

我国教育收益率性别、地区差异几何?

李俊璋, 蒲文柔

(西南大学 经济管理学院, 重庆, 400700)

[摘要] 基于明瑟方程测度了全国居民教育收益率, 同时利用分位数回归比较了教育收益率的性别和地区差异。在纠正了样本选择偏差和教育变量内生性后, 教育对全国居民收入有一定的促进作用; 分位数回归发现绝大多数分位数上女性收益率更高, 沿海教育收益率在中高、高分位数上高于内陆, 教育收益率在性别、地区上存在显著差异。应重视女性教育问题, 给予女性一定政策倾斜, 缩小男女性教育收益差距; 同时应加大对内陆的资源政策扶助和基础交通建设, 促进区域协同均衡发展。

[关键词] 教育收益率; 工具变量; 样本选择; 分位数回归

[中图分类号] F240 **[文献标识码]** A **[文章编号]** 2096-3300 (2019) 04-0069-08

2019年政府工作报告中, 李克强总理提出要发展更加公平更有质量的教育, 要推进高中阶段教育的普及以及一流大学一流学科的建设。对国家来说, 教育是发展、进步之基, 而对个人来说, 教育则能使收入更高、生活更好。但柏拉图持有教育无用论的观点。筛选假设理论认为, 人接受教育的程度越高, 拥有的文凭越高, 越有可能在劳动力市场获得高收入; 工作竞争理论指出, 拥有较高教育程度的人在劳动市场上更有优势; 在经济学理论中, 人都是理性的, 为了追求利益最大化会尽可能多地接受教育, 使自己进入劳动力市场时有更多的筹码, 以获得最大收入。这些理论观点都肯定教育的作用。

《中国民生发展报告 2017》揭示出目前我国教育程度与个体收入不对等的现状, 同时沿海内陆之间、不同性别之间的教育回报率存在差距, 配偶的受教育程度也会对个体教育收益造成部分影响。那么, 教育对个人是否重要? 地区、性别之间教育的投资收益率真的存在差距么? 本文对我国教育收益

率进行测算并且对地区、性别之间的教育收益率进行比较分析。

一、文献综述

已有的关于教育收益率的文献大体可以分为以下几类: 一是整体教育收益率的测算。如诸建芳等^[1]估计出国有、集体、三资企业全体职工样本基础教育收益率为 1.8%, 专业教育收益率为 3.0%; 侯风云^[2]估计出全样本农村教育的收益率为 3.655%; 罗楚亮^[3]对城镇居民教育收益率进行了研究, 根据控制了性别与党员身份的扩展明瑟模型得到的教育收益率为 6.76%。二是研究教育收益率的差异, 比如性别差异、城乡差异、区域差异、基础教育与职业教育差异。张永丽等^[4]研究了贫困地区农村男、女性劳动力在不同阶段的教育回报, 发现教育收益率存在显著性别差异; 闫琦等^[5]通过对相关的文献进行整理分析观察到教育、教育扩展对女性的影响要高于男性, 并且提升女性的受教育水平

收稿日期: 2019-04-10

作者简介: 李俊璋 (1995-), 男, 重庆永川人, 硕士研究生, 研究方向: 劳动经济学;

蒲文柔 (1995-), 女, 四川遂宁人, 硕士研究生, 研究方向: 金融学。

有利于缩小性别工资差异; 张抗私等^[6]基于性别、院校视角对大学生毕业工资方程进行了估计, 对性别差异进行分解, 得到性别歧视非常严重的结论; 张琦等^{[7]73-75}基于大学生就业地区的选择分析, 探讨了我国不同地区教育收益率, 认为劳动力市场化地区差异、经济发展水平导致了教育收益率的差距; 杨汶鑫等^[8]运用分位数回归方法研究地域分布对不同收入水平的个人经济收入的影响, 注意到从低收入人群到高收入人群, 教育收益率不断增大, 而地域差异对个人收入的影响显著增强; 鞠高升等^{[9]40-45}对中国教育的明瑟收益率的男女差异进行了检验, 发现女性教育收益率高于男性; 颜敏^{[10]37-49}估算了农村高中教育溢价并对比了职业高中和普通高中收益率的差别; 岳昌君^{[11]135-150}则对城镇职工的收入影响因素进行了回归分析, 发现教育并不是影响个人收入的唯一因素, 行业差异和地区差异等因素同样会对个人收入有显著影响。三是关于教育收益率变动的研究。郭小弦等^[12]研究了新中国成立、“文革”时期、改革初期、深入改革时期四个历史阶段的教育回报率变动趋势, 得出了我国城市居民教育回报率变动趋势为先下降后上升的结论; 李实等^[13]对1990-1999年期间城镇个人收益率的动态变化进行了经验估计, 发现个人教育收益率是逐年上升的。

