

教育和培训对中国城镇劳动力就业的影响

——基于 CLHLS 数据的经验研究

顾和军¹, 刘云平²

(1. 南京信息工程大学 经济与管理学院, 江苏 南京 210044;

2. 南京信息工程大学 信息与控制学院, 江苏 南京 210044)

摘要: 本文使用“中国老年人健康长寿影响因素调查”(CLHLS)数据,分别运用 Probit 和“倾向评分匹配”(PSM)方法测度了教育和职业培训对城镇劳动力就业的影响。研究结果表明,教育水平越高,劳动参与率越高,具有初中或者高中及以上学历的个体,相比于仅有小学学历的个体劳动参与率有较大的提高;培训对于中国城镇劳动力的就业有显著的正面影响,参加培训会将城镇劳动力的劳动参与率提高 11.4%,工作年限与职业培训之间呈现出一种“倒 U 型”变化趋势。

关键词: 城镇劳动力; 就业; 教育; 培训; PSM

中图分类号: F241.4 **文献标识码:** A **文章编号:** 1000-4149(2013)01-0076-07

The Impact of Education and Training on Labor Force Participation in Urban China: The Empirical Study Based on the CLHLS Data

GU He-jun¹, LIU Yun-ping²

(1. School of Economy and Management, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Using the “Chinese Longitudinal Healthy Longevity Survey” (CLHLS) data, this paper analyzes the impact of education and vocational training on the employment of urban workers with probit and PSM model. The result shows that the higher level of the education, the higher employment of workers. Compared with the individual with primary education, the individual who have junior or senior or above education, the probability of individual participation in employment will be increased; Training has a significant positive impact on employment, it will increase the probability of employment of urban labor force by 11.4%. Years of work and occupation training showed an inverted “u” change trend.

收稿日期: 2012-03-23; 修订日期: 2012-06-05

基金项目: 全国教育科学教育部青年专项课题“江苏农村劳动力转移培训有效模式研究”(EJA100434)。

作者简介: 顾和军(1981-),女,南京信息工程大学副教授,管理学博士。研究方向: 农业经济学、性别经济学。

Keywords: urban labor force; employment; education; training; PSM

一、引言

按照经典的人力资本理论,劳动者对于正规教育、培训、迁移和健康的投资最终形成个人的“人力资本”,作为一种人格化的知识和技能,将使拥有者在劳动力市场上获得更高的回报。对于个体来说,个人对教育、培训等的投资会提高其自身的能力和素质,使之成为具有更高生产率的劳动者,同时还具有资源配置能力,可以提高个人对各种信息和机会的分析判断能力,使资源得到更有效的配置^[1],因而教育、培训被认为是决定个人职业地位获得的一个非常重要的因素^[2]。

然而,教育和培训能够提高个体劳动参与率的假说并没有通过普遍的实证检验,经验研究结果因实施对象的不同而存在较大差异:有的研究认为教育和培训能够显著提高农村迁移劳动力选择成为工资收入者的概率^[3];另外的一些研究则认为很多有较高文化程度的下岗人员找工作并不是更容易,或者找到工作后的工资收入并不比文化水平低的人高,人力资本在再就业过程中没有发挥积极作用,在这个过程中,甚至出现了“人力资本失灵”的现象^[4-5]。

培训对于个体就业影响的传统经验研究是基于将培训作为个体就业方程的一个解释变量,应用截面数据通过普通最小二乘法估算培训对就业的影响。但这种估算方法不可避免地会带来偏差^[6],为了得到准确的培训的效果,需要尽可能地将阻碍分析的因素剔除掉,应用社会实验方法是一个有效途径。尽管社会实验方法具有明显的优势,但对数据的要求较高而使其受到限制。近年来,倾向评分匹配法成为经济学界逐渐采用的非实验方法。顾名思义,它分为以下两步,倾向评分和匹配。该方法对于一些没有采用实验方法区分实验组和参照组的数据采用了一种近似于实验的方法,尽可能的产生出随机分组,通过对受试组和参照组特征相似个体的比较来估算公共政策的作用效果^[7]。

