

对低收入家庭子女大学收益的观察^{*}

袁 诚 张 磊

内容提要:本文使用 2002 年 CHIP 数据,研究了家庭收入对子女大学收益率的影响。基于教育收益存在异质性的假设,本文通过 Roy 模型对子女的能力偏差进行了纠正,并计算出每个子女潜在的大学毕业收入和高中毕业收入,以及因人而异的大学收益率。本文观察到子女的大学收益以及收益率受到家庭收入和背景的影响,低收入家庭子女的收益和收益率明显低于中高收入家庭子女。在此基础上,本文讨论了这种低收益率的原因;低收益率对助学贷款政策效率和可持续性的不利影响,以及提高低收入家庭子女大学收益率的必要性与途径。

关键词:大学收益率 异质性 选择偏差 收入的代际传递 助学贷款

一、引言

教育尤其是高等教育,在理论上被认为是降低社会不公平和减少贫困代际传递的有效手段。因此,世界上几乎所有的国家都会对高等教育给予财政支持。而助学贷款政策是政府为帮助低收入家庭子女完成大学学业所提供的补贴。在我国,政府要求所有公办全日制普通高等学校均要落实国家助学贷款政策,帮助贫困学生克服受教育过程中的信贷约束,从而增加他们在社会和收入阶层中正向流动的机会。所以,对低收入家庭子女大学收益率的了解和观察,对于帮助我们进一步了解高等教育在减少收入不平等的代际传递中的作用,客观评价助学贷款政策的效率和可持续性,有着重要的意义。

本篇论文运用 CHIP2002 年的数据,将 Roy 模型用于对异质性教育收益率的观察,并将这种观察运用于对助学贷款政策的评估与分析。我们发现,在进行了能力偏差的校正之后,子女的大学收益以及收益率都受到家庭收入和背景的影响,低收入家庭子女的收益和收益率明显低于中高收入家庭子女。我们还发现,低收入家庭子女在大学毕业之后,不仅偿还助学贷款

存在困难,而且在偿还期限内,他们的会计收益率还低于助学贷款利率,这将对助学贷款政策的实施造成不利的影响。对此,我们论证了提高低收入家庭子女大学收益率的必要性和可行途径。

我们的贡献主要在于两个方面:首先,我们发现低收入家庭子女的大学收益率并不高于中高收入家庭的子女。长期以来,经济学家们都相信,低收入家庭的教育投资由于信贷约束的存在而受到抑制,因此他们的教育投入低于最优水平,其子女的教育收益率应该远远高于高收入家庭子女的教育收益率(Becker and Tomes,1986)。但是这样的判断没有考虑教育、能力以及收入在代际间的传递,因此与现实中贫困学生的高等教育结果并不完全一致。父母的收入水平不仅影响子女的教育水平,而且也影响他们的教育结果。特别是在发展中国家,贫困学生的就业和毕业收入情况仍然受到自身家庭的社会和经济背景的显著影响,较高的教育收益率对于他们而言是一个过于乐观的判断。

第二,我们的研究对于正确认识和应对我国目前较高的助学贷款违约率提供了经验依据。目前我国的助学贷款还款违约率已经高达 30%,在有些省份甚至达到 60%以上。如果相信贫困学生拥有较高的教育收

^{*} 袁诚,北京大学经济学院,邮政编码:100871,电子信箱:yc@pku.edu.cn;张磊,Clemson 大学经济系,电子信箱:zlei@clemson.edu。同时感谢北京师范大学李实教授提供的数据和支持;同时感谢北京大学于平在本文写作过程中的助研工作,以及作者的讨论。本项研究得到了 2006 年国家教育部人文社会科学研究青年项目“国家助学贷款的可持续性与公平性研究”(EFA060202)的资助。感谢匿名审稿人的意见和建议,但文责自负。

益率,那么对这个现象的唯一解释就是这些接受贷款的学生缺乏基本的诚实和信用。显然,这样的解释已经受到了越来越多的质疑。我们所发现的低收入家庭子女的低大学回报率,成为解释较高的还款违约率,评估助学贷款政策风险的重要依据。

二、文献回顾:异质性教育收益率及研究方法

家庭或者父母收入对于子女受教育状态的影响,已经被很多的计量经济学家所关注。同时,父母收入对子女收入水平的影响,也在大量的收入代际传递的文献中得到了充分的研究。但是,父母收入对于子女教育收益率的影响,在教育回报的研究中,却非常少见。这一方面是因为人们在过去没有充分认识到这个问题的意义,另一方面则是因为在传统的教育回报研究中,缺乏对不同的家庭背景下的收益率变化进行有效观察的工具。

在教育回报的研究中,明瑟方程是计量经济学家最常采用的估计方程:

$$\ln y_i = \beta_s S_i + X_i + u_i \quad (1)$$

其中, y 表示个人的工资收入, S 为个人的教育水平, X 为有关个人其他特征的控制变量, u 为误差项。 β_s 为教育回报率,是研究者感兴趣的参数。基于 Griliches (1977) 劳动力市场中“效率单位”的观点:人力资本是同质的,每个人所面对的教育回报率 β_s 是一样的。在经验分析中,由于影响因变量工资水平的能力变量不可观测,但它又与教育水平相关。为了得到对 β_s 的一致估计,长期以来,计量经济学家投入了大量的研究,来寻找合适的工具变量,消除因能力变量的遗漏所导致的教育内生性,以及 OLS 对教育回报高估的偏差。

所有的工具变量研究都给出了各自不同的收益率估计结果,而且这些结果全都比 OLS 结果来得更高,这与人们寻找工具的初衷完全相反。对此,一个合理的解释就是,教育的回报存在明显的异质性。由于工具变量只能识别对该变量作出反应的群体的回报,因此,工具的不同就会导致不同的估计结果。进一步,由于这些群体往往都具有较高的能力,从而就得到了更高的教育回报估计。

工具变量的经验研究间接地证明了不同群体之间的教育回报存在较大的差异,同时,也有理论研究证明,在家庭背景因素中,家庭收入是个人教育结果和收入能力存在差别的最主要原因(Han and Mulligan, 2001)。这促使经济学家们进一步思考,父母的教育、收入状况不同是否会导致子女的教育回报率存在差别?也就是说,在明瑟方程中,教育项的斜率系数是否