现有研究在方法以及视角方面仍有一些不足之处。从方法上来说, 学者们多使用明瑟模型, 采用OLS最小二乘法研究教育收益率, 但教育变量的内生性会导致估计结果的偏误; 同时, 由于居民收入大多呈右偏分布, 使用最小二乘法可能会导致估计结果比较片面、不准确, 不能反映条件分布的全貌, 也会使估计结果产生偏误进而影响最终结论。从研究视角来看, 大多数学者在进行地区差异性研究时, 从整体上进行比较, 较少对各样本内部的差异进行细致、全面的对比。本文利用CGSS2012-2015(不含CGSS2014)共3年数据, 基于明瑟方程, 采用heckman两步法以及工具变量法来测算我国居民教育收益率, 同时纠正OLS可能存在的样本选择偏差以及缓解教育变量内生性问题, 使测算出的教育收益率数值更为准确; 然后采用分位数回归的方法对教育收益率的性别、地区(沿海内陆的对比)差异进行更为细致的研究。

二、数据以及统计性描述

(一) 模型介绍

1. 明瑟模型。采用大多数学者使用的明瑟模型来研究教育收益率以及研究教育收益率的性别、地区差异。为了达到研究教育异质性的目的, 在原始的明瑟模型中加入性别等差异因素, 得到扩展的明瑟模型:

$$\ln y = \alpha_1 + \alpha_2 \text{edu} + \alpha_3 \text{ex} + \alpha_4 \text{ex}^2 + \sum \psi_i c_i + \varepsilon \quad (1)$$

式中, c 为控制变量, 包含性别变量、民族、沿海地区虚拟变量 sea ^①。被解释变量 $\ln y$ 为研究对象职业性月收入的对数, edu 、 ex 、 ex^2 依次为受教育年限、工作年限及其平方。更进一步, 为了控制不同教育质量所带来的教育收益率的差异, 在模型(1)的基础上将教育变量换成 D_1 初中、 D_2 高中、 D_3 大专及以上学历虚拟变量(不设置小学虚拟变量的原因是为了避免多重共线性)得到模型(2):

$$\ln y = \alpha_1 + \alpha_2 D_1 \text{edu} + \alpha_3 D_2 \text{edu} + \alpha_4 D_3 \text{edu} + \alpha_5 \text{ex} + \alpha_6 \text{ex}^2 + \sum \psi_i c_i + \varepsilon \quad (2)$$

2. heckman 样本选择模型。明瑟模型为线性模型, 由于本文仅仅采用了三年的数据, 这些数据远远不能覆盖全国的劳动者, 因此模型中的被解释变量 $\ln y$ 会出现断尾的问题, 如果采用大多数学者所使用的OLS法对明瑟模型进行回归, 会导致估计的结果不满足一致性。对工资性收入这个被解释变量来说, 只有个体选择参加劳动才可以获得劳动性收入即工资, 所以决定劳动者是否参加劳动以获取收入的变量也与被解释变量 $\ln y$ 有关, 此时就产生了“样本选择”的问题, 而“样本选择”将会导致“选择性偏差”。heckman^[14]提出了“两步估计法”来纠正选择性偏差。结合本文所采用的数据库自身的特点, 对模型设定如下:

$$\ln y = \begin{cases} \text{可观测, 若 } z_i = 1 \\ \text{不可观测, 若 } z_i = 0 \end{cases} \quad z_i = \begin{cases} 1, & \text{若 } z_i^* > 0 \\ 0, & \text{若 } z_i^* \leq 0 \end{cases} \quad z_i^* = w_i' \delta + \mu_i \quad (3)$$

其中, z_i 为个体是否参与劳动的二元变量, 取值为1或0; 若居民参与劳动, 有工资性收入, 则为1; 居民不参与劳动, 无工资性收入则为0。 z_i^* 为不