上述研究不同程度地丰富了教育和培训对于劳动力就业选择的研究,为我们的研究提供了方法和思路上的借鉴。但现有研究更多地关注农村劳动力特别是农民工,深入研究教育和培训在城镇劳动力就业决定中所起作用的文献寥寥无几。对于这样一个在实践中尚存争议的问题,我们需要更多经验研究的诠释,因此,本文试图利用“中国老年人健康长寿影响因素调查”(CLHLS)数据估算教育和培训对于中国城镇劳动力就业决定的作用。

二、计量模型、估计方法和数据

1. 模型与估计方法

研究教育和培训对就业的影响,传统的方法是采用 Probit 模型,其具体形式如下:

$$P(\text{Occupat} = 1) = F(\gamma D_i + \beta X_i + U_i) \quad (1)$$

其中, P 表示实现就业的概率, $\text{Occupat} = 1$ 在这里表示实现就业, F 是标准正态分布函数, γ 、 β 是待估参数, D_i 是反映教育水平或培训状况的虚拟变量, X_i 是影响就业的其他变量向量, U_i 是期望为零的扰动项。

模型(1)假定个体是否具有初中或高中以上学历以及是否参加培训是外生决定或随机挑选的,而且教育和培训对就业的影响对所有参加者都是相同的,事实上,理性人是否接受更高的教育、是否参加培训存在自选择或被选择的情况,具有较高学历与较低学历群体、参加培训的人群与未参加的人群之间可能存在异质性,亦即受教育前或培训前这两类群体在个人特征方面已存在较为明显的差异,因而直接用模型(1)来估计教育和培训对就业的影响会造成一定程度的偏差。表1汇报了处理组、控制组及样本总体个人特征的统计性描述。

从表1看,处理组和控制组之间确实存在一定差异。处理组的劳动力更年轻,平均年龄比控制组要小4岁多,处理组有更多的男性,同时文化程度更高一些,从受教育程度来看,处理组受过高中及以上教育的比例达到72.4%,而控制组的这一比例仅为41.6%;处理组仅有小学文化的比例占

9.2%，而控制组这一比例为 22.4%，处理组的文化程度明显高于控制组。受访者本人的工作经验和自评健康状况没有明显差异，但从其配偶的受教育程度及自评健康状况来看，处理组要明显优于控制组。此外，处理组有 62.1% 的人处于广东、上海、浙江、江苏这类发达地区，控制组的这一比例为 48.2%，而显然发达地区的就业机会要多于欠发达地区。处理组和控制组的可观测变量的显著差异从一个侧面反映了参加培训人员与随机挑选的大不一样，如果我们忽视这种样本选择性，单纯对参加培训人群与未参加培训人群进行回归分析必然导致有偏估计。

表 1 处理组和控制组个人特征的统计性描述

变量名	处理组	控制组	样本总体
工作	0.828	0.524	0.534
培训	1	0	0.080
年龄	49	53.362	53.274
年龄平方	2447.046	2919.866	2910.808
工作经验	17.528	17.092	17.239
经验平方	409.944	451.410	451.379
性别（男性 = 1）	0.713	0.669	0.666
是否小学（是 = 1）	0.092	0.224	0.202
是否初中（是 = 1）	0.184	0.360	0.328
是否高中及以上（是 = 1）	0.724	0.416	0.417
自评健康（非常好、好 = 1）	0.793	0.771	0.771
配偶自评健康状况（非常好、好 = 1）	0.813	0.748	0.752
配偶教育水平（不识字 = 1，小学 = 2，初中 = 3，高中及以上 = 4）	3.400	2.921	2.912
是否广东、上海、浙江、江苏这类发达地区（是 = 1）	0.621	0.482	0.493

注：参加培训的为处理组，未参加培训的为控制组。为了节约篇幅，在这里没有汇报反映教育程度处理组和控制组个人特征的统计性描述。

数据来源：CLHLS（2005）。

为此，我们采用“倾向评分匹配”（PSM）法，该方法的核心思想是在参加组和非参加组不是随机选择而导致估计结果有偏的情况下，根据影响接受教育或培训的可观察特征找出与处理组最为相似的参加组个体。这样既可以去除不可观测因素带来的内生性问题，又可以控制住一些可观察的异质性因素，消除由于非随机试验中自选择带来的内生性问题。

