

谁从高等教育扩招中获益更多？

——基于广义罗伊模型的实证证据

刘生龙 郑世林*

内容提要 教育回报是个人以及社会最优教育决定的重要组成部分，传统的估计方法如工具变量法、回归断点设计和双胞胎设计等主要都是为了解决教育的内生性问题。然而，当教育回报存在根本的异质性，即教育回报与教育选择本身相关时，传统的估计方法便无法刻画这种不可观测的异质性。本文利用广义罗伊模型估计了大学教育的边际处理效应（MTE），研究证实，中国大学教育的回报率与大学教育选择之间存在相关关系。半参数局部工具变量（LIV）估计结果表明，随着个人大学选择的心理成本增加，大学教育的MTE单调增加。本文的研究结论表明，大学教育回报越低的人，越是倾向于选择接受大学教育。这种逆向选择模式主要由无法观测的家庭背景特征所导致：家庭背景越差的个人接受大学教育的心理成本越高，但是一旦上了大学，其回报率越高。

关键词 教育 广义罗伊模型 边际处理效应 倾向得分 半参数局部工具变量

一 引言

教育回报是个人以及社会最优教育决定的重要组成部分。由于接受教育是有成本的，对于个体来讲，个人在选择学校教育的年限时，主要取决于额外增加一年教育所

* 刘生龙，清华大学公共管理学院、清华大学国情研究院，电子邮箱：liu_s_long11@163.com；郑世林，北京大学国家发展研究院，电子邮箱：shilinzheng@126.com。作者感谢中共中央宣传部全国哲学社会科学规划办公室清华大学国情研究院国家高端智库建设专项（20155010298）和国家自然科学基金面上项目“互联网基础设施对中国经济发展及公民政治参与的影响”（71573272）的资助。

获得的边际收益与边际成本之间的差异。如果收益大于成本，那么选择继续增加教育年限，直到收益与成本相等为止。对于政策层面来说，社会平均的边际教育回报既可以决定教育的投入水平，也可以决定法定的教育年限。

经典的明瑟方程中，由于选择性偏差的存在，核心解释变量与残差项之间存在相关关系，这种相关关系使得核心解释变量的最小二乘（OLS）估计是有偏的。具体来说，无法观测的个人能力越高，上大学的概率也就越高，而能力越高的人也更容易获得更高的收入。简单比较上大学和未上大学人群的工资差异，会由于能力变量的遗漏而高估大学教育的回报率。当然，由于测量误差的存在也有可能使得高等教育的回报率低估。但是，这种内生性偏差可以通过一系列传统的方法加以克服，比如工具变量法（Angrist & Krueger, 1991；Card, 1993；Harmon & Walker, 1995）、双胞胎设计（Ashenfelter & Rouse, 1998；Li et al., 2012）、断点回归设计（Zhong, 2015；刘生龙等, 2016）。

然而，传统的工具变量法都假定大学教育回报率是确定的，即估计值不会随着教育的差异而发生变化。Heckman et al. (2010) 提出一些方法，针对传统估计方法的假设进行检验。他们的研究发现大学教育回报率并不是确定的，而是一个随机变量。由于不同的个体在选择是否上大学时心理成本是不一样的，比如说家庭条件更好的人群在上大学时心理成本明显要低于家庭条件更差的，这是因为后者的大学教育学费对他们的家庭来说很可能就是一笔沉重的经济负担。此外，所处教育环境更差的个体在通过高考上大学时的痛苦程度要更高一些，这是因为教育条件越差的地方，如果个体要想考上大学就得花更多的时间和精力来应对高考。如果这种上大学的心理成本与大学教育回报之间并不相互独立，教育回报在不同的人群中就会出现根本的异质性（essential heterogeneity）。在这种情况下，传统的工具变量法、双胞胎设计和回归断点设计方法便无法刻画大学教育在不同人群中由大学选择动机不同而导致的异质性回报问题^①。

事实上，教育的异质性回报问题早就在一些文献中得到体现，在工具变量满足一系列假设条件下，Imbens & Angrist (1994) 证明工具变量估计可以识别局部平均处理效应（local average treatment effect, LATE），该效应可以用来测量由工具变量的变化导

^① 需要强调的是，本文的教育异质性回报指的是大学教育回报率 β 与大学教育选择动机是相关的，而大学教育选择动机是不可观测的，因而本文涉及到的大学教育回报的异质性是不可观测的异质性。传统的工具变量估计、双胞胎设计和回归断点设计方法也可以通过将样本人群分成城乡、男女等不同人群进行参数估计，根据不同人群的不同回报率来检测教育的异质性回报，但是这些异质性回报都是可观测的异质性，因而不同于本文中所提到的异质性回报。

致上大学的那部分人，即顺从者（compliers）的大学教育回报率^①。Heckman（2010）认为 LATE 既有可能可以回答，也有可能不能回答与政策相关的问题，在大部分情况下都不能回答。这是因为 LATE 估计具有较高的内部有效性，但是不一定满足外部有效性。对于顺从者来说，工具变量估计的大学教育回报率是无偏的，但是对于非顺从者来说^②，大学教育的回报率是未知的，而且他们的大学教育回报率与顺从者的大学教育回报率往往是不相同的。如果未来的政策变化导致现在的非顺从者成为顺从者，由于两组人群的大学教育回报率存在明显的异质性，由当前顺从者得出来的大学教育回报的估计结论往往不能够用来推断未来政策变化的效果。

为了解决大学教育由选择性动机导致的异质性回报问题，Heckman & Vytlacil（2005）和 Heckman（2010）提出边际处理效应（marginal treatment effect, MTE）方法，该方法基于广义罗伊模型（generalized Roy model）估计大学教育回报率如何随着大学选择意愿的不同而变化。借鉴该方法，本文将对中国大学教育在不同人群不可观测的异质性回报进行测算。到目前为止，MTE 的应用在文献中还非常少。Carneiro et al.（2011）对美国劳动力市场教育回报进行了估计；Nybom（2014）利用瑞典的数据对大学教育一生的收入回报进行了估计；Basu（2014）利用美国的医疗数据检验以人为中心的癌症治疗对治疗效果的影响。从作者收集到的文献来看，尚未发现有人利用中国的数据基于该方法对大学教育不可观测的异质性回报进行估计。

基于广义罗伊模型，本文利用中国综合社会调查 2010 年、2011 年、2012 年和 2013 年的数据，分别采用参数估计方法即极大似然估计（MLE）和非参数估计方法即半参数局部工具变量（local instrumental variable, LIV）方法对中国大学教育的 MTE 进行了估计。实证结果发现大学教育回报率在不同人群之中具有明显的异质性，个体选择大学教育的心理成本越高（意愿越低），大学教育的回报率越高。除了可以刻画由选择动机导致的异质性回报之外，MTE 估计的另一个比较大的优点在于可以通过对其进行加权平均得到大学教育的平均处理效应（average treatment effect, ATE），处理组人群大学教育的平均处理效应（average treatment effect of treated, ATT）和非处理组人群大学教育平均处理效应（average treatment effect of untreated, ATUT）。无论是 MLE 方法，