可观测的潜变量。假设 u_i 服从正态分布,则 z_i 为 probit 模型,所以 $P(z_i = 1 | w_i) = \Phi(w_i' \lambda)$, 然后对观测样本取其期望为 $E(\ln y | \ln y \text{ 可观测}) = x_i' \alpha + \rho \sigma_\varepsilon \lambda(-w_i' \delta)$, 此时,由于存在非线性项,直接使用 OLS 将导致非线性项的遗漏,明瑟方程里的自变量 x_i 与影响潜变量 z_i^* 的 w_i 存在相关关系,将导致 OLS 估计不具有一致性。然后对明瑟方程里的某个解释变量如教育变量 edu 取边际效应可以得到:

$$\frac{\partial E(\ln y | z_i^* > 0)}{\partial \text{edu}} = \alpha_2 + \rho \sigma_\varepsilon \frac{\partial \lambda(-w_i' \delta)}{\partial \text{edu}} \quad (4)$$

显然,如果想要得到准确的参数估计,就需要知道 $\lambda(-w_i' \delta)$, 而要知道 $\lambda(-w_i' \delta)$, 只需要知道 δ 。知道了 δ 后将 $\lambda(-w_i' \delta)$ 代入原回归方程 (1) 进行最小二乘回归就能得到一致的估计。

3. 分位数回归模型。OLS 最小二乘法是一种利用条件期望 $E(y|x)$ 来反应条件分布 $y|x$ 集中趋势的均值回归。当条件分布 $y|x$ 为不对称分布时,条件期望 $E(y|x)$ 不能很好地反映条件分布的全貌。同时,由于 OLS 利用的是残差平方和来最小化目标函数,容易受到极端值的影响,故利用最小二乘法对教育收益率进行估计得到的结果并不准确。而 Koenker and Bassett^[15] 提出的 QR, 利用残差绝对值得加权平均为最小化的目标函数,较为稳健,不易受极端值影响且可以利用条件分布的若干重要分位数,能对被解释变量 $\ln y$ 的条件分布有更充分的认识。

对于线性回归模型,假设条件分布 $y|x$ 的总体 q 分位数 $y_q(x)$ 为 x 的线性函数,可以写成:

$$y_q(x_i) = x_i' \alpha_q \quad (5)$$

上式中, α_q 为“ q 分位数回归系数”,它的估计量 $\hat{\alpha}_q$ 可以通过下式来得到:

$$\min_{\alpha_q} \sum_{i: y_i \geq x_i' \alpha_q} q |y_i - x_i' \alpha_q| + \sum_{i: y_i < x_i' \alpha_q} (1-q) |y_i - x_i' \alpha_q| \quad (6)$$

式子 (2) 与式子 (5) 结合,得到式子 (7):

$$q^p(y) = \alpha_1^{(f)} + \alpha_2^{(f)} \text{edu} + \alpha_3^{(f)} \text{ex} + \alpha_4^{(f)} \text{ex}^2 + \sum \psi_i^{(f)} c_i + \varepsilon \quad (7)$$

式 (7) 中, p 为居民收入的分位点,收入 y 的变化使得 (7) 中的解释变量发生变动。

(二) 数据与变量说明

1. 数据来源。使用中国人民大学收集的中国综合社会调查 (CGSS) 三年的数据库,分别是 CGSS2012、CGSS2013、CGSS2015。经过筛选后得到沿海、内陆地区共 3 个直辖市、24 个非直辖市容量为 9 518 个研究对象的混合截面数据。其中,男性、女性个体数分别为 5 294、4 224 个;沿海、内陆地区个体数分别为 4 062、5 456 个。本文研究的目标群体是接受过小学及以上教育且教育状态为毕业的人群。对于不同年份数据的处理,采用了控制年份的办法来控制年份差异,得到更为科学、严谨的结果。

