在引导和促进新农民工构建城市社会网络上有所作为,建立一个以政府为主导、以企业为基石、以社区为纽带的多元支持系统,政府可采取措施鼓励社区和社会机构接纳新农民工参加各种社会活动,通过人际传播与交流,从中吸收新的网络元素,扩大自身网络规模,为新农民工在城市的社会资本投资和积累创造条件。

本文的不足之处在于研究的数据年份相对较早,初次外出的新农民工收入水平不高,因此文中并未展示务工者收入差异的绝对数,而是专注于收入差异的相对数。另外,需要指出的是,受篇幅限制,本文没有研究协变量的交互效应及数值型变量的高阶项效应,也没有与其他匹配方法的结果进行对比分析,因而难以体现熵均衡法的优势,这些将有待今后进一步分析与探讨。

参考文献:

1. 边燕杰、张文宏(2001):《经济体制、社会网络与职业流动》,《中国社会科学》,第2期。
2. 费孝通(1997):《江村农民生活及其变迁》,敦煌文艺出版社。
3. 李根强等(2016):《人力资本、社会资本与农民工工资差异》,《华中农业大学学报(社会科学版)》,第2期。
4. 宋月萍、马腾(2015):《同乡会对农民工劳动收入的影响》,《中国人民大学学报》,第2期。
5. 王春超、周先波(2013):《社会资本能影响农民工收入吗?——基于有序响应收入模型的估计和检验》,《管理世界》,第9期。
6. 吴愈晓(2011):《社会关系、初职获得方式与职业流动》,《社会学研究》,第5期。
7. 叶静怡、武玲蔚(2014):《社会资本与进城务工人员工资水平——资源测量与因果识别》,《经济学(季刊)》,第4期。
8. 张春泥、谢宇(2013):《同乡的力量:同乡聚集对农民工工资收入的影响》,《社会》,第1期。
9. 张顺、郭小弦(2011):《社会网络资源及其收入效应研究——基于分位回归模型分析》,《社会》,第1期。
10. 章元等(2008):《社会网络与工资水平——基于农民工样本的实证分析》,《世界经济文汇》,第6期。
11. 章元、陆铭(2009):《社会网络是否有助于提高农民工的工资水平?》,《管理世界》,第3期。
12. 赵延东、王奋宇(2002):《城乡流动人口的经济地位获得及决定因素》,《中国人口科学》,第4期。
13. Bian Y.(1997),Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China. *American Sociological Review*. 62(3):366-385.
14. Diamond A.,Jasjeet S.S.(2013),Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *The Review of Economics and Statistics*. 95(3):932-945.
15. Granovetter M.S.(1973),The Strength of Weak Ties. *American Journal of Sociology*. 78(6):1360-1380.
16. Hainmueller J.(2012),Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*. 20(1):25-46.
17. Kullback S.(1959),Information Theory and Statistics, John Wiley & Sons, Inc.
18. Lin N.(1999):Social Networks and Status Attainment. *Annual Review of Sociology*. 25:467-487.
19. Mouw T.(2003),Social Capital and Finding a Job: Do Contacts Matter?. *American Sociological Review*. 68(6):868-898.
20. Read T.R.C.,Cressie N.A.C.(1988),*Goodness-of-fit Statistics for Discrete Multivariate Data*,Springer-Verlag.
21. Rosenbaum, P.R.,Rubin D.B.(1983),The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 70(1):41-55.

(责任编辑:李玉柱)