① 这些假设条件包括条件独立性假设、排除性假设、第一阶段假设和单调性假设，具体可以参看 Imbens & Angrist（1994）。

② 包括总是接受者（always-takers）和从不接受者（never-takers），有关这些定义可以参看 Imbens & Angrist（1994）和 Abadie（2003）。

还是半参数 LIV 方法，基于 MTE 计算出来的大学教育的平均处理效应都显著为正，这一估计结果与传统的工具变量法得出的结论比较类似。但是基于 MTE 估计出来的 ATE 明显低于传统工具变量法得出的估计结果（即 LATE），说明如果用传统的工具变量估计出来的大学教育回报率会高估所有人群的平均回报率。

与已有文献相比，本文的边际贡献在于：首先，本文利用中国数据基于广义罗伊模型估计大学教育的 MTE，估计结果表明随着个人选择大学教育的意愿的不同，不同大学选择动机的个人大学教育回报率存在根本上的异质性，该异质性基于传统的工具变量估计是无法刻画的。其次，本文基于 MTE 估计大学教育的 ATE、ATT、ATUT，并将这些估计值与 LATE 和 OLS 估计值进行比较，结果发现 OLS 严重低估大学教育回报率，LATE 明显高估大学教育回报率。ATUT 明显高于 ATT，再次说明大学教育在不同人群中的异质性回报。再次，本文分别用参数方法 MLE 和半参数方法 LIV 对大学教育的 MTE 进行参数估计和模拟，并对两种估计方法进行了比较，结果发现 MLE 估计的前提条件，即联合正态分布假设并没有得到满足，而半参数 LIV 估计形式更加灵活，估计结果也更加稳健。最后，本文的研究发现随着上大学的心理成本增加，大学教育回报率上升，大学教育对家庭背景更差的群体回报率也更高。这就意味着大学教育扩招使得更多的平民家庭子弟有机会接受大学教育，而这部分群体从大学教育扩招中获得的教育回报比家庭背景更好的群体要更高。因此，大学教育扩招从这个意义上来说有助于促进收入公平。

二 实证策略

（一）广义罗伊模型

本文基于广义罗伊模型进行实证分析，该模型是用于政策分析的一个基本的选择理论框架（Heckman & Vytlačil, 2007）^①。令 Y_1 为个体如果获得大学教育的潜在对数工资， Y_0 为个体如果没有获得大学教育的潜在对数工资。与 Carneiro et al. (2011) 和 Eisenhauer et al. (2015) 一样，潜在的收入可以定义如下：

$$Y_D = \mu_D(X) + U_D \quad (1)$$

这里的 D 分别为 1 和 0，分别表示上大学和未上大学。其中， $\mu_1(X) = E(Y_1 | X = x)$ ， $\mu_0(X) = E[Y_0 | X = x]$ 。对于个体而言，不可能同时观测到其参加大学和不参加大学的结果变量，可以观测到的结果变量为：

^① 广义罗伊模型最早的文献可以参看 Roy (1951)、Quandt (1958)。

$$Y = DY_1 + (1 - D)Y_0 = Y_0 + D(Y_1 - Y_0) \quad (2)$$

由于对于个体而言，上大学的处理效应无法观测得到，一般需要更多的样本来估计大学教育平均的处理效应（ATE），即 $E(Y_1 - Y_0 | X = x)$ 。然而，由于选择偏差的存在，通过估计处理组（上大学的人群）与非处理组（未上大学的人群）之间平均的工资差异并不能准确给出大学教育的处理效应。由于方程（2）式中的 X 不一定与 (U_1, U_0) 相互独立，于是大学教育的处理效应为 $Y_1 - Y_0 = \beta = E(\mu_1(X) - \mu_0(X) + U_1 - U_0)$ 。在给定 $X = x$ 时，上大学的平均处理效应为 $\bar{\beta}(x) = E(\beta | X = x) = \mu_1(X) - \mu_0(X)$ 。而在给定 $X = x$ 时，那些已经选择了上大学的个体的大学教育平均的处理效应（ATT）为 $E(\beta | X = x, D = 1) = \bar{\beta}(x) + E(U_1 - U_0 | X = x, D = 1)$ 。

为了解决由选择性导致的估计偏误，微观经济计量文献通常基于排除性约束工具变量法对处理效应进行估计。也就是找到一个工具变量 Z ，该工具变量不包含在 X 中，并且与处理变量 D 相互独立。在最简单的情形下，如果处理效应 β 是一个常数，即对于每一个个体而言，在给定 $X = x$ 时，对于每一个个体来说处理效应都是一样的，那么根据工具变量估计出来的 ATE 和 ATT 都是一样的。另外，虽然处理效应存在异质性，但是在处理组个体和非处理组个体之间是完全随机分布的，在这种情况下，工具变量估计出来的 ATE 和 ATT 也是一样的。

然而在实践中 β 在给定 $X = x$ 时往往取决于是否受到处理，这种情形被 Heckman et al. (2006) 称为根本的异质性。Heckman et al. (2010) 建议如果研究者关心的是 ATE 或者与政策相关的处理效应（policy relevant treatment effect, PRTE），在存在根本上的异质性时就不要使用传统的工具变量法进行参数估计。而为了识别处理效应并且与已有的关于处理效应的文献保持一致，Heckman & Vytlacil (2005) 和 Caneriro et al. (2011) 提出用一个标准的隐性变量（latent variable）离散选择模型来代表个体决定是否上大学。令 I_D 代表个体上大学的净收益，该净收益取决于可观测到的变量 Z 和不可观察到的变量 V ：

$$I_D = \mu_D(Z) - V \quad (3)$$

这里当 $\mu_D(Z) \geq V$ 时，个体选择上大学（ $D = 1$ ）；反之，当 $\mu_D(Z) < V$ 时，个体选择不上大学（ $D = 0$ ）。方程式（1）~（3）就构成了广义罗伊模型，在该模型中， (Z, X) 是可以观测的，而 (U_0, U_1, V) 是不可观测的。这里的 Z 向量包含 X 的部分或者全部变量，同时 Z 向量也包含方程（1）式中所不包含的变量，即排除性的工具变量。这里的 V 被假定为连续的随机变量，其分布函数 F_V 是严格单调递增的，且 V 一般地取决于 U_1 和 U_0 。在给定 X 条件下， Z 与 (U_0, U_1, V) 是相互独立的，即 $(U_0, U_1, V) \perp Z | X$ 。

定义 $P(Z)$ 为个体在给定 Z 的条件下接受大学教育的概率, 即 $P(Z) = \Pr(D = 1 | Z) = F_{U_D | X}(\mu_D(Z))$, 这里的 $F_{U_D | X}(\cdot)$ 表示在给定 X 条件下 U_D 的分布。这里的 $P(Z)$ 被称为倾向得分。这里假定 $U_D = F_V(V)$ 服从 0 到 1 之间的均匀分布, 即 $U_D \sim \text{Uni}[0, 1]$ 。这样, 不同的 U_D 值代表 V 不同的分位数。由于 V 以负值进入方程式 (3) 中, 因此, 它反映了上大学的一种成本, 其数值越大则这种成本越高。