就教育和培训而言，每个个体 i 存在两种潜在的结果 Y_{1i} 和 Y_{0i} ， Y_{1i} 是个体接受了教育或参加了培训的结果， Y_{0i} 表示没有接受教育或参加培训的结果，这样 $Y_{1i} - Y_{0i}$ 就是处理效应，平均处理效应可以通过下式来表达：

$$ATT = E(Y_{1i} | Z_{enroll_i} = 1) - E(Y_{0i} | Z_{enroll_i} = 0) \quad (2)$$

如果以一系列可观测个体特征 Z 为条件的均值相等，接受教育或培训就可以看作一种随机行为。其中第一项可以通过处理组加以估计，而第二项则通过经配比的对照组估计得出。

这种方法的计算步骤是，首先构建一个培训决策的方程，这个方程的自变量是影响个体教育或培训决策的各种家庭和个体特征；其次，运行 `psctest` 以检验处理组和控制组的“平衡属性”是否具有相同的平均倾向评分；第三，进行匹配，匹配过程中的权重函数采用核配比法（Kernel matching）。该方法对于一些没有采用实验方法区分实验组和参照组的数据采用了一种近似于实验的方法，尽可能的产生出随机分组。在 Stata 软件中倾向分值和匹配分别通过 `pscore` 命令和 `psm` 命令来完成。

2. 数据来源

本文使用 CLHLS 2005 年调查的数据。该调查是目前中国最大的关于老年人口研究的微观面板数据。调查始于 1998 年，覆盖了中国 22 个省市的 631 个县级行政区，具有广泛的全国代表性。调查收集了受访者人口学、社会经济、疾病、健康等翔实的信息，通过了关于准确性、可靠性、一致性、随机性的系统性测试，数据质量良好。

本研究所用数据为其中一个子样本，是 2005 年的子女调查数据，共有 2911 个子女接受采访，剔除农村样本，城市和城镇的样本共 1145 个，由于本文研究的内容是职业培训对城镇劳动力就业的影响，因而剔除职业培训信息缺失的样本，最后的样本数量为 1084 个。其中参加培训的观察值有 87 个，未参加的观察值是 997 个。

三、实证结果与解析

1. Probit、Logit 估算结果

表 2 第 2、3 列报告了教育对城镇劳动力就业影响的 Probit 和 Logit 回归估计结果。在其他因素保持一致的情况下，教育水平越高参与就业的概率越高，反映初中水平和高中水平变量的回归系数分别为 0.1429 和 0.2298、0.4330 和 0.7044，但前者在统计上并不显著，而后者在 1% 的水平上显著，说明接受高中及以上教育会显著提高个体参与就业的概率，但接受初中教育并不能显著提高个体参与就业的概率。反映培训变量的回归系数为 0.3679 和 0.6048，并且在 5% 的水平上显著，但这一结果存在高估，此外，反映性别变量的回归系数为 0.4277 和 0.6964，并且在 1% 的水平上显著，说明男性参与就业的概率要显著高于女性，这也在一定程度上体现我国“男主外，女主内”的社会传统。

在估计回归系数的基础上，我们估计了教育水平对我国城镇劳动力就业影响的边际效应，这是我们更为关注的问题。表 2 的第 4 列报告了我国城镇劳动力就业 Probit 边际估计结果。统计上显著的变量主要有初中文化、高中及以上文化程度、健康水平、性别，在控制其他变量之后，相比于仅有小学及以下文化程度的个体而言，具有初中或者高中及以上学历的个体，参与就业的概率将分别提高 5.52% 和 16.66%，由于没有控制能力的作用，在这里教育的影响也有可能被高估；健康水平越高参与就业的概率越高，“自评健康状况很好”、“好”的个体比“自评健康状况一般”、“差”或“非常差”的个体参与就业的概率高 3.55%；培训的边际影响为 13.71%，这一结果有可能存在高估；此外，男性参与就业的概率要高于女性，边际影响是 16.94%。