会随着家庭背景的不同而有所不同。为了回答这个问题,就必须在模型设定中放弃对教育回报率一成不变的假设,并需要恰当的方法对此进行观察。

最容易想到的办法就是将明瑟方程中所包含的教育项进行非线性化。Altonji and Dunn (1996) 通过在明瑟方程中引入父母的教育水平与子女教育水平的交叉项,运用同胞子女数据,发现父母的教育水平对子女的教育回报存在一定的影响,而且母亲的影响更为显著。

分位数回归方法(Quantile Regression)也被用来观察在不同的收入分布点上,教育回报率的变化。当家庭收入与子女收入呈现出较强的相关性时,分位数回归对教育回报率的估计结果可以间接地解读为父母收入对子女教育回报的影响。例如,Valero-Gil and Tijerina-Guajardo (2002) 就运用分位数回归,发现在墨西哥,家庭收入越高的子女,其大学教育的回报率也越高,这个结果显然不利于高等教育在降低收入不平等中的作用。

在以 Roy (1951), Willis and Rosen (1979) 和 Willis (1986) 为代表的研究中,提出了另一种方式来研究异质的教育回报率。他们采用的是教育和劳动力市场中的“自选择”和“比较优势”的分析工具,认为每个人都遵循“比较优势”的原则来选择他的受教育水平。在这样的分析框架下,教育是个人内生选择的结果,教育回报不仅因人而异,而且还依赖于人们的受教育状态 S 。我们可以用随机系数表示教育的异质回报:

$$\ln y_i = \beta_{si} S_i + X_i + u_i \quad (2)$$

与明瑟方程(1)不同的是,在方程(2)中,教育回报率系数不再是常数 β_s ,而成为随机系数 β_{si} , β_{si} 将随个人 i 的变化而变化。教育水平 S_i 的选择,既取决于个人和家庭背景,也取决于个人对于预期收入的比较。如果只考虑大学和高中两种教育水平的话,即 $S=1$ 表示大学毕业, $S=0$ 表示高中毕业,那么选择上大学的人应该预期自己上大学的净效用高于不上大学的净效用,而选择不上大学的人则反之。在这样的假设下, β_{si} 不仅与随机扰动项 u_i 相关,而且与大学收益率 β_{si} 相关。

Card (1995) 将 β_{si} 分解为: $\beta_{si} = d_i - k$, d_i 解释为个人能力,是导致收益率出现异质性的原因。而个人能力的差异有相当一部分来自遗传的因素,因此有能力的父母不仅会把自己的能力遗传给子女,而且还能通过他们所获得的物质财富和社会资源,帮助子女在受教育和工作中得到更好的机会。所以,我们有理由相信,父母的收入水平对子女的大学收益率有着正向的影响。

如果能够同时观察到每个人选择两种受教育水平

时的结果 y_{1i} 和 y_{0i} , 那么就可以通过求上述两者之差, 方便地得到个人的教育回报。但是对于随机抽样的个人数据, 由于我们只能看到他(她)在实际的受教育水平下的收入 $y_{1i}|S=1$ 或者 $y_{0i}|S=0$, 而观察不到 $y_{1i}|S=0$ 或者 $y_{0i}|S=1$ 。在计量分析中, Treatment Effect 模型成为解决这类“反事实缺失”(Counterfactual Missing)问题, 估计 τ_i 的主要方法, 其中, local treatment effect (LATE)、matching 以及 marginal treatment effect (MTE) 方法在近来的异质教育回报率的估计中逐渐得以应用。特别是由 Björklund and Moffitt (1987) 和 Heckman and Vytlacil (1999, 2000) 创立以及发展起来的 MTE 模型可以很好地刻画出教育收益率如何随着个人特征的不同而变化。

而 Roy 模型正是 MTE 参数估计的基础。通过引入合适的工具变量 Z , 将每个人的大学教育边际回报率定义为 $MTE_i = E(\ln y_{1i} - \ln y_{0i} | X = x, P(Z) = p)$, 其中 p 为上大学的概率。当 MTE_i 被估计出来后, 将 MTE_i 在父母收入取值所确定的 P 区域进行积分, 从而得到来自不同收入水平家庭子女的大学教育收益率。

三、数据说明和计量方法

1. 数据说明

本文的数据来自于 2002 年国家统计局城调总队和中国社会科学院经济研究所对于城市居民生活的入户调查数据(CHIP), 样本来自于全国 12 个地区 77 个城市, 总共包含 6835 个家庭 20632 个个体。数据库包含丰富的个人信息和家庭背景信息。

本文主要测算父母收入对于子女教育回报率的影响, 根据研究目的, 本文将 CHIP 数据按照子女和父母的关系, 以家庭为单位重新进行了整理。子女样本的选择按照以下原则, 即受教育水平为高中(含中专)及高中以上, 工作状态为工作(不含退休), 收入信息没有缺失, 且父母双方至少有一位健在。这样得到的每个子女的信息不仅包括他(她)本人的相关信息, 而且还包括其父母的相关信息。在对原有的 6835 个家庭样本筛选的过程中, 我们把三代人家庭拆分为两个两代人家庭。最后得到 1295 个家庭样本, 共有 1403 个包含父母信息的子女样本。

在我们的研究中, 父母收入是一个非常重要的解释变量。但在我们的样本中, 父亲群体与母亲群体分别有 35.14% 和 51.56% 处于退休状态。在职父母的收入与离退休父母的收入存在不可比性, 特别是, 退休父母的收入基本上不能反映这些父母的工作能力, 而且退休收入的差别远远小于工作收入的差别, 所以退休工资与工作时的工资收入分布有很大的不同。如

果把这部分样本去掉, 将大大减少我们的观察对象。为了减少这种不可比性所造成的影响, 我们把退休父母的收入进行了标准化处理。

对退休父母标准化处理的方式如下: 首先根据每位退休父母的收入(对数形式), 求出他们在退休收入总体正态分布假设下的累积百分比排位; 然后根据退休父母收入的百分比排位, 将他们的收入转化为与没有退休的总体具有相同均值和标准差的正态分布的水平值。由于男性和女性的工资水平也存在显著的差异, 因此, 在进行标准化处理时, 将退休父亲的收入和退休母亲的收入分别进行处理。