(二) 边际处理效应 (MTE)

最早提出 MTE 参数估计的文献为 Willis & Rosen (1979) 和 Bjorklund & Moffitt (1987), Heckman & Vytlačil (1999, 2005, 2007) 对其进行了扩展。根据这些文献, MTE 的定义如下:

$$\text{MTE}(x, u_D) = E(Y_1 - Y_0 | X = x, U_D = u_D) \quad (4)$$

其意义为在给定个人的特征向量满足 $X = x$ 和 $U_D = u_D$ 时, 个人接受大学教育的回报率。前面已经指出 U_D 服从单位均匀分布, 因此 MTE 在不同的 u_D 上的估计值实际上反映的就是个人的大学教育回报如何随着个人大学就读意愿的不同而变化。

Heckman & Vytlačil (1999, 2005, 2007) 提出局部工具变量方法 (local instrumental variable, LIV) 对 MTE 进行估计。由于当 $P(Z) = u_D$ 时, 个人选择上大学和不上大学的意愿是相同的, MTE 可以通过个人的条件期望收入对上大学的概率进行求导来获得:

$$\text{MTE}(x, p) = \frac{\partial E(Y | X = x, P(Z) = p)}{\partial p} \quad (5)$$

由方程 (2) 式可以得到:

$$\begin{aligned} E(Y | X = x, P(Z) = p) \\ = E(Y_0 | X = x, P(Z) = p) + E(Y_1 - Y_0 | X = x, D = 1, P(Z) = p)p \end{aligned} \quad (6)$$

很明显, MTE 只能够在 $P(Z | X)$ 的支撑集上进行估计。在条件支撑集 ($P(Z | X)$) 范围内, 根据 (5) 式就可以对所有的 u_D 边际上个人在选择上大学或者不上大学的心理成本相同时的大学教育回报率进行估计。

边际处理效应有两个方面的明显优势, 首先是它可以对不同人群大学教育的异质性回报进行参数估计, 即它可以估计大学回报率如何随着 X 或者 u_D 的不同而发生变化。通过估计不同 u_D 下的 MTE, 有可能识别在任何一个边际上的大学教育回报率。Heckman & Vytlačil (1999, 2005, 2007) 的研究表明所有的其他处理效应, 包括 ATE、ATT 和 LATE 等, 参数估计都可以表示成 MTE 不同的加权平均值。也就是说, 如果内生性导致其他一些处理效应不能直接估计时, 可以通过 MTE 对其他处理效应进行估

计。具体来说，ATE、ATT、ATUT、LATE 和 OLS 可以通过如下方程式进行估计：

$$\Delta_j(x) = \int_0^1 \text{MTE}(x, u_D) h_j(x, u_D) d u_D \quad (7)$$

这里，当 $h_j(x, u_D) = 1$ 时， $\Delta_j(x)$ 的估计值是 ATE；当 $h_j(x, u_D) = [\int_{u_D}^1 f(p) X = x) dp] \cdot \frac{1}{E(P | X = x)}$ 时， $\Delta_j(x)$ 的估计值是 ATT；当 $h_j(x, u_D) = [\int_0^{u_D} f(p) X = x) dp] \cdot \frac{1}{E[(1 - P) | X = x]}$ 时， $\Delta_j(x)$ 的估计值是 ATUT；当 $h_j(x, u_D) = [\int_{u_D}^1 (p - E(p | X = x)) f(p | X = x) dp] \cdot \frac{1}{\text{Var}(p) | X = x}$ 时， $\Delta_j(x)$ 的估计值是 LATE；当 $h_j(x, u_D) = \frac{E[U_1 | X = x, U_D = u_D] h_1(x, u_D) - (U_0 | X = x, U_D = u_D) h_0(x, u_D)}{\text{MTE}(x, u_D)}$ 时， $\Delta_j(x)$

的估计值是 OLS。

(三) MTE 的参数估计与非参数估计

与现有实证文献一样，本文假定 $u_D(x)$ 和 $u_D(Z)$ 都是线性的：即 $u_1(X) = X\alpha_1$ ， $u_0(X) = X\alpha_0$ ， $u_D(Z) = Z\beta$ 。在这种情况下：

$$E(Y | X = x, P(Z) = p) = X\alpha_0 + px[\alpha_1 - \alpha_0] + K(p) \quad (8)$$

由 (6) 式可知，这里的 $K(p) = E(U_1 - U_0 | D = 1, P(Z) = p)$ 。

现有的文献对 MTE 已经提出了多种估计方法，首先是极大似然估计。极大似然估计是一种参数估计方法，它假定 (U_0, U_1, V) 满足联合正态分布 (jointly normally distributed)，即 $(U_0, U_1, V) \sim N(0, \delta)$ 而且与 (X, Z) 是相互独立的。遵循传统的离散选择模型，这里将 V 的方差正则化为 1，于是：

$$\begin{aligned} \text{MTE}(x, u_D) &= x(\alpha_1 - \alpha_0) + E(U_1 - U_0 | U_D = u_D) \\ &= x(\alpha_1 - \alpha_0) + E(U_1 - U_0 | V = \Phi^{-1}(u_D)) \\ &= x(\alpha_1 - \alpha_0) + (\sigma_{1v} - \sigma_{0v}) \Phi^{-1}(u_D) \end{aligned} \quad (9)$$

这里 $\sigma_{1v} = \text{Cov}(U_1, V)$ ， $\sigma_{0v} = \text{Cov}(U_0, V)$ ， $\Phi^{-1}(\cdot)$ 是标准正态累积分布函数的逆。通过极大似然估计可以将参数 $(\alpha_1, \alpha_0, \sigma_{1v}, \sigma_{0v})$ 估计出来，然后将估计出来的参数代入 (9) 式中就可以对 MTE 进行估计。

另一种参数估计的方法是多项式估计，Basu et al. (2007) 曾提出一种灵活的多项式近似方法对 $K(p)$ 进行模拟，即对如下式子进行估计：

$$E(Y | X = x, P(Z) = p) = X\alpha_0 + px[\alpha_1 - \alpha_0] + \sum_{j=1}^k \delta_j p^j \quad (10)$$

根据 OLS 可以对 $(\alpha_1, \alpha_0, \delta_j)$ 进行参数估计, 然后代入 (10) 式, 将 (10) 式对 P 进行求导即可得出 MTE 的估计值。

$$\widehat{\text{MTE}} = x(\widehat{\alpha}_1 - \widehat{\alpha}_0) + \sum_{j=1}^k \widehat{\delta}_j j p^{j-1} \quad (11)$$