2. 变量选取与处理。自变量和因变量的选择参照孙婉琳^[16] 的研究。模型中所有变量选取与描述如表 1 所示。因变量选择工资性月收入 y , 用劳动者上一年工资性总收入除以 12 得到,为了增加有效样本数量,对每个个体的月收入 y 都加了 1, 从而使其能够取到自然对数,起到增加样本数量的作用; 由于测算的收入为工资性收入,结合我国相关退休政策,剔除工龄在 42 年以上的个体 (劳动者一般 18 岁进入劳动市场,60 岁退休,工龄最高只有 42 年); 只选取了小学及以上学历的个体,所以将教育阶段分为小学、初中、高中、大专及以上四个阶段,这四个阶段的教育变量 edu 分别赋值 6、9、12、16; 同时,为了探求不同教育阶段教育收益率的具体差别同时为了避免模型存在的多重共线性问题,还将教育变量 edu 删去加入了三个虚拟变量 $D1$ 、 $D2$ 、 $D3$ (名称及描述见表 1) 以表示四种不同的受教育程度,若个体受教育程度为小学,则三个虚拟变量都取 0,若个体受教育程度为初中,则初中教育虚拟变量 $D1$ 取 1,其余两个虚拟变量取 0,高中、大专及以上受教育程度的取值方式同初中。其余变量见表 1 中变量的描述。

三、实证结果与数据分析

(一) 全国教育收益率的估计

本文所有的实证结果均由 stata12.0 完成。表 2 报告了模型 1、模型 2 分别采用 OLS 最小二乘法、heckman 两步法、工具变量 heckman 两步法的回归结果。

表1 变量的设置与描述

Tab.1 Setting and description of variables

变量名称	变量描述	变量名称	变量描述
月收入工资的对数 $\ln y$	月收入工资的自然对数	配偶政治面貌 spolitics	配偶为党员 = 1; 配偶非党员 = 0
工作年限 ex	年龄减去教育年限再减去 6	家庭总人数 tfnumber	家庭所有成员的个数 / 个
工龄的平方 ex^2	年龄减去教育年限再减去 6 然后 再平方	配偶总收入 sincome	上一年配偶总收入 / 元
受教育年限 edu	根据不同受教育程度分别赋值	家庭是否拥有汽车 car	拥有 = 1; 无 = 0
性别 gender	男性 = 1; 女性 = 0	初中虚拟变量 $D1$	初中 = 1; 非初中 = 0
沿海 sea	个体处在沿海 = 1; 其他 = 0	高中虚拟变量 $D2$	高中 = 1; 非高中 = 0
住房套内面积 housearea	全家住房套内面积 / m^2	大专及以上学历虚拟变量 $D3$	大专及以上学历 = 1; 非大专及以上学历 = 0
住房数量 housenumber	全家拥有房产的数量 / 个	是否拥有工资性收入 z	有 = 1; 无 = 0

模型 1 最小二乘法 OLS 的估计结果中, edu 教育变量系数为 0.075 且非常显著, 也就是每增加一年的教育, 能够使得收入增加 7.5%, 而从 heckman 两步法第一步 probit 回归的结果可以看出, 影响劳动供给的因素有很多, 如配偶的政治面貌、家庭成员总数等。模型 1 中 heckman 两步法的第二步回归中 λ 值是显著的, 表明样本选择偏差的确存在。变量 edu 的估计系数变为 9.1%, 同样在 1% 水平上显著。

模型 2 探求不同教育质量所带来教育收益的差异。最小二乘法回归结果中三个虚拟变量的值分别为 0.138、0.347、0.729, 且均在 1% 水平上显著, 可以看出, 与仅接受了小学教育的个体相比, 接受初中、高中、大学教育的劳动者收入更高, 且个体收入随着劳动者教育层次的提升而增加, 增加幅度越来越大。根据模型 2 的 probit 回归结果, 性别、工龄等个体特征以及家庭成员、配偶收入都会显著影响个体劳动的参与。样本选择模型的第二步回归结果中三个教育虚拟变量的值均在 1% 水平上显著, 其值分别为 0.139、0.375、0.872。虽然各参数与 OLS 存在差异, 但整体趋势是完全相同的, 这再次证明了 OLS 所得结果的稳健性。

颜敏^{[10]40}认为明瑟方程中教育变量存在内生性, 这会带来估计结果有偏且非一致。为了克服教育变量内生性带来的问题, 采用了工具变量法。郭东梅

等^[17]已经证明个体配偶的受教育年限是内生变量受教育程度的强工具变量。因此本文选择配偶受教育程度作为劳动者受教育变量的工具变量, 同时采用两步最小二乘法来试图克服内生性所带来的问题。在使用了配偶受教育年数作为工具变量后, 模型 1、2 中的 λ 值依然显著, 再次说明样本存在选择偏误。模型 1 的教育变量的系数为 0.167。模型 2 各教育虚拟变量也发生了改变, 分别为 0.617、0.780、1.779, 且均在 1% 的显著性水平上显著。