我们看到, 家庭收入对子女的教育水平有着非常明显的影响, 按照父亲收入由低向高累积 30% 和 70% 的排位分界, 把家庭收入分为低、中、高三等, 其子女上大学的比例分别为 34.91%、43.09% 和 61.14%, 呈现出明显的随收入增加而上升的趋势。不仅如此, 子女的收入, 不论是高中毕业还是大学毕业, 都表现出随家庭收入增加而增加的趋势, 子女收入与父亲收入之间的相关系数达到 25%, 为显著的正向相关关系。特别是, 对于大学毕业生, 高收入家庭子女的收入要比低收入家庭子女高出 50%。

在大学毕业和高中毕业的收入对比中, 我们发现, 前者的平均年收入要比后者高出 3000 元, 他们所在的职位和行业也都要优于后者。不仅如此, 他们父亲的年收入也要比对比组的父亲多出近 3000 元; 他们所就读的高中以及高中成绩也都要好于高中毕业生。这表明, 造成他们收入差异的, 不仅仅是受教育水平的不同, 而且还与这两个群体的个人特征与家庭背景存在着很大的关系。所以在计算大学收益率时, 需要克服这些不同所导致的偏差。

表 1 给出了样本子女的主要特征, 包括个人特征和家庭背景的基本描述信息。变量包括: 性别、工作时间(工龄)、职位性质是否为固定或长期合同、职位类别(管理层、技术职位、普通职位)、所在行业(第一类行业为垄断行业、第二类为普通生产性行业、第三类为普通

我们认为, 虽然退休父母的收入绝对值不能与他们的能力相互对应, 但是其相对排位仍然可以反映出这些退休父母的能力排位。有理由相信, 退休收入较高的父母, 他们在退休之前的收入也应该较高, 他们的收入排位在退休前后没有变化。事实上, 我们后面的经验结论并不受标准化处理的影响。

表 1 子女特征的描述性统计

变量		高中毕业	大学毕业
		均值或比重	均值或比重
年 收 入	平均	9956.024 (9971.462)	12977.720 (9764.205)
	高收入家庭	11948.35 (9868.954)	15430.06 (10828.03)
	中收入家庭	10583.71 (11858.27)	12229 (8924.291)
	低收入家庭	7964.651 (6790.365)	10047.56 (8065.303)
男性		0.609	0.534
年龄		28.349 (7.55)	28.348 (6.294)
工作时间		15.197 (102.682)	11.523 (78.780)
固定或长期合同		0.495	0.664
管理职位		0.172	0.759
技术职位		0.297	0.701
普通职位		0.686	0.223
第一类产业		0.243	0.331
第二类产业		0.214	0.133
第三类产业		0.543	0.536
第一份工作寻找时间		8.674 (18.686)	4.376 (11.996)
父 亲 收 入	平均	11667.940 (8619.469)	14496.850 (9089.778)
	高收入家庭	23280.29 (12206.82)	22531.51 (8573.337)
	中收入家庭	11367.78 (2066.798)	12063.5 (2083.099)
	低收入家庭	6215.8 (1638.79)	6366.232 (1764.12)
母亲受教育年限		8.292 (3.512)	9.655 (3.378)
父亲职位较高		0.432	0.644
重点高中就读		0.170	0.457
非重点高中就读		0.354	0.376
中专等其他学校		0.476	0.167
高中成绩		2.295(0.741)	2.842(0.749)
观测个数		757	646

注:括号中的值为标准差。

服务性行业)、所在地区、第一份工作的寻找时间,以及父亲的收入、母亲的受教育年限、父亲的职位是否为中高层、曾经就读高中的类别(重点高中、普通高中、职高和中专等其他学校)、高中成绩排名。

2. 方法说明

本文的计量方法主要运用 Roy 模型来估计每个子女的大学毕业收益率,然后根据其父母收入水平的划分,对低收入与高收入家庭子女的收益率进行比较。除此之外,我们还参考了引入交叉项的线性回归结果。在引入大学教育与父亲收入交叉项的线性回归中,我们允许教育回报率随父亲收入的不同而不同。

采用 Roy 模型进行大学毕业收益率的估计主要基于以下三个理由:首先,Roy 模型可以在一定程度上消除由于能力高低所导致的“自选择”偏差;更重要的是,在 Roy 模型中,教育收益率将随每个人的个人特征与背景的变化而变化;最后,通过 Roy 模型,我们可以看到父母

的收入在影响子女教育收益率中的传递机制。

在 Roy 模型中,子女的教育水平由下面的选择方程决定:

$$S = \begin{cases} 1 & \text{if } S^* > 0 \\ 0 & \text{if } S^* \leq 0 \end{cases}$$

其中, S^* 为隐性变量,表示上大学的意愿或者可能性,被设定为个人特征 Z 的线性函数:

$$S^* = Z +$$

这里的 u 代表观察不到的随机扰动项,通常和上大学的成本或付出有关,包括物质成本和心理成本,它的标准差记为 σ 。

高中毕业和大学毕业的收入水平分别由下面的两个水平方程式来表示:

$$\ln y_0 = \mu_0 + u_0 = X + u_0$$

$$\ln y_1 = \mu_1 + u_1 = X + u_1$$

其中, y_0 和 y_1 分别表示高中毕业生和大学毕业生的潜在或理论收入, μ 为收入对数的均值,被设定为个人特征 X 的线性函数,并且对于学历不同的人群, X 对收入的边际影响也是不同的,这反映了高等教育的作用,对此我们分别用参数 β_0 和 β_1 来表示。在理论上,对于每个观测子女,都存在 y_0 和 y_1 的结果,但是,在实际中,我们只能观察到其中的一个结果。我们所观察

到的样本结果 y 取决于:

$$y = \begin{cases} y_0 & \text{if } S^* \leq 0 \text{ 或 } S = 0 \\ y_1 & \text{if } S^* > 0 \text{ 或 } S = 1 \end{cases}$$