如前所述，处理组和控制组是存在差异的两个群体，因而用 Probit 模型估算的结果显然是有偏的，有可能会高估教育和培训的作用，下面将用 PSM 方法估算教育和培训对就业的影响。

2. PSM 估算结果

根据潜在收益的思想，可以通过对微观个体接受教育、培训与没有接受教育、培训进行绩效比较来评价培训政策的作用效果。但微观个体这两种状态是相互排斥的，不可能同时观察到“接受教育、培训”和“未接受教育、培训”时的效果。因此，我们首先计算处理组和控制组观察值的倾向分值，以是否参加教育、培训为因变量，以与其显著相关的个体特征变量为自变量，建立 Probit 模型，通过估计该模型，我们可以获得各变量的倾向分值，表 3 是 Probit 模型的估计结果。

表 2 教育和培训对城镇劳动力就业影响的 Probit、Logit 估计结果

变量	Probit 模型	Logit 模型	边际效应 (Probit)
培训	0.3679 **	0.6048 **	0.1371 *
年龄	-0.1130	-0.1912	-0.0440
年龄平方	0.00092	0.0016	0.00036
经验	0.0055	0.0091	0.0022
经验平方	-0.0001	-0.0001	-0.0000
性别	0.4277 ***	0.6964 ***	0.1694 *
小学文化	—	—	—
初中文化	0.1429	0.2298	0.0552 *
高中及以上文化程度	0.4330 ***	0.7044 ***	0.1666 *
健康水平	0.0909	0.1394	0.0355 *
配偶健康水平	0.0025	0.0099	0.0005
配偶教育水平	0.0489	0.0778	0.0190
是否发达地区	-0.0266	-0.0391	-0.0104
常数项	2.6936	4.5706	
Pseudo R ²	0.0412	0.0413	
观测值数量	584	584	

注：* 指在 10% 水平上显著，** 表示在 5% 水平上显著，*** 表示在 1% 水平上显著。
数据来源：CLHLS (2005)。

从表 3 可以看出，影响城镇劳动力培训决策的因素主要有经验，配偶的教育水平，是否处于上海、浙江、江苏、广东等发达地区。受教育程度对城镇劳动力接受培训决策的影响不确定，并且统计上不显著，相反其配偶的教育水平对其本人有显著的正向影响，配偶教育水平每提高 1 个单位，个体接受培训的概率将提高 35.2%，并且在 5% 的水平上显著，可能的解释是在面临一些家庭决策时，配偶代替个体本人在做决定。

工作年限越长参与培训的概率越高，但达到一定年限后，参与的概率会下降，这一点可以从经验平方项的系数为负看出，一方面可能是因为工作达到一定年限之后技术上已经炉火纯青，另一方面工作到一定年限往上升迁的空间越来越小，个体对于自身的要求也会放松，或者已经进入管理层，技术上的要求反而有所下降，并且一次项和二次项的系数分别在 5% 和 10% 的水平上显著，说明拐点确实存在。此外，上海、浙江、江苏、广东这几个地区的个体参与培训的概率在 5% 的水平上显著高于辽宁、福建、山东、广西地区的个体。发达地区的就业和培训机会更多，与此同时对劳动力的要求也更高，决定了劳动力参加培训的概率更高。

根据处理组和控制组的倾向分值，我们采用核配比法（Kernel matching）进行估计，用以发现处理组与控制组之间就业概率的变化，表 4 为估计结果。对于没有经过配比的样本而言，参加教育和培训会将城镇劳动力就业的概率分别提高 18.67% 和 15.9%，并且 T 值达到 3.68 和 2.50，意味着在统计上显著；经过配比后，接受教育和培训会将城镇劳动力就业的概率分别提高 16.2% 和 11.4%，T 值达到 2.79 和 1.79，意味着在统计上显著，表明接受高中及以上教育和参加培训会显著提高城镇劳动力的就业率。