还有一种方法就是直接对 $K(p)$ 进行非参数估计。在 MTE 估计实践中, 由于 X 包含的变量很多, 要想直接通过非参数方法对 MTE 进行估计是非常困难的^①。因此, 在进行参数估计时, 还需要增加一些假设条件, 即 (U_0, U_1, V) 与 (X, Z) 是相互独立的, 这一假设条件要强于 $(U_0, U_1, V) \perp Z/X$ 。这是同时存在工具变量和选择模型时标准的假设前提, 在这个假设前提下, MTE 在 X 和 u_D 上是加性可分的 (additively separable), 这样可以通过半参数方法对 MTE 进行估计。Caneriro et al. (2011) 和 Nybom (2014) 采用 Robinson (1988) 提出的半参数 LIV 估计方法对 MTE 进行估计, 本文也使用这种估计方法。MTE 的一个局限就是只能够在 P 的支撑集上进行识别, 然而由于计算出来的 P 的支撑集往往覆盖 $Uni [0, 1]$ 上很大的范围, 因此即使在 P 的支撑集上计算出来的 MTE 的信息量就已经很大了, 很多研究者认为它所包含的信息要明显优于标准的 LATE 所包含的信息^②。

三 数据、变量及其描述性统计

(一) 数据

本文使用的数据是中国综合社会调查 (Chinese General Social Survey, CGSS) 2010 年、2011 年、2012 年和 2013 年的数据。总体上, CGSS 的原始样本 2010 年为 11783 人、2011 年为 5620 人、2012 年为 11765 人、2013 年为 11483 人^③。被调查者的年龄在 18 岁以上, 调查的信息包括个人基本特征、教育、移民、收入和社会关系等各个方面。由于本文在估计大学教育的边际处理效应时使用的排除性工具变量之一是否受到 1999 年大学教育扩招的影响, 而中国学生参加高考的年龄普遍为 18 岁, 1981 年之后出生的人会受到 1999 年高等教育扩招的影响, 之前的人群则不会受到扩招的影

① 即人们常说的维度诅咒 (curse of dimension)。

② 正是在这个意义上, 针对 Imbens (2010) 年提出的“好的 LATE 比什么都没有要强”, Carneiro et al. (2011) 指出“好的 MTE 要强于 LATE”。

③ CGSS 数据由中国人民大学中国调查与数据中心负责执行, 有关数据的详细介绍可以查看网址: <http://www.chinagss.org/>。

响，本文将样本集中于1981年前后各10年出生的人群。由于本文要估计大学教育的边际处理效应，实证研究时存在是否接受大学教育的一个选择性质量问题。只有高中及以上学历者才存在上不上大学的选择性问题，高中以下学历者不存在上不上大学的选择性问题，因为这部分人是大学教育的从不接受者（never takers）。与Carneiro et al. (2011)一样，初中及以下教育程度者也从样本中排除。最终，本文进入回归分析最终的有效观测值为4692人。

（二）变量

本文的变量包括三类，分别是被解释变量（ Y ）、解释变量（ X ）和排除性的工具变量（ $Z \setminus X$ ）。本文的被解释变量是对数收入（ $\ln income$ ），为了使各年份的收入具有可比性，我们根据各年份的消费者物价指数（CPI）对个人收入按照2010年的不变价格进行平减^①。核心解释变量是是否受到了大专及以上教育，如果是大专、本科或者研究生以上教育者，那么 $D=1$ ，高中教育程度者 $D=0$ 。

其余解释变量包括年龄（ age ）及年龄的平方项。CGSS还给出了家庭成员信息，根据该信息可以计算出个体的家庭规模（ $family$ ）。此外，解释变量还包含一系列虚拟变量，分别为是否单身（ $single$ ），已婚有配偶或者再婚有配偶的个体取值为0，其余取值为1；是否跨省移民（ $immigrant$ ），5年前居住省份与现居住省份不一样取值为1，其余取值为0；是否为男性（ $gender$ ），男性取值1，女性取值0；是否农村户口（ $rural$ ），农村户口取值1，非农户口取值0；是否为汉族（ $race$ ），汉族取值1，其他民族取值0；14岁时父亲是否正在工作（ $fjob14$ ），在工作取值1，否则取值0；14岁时母亲是否正在工作（ $mjob14$ ），在工作取值1，否则取值0。最后，为了控制不同省份和调查年份对个人收入的影响，解释变量还包含所有的省份虚拟变量和调查年份虚拟变量。

本文中排除性的工具变量（ $Z \setminus X$ ）共有两个。第一个是是否出生在1981年之后（ $expansion$ ），如果是，那么他们会受到1999年高等教育扩招的影响，取值为1，1981年之前出生的人取值为0。Zhong（2015）也使用了高等教育扩招作为工具变量估计了大学教育对健康的影响。第二个排除性工具变量是配偶的受教育年限（ $mschool$ ），由于配偶的受教育年限会直接影响个体的受教育年限，但不会直接影响其收入水平，因此也有一些文献采用配偶的受教育年限作为教育的工具变量（Chen & Hamori, 2009）。CGSS给出了受访者配偶的受教育程度，根据受教育程度我们可以给定受访者母亲的受

^① 各年份CPI来自《中国统计年鉴》（2014年）。

教育年限,分别是未上学0年、小学6年、初中9年、高中12年、中专13年、大专15年、大学16年、研究生及以上19年。

(三) 变量的描述性统计

表1给出了所有变量的描述性统计。可以看到,获得大学教育的个体平均对数收入是10.44,比未获得大学教育的对数收入要高出0.48,两者之间的差异在1%的显著性水平下通过检验。本文样本中,获得大学教育的人群比重达到了60%,这就意味着在参加高考的人群当中有一半以上的人都可以获得大专及以上学历的机会,在一定程度上反映了1999年高等教育扩招之后,中国大专及以上学历获得者迅速增加的情形。由于获得大学教育的人群与未获得大学教育的人群的其他解释变量也存在显著差异,仅仅比较获得大学教育和未获得大学教育之间人群的收入差距并不一定就是大学教育对收入的因果影响,因此必须进行更加细致的实证检验。

表1 变量的描述性统计

变量	所有样本(n=4692)		D=1(n=2795)		D=0(n=1897)	
	均值	方差	均值	方差	均值	方差
被解释变量(Y)						
<i>lnincome</i>	10.24	0.86	10.44	0.81	9.96	0.86
解释变量(X)						
<i>D</i>	0.60	0.49				
<i>age</i>	31.22	5.31	31.07	5.07	31.44	5.65
<i>family</i>	3.42	1.31	3.24	1.23	3.68	1.38
<i>single</i>	0.72	0.45	0.69	0.46	0.76	0.43
<i>fjob14</i>	0.90	0.30	0.90	0.30	0.90	0.30
<i>mjob14</i>	0.76	0.42	0.77	0.42	0.75	0.43
<i>immigrant</i>	0.78	0.42	0.78	0.41	0.77	0.42
<i>rural</i>	0.24	0.43	0.12	0.33	0.42	0.49
<i>gender</i>	0.54	0.50	0.52	0.50	0.57	0.49
<i>race</i>	0.94	0.24	0.94	0.24	0.93	0.25
排除性工具变量(Z\X)						
<i>expansion</i>	0.50	0.50	0.52	0.50	0.46	0.50
<i>mschool</i>	7.48	4.15	8.32	4.08	6.25	3.95