需要说明的是,影响子女受教育水平的个人特征变量 Z 通常会包含影响他们收入水平的个人特征变量 X ,而那些影响 S 但不影响 y 的变量成为模型识别的

地区分类按照 2002 年底各地区人均国内生产总值水平,第一类地区包括安徽、四川、云南、甘肃;第二类地区包括山西、河南、重庆;第三类地区包括湖北;第四类地区包括辽宁、江苏、广东;第五类地区包括北京。

如果父亲的收入信息缺失,则用母亲的收入代替,这样的观测对象有 205 个。

如果母亲受教育年限的信息缺失,则用父亲的代替,这样的观测对象有 154 个。

高中成绩前 20%、40%、60%、80%,我们分别以 4 分、3 分、2 分和 1 分计算。

“排除约束”(exclusion restrictions),有时也被称为工具变量。

与标准意义上的明瑟教育收益率的研究不同的是,在我们的水平方程中,把收入代际传递的因素考虑进来了。我们认为子女的收入水平主要取决于子女本身的人力资本状况,而人力资本状况非常依赖于其父母的收入水平和人力资本状况。因此,父母的收入状况成为子女收入方程中的一个解释变量。收入的代际传递可能有以下原因:一是父母收入高,能够为子女提供较好的生活条件,甚至可以有条件帮助子女找到好的工作,这是收入的直接影响;二是父母收入高,会增加在子女教育方面的投资,提高了子女的受教育水平,而教育作为人力资本投资的方式,会增加子女未来获得更好收入的机会;三是父母收入高可能由于本身其个人能力强或是家庭背景好,这一点也可以传递给子女,子女的能力强或者家庭背景好,有机会获得更好的收入。

三个方程中的随机扰动项被认为服从联合正态分布,即

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

出于模型估计和 方差的不可识别性,我们将选择方程中的方差 标准化为 1,这样有:

$$u_0 = \frac{0}{\sigma_0} + 0 = 0 + 0, u_1 = \frac{1}{\sigma_1} + 1 = 1 + 1, 0, 1 \sim N(0,1), \text{且独立于 } \sigma_0$$

和 分别为 u_0 和 u_1 与 的相关系数。当 $S = 1$ 时, y_1 的条件期望成为断尾情形下的条件期望,运用断尾变量的条件期望结果得到:

$$E(\ln y_1 | S = 1) = E(X + u_1 | Z + > 0) = X + E\left(\frac{1}{\sigma_1} + 1 \mid > - Z\right) = X + 1 E\left(\frac{\phi(-Z/\sigma_1)}{1 - (-Z/\sigma_1)}\right) = X + 1 \cdot (Z/\sigma_1) \quad (1)$$

记 $v = Z/\sigma_1$, 上式可写为: $E(\ln y_1 | S = 1) = X + 1 \cdot (v)$ 其中, $(v) = \phi(v)/\Phi(v)$ 称为 Inverse Mills Ratio, $\phi(v)$ 和 $\Phi(v)$ 分别为标准正态分布的密度函数和累积分布函数。由于上式中的 $y_1 | S = 1$ 是可观测的,所以该条件期望成为一个可以估计的经验方程,它意味着我们可以运用上大学的子女的收入资料,对 进行估计。而要得到 的一致估计量,在回归方程中需要加入回归变量 (v) ,来控制由于能力变量遗漏导致的“自选择”偏差。 (v) 的值可以通过对是否接受高等教育的

状态:

$$S = \begin{cases} 0 & \text{if } Z \leq 0 \\ 1 & \text{if } Z > 0 \end{cases}$$

进行 Probit 估计,得到 $\hat{\alpha}$ 之后,计算 (v) 的拟合值,然后代入到(1)式,估计 和 σ_1 。这种估计方法称为 Heckman 两阶段估计。

对于 $S = 0$ 的情形, $\ln y_1$ 的条件期望为:

$$E(\ln y_1 | S = 0) = E(X + u_1 | Z + < 0) = X + E(u_1 | < - Z) = X - \sigma_1 \cdot (-v) \quad (2)$$

把高中毕业子女的有关信息代入(2)式,则可以得到他们潜在的大学毕业的期望收入。在估计 的过程中,方法是类似的。用于估计的经验方程为:

$$E(\ln y_0 | S = 0) = E(X + u_0 | Z + < 0) = X - \sigma_0 \cdot (-v) \quad (3)$$

上大学子女的潜在高中毕业收入为:

$$E(\ln y_0 | S = 1) = E(X + u_0 | Z + > 0) = X + \sigma_0 \cdot (v)$$

这样,对于大学毕业的子女,他们的大学收益率为:

$$E(\ln y_1 - \ln y_0 | S = 1) = E(\ln y_1 | S = 1) - E(\ln y_0 | S = 1) = X + \sigma_1 \cdot (v) - [X + \sigma_0 \cdot (v)] = X(\sigma_1 - \sigma_0) + (\sigma_1 - \sigma_0) \cdot (v)$$

高中毕业子女的大学收益率为:

$$E(\ln y_1 - \ln y_0 | S = 0) = X(\sigma_1 - \sigma_0) - (\sigma_1 - \sigma_0) \cdot (-v)$$

根据 Roy 模型的设定,每个人都拥有自己的大学收益率水平: $sr_i = \ln y_{i1} - \ln y_{i0}$,这个结果对于每个人是不一样的,而且通过 Roy 模型,我们能够计算的期望收益率是一个条件收益率,它取决于子女的受教育状态。

当估计出参数 和 之后,我们就可以计算出每个观测对象的大学期望收益率。可以看到,不论对于大学毕业生还是高中毕业生,大学收益率中很大一部分取决于 $X(\sigma_1 - \sigma_0)$,即大学毕业与高中毕业收入方程的系数之差与其个人特征的乘积。这可以解释为,大学教育改变了个人素质和特征对收入的贡献程度。例如,同样在国企工作,性别、地区等所有的个人特征都一样,但由于学历不同,这些因素对收入的影响是不同的。

同时,我们也可以运用 y 的分布性质,进行偏极大似然估计(Partial ML)。在偏极大似然估计中,所建立的似然函数是以 $S = 1$ 为条件的,因此所使用的只是部分观察到的 y_1 样本。