(二) 教育收益率性别、地区差异性研究

如图 1 所示, 从月收入变量 y 的正态密度函数和核密度函数图可以明显看出收入分布呈现右偏的趋势, 因此, 使用不受分布影响的分位数回归 QR 来探讨教育收益率的差异是十分合理的。拟建立以下两个模型来对异质性进行研究:

$$\ln y = \alpha_1 + \alpha_2 \text{edu}_i + \alpha_3 \text{ex} + \alpha_4 \text{ex}^2 + \psi_1 \text{gender} + \psi_2 \text{nation} + \nu_1 \text{gender} * \text{edu}_i + \varepsilon_i \quad (8)$$

$$\ln y = \alpha_1 + \alpha_2 \text{edu}_i + \alpha_3 \text{ex} + \alpha_4 \text{ex}^2 + \psi_1 \text{gender} + \psi_2 \text{nation} + \psi_3 \text{sea} + \nu_1 \text{sea} * \text{edu}_i + \varepsilon_i \quad (9)$$

在以上两个模型中, 分别引入了需要研究的控制变量与教育变量的乘积, 当对这些模型做回归时, 最后得到的 edu 的系数 α_2 为基准变量的系数, 乘积项的系数则为非基准变量与基准变量的差额。回归结果如表 3 所示 (均采用 bootstrap 自助法经 100 次抽样得到)。

表2 样本选择模型、工具变量样本选择模型回归结果

Tab. 2 Regression results of sample selection model and instrumental variable sample selection model

变量	模型1 OLS	模型1 样本选择		工具变量	模型2 OLS	模型2 样本选择		工具变量
		第一阶段	第二阶段			第一阶段	第二阶段	
gender	0.452*** (27.700)	0.562*** (6.877)	0.564*** (18.735)	0.826*** (20.503)	0.458*** (28.023)	0.576*** (6.967)	0.578*** (20.217)	0.789*** (18.308)
nation	0.295*** (9.559)	-0.060 (-0.434)	0.271*** (7.980)	0.229*** (6.532)	0.299*** (9.723)	-0.043 (-0.309)	0.277*** (8.168)	0.235*** (6.607)
ex	0.028*** (7.217)	0.059*** (3.646)	0.044*** (8.917)	0.087*** (13.176)	0.030*** (7.724)	0.071*** (4.230)	0.047*** (9.773)	0.086*** (12.098)
ex ²	-0.001*** (-11.336)	-0.001*** (-4.213)	-0.001*** (-11.973)	-0.002*** (-15.415)	-0.001*** (-11.756)	-0.002*** (-4.727)	-0.001*** (-12.828)	-0.002*** (-13.803)
edu	0.075*** (25.219)	0.056*** (4.025)	0.091*** (23.322)	0.167*** (19.587)	-	-	-	-
sea	0.416*** (24.426)	0.053 (0.678)	0.430*** (23.517)	0.399*** (21.022)	0.418*** (24.542)	0.060 (0.764)	0.433*** (23.733)	0.396*** (19.914)
spolitical	-	0.222* (1.764)	-	-	-	0.210 (1.631)	-	-
familynumber	-	-0.056** (-2.056)	-	-	-	-0.052* (-1.890)	-	-
housearea	-	-0.000 (-0.769)	-	-	-	-0.000 (-0.798)	-	-
housenumber	-	0.097 (1.418)	-	-	-	0.104 (1.509)	-	-
sincome_r	-	-0.000004** (-2.235)	-	-	-	-0.000005** (-2.532)	-	-
car	-	-0.0003 (-0.003)	-	-	-	-0.018 (-0.190)	-	-
lambda	-	-	2.916*** (6.312)	8.360*** (11.654)	-	-	3.054*** (7.210)	7.641*** (10.514)
D1	-	-	-	-	0.138*** (5.494)	-0.022 (-0.225)	0.139*** (4.878)	0.617*** (4.023)
D2	-	-	-	-	0.347*** (12.564)	0.069 (0.616)	0.375*** (12.222)	0.780*** (7.821)
D3	-	-	-	-	0.729*** (22.99)	0.734*** (4.314)	0.872*** (22.314)	1.779*** (13.55)
_cons	5.580*** (83.914)	1.001*** (3.126)	5.072*** (46.545)	3.462*** (17.784)	6.062*** (102.924)	1.297*** (4.513)	5.657*** (65.243)	4.470*** (26.117)
N	9 364	7 930	7 776	7 776	9 364	7 930	7 776	7 776
R ²	0.363		0.368	0.337	0.364		0.370	0.322