资料来源:根据中国综合社会调查(CGSS)2010-2013年数据计算得到。

四 估计结果

（一）OLS 与 2SLS 估计

尽管本文感兴趣的是大学教育的异质性的处理效应，但是作为对照基准，本文首先给出大学教育回报的 OLS 估计和传统的 2SLS 估计。表 2 第（1）列 OLS 估计结果表明获得高等教育的人群比未获得高等教育的人群收入要高出 30.7%，由于本文中大专及以上学历者的平均受教育年限比高中学历者多出 4 年，因此 OLS 估计的大学教育每年的回报率仅为 7.68%。第（2）列用 1999 年的大学扩招作为工具变量通过 2SLS 方法估计大学教育的回报率，结果发现大学教育的总体回报为 103.8%，年均回报率为 25.95%。第（3）列用配偶的受教育年限作为工具变量，结果发现大学教育的回报率达到了 126.7%，大学教育的年均回报率达到了 31.68%。而当同时引入大学扩招和配偶的受教育年限两个工具变量时，我们发现大学教育的回报率为 124.3%，年均回报率为 31.08%。表 2 的最后一行给出了使用工具变量时，工具变量有效性的第一阶段 F 统计值，可以看到所有工具变量设定下，F 统计值都大于 10，根据拇指法则（rule of thumb），不存在弱工具变量的问题。第（4）列因为同时引入两个工具变量，因此给出工具变量过渡约束识别的 Sargan 统计检验值，结果发现不存在过度约束识别问题。因此，工具变量在所有设定下都是有效的。

第（2）列和第（3）列用了同样的控制变量和同样的观测值，但是仅仅由于工具变量的不同，估计的系数就出现了明显的不同，一个重要的原因就是不同的工具变量所影响的人群不同，而不同的人群大学教育的回报率很可能也不一样。由于工具变量并不能清楚地定义谁是顺从者（compliers），此时利用 MTE 来估计不同人群大学教育的回报率显然更加有意义（Carneiro et al., 2011）。

表 2 大学教育回报率的 OLS 估计和 IV 估计

	(1) OLS	(2) IV1	(3) IV2	(4) IV3
<i>D</i>	0.307 *** (0.023)	1.038 ** (0.532)	1.267 *** (0.216)	1.243 *** (0.204)
控制变量	Yes	Yes	Yes	Yes

续表

	(1) OLS	(2) IV1	(3) IV2	(4) IV3
观测值数	4692	4692	4692	4692
First-stage F statistic		10.76	72.25	40.80
Sargan statistic (p-value)				0.702

注：括号中的数值是标准误；*、** 和 *** 分别表示在 10%、5% 和 1% 的显著性水平下通过检验；所有估计包括省份和观察年份虚拟变量。

资料来源：根据中国综合社会调查（CGSS）2010 - 2013 年数据计算得到。

（二）MTE 的参数估计结果

本文首先采用参数估计方法即 MLE 方法对中国大学教育的 MTE 进行估计。基于方程式 (9)，表 3 报告了 α_1 、 α_0 和 β 的估计。第 (1) 列报告的是处理组人群的估计，即 α_1 。第 (2) 列报告的是非处理人群的估计，即 α_0 。第 (3) 列报告的是第一阶段的估计，即 β 。从第 (3) 列可以看到，两个工具变量，高等教育扩张与配偶的教育年限都对获得大学教育的概率产生显著正向的影响。由于工具变量都在 1% 的显著性水平下通过了检验，再次说明不存在弱工具变量问题。此外，家庭人口规模对大学教育获得产生显著的负向影响，说明子女越多的家庭，上大学的概率就越低，子女数量和质量之间存在着取舍 (trade-off) 关系。14 岁时母亲是否有工作、是否跨省移民、是否农村户口对大学教育获得产生显著的影响，说明家庭背景对上大学与否产生显著的影响，很明显，家庭背景越好的子女上大学的概率越高。

表 3 正态转换回归模型的 MLE 估计

	$\mu_1(X)$ (1)	$\mu_0(X)$ (2)	$\mu_Z(X)$ (3)
<i>age</i>	0.361 *** (0.037)	0.094 ** (0.037)	
<i>Square of age</i>	-0.005 *** (0.001)	-0.001 ** (0.001)	
<i>family</i>	-0.018 (0.012)	-0.047 *** (0.015)	-0.098 *** (0.016)
<i>single</i>	0.090 ** (0.037)	0.130 ** (0.055)	0.030 (0.052)
<i>fjob14</i>	0.070 (0.052)	-0.018 (0.064)	0.052 (0.076)

续表

	$\mu_1(X)$ (1)	$\mu_0(X)$ (2)	$\mu_Z(X)$ (3)
<i>mjob14</i>	-0.119 *** (0.034)	-0.172 *** (0.048)	0.159 *** (0.052)
<i>immigrant</i>	0.027 (0.040)	0.073 (0.065)	0.134 ** (0.067)
<i>rural</i>	0.342 *** (0.080)	0.107 (0.099)	-0.916 *** (0.050)
<i>gender</i>	0.307 *** (0.027)	0.415 *** (0.038)	-0.093 ** (0.040)
<i>race</i>	0.015 (0.064)	0.244 ** (0.098)	-0.005 (0.095)
Instruments (<i>Z</i>)			
<i>expansion</i>			0.217 *** (0.045)
<i>mschool</i>			0.044 *** (0.005)
观测值数	4692	4692	4692

注：括号中的数值是标准误；*、**和***分别表示在10%、5%和1%的显著性水平下通过检验；所有估计包括省份和观察年份虚拟变量。

资料来源：根据中国综合社会调查（CGSS）2010-2013年数据计算得到。

MTE估计的一个重要检验就是看高等教育的处理效应是否随着高等教育的选择机的不同而发生变动。一个简单的检验法则就是检验 $\mu_1(X)$ 与 $\mu_0(X)$ 的逆米尔斯 σ_{1v} 和 σ_{0v} 之间的差异是否为0，如果两者之间的差异没有通过显著性检验，那么MTE估计值的斜率不会随着高等教育的选择而发生变化，在这种情况下，MTE的估计值其实就是ATE，反之，MTE估计值的斜率随着高等教育的选择发生变化，MTE的估计值不能够简单地等同于ATE。MLE的估计结果表明 $\sigma_{1v} - \sigma_{0v} = 0.350$ ，其标准误是0.179，估计值在10%的显著性水平下通过了检验。说明MTE估计值随着高等教育的选择会存在根本的异质性，在这种情况下，仅仅估计ATE和LATE不能完整地刻画高等教育在不同人群中异质性的处理效应。

基于方程（9）式估计出来的MTE在图1中进行刻画，其中阴影部分是基于bootstrap方法计算出来的95%置信区间。从图1可以看到，在不同的高等教育选择 U_D

上，大学教育的处理效应存在着明显的异质性^①。在 U_D 小于 0.13 时，MTE 的估计值虽然大于 0，但是并没有通过显著性检验。当 U_D 大于或等于 0.14 以后，MTE 的估计值开始显著为正，而且随着 U_D 的增加，大学教育的 MTE 估计值逐渐增大，从 0.609 增加到 1.801。