表 2 加入“父亲收入与大学教育交叉项”的 OLS 估计结果

变量		方程 1	方程 2	方程 3
是否大学毕业		0.261 (0.000)	- 1.639 (0.00)	- 1.648 (0.000)
父亲收入与大学交叉项			0.200 (0.000)	0.197 (0.000)
高中学习成绩				0.075 (0.001)
性别		0.147 (0.000)	0.142 (0.000)	0.149 (0.000)
工作时间		0.000 (0.043)	0.001 (0.016)	0.001 (0.017)
工作时间平方		0.000 (0.154)	0.000 (0.118)	0.000 (0.110)
调整 R^2		0.1878	0.2	0.2664
收益 率	低收入家庭	0.261	0.095	0.060
	中收入家庭	0.261	0.241	0.204
	高收入家庭	0.261	0.368	0.329

四、经验结果

1. 初步的 OLS 观察

为了对大学收益率的异质性有一个初步的直观的了解,我们首先进行包含“父亲收入与大学教育交叉项”的明瑟方程 OLS 回归。

表 2 给出的是明瑟方程中 s 的估计值,括号中为 p 值。在方程 1 中,解释变量除了 s 这一虚拟变量,用于表示是否上过大学之外,还包括子女的自身条件:工作时间、工作时间平方,以及性别;方程 2 在方程 1 的基础上,增加了“父亲收入与大学教育交叉项”;方程 3 则在方程 2 的基础上增加了“高中学习成绩”来控制子女的能力。

表 2 的回归结果显示,当加入交叉项后,“是否大学毕业”对子女收入的单独影响为负,但是这并不表明大学的教育回报是负的,因为真正的大学收益率是由这两个变量的共同影响决定的。交叉项的影响在两个方程中都为正,且高度显著,这表明子女的大学收益率存在某种因家庭收入而异的特征。

表 2 下方的“大学收益率”是根据三类收入家庭的各自收入均值与“是否大学毕业”和“父亲收入与大学交叉项”估计系数计算出来的。可以看出,随着家庭收入的增加,子女的大学收益率是上升的。同时,比较方程 2 和方程 3,当增加了控制能力的“高中学习成绩”时,各类收入家庭的子女大学收益率的估计值是下降的,这表明,在教育回报的研究中,确实存在由于遗漏能力变量,收益率被高估的问题。

一般而言,能力的高低与家庭收入状况有着密切的联系,高收入家庭的子女很有可能拥有较高的能力,因此当加入“高中学习成绩”来控制子女的能力时,高收入家庭子女的教育收益率应该下降的幅度最大,低

收入家庭子女则最小。在表 2 的结果中,高中低三类家庭子女的收益分别下降了 0.039,0.037,0.035,下降的幅度非常接近,这种同质性正是线性模型设定的局限所在。

2. Roy 模型的估计结果

上面的 OLS 结果已经反映出子女的大学回报率受到家庭收入高低的影响,但是线性模型是建立在教育回报同质性的假设之上的,即使加入了“父亲收入

与大学交叉项”,OLS 的估计结果仍然不能够精确地反映子女的教育收益如何受到家庭收入的影响。因此,在这里我们将采用 Roy 模型来估计大学回报率,以期得到更加精确的结果,为进一步来探讨教育回报异质性存在的原因以及政策分析提供依据。

在我们的 Roy 模型中,影响子女收入的水平方程中的解释变量 X 包括:工作时间及其平方、职位性质是否为固定或长期合同、职位类别、所在行业、所在地区,以及性别,此外还加入了父亲的收入对数和第一份工作的寻找时间。加入父亲的收入,一方面是出于对收入的代际传递影响考察的需要;另一方面,父亲的收入在一定程度上可以作为子女的能力代理,从而对子女的能力进行控制。加入第一份工作的寻找时间的理由在于,寻找工作的时间长短可以直接反映出子女在工作市场上的能力和表现,从而可以进一步控制能力偏差。

在影响子女接受高等教育的解释变量 Z 中,除了包括性别、父亲收入、地区之外,还包括子女 18 岁时所在省份的大学入学机会、曾经就读高中的类别、高中成绩排名、母亲的受教育年限、父亲的职位是否为中高层,这五个“排除约束”是收入水平方程所不包含的,它们对子女的上大学概率有着明显的影响,但并不直接影响子女的收入水平。

在估计的过程中,我们分别运用上大学和没上大学的子女样本对方程(1)和(3)进行估计,得到 $\hat{\alpha}$ 、 $\hat{\beta}$ 和 $\hat{\gamma}$

我们也试图加入“母亲的受教育年限”来控制子女的能力,但由于该变量与子女是否接受大学教育相关性很高,达到 30%,如果加入将导致严重的多重共线性。在后面的 Roy 模型中,我们把该变量放在选择方程中了。

资料来源:李立峰,2007:《各省高考机会排名及 1977 年以来各省录取分数线之变化》,高校信息网, [http://www.rankedu.com/? dp=bbsthead+4551.html](http://www.rankedu.com/?dp=bbsthead+4551.html)。

以及 $\hat{\alpha}_1$ 、 $\hat{\alpha}_0$ 和 $\hat{\alpha}$ 。表3给出了极大似然估计结果。

通过表3的结果可以看到,所有的解释变量都表现出了和预期相一致的符号。在对子女(不论是大学毕业还高中毕业)收入水平的影响中,子女的工龄、固定或长期合同、男性以及父亲收入对数都有着正向的显著影响;技术职位(相对于管理层)、非垄断的普通行业以及寻找第一份工作的时间都有着负向的显著影响。同样的解释变量对于大学毕业生和高中毕业生的收入影响均存在差异,特别是职位与行业的回归系数,对于高中毕业生不仅数值很小,而且非常不显著,与大学毕业生的结果显著不同。

在对子女上大学的概率($Pr(S=1)$)影响中,我们发现,女性、就读于重点高中、高中成绩靠前、父亲的收入高、母亲的受教育时间长、升学率高,对于上大学有着明显积极的影响,而这些因素对于不上大学的概率($Pr(S=0)$)影响则是反向的。