注: ***, **, * 分别表示在 1%、5% 和 10% 的水平上显著, 括号内为 t 值; 年份虚拟变量限于篇幅未展示, 但年份已经得到控制; heckman 两步法以及 IV-heckman 法的第一步相同, 均是对被解释变量 z 进行 probit 回归, 工具变量那一列是 IV-heckman 法第二步 2sls 的回归结果。

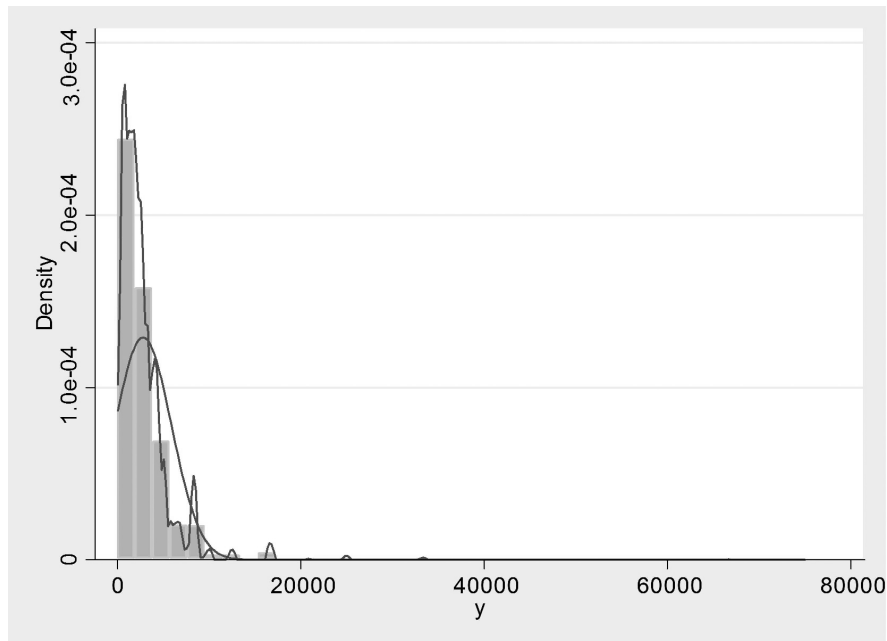


图1 月收入直方图、核密度函数、正态密度函数图

Fig. 1 Monthly income Y distribution histogram , kernel density function and normal density function

表3 教育回报率差异回归表

Tab. 3 Regression of differences in educational return rate

变量	QR05	QR25	QR50	QR75	QR95
edu	0.162*** (0.0118)	0.138*** (0.00436)	0.112*** (0.00445)	0.0900*** (0.00455)	0.0761*** (0.00891)
edugender	-0.0591*** (0.0146)	-0.0533*** (0.00538)	-0.0441*** (0.00551)	-0.0305*** (0.00571)	-0.0151 (0.0118)
edu	0.155*** (0.00838)	0.122*** (0.00404)	0.0870*** (0.00435)	0.0684*** (0.00424)	0.0471*** (0.00806)
seaedu	0.00716 (0.0130)	-0.00893 (0.00588)	0.00982 (0.00617)	0.0233*** (0.00601)	0.0512*** (0.0120)

注: ***, **, * 分别表示在 1%、5% 和 10% 的水平上显著, 括号内为 t 值。

1. 性别差异分析。从表 3 可以看出, 除了在 95% 的分位数上, 其他分位数上女性教育收益率均大于男性, 即除了在高收入人群中男女性教育收益率不存在显著差异外, 其他收入水平上女性教育收益率均高于男性。且随着分位数的提高, 教育收益率在性别上的差异逐渐缩小。造成在绝大多数分位点上女性教育收益率较高的原因可以用刘泽云^[18]的观点来解释: 职工的教育水平与其面临的工资性别歧视程度呈反向关系, 即低教育水平职工中的工资性别歧视程度更高, 结果导致不同教育水平的女性

之间的工资差异比男性更大, 从而女性的教育收益率更高。因此女性每多接受一年教育, 所得到的收益会大于男性。其次是因为样本有失代表性 (由于部分女性对配偶的依赖性), 所以导致调查中进入劳动力市场的女性学历都较高^[19], 从而导致女性教育收益率在大多数分位数上比男性高。