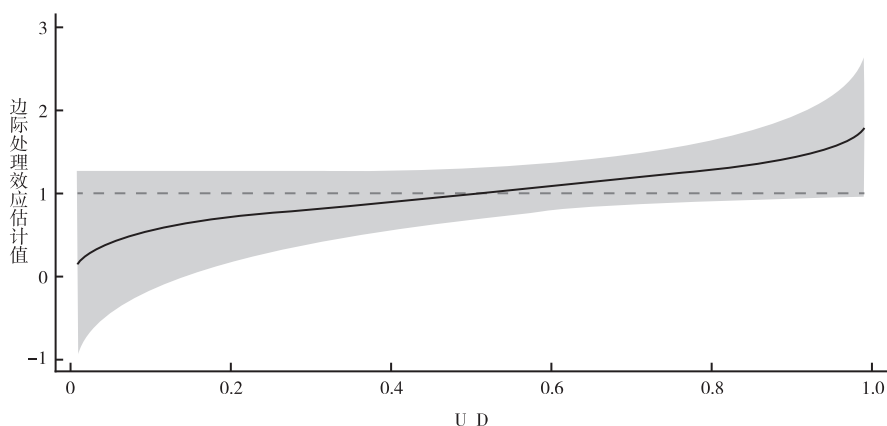


图 1 基于正态选择模型的 MTE 估计^②

资料来源：根据中国综合社会调查（CGSS）2010 - 2013 年数据计算得到。

前面已经提到过， U_D 越大意味着更多的无法观测的心理成本因素使得个体不愿意选择大学教育，也就是说，本文的研究结果表明个体选择大学教育的意愿越低，大学教育的回报率更高。一个可能的解释是对于那些选择大学教育心理成本更高的个体而言，家庭背景往往比较差，上大学的学费往往会带来很大的经济压力，因此上大学的意愿并不高。但是，这部分人一旦接受大学教育使他们从原有的家庭环境中走出来，比如说从农村走入城市工作，此时大学教育的回报率会非常高。而对于那些选择大学教育心理成本更低的人而言，家庭条件往往更好，即使不上大学这部分人往往也能够找到体面的工作。因此，即使上了大学之后，大学教育的回报率并不会比家庭背景更差的人高。尽管中国的大学教育回报率在不同人群中表现出根本的异质性，但中国大

① 为了准确估计 MTE 的置信区间，本文 bootstrap 的次数为 200。

② 图 1 中刻画的是如下公式： $\Delta^{MTE}(x, u_D) = x(\alpha_1 - \alpha_0) + (\sigma_{1v} - \sigma_{0v})\Phi^{-1}(u_D)$ ，这里的 X 包含表 1 中所有的解释变量，以及省份和观察年份虚拟变量。阴影部分使用 bootstrap 重复 200 次计算出来的 95% 的置信区间。

学教育的 MTE 在绝大部分人群中都显著为正，与前面 OLS 和 IV 估计结果以及先前已有的文献一样，再次表明大学教育对收入增加具有重要的意义。

MTE 的参数估计的另一种方法就是多项式估计，具体来说就是对方程式 (10) 进行估计，然后对 p 求导数之后得到 MTE 的估计值，即方程式 (11)。作为对 MLE 估计的一个基本对应，本文基于三阶多项式模型对方程 (10) 式进行估计。在三阶多项式估计中，如果 p 的二次项和三次项前面的系数为 0，那么方程式 (11) 估计出来的 MTE 就是一个常数，此时，教育的边际处理效应不随选择意愿的变化而变化^①。利用本文的样本数据估计得到的 p 的二次方前面的系数虽然没有通过显著性检验，但是 p 的三次方前面的系数显著为负。他们联合概率为 0 的卡方统计值的 p 值为 0.0335，在 5% 的显著性水平下通过检验，说明参数多项式的 MTE 估计并不是一个常数，再一次说明对于不同人群来说，教育的回报存在根本的异质性。

(三) MTE 的半参数局部工具变量 (LIV) 估计

参数方法对 MTE 估计存在一定的局限性，MLE 假定 (U_0, U_1, V) 满足联合正态分布，这一假设前提往往不能够得到满足，参数多项式估计结果通常随着多项式阶数的不同而出现较大的差异。Carneiro et al. (2011) 提出了一种更加稳健的半参数估计方法，即半参数局部工具变量 (locally instrumental variable, LIV) 方法来对 MTE 进行估计。相较于参数估计方法，半参数 LIV 方法被认为是更加灵活、更加稳健的估计方法。从目前较少量的 MTE 研究文献来看，研究者普遍认为半参数 LIV 方法在 MTE 估计上具有独特的优势。

本文前面已经指出 MTE 的半参数估计一个局限就是只能够在 p 的支撑集上进行识别，因此在进行半参数 LIV 估计时需要首先对 p 的支撑集进行估计。图 2 给出了大学教育和非大学教育的倾向得分分布情况，可以看到大学教育和非大学教育的倾向得分有较大范围共同支撑集，该支撑集所包含的范围从 0.05 到 0.95。一个大的共同支撑集意味着，在支撑集下面得到的半参数 LIV 估计结果几乎可以代表所有样本情况下的估计结果。

前面已经指出，方程式 (8) 可以直接通过部分线性回归 (partially linear regression) 方法进行估计。本文分两步对该过程进行估计，第一步基于概率模型

^① 由于 MTE 估计的主要方法是 MLE 估计和半参数局部工具变量 (LIV) 的估计，本文没有在这里列出 MTE 参数多项式的估计结果，读者如果对参数多项式估计结果感兴趣可以向作者索取。

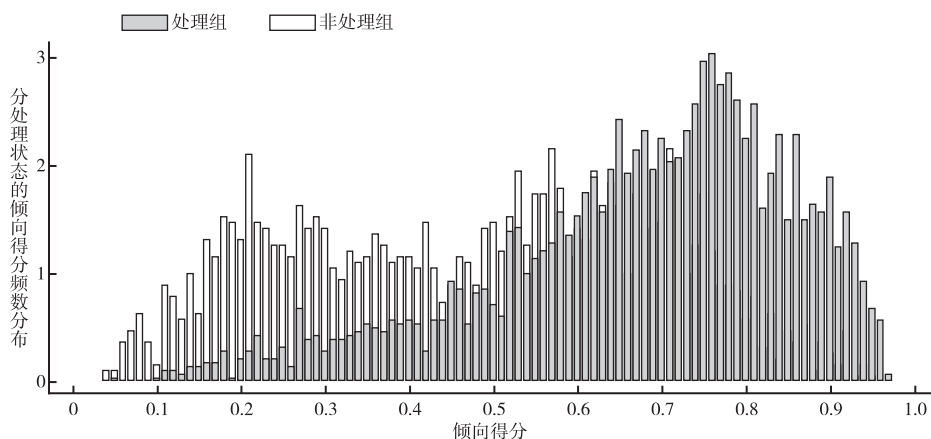


图 2 基于大学教育获得与否的倾向得分分布情况

资料来源：根据中国综合社会调查（CGSS）2010-2013 年数据计算得到。

(probit model) 对 $p(Z)$ 进行估计，然后利用估计的 $p(Z)$ 对 α_1 和 α_0 进行估计^①。第二步，本文利用 Robinson (1988) 提出的部分线性回归方法进行估计，估计的结果如表 4 所示。