表 3 倾向分值的 Probit 估计（因变量分别是教育、培训）

变量名	培训		教育	
	系数	z	系数	z
培训			-0.058	-0.39
年龄	0.226	1.47	0.258	2.30
年龄平方	-0.002	-1.52	-0.003	-2.41
经验	0.054	2.25	-0.028	-1.56
经验平方	-0.001	-1.76	0.001	1.27
性别	0.207	1.15	0.274	1.65
是否小学	—	—	—	—
是否初中	-0.215	-1.04	—	—
是否高中	-0.064	-0.34	—	—
自评健康	-0.026	-0.12	-0.105	-0.53
配偶自评健康	0.207	0.96	0.094	0.51
配偶教育水平	0.352	3.83	0.283	3.86
是否发达地区	0.345	2.30	0.165	1.20
常数项	-8.441	-2.24	-6.4667	-2.33

数据来源：CLHLS（2005）。

表 4 接受教育、培训个体就业情况的变化：PSM 估计

方法	样本	处理组（1）	控制组（2）	差分（3） = （1） - （2）	T 值
核配比法 （Kernel matching）	教育	未匹配	0.6626	0.4759	0.1867
		匹配	0.6626	0.5006	0.1620
	培训	未匹配	0.721	0.562	0.159
		匹配	0.721	0.606	0.114

数据来源：CLHLS（2005）。

此外，要保证估计结果的可靠性，必须确保估计式满足平衡特征，为了检验表 4 结论的可靠性，我们针对处理组和控制组进行了平衡性检验^①，表 5 汇报了处理组和控制组平衡性属性的检验结果。其中，第 3 栏表示处理组与控制组在配比前后的样本均值，第 4 栏为配比前后的标准偏误（standardized error，用%表示），即处理组与控制组的样本均值之差与该两组样本方差平均值的平方根之比，第 5 栏表示配比之后标准偏误绝对值所减少的百分比，第 6 栏和第 7 栏分别是处理组与控制组样本均值是否相等检验的 T 值与 P 值。

从表 5 可以看出，所有变量的标准误差绝对值均有不同程度的减少，其中减少幅度最大的是年龄的

① 检验平衡属性使用的是 Stata 中的 pstest 命令。

平方，其偏误减少 96.4%，减少最少的是经验的平方，其偏误仅减少 19%，但无论是年龄、年龄平方、工作经验、工作经验平方、性别、教育水平、健康及配偶教育水平、健康，这些变量匹配后处理组与控制组的差异在统计上都是高度不显著。由此可见，经过 PSM 配比后，处理组控制组样本的个体差异基本得以消除，从而平衡性检验获得通过，说明表 4 的结论是可靠的。

表 5 处理组与控制组的平衡性检验结果

变量	样本	均值		标准偏误 (%)	标准误绝对值 减少 (%)	T 值	P 值
		处理组	对照组				
年龄	未匹配	47.632	49.878	-32.9	96.0	-2.35	0.019
	匹配	47.632	47.722	-1.3		-0.09	0.931
年龄平方	未匹配	2304.500	2544.700	-34.9	96.4	-2.46	0.014
	匹配	2304.500	2313.100	-1.3		-0.09	0.932
经验	未匹配	18.221	17.193	9.0	42.1	0.64	0.522
	匹配	18.221	17.625	5.2		0.34	0.732
经验平方	未匹配	431.690	456.190	-5.2	19.0	-0.35	0.723
	匹配	431.690	411.840	4.2		0.30	0.763
性别	未匹配	0.794	0.777	4.2	72.5	0.32	0.749
	匹配	0.794	0.799	-1.1		-0.07	0.946
是否小学	未匹配	0.191	0.207	-4.0	83.8	-0.31	0.760
	匹配	0.191	0.194	-0.6		-0.04	0.970
是否初中	未匹配	0.265	0.321	-12.3	90.2	-0.93	0.351
	匹配	0.265	0.259	1.2		0.07	0.942
是否高中	未匹配	0.500	0.422	15.6	85.6	1.21	0.226
	匹配	0.500	0.511	-2.2		-0.13	0.897
自评健康	未匹配	0.868	0.831	10.3	96.1	0.77	0.442
	匹配	0.868	0.869	-0.4		-0.02	0.980
配偶自评健康	未匹配	0.868	0.783	22.4	87.36	1.62	0.106
	匹配	0.868	0.857	2.8		0.18	0.860
配偶教育水平	未匹配	3.485	2.912	64.4	91.5	4.54	0.000
	匹配	3.485	3.436	5.5		0.37	0.712
是否发达地区	未匹配	0.647	0.482	33.6	96.9	2.56	0.011
	匹配	0.647	0.652	-1.0		-0.06	0.950