2. 内陆、沿海差异分析。参照岳昌君^{[11]142}对于沿海内陆的划分标准, 将沿海定义为东部地区与京津沪地区, 内陆则定义为中部与西部地区。表 3 中所示, 在中、中低、低分位数上, 两者教育收益率

差距并不显著;在中高、高分位数上,沿海与内陆的差距分别为2.33%、5.12%。也可以说在中高、高收入群体,内陆地区教育收益率均低于沿海,且分位数越往右端走,内陆、沿海差距绝对值越大,在95%的分位数上差距达到最大。可以从以下两个方面来解释这种现象。一方面,由于沿海地区相比内陆具有极佳的地理优势,因此在招商引资、对外开放方面有巨大优势。由于沿海地区人数众多,劳动力数量巨大,同时改革开放初期,我国工人工资相比国外较低,所以大量外商将工厂建在沿海地区,这也使得沿海地区产业结构更加丰富、更高级,而一定的产业结构需要对应一定的资源配置结构,劳动力这种重要的人力资源与产业结构相适应,形成了一定的就业结构。产业结构的变化使各产业劳动力配置发生相应变化,体现在劳动力职业变化上。沿海地区产业升级与技术创新加快会促进职业的更新与升级,这对工人的技术以及受教育水平提出了更高的要求。只有掌握了更多知识和技能的劳动者才能胜任新的岗位和工作,当受过更多教育的人才出现在工作岗位上,就能得到更高的收入。更高的收入吸引更多的中高、高端人才聚集在沿海地区,而由于人才的聚集,反过来又促进了沿海地区的进一步发展。另一方面,随着改革开放,中国经济和社会发展日新月异。为了适应日益增长的GDP,工人的工资得到极大幅度的提高,而工资对企业来说是成本,成本的提高会使企业家利润减少,使得企业家将工厂迁离沿海地区进入内陆,但由于大多数中、高端人才都在沿海地区,因此迁入内陆的大多是技术水平不太高的产业;虽然沿海地区就业机会很多,但是聚集的人才更多,导致无论对于企业还是个人来说,竞争压力过大,一部分技术知识水平不是那么高的劳动者以及竞争力不那么强的企业就倾向于往压力更小的内陆迁移。随着沿海地区中、低端技术性人才以及工厂往内陆迁移,最终导致中低收入人群教育收益率在内陆大致与沿海相等的结果。正如张琦^{[7]75}所说“经济发达地区和城市因拥有相对成熟的劳动力市场和较高的发展水平,有利于实现更高的教育收益率”。这可能就是沿海、内陆在低、中低、中与中高、高分位数差异迥然不同的原因。

四、结论以及政策建议

本文研究结果表明,教育对个人收入的确有一定的促进作用,且更高的受教育水平带来更高的收入。男女性教育收益率存在显著差异且女性教育收益率在绝大多数分位数上均大于男性(除了95百分位)。随着分位数的提高,男女差异正在逐渐缩小,到了高分位数,两者已不存在显著差距。这可能为解决男女工资性别歧视问题提供了一个新的解决办法。沿海、内陆地区在中分位数及以下不存在显著差别,而在中分位数上沿海地区教育收益率显著高于内地。进一步还观察到,两者差距随着分位数增加而变大。个体能力、生产率、劳动市场发育程度异质性造成沿海内陆教育收益率的差距。

政府应通过各种渠道大力宣传教育的作用,帮助民众树立科学理性认知,深刻认识教育的重要性,从而起到重视教育的作用。

消除工资性别歧视,鼓励女性接受教育。首先,应呼吁社会重视女性的教育情况,营造重视女性教育的社会环境;其次,政府应出台相关政策,制定相关法规,切实保障女性群体的就学率;最后,还应当鼓励女性接受更高层次的教育并给予女性一定程度的政策倾斜,从而减少性别歧视。

缩小沿海内陆教育差距,实现沿海内陆教育均衡发展。要促进教育均衡发展,最主要的是提高内陆的经济水平,扩大内陆对外交流。国家修建的贯穿沿海内陆的高铁对于沿海内陆差距的缩小有极大的作用。在此基础上,政府还应加大基础设施建设,打造内陆与沿海以及国外多维、高效的连接方式;同时,正如习近平总书记在重庆考察时提出的新一轮西部大开发政策所言,国家应加大对内陆的政策、资源倾斜以及高新技术的扶持,使内陆留得优秀的劳动者,从而缩小内陆沿海在高端人才上的差距,实现海陆经济、教育并驾齐驱共同发展。