半参数 LIV 估计在非参数估计的部分由于在 $p(Z)$ 的首末两端观察值在共同支撑集上的数量太小容易导致结果不精确，与 Carneiro et al. (2011) 一样，本文集中于倾向得分数值位于 5% ~ 95% 分位上的观测值进行估计。在估计时，本文还根据局部多项式回归的拇指法则确定的最优窗宽为 0.08，表 4 给出了 α_1 和 α_0 的估计结果。接下来考虑估计 $K(P)$ ，由方程式 (8) 可知：

$$E(Y - X\alpha_0 - PX[\alpha_1 - \alpha_0])P = K(P)$$

利用 $Y - X\hat{\alpha}_0 - \hat{P}X[\hat{\alpha}_1 - \hat{\alpha}_0]$ 对 \hat{P} 进行局部多项式回归就可以得到 $K(P)$ 及其对 P 的偏导数 $\partial K(P) / \partial \hat{P}$ 。最后，得到的半参数 LIV 估计的 MTE 为 $X[\hat{\alpha}_1 - \hat{\alpha}_0] + \partial K(\hat{P}) / \partial \hat{P}$ 如图 3 所示。

可以看到，基于半参数 LIV 估计的结果与 MLE 估计的结果比较相似，随着 U_d 增加，MTE 逐渐增大。不过与 MLE 相比较，LIV 估计的 MTE 增长幅度没有 MLE 估计的那样大。在 U_d 小于 0.40 之前，MTE 没有显著不为 0，自 U_d 大于或等于 0.40 以后，MTE 的估计值都显著大于 0，从 0.53 增加到 3.36。

① 基于 logit 回归模型的估计结果与本文报告的回归结果基本相同。

表 4 半参数 LIV 估计

	$\mu_0(X)$ (1)	$\mu_1(X) - \mu_0(X)$ (2)
<i>age</i>	0.141 (0.093)	0.156 (0.165)
<i>Square of age</i>	-0.003* (0.001)	-0.002 (0.003)
<i>family</i>	-0.043 (0.038)	0.036 (0.065)
<i>single</i>	0.142 (0.131)	-0.038 (0.218)
<i>fjob14</i>	-0.086 (0.171)	0.228 (0.282)
<i>mjob14</i>	-0.228* (0.137)	0.147** (0.075)
<i>immigrant</i>	-0.075 (0.169)	0.128 (0.282)
<i>rural</i>	-0.125 (0.287)	0.628 (0.525)
<i>gender</i>	0.523*** (0.082)	-0.258** (0.128)
<i>race</i>	0.436* (0.263)	-0.546 (0.411)
观测值数	4692	4692

注：括号中的数值是标准误；*、**和***分别表示在10%、5%和1%的显著性水平下通过检验；所有估计包括省份和观察年份虚拟变量。

资料来源：根据中国综合社会调查（CGSS）2010-2013年数据计算得到。

图3中刻画的公式为： $X[\hat{\alpha}_1 - \hat{\alpha}_0] + \partial \hat{K}(\hat{P}) / \partial \hat{P}$ 。这里的 $\hat{K}(\hat{P})$ 是基于对方程式(8)中的 $K(p)$ 基于非参数的估计，基于拇指法则确定的窗宽为0.08。这里的X包含表1中所有的解释变量，以及省份和观察年份虚拟变量。阴影部分使用bootstrap重复200次计算出来的95%的置信区间。

（四）半参数 LIV 估计稳健性检验

到目前为止，半参数 LIV 估计被认为是 MTE 估计相对来说最稳健的估计方法，为了对其稳健性进行检验，这里在不同的设定下再次对方程式(8)进行半参数估计。具体来说，将样本中中途辍学的人群排除掉^①，将样本精炼至保留高中和大学以上正常毕业的那部分人群。另外改变非参数估计的最优窗宽，将窗宽设定为0.16。

① “精炼样本”是指排除中途辍学的人群之后的样本。

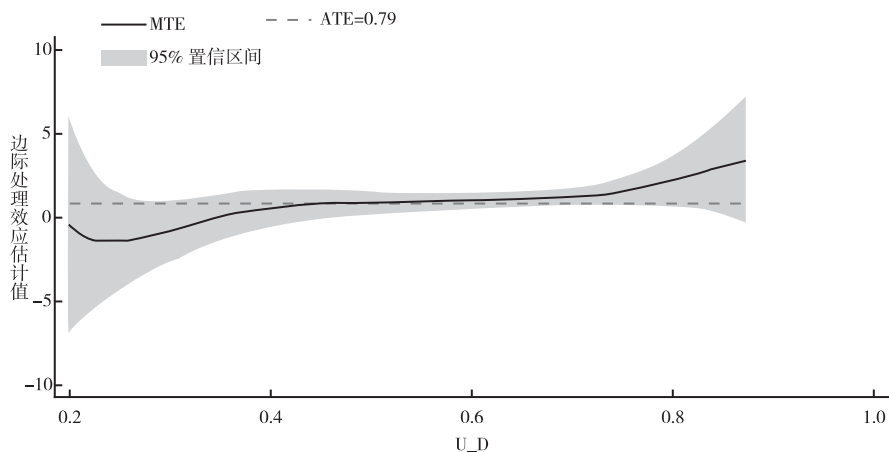


图 3 基于半参数 LIV 估计的 MTE

资料来源：根据中国综合社会调查（CGSS）2010 - 2013 年数据计算得到。

表 5 和图 3 给出了稳健性的检验结果，从中可以看到当窗宽从 0.08 变为 0.16，以及非正常毕业的样本被排除之后，教育的边际效应估计结果并没有发生很大变化。随着 U_D 增加，MTE 仍然表现出逐渐增大的趋势。

除了利用精炼样本和改变窗宽进行半参数估计的稳健性检验之外，与 Cornelissen et al. (2018) 一样，本文还通过参数多项式估计进行稳健性检验^①。分别选择二阶多项式和四阶多项式进行 MTE 估计，估计的结果仍然发现大学教育的边际回报率在不同的人群上具有明显的异质性，而且 MTE 随着 U_D 增加而变大。所有的估计结果都表明随着 U_D 增加，大学教育回报率增加。

表 5 半参数 LIV 的稳健性检验

	$\mu_0(X)$ (1)	$\mu_1(X) - \mu_0(X)$ (2)
<i>age</i>	0.122 (0.109)	0.184 (0.178)
<i>Square of age</i>	-0.002 (0.002)	-0.002 (0.003)