注：教育的处理组与控制组同样通过了平衡性检验，为了节约篇幅，本文在这里没有汇报相关结果。
数据来源：CLHLS（2005）。

四、基本结论及政策含义

本文利用 CLHLS 数据分别运用 Probit 和 PSM 模型估算了教育、培训对于中国城镇劳动力就业的影响，研究结果表明：一方面，教育水平越高，个体的劳动参与率越高，接受高中及以上教育会显著提高个体的劳动参与率。在估计回归系数的基础上，我们估计了教育水平对我国城镇劳动力就业影响的边际效应，在控制其他变量之后，相比于仅有小学及以下文化程度的个体而言，具有初中或者高中及以上学历的个体，劳动参与的概率将分别提高 5.52% 和 16.66%，PSM 方法估算的结果甚至更高，高中文化程度将会使劳动参与率提高 18.67%。

另一方面，培训有利于提高我国城镇劳动力劳动参与的概率。基于 Probit 模型边际效应的分析结果表明，培训对就业的边际影响为 13.71%，但由于处理组和控制组的可观测变量的显著差异从一个侧面反映了参加培训人员与随机挑选的大不一样，如果我们忽视这种样本选择性，单纯对参加培训人群与未参加培训人群进行回归分析必然导致有偏估计；为了控制处理组与控制组的差异，本文使用了 PSM 模型来估计培训对就业的影响，经过配比后，参加培训会使城镇劳动力的劳动参与率提高 11.4%。

基于以上结论，我们认为在促进中国城镇劳动力就业的政策选择中，应着重从以下两个方面

入手。

一是要加强市场研究，实现培训供需对接。与政府的不遗余力和热情形成对比的是，城镇劳动力参与培训的积极性并不高，在我们的样本中仅有 8.03% 的城镇劳动力参与了培训，这种情况在其他地区同样存在，因此，今后要加大对城镇劳动力特别是下岗职工培训市场需求的研究，推进符合现代工业和城市建设发展趋势、适应当地工业发展、城镇劳动力喜欢、符合劳动力需要的培训项目，并以此确定培训的对象、内容和渠道等。

二是要对城镇劳动力进行有针对性的培训。针对城镇劳动力人力资本不断贬值的现状，有效的解决方式是对其进行有针对性的职业培训，加快其知识更新，使他们能快速适应新的就业市场需求，在这一过程中，政府应积极提供高质量、易于获得的职业培训，使城镇劳动力的人力资本可以得到及时更新。对已有的培训项目进行科学评估，加强培训效果的评估，以掌握哪些人群最需要培训，并且最能够获益。

本研究的重要意义和启示在于，个体的教育水平往往很难改变，但可以通过后天的职业技能培训促进个体的就业。只是对于哪类技能培训更有利于促进个体就业，以及政府是否应该干预，如何干预，干预的效率问题，我们仍知之甚少，这也是下一步研究的重要方向。

参考文献:

- [1] 赵延东. 人力资本、再就业与劳动力市场建设 [J]. 中国人口科学, 2003, (5) .
- [2] 丁煜. 下岗失业人员的再就业培训: 效用与局限性——从人力资本理论的分析视角 [J]. 市场与人口分析, 2005, (6) .
- [3] 王德文, 蔡昉, 张国庆. 农村迁移劳动力就业与工资决定: 教育与培训的重要性 [J]. 经济学 (季刊), 2008, (4) .
- [4] 同 [2] .
- [5] 同 [3] .
- [6] Meyer B. Natural and Quasi-Experiments in Economics [J]. Journal of Business and Economics Statistics, 1995, (13) .
- [7] Dehejia R. S. Wahba. Causal Effects in Nonexperimental Studies Reevaluating the Evaluation of Training Programs [J]. American Statistical Association, 1999, 94.

[责任编辑 方 志]