注释:

①之所以不添加城乡以及职业与基础教育虚拟变量是因为城乡、职业基础教育间的差异并不是本文研究的重点;不加入个人能力对收入的影响比如IQ等变量作为控制变量是因为CGSS数据中没有可以反映个人能力的变量。同时,收入受各种可观测与不可观测因素的影响,想要控制住所有影响劳动者收入的变量是几乎不可能的事情。

参考文献:

- [1] 诸建芳,王伯庆,恩斯特·使君多福. 中国人力资本投资的个人收益率研究[J]. 经济研究,1995(12):55-63.
- [2] 侯风云. 中国农村人力资本收益率研究[J]. 经济研究,2004(12):75-84.
- [3] 罗楚亮. 城镇居民教育收益率及其分布特征[J]. 经济研究,2007(6):119-130.
- [4] 张永丽,李青原,郭世慧. 贫困地区农村教育收益率的性别差异——基于PSM模型的计量分析[J]. 中国农村经济,2018(9):110-130.
- [5] 闫琦,孟大虎,孙永强. 教育、教育扩展与性别工资差异——一个文献综述[J]. 教育经济评论,2017,2(6):115-128.
- [6] 张抗私,刘翠花. 大学毕业生性别工资差异的实证研究[J]. 经济与管理研究,2017,38(9):84-94.
- [7] 张琦,吴克明. 大学生就业地区选择分析——基于教育收益的视角[J]. 中国高校科技,2017(Z1):73-75.
- [8] 杨汶鑫,王小舟. 中国国民的收入分布与教育收益率探析[J]. 湖南人文科技学院学报,2017,34(5):115-119.
- [9] 陈良焜,鞠高升. 教育明瑟收益率性别差异的实证分析[J]. 北京大学教育评论,2004(3):40-45.
- [10] 颜敏. 技能高中还是普通高中?——中国农村学生的教育选择[J]. 中国农村经济,2012(9):37-49.
- [11] 岳昌君. 教育对个人收入差异的影响[J]. 经济学(季刊),2004(S1):135-150.
- [12] 郭小弦,张顺. 中国城市居民教育收益率的变动趋势及其收入分配效应——基于分位回归模型的分析[J]. 复旦教育论坛,2014,12(3):51-56.
- [13] 李实,丁赛. 中国城镇教育收益率的长期变动趋势[J]. 中国社会科学,2003(6):58-72,206.
- [14] HECKMAN J J. Sample selection bias as a specification error[J]. Econometrica,1949(47):153-162.
- [15] KOENKR R, BASSETT G. Regression quantiles [J]. Econometrica,1978,46(1):33-50.
- [16] 孙婉琳. 城乡居民普通教育和职业教育收益率的实证研究[D]. 昆明:云南财经大学,2018.
- [17] 郭冬梅,胡毅,林建浩. 我国正规就业者的教育收益率[J]. 统计研究,2014,31(8):19-23.
- [18] 刘泽云. 女性教育收益率为何高于男性?——基于工资性别歧视的分析[J]. 经济科学,2008(2):119-128.
- [19] 王佳琪. 我国教育收益率测算[D]. 广州:暨南大学,2016.

Gender and Regional Differences of Educational Return Rate in China

LI Junzhang, PU Wenrou

(School of Economics and Management, Southwest University, Chongqing 400700, China)

Abstract: In this paper, the educational return rate of national residents was measured by using the unique equation, and the gender and regional differences in educational return rate were compared by using the method of quantile regression. The results show that after correcting the sample selection bias and the endogenous nature of education variables, education has a certain effect on the national residents' income. According to the quantile regression, most of the female students have higher return rate, the coastal education return rate is higher than the inland education return rate, and there are significant differences in the educational return rate between sexes and regions. We should attach importance to female education, give women a certain policy preference, and narrow the gap between male and female education benefits; at the same time, we should increase policy support for inland resources and infrastructure transportation construction to promote a coordinated and balanced development of land and sea.

Key words: educational return rate; instrumental variables; sample selection; quantile regression

(责任编辑: 杨成平)