在这些系数的估计结果中,我们对于 (ν) 和 $(-\nu)$

的系数 $\hat{\alpha}_1$ 和 $\hat{\alpha}_0$ 也非常关心,因为它们符号和大小能够反映出“自选择效应”的方向和大小。对于高中毕业生, $(-\nu)$ 的系数为0.338,并且在5%的水平上显著。这表明高中毕业生“选择”不上大学,在收入的获取中得到了“比较优势”的正向回报,因此他们的自选择效应是正的。对于大学毕业生而言, $\hat{\alpha}_1$ 符号为负,但是很不显著,这表明,当我们加入控制能力的变量(父亲收入、寻找第一份工作的时间以及高中成绩)时,已经可以很好地解释大学生的收入回报;同时,他们的水平方程中的误差项与选择方程误差项之间相关性变得很弱。

考虑到 $\hat{\alpha}_1$ 和 $\hat{\alpha}_0$ 的估计结果,以及我们在前面所推导的 $E(\ln y_0 | S)$ 与 $E(\ln y_1 | S)$ 表达式,可以看出,相对于大学毕业生的收入,高中毕业生的理论收入 y_0 在Roy模型的估计中被上调了,而 y_1 则下降了。表4给出了根据Roy模型的估计结果,按照父亲收入30%和70%排名分组,所计算的大学收益。

表3 Roy模型参数估计结果

解释变量		被解释变量			
		大学毕业 收入对数 $\ln y_1$	大学毕业 = 1 $Pr(S=1)$	高中毕业 收入对数 $\ln y_0$	高中毕业 = 1 $Pr(S=0)$
工作背景	工作时间	0.001 (0.007)		0.000 (0.537)	
	工作时间平方	0.000 (0.233)		0.000 (0.363)	
	固定或长期合同	0.133 (0.015)		0.186 (0.000)	
	技术职位	- 0.304 (0.001)		- 0.102 (0.549)	
	普通职位	- 0.421 (0.000)		- 0.204 (0.224)	
	第二产业	- 0.160 (0.037)		0.001 (0.988)	
	第三产业	- 0.096 (0.074)		- 0.030 (0.600)	
家庭背景	男性	0.158 (0.001)	- 0.243 (0.002)	0.094 (0.044)	0.216 (0.007)
	父亲收入对数	0.176 (0.000)	0.295 (0.000)	0.107 (0.003)	- 0.259 (0.000)
	母亲受教育年限		0.037 (0.002)		- 0.034 (0.005)
能力变量	父亲职位较高		0.230 (0.007)		- 0.212 (0.012)
	第一份工作寻找时间	- 0.006 (0.003)		- 0.004 (0.002)	
	非重点高中就读		- 0.354 (0.060)		0.344 (0.000)
	中专等其他学校		- 1.166 (0.000)	1.142 (0.000)	
	高中成绩	0.056 (0.261)	0.512 (0.000)	- 0.028 (0.463)	- 0.527 (0.000)
	18岁时大学入学机会		0.322 (0.002)		- 0.346 (0.001)
	(ν)	- 0.138	(0.469)		
	$(-\nu)$			0.338 (0.042)	
	常数项	7.472 (0.000)	- 4.362 (0.000)	7.744 (0.000)	4.706 (0.000)
	观测个数	757	1403	646	1403

说明:(1)括号中的值为p值;(2) (ν) 和 $(-\nu)$ 系数的p值是根据 $H_0: = 0$ 的卡方检验计算的;(3) $(-\nu)$ 的系数是根据 $Pr(S=0)$ 估计出来的,因此实际的 $\hat{\alpha}_0$ 等于负的 $(-\nu)$ 系数;(4)收入估计方程包含地区变量。

表 4 家庭收入与子女大学收益分组表

		低收入家庭	中收入家庭	高收入家庭	平均/总和
全部样本	ln y ₁	8.930 (0.327)	9.190 (0.342)	9.403 (0.372)	9.176 (0.390)
	ln y ₀	8.718 (0.333)	8.931 (0.381)	9.064 (0.362)	8.907 (0.372)
	大学收益率	0.212 (0.233)	0.260 (0.218)	0.339 (0.204)	0.269 (0.224)
	大学收益	1534	2389	3802	2559
	观测个数	421	558	424	1403
上大学样本	ln y ₁	8.948 (0.316)	9.174 (0.337)	9.391 (0.388)	9.207 (0.391)
	ln y ₀	8.512 (0.285)	8.726 (0.296)	8.927 (0.346)	8.755 (0.346)
	大学收益率	0.436 (0.177)	0.449 (0.139)	0.464 (0.145)	0.452 (0.149)
	大学收益	2927	3942	4984	4085
	观测个数	151	241	254	646

注:括号中的数值为均值的标准差。

从全部样本的计算结果来看,子女的大学收益率为 0.269,其中上大学群体的收益率为 0.452,高中毕业群体的收益率为 0.113。中等收入家庭子女的大学回报率要比低收入家庭的子女约高 23%,高收入家庭的子女则约高出 60%。当然,如果只是考察上大学的群体,他们之间收益率的差别会有明显的缩小,差距分别降为 3%和 6%。尽管如此,他们收入绝对值之间的差别没有发生明显的改变。对于全部样本,低收入家庭子女的大学年收入要比中等收入家庭的子女少 2407 元,比高收入家庭的子女少 5020 元;而对于大学生样本,上述两个差额则变为 2118 元和 4818 元。

在前面的 Roy 模型估计中,父亲的收入进入到水平方程中,而我们对收益率的分组考察又是以父亲的收入为依据进行的,因此这在一定程度上会强化父亲收入对子女收入的影响。为了客观评价父亲收入对于子女大学收益的影响,我们同时也估计了收入方程不含父亲收入的 Roy 模型,在这个模型设定中,“父亲收入”仅仅进入选择方程,影响子女的上大学概率,对于上大学样本得到下面的结果:

表 5 父亲收入仅作为工具变量的子女收入估计结果

	低收入家庭	中收入家庭	高收入家庭
ln y ₁	8.879 (0.354)	9.074 (0.404)	9.269 (0.427)
ln y ₀	8.391 (0.302)	8.567 (0.336)	8.733 (0.360)
大学收益率	0.488 (0.160)	0.507 (0.135)	0.536 (0.138)
大学收益	3031	3893	4021

注:括号中的数值为均值的标准差。

这时,计算出来的低收入家庭大学毕业生的大学收益率仍然明显低于中高收入家庭的大学毕业生,年收入差距分别达到 1778 元和 2958 元,尽管低于前面所计

算的 2118 元和 4818 元,但差距仍然是明显的。这表明,父母收入对于大学生子女的收入差距影响是非常显著的。同时,如果把父亲收入从子女的收入方程中去掉,会得到高度显著的负的自选择效应,这有悖于 Roy 模型的基本假设。所以,我们在收入方程中包含父亲的收入,是非常必要的。

五、结果解释与政策讨论

1. 对低收入家庭子女低收益率的进一步分析

根据 Roy 模型,大学毕业子女的大学收益率为:

$$E(\ln y_1 - \ln y_0 | S = 1) = X(\beta_1 - \beta_0) + (\beta_1 - \beta_0) \cdot (\nu)$$

高中毕业子女的大学收益率为:

$$E(\ln y_1 - \ln y_0 | S = 0) = X(\beta_1 - \beta_0) - (\beta_1 - \beta_0) \cdot (-\nu)$$

因此,在收益率的计算中,收益率的大小在很大程度上取决于第一项 $X(\beta_1 - \beta_0)$ 。由于 $(\beta_1 - \beta_0)$ 对于每个子女都是一样的,那么收益率的不同就来自于个人特征 X 的不同。在解释变量 X 中,有相当多的变量和子女的工作职位、行业、找工作的时间有关,而低收入家庭的子女,即使在上大学之后,在职业市场上并没有取得更多的优势,他们可能找工作的时间较长,所在的职位较低,工作的行业盈利性较差,这直接导致了他们较低的收益率。同时,解释变量 X 还包括一些代表能力的变量,如高中成绩和父母收入,这些变量的取值较低也直接导致了低收入家庭子女的大学收益率偏低。

收益率表达式中的第二项称为“选择性偏差”,表 6 给出了对选择性偏差的分析结果:

表 6 的结果表明,对于全部样本,选择性偏差对于三类家庭的子女都是很显著的,因此它对收益率的差异化影响很小。但对于上大学的样本而言,低收入家庭子女的选择性偏差是正的,而且高于中高收入家庭。这表明,在 Roy 模型的估计中,低收入家庭大学生子女的收益率要比没有经过偏差校正的结果更高,即使这样,他们的大学收益率仍然偏低,这说明,在职业市场上,低收入家庭子女的就就业状况差于中高收入家

大学的质量和排名也会影响到子女的大学收益,但是这个变量我们无法进行控制。

庭是导致前者收入及收益率偏低的主要原因。

表 6 选择性偏差的分组结果

	全部样本			上大学样本		
	低收入家庭	中收入家庭	高收入家庭	低收入家庭	中收入家庭	高收入家庭
选择性偏差	0.005 (0.139)	- 0.006 (0.147)	0.004 (0.141)	0.166 (0.083)	0.138 (0.080)	0.103 (0.069)

注：括号中的数值为均值的标准差。

此外，选择性偏差代表的是没有观察到的变量对大学收益率的影响，对于上大学的子女，他们的能力应该服从相同的分布，而低收入家庭大学生子女的选择性偏差较高的结果表明，这部分群体相对而言具有更高的能力，这意味着上大学的机会对于不同家庭经济条件的子女是不平等的。

2. OLS 与 Roy 模型结果的含义和差异

我们观察到，Roy 模型所估计的大学回报率都远远高于 OLS 的估计结果，其原因在于，由于模型设定的不同，这两个模型所估计出的大学收益率是两个完全不同的参数。

在 OLS 估计中，高中毕业与大学毕业的理论收入被假定为：

$$\ln y_0 = \alpha_0 + X + u_0, \ln y_1 = \alpha_1 + X + u_1$$

在上面两个方程中，所有解释变量的边际收入都是相同的，只有截距项不同。因此，通过加入 s 这一虚拟变量，所得到的大学收益率为以上两个方程的截距项之差 ϕ 。而在 Roy 模型中，大学收益率为 $X(\beta_1 - \beta_0)$ ，其中不仅包含了高中毕业与大学毕业的理论收入方程的截距项之差，而且还包括了所有解释变量的边际影响之差。所以，Roy 模型在大学收益率要大于线性模型中的大学收益率，而且从模型设定的角度而言，Roy 模型所附加的约束更少，从而对教育回报的定义更加具有一般性。

3. 对助学贷款政策的考察

根据国家助学贷款政策 2007 年的最新规定，助学贷款的借款额每年不超过 6000 元，最迟在毕业后六年内还清。假设贫困学生在上大学期间，每年借款 6000 元支付学费和生活费，按照现行 6.12% 的贷款利率计算，如果从毕业后第三年开始还贷，每月需要偿还 565 元。前面的描述统计显示，低收入家庭大学毕业子女在 2002 年的平均年收入为 10048 元，按照 3% 的年利率计算，相当于 2007 年 11600 元，即平均每月接近 1000 元，月收入中有一半以上的钱需要偿还贷款，这无疑是很重的负担。

再从收益率的角度来看，我们按照舒尔茨大学教育收益率公式：大学收益率 = (大学毕业平均年收入 -

高中毕业平均年收入)/大学教育投资成本。根据前面 Roy 模型计算的大学生样本的收入结果，四年大学教育成本取 40000 元，即年均 1 万元，我们重新计算了大学教育的会计收益率，其结果如下：低收入家庭的大学生子女的收益率为 7.32%，中收入家庭子女为 9.86%，高收入家庭子女为 12.46%。如果大学教育成本超过 47827 元，那么低收入家庭子女在毕业六年内的大学收益率要低于他们所面对的贷款年利率 6.12%，从这个意义上说，在短期内，贫困学生不仅偿还贷款存在困难，而且在收益上有可能也是不合算的。

要解决这个难题只有两条途径：一是加大财政贴息的力度，放宽还款的年限；另一条则是提高贫困家庭子女的大学收益率。而后者显然是根本的出路。

4. 提高低收入家庭子女大学收益率的必要性与途径

提高低收入家庭子女大学收益率不仅是保证助学贷款政策持续性的重要前提，也是助学贷款政策实施的理由所在。对于贫困家庭的子女而言，只要收益率为正，上大学总是要优于不上大学，即使毕业初期无力偿还助学贷款，因为作为教育投资的固定成本，从长期来看是可以忽略不计的。但是，对于政府而言，如果贫困子女的大学收益率很低，那么从收入分配的角度来讲，这一部分的财政资助就应该投入到收益率更高的基础教育或职业教育上。我们在前面的分析表明，低收入家庭子女的大学收益率偏低，不仅已经影响到他们的助学贷款偿还能力，而且也影响到助学贷款政策的实施效果。因此，提高这部分群体的收益率是一个需要引起重视的问题。

如何提高他们的收益率，我们认为应该从以下两个方面考虑。首先，尽可能消除就业市场中的不平等，对贫困大学生给予更多的就业指导和帮助，这实际上是助学贷款政策的进一步延伸。第二，通过对 30 岁上下两个年龄段的大学生子女的收益率进行比较，我们看到，贫困大学生收益率偏低是伴随全体大学生收益率下降同时出现的现象，因此，普遍提高高等教育的教学质量，改善大学教育对学生能力培养的效果，是提高大学收益率的必由之路，也是使贫困学生真正受益的

如果一毕业就偿还贷款，每月则要偿还 377 元，这个数额接近月收入的三分之一。

与世界上的其他国家相比，我国助学贷款的还款年限是最短的。例如，美国为 10 年，最长可达 30 年；日本为 20 年；英国为 5 年，但可以延至 7 年，实行的是无息贷款；澳大利亚和加拿大则根据毕业生的实际收入情况安排还款年限。

小于 30 岁的低中高收入家庭的大学生子女的收益率分别为 0.402、0.425、0.451，所对应的大于 30 岁的群体收益率分别为 0.491、0.502、0.502。

可取之路。

5. 对样本偏差的一个说明

本文所用到的样本具有较强的特殊性,即样本中的子女是与自己的父母生活在一起的。这表明,这部分群体具有很弱的流动性,他们即使上了大学,也仍然是回到父母的身边,因此,父母以及家庭对他们的影响很大,这造成不同家庭之间的大学收益率的差异被放大了。在我们国家,上大学是人才流动的一个主要原因,有相当一部分大学生在毕业之后,没有回到原籍,特别是来自贫困地区和贫困家庭的学生,他们的工作状况和收入水平可能会较少地受到家庭和代际传递的影响。但这是一个有待观察和证实的问题,还需要进一步的研究。

参考文献

李雪松、詹姆斯·赫克曼, 2004:《选择偏差、比较优势与教育的异质性回报:基于中国微观数据的实证研究》,《经济研究》第4期。

J. Altonji and T. Dunn, 1996, "The Effects of Family Characteristics on the Return to Education", *The Review of Economics and Statistics*, Vol. 78(4): 692—704.

A. Björklund and R. Moffitt, 1987, "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models", *The Review of Economics and Statistics*, Vol. 69(1): 42—49.

G. Becker and N. Tomes, 1986, "Human Capital and the Rise and Fall of Families", *Journal of Labor Economics*, Vol. 4(3): S1—S39.

D. Card, 1995, "Earnings, Schooling, and Ability Revisited", *Research in Labor Economics*, vol. 14: 23—48.

P. Carneiro, J. Heckman, and E. Vytlačil, 2002, "Estimating The Return to Education When It Varies Among Individuals," working paper at www.econ.upenn.edu/cgi-bin/mecon/bin/serve.cgi/Epirical120520020

E. Eide and M. Showalter, 1999, "Factors Affecting the Transmission of Earnings across Generations: A Quantile Regression", *The Journal of Human Resources*, Vol. 34(2): 253—267.

D. Gnther, 2000, "Alternative Estimates of the Effect of Schooling on Earnings", *The Review of Economics and Statistics*, Vol. 82(1): 103—116.

Z. Griliches, 1977, "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, Vol. 45(1): 1—22.

J. Heckman and E. Vytlačil, 1999, "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects", *Proceedings of the National Academy of Sciences*, Vol. 96:4730—4734.

J. Heckman and E. Vytlačil, 2000, "Local Instrumental Variables," in C. Hsiao, K. Morimune, and J. Powells, (eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, (Cambridge: Cambridge University Press, 2000): 1—46.

S. Han and C. Mulligan, 2001, "Human Capital, Heterogeneity and Estimated Degrees of Intergenerational Mobility," *The Economic Journal*, Vol. 111(470): 207—243.

P. Taubman, 1989, "Role of Parental Income in Educational Attainment", *The American Economic Review*, Vol. 79(2): 57—61.

A. Roy, 1951, "Some Thoughts on the Distribution of Earnings", *Oxford Economics Paper*, vol. 3(2): 135—146.

J. Valero-Gil and J. Tijerina-Guajardo, 2002, "Effects of Education on the Intergenerational Transmission of Labor Income in Mexico", *Eastern Economic Journal*, Vol. 28(3): 381—392.

R. Willis and S. Rosen, 1979, "Education and Self-Selection," *The Journal of Political Economy*, Vol. 87(5): S7—S36.

R. Willis, 1986, "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions", in *Handbook of Labor Economics*, Vol. I, O. Ashenfelter and R. Layard (eds.), North-Holland.

Parental Income and Return to College Education in Urban China

Yuan Cheng and Zhang Lei

(School of Economics, Peking University; School of Economics, Clemson University)

Abstract: In this paper, we examine the influence of parental income on the return to college education in a framework of heterogeneous returns to education among individuals. The Roy model is employed to correct the ability bias and to estimate the potential incomes from obtaining a college education and a high school education for each individual. We use the 2002 CHIP data and find strong evidence that the return to college education increases with parental income. Based on these findings, we discuss the efficiency and sustainability of National College Loan Policy and the possible measures that government may take to increase the return to college education for students from low-income families.

Key Words: Return to College; Heterogeneity; Selection Bias; Intergenerational Transmission of Income; College Loan

JEL Classification: C300, I200, J300

(责任编辑:金成武)(校对:王利娜)