① 限于篇幅以及本文中的图表过多，不同排除性工具变量选择的估计结果不再一一列出，读者如果感兴趣可以向作者索取。

续表

	$\mu_0(X)$ (1)	$\mu_1(X) - \mu_0(X)$ (2)
<i>family</i>	0.011 (0.033)	-0.034 (0.059)
<i>single</i>	0.148 (0.115)	-0.054 (0.172)
<i>fjob14</i>	-0.072 (0.220)	0.204 (0.331)
<i>mjob14</i>	-0.274 ** (0.124)	0.224 (0.199)
<i>immigrant</i>	-0.154 (0.157)	0.245 (0.230)
<i>rural</i>	0.142 (0.173)	0.146 (0.329)
<i>gender</i>	0.528 *** (0.115)	-0.271 (0.176)
<i>race</i>	0.398 (0.265)	-0.507 (0.406)
观测值数	4509	4509

注：括号中的数值是标准误；*、**和***分别表示在10%、5%和1%的显著性水平下通过检验；所有估计包括省份和观测年份虚拟变量。

资料来源：根据中国综合社会调查（CGSS）2010-2013年数据计算得到。

半参数估计结果再次证明，大学教育回报随着大学选择的心理成本增加而增加。前面第一阶段的研究表明，越是家庭背景差的个体选择上大学的心理成本越高，而这部分人群上大学之后的回报率往往更高。1999年中国高等教育扩招之后，大学教育逐渐从精英教育走向平民教育，这其实就意味着高等教育扩招使得更多家庭背景较差的受访者有机会接受高等教育，而这部分人的高等教育回报率往往更高。因此，从这个意义上来说，高等教育扩招有助于促使更低收入层次的人群的收入增长，从而实现收入公平的效果。

图4中刻画的公式为： $X[\hat{\alpha}_1 - \hat{\alpha}_0] + \partial \hat{K}(\hat{P}) / \partial \hat{P}$ 。这里的 $\hat{K}(\hat{P})$ 是基于对方程式(8)中的 $K(p)$ 基于非参数的估计。这里的X包含表1中所有的解释变量，以及省份和观察年份虚拟变量。阴影部分使用bootstrap重复200次计算出来的95%的置信区间。与图3所不同的地方在于非参数部分估计的窗宽人为设定为0.16，估计样本中所有学历者没有正常毕业的样本都在回归中进行了排除。

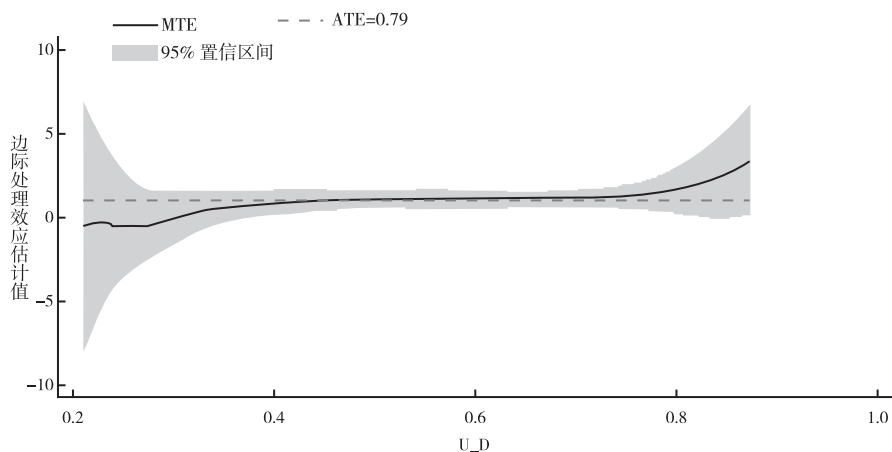


图 4 半参数 LIV 估计 MTE 的稳健性检验

资料来源：根据中国综合社会调查（CGSS）2010-2013 年数据计算得到。

五 不同处理效应之间的比较

由方程式（7）知道，ATE、ATT、ATUT、LATE 和 OLS 估计都可以看成是 MTE 的加权平均值。当不存在根本的异质性时，ATE、ATT、ATUT、LATE 是相等的，OLS 估计由于无法克服教育选择的内生性问题从事导致估计是有偏的。表 6 给出了基于方程式（7）计算出来的教育的各类处理效应。

从表 6 的最末一行可以看到，不论是在所有样本，还是在精炼样本条件下，OLS 估计值都未发生太大的变化，大学教育的总回报率为 30%，年均回报率为 7.5%。与先前大多数文献一样，OLS 估计的教育回报率明显偏低（Harmon & Walker, 1995；Card, 1999；Chen & Hamori, 2009）。基于 IV 法估计的 LATE 远高于 OLS 估计值，而且在全样本下和在精炼样本下也非常接近，大学教育的总回报率达到 124.3%，年均回报率为 31.08%。LATE 最大的问题是仅仅能够反映顺从者（compliers）的教育回报率，而不能反映所有人群的教育回报率（Imbens & Angrist, 1994；Abadie, 2003）^①。

在全样本下可以比较 MLE 估计值和半参数 LIV 估计值。首先是对 ATE 的估计，可以看到 MLE 和半参数 LIV 的估计值分别为 98.7% 和 78.9%，这两个估计值均明显低于

① 当工具变量所影响的人群（compliers）比重足够大时，LATE 估计值越来越接近 ATE（Oreopoulos, 2006）。

LATE 估计值，说明如果用 LATE 的估计值来代替所有人的教育回报率会明显高估。进一步的分析还可以看到，MLE 估计的 ATE 明显高于半参数 LIV 的估计，前者年均大学教育回报率达到 24.68%，后者为 19.73%。MLE 和半参数 LIV 估计出来的未上大学的人群的平均处理效应 (ATUT) 要高于 ATE，而估计出来的上大学人群的平均处理效应 (ATT) 要低于 ATE，说明那些未上大学的人群大学教育的回报率要更高一些。不过简单的比较会发现，与 MLE 估计结果相比，半参数 LIV 估计的 ATUT 要更高一些，而估计的 ATT 要明显低一些，而且 ATT 还没能通过显著性检验。值得注意的是，MLE 是参数估计，其前提假设是 (U_0, U_1, V) 满足联合正态分布，但是由于本文中 σ_{1v} 与 σ_{0v} 不相等，因此联合正态分布的条件并没有得到满足，MLE 估计只能作为一个参照。

前面已经指出，半参数 LIV 估计是一种更加稳健的估计方法，但是仅仅只能够在 $P(Z)$ 共同的支撑集上有效识别。从这个意义上来说，本文中利用半参数 LIV 估计的 ATE、ATT 和 ATUT 只能算是各自的近似值 (Heckman & Vytlacil, 2005; Carneiro et al., 2011)，但是，由于本文中 $P(Z)$ 共同的支撑集足够大，因此近似值就已经具有足够的说明意义 (Carneiro et al., 2011)。

基于半参数 LIV 估计，没有受过大学教育的人群的大学教育总体回报率 (ATUT) 为 157.1%，年均回报率达到 39.28%，而受过大学教育的人群的大学教育回报率 (ATUT) 没有通过显著性检验，ATT 估计值明显高于 ATUT，两者之间的差异在 1% 的显著性水平下通过显著性检验。事实上，MLE 和精炼样本下的半参数 LIV 估计的 ATT 与 ATUT 之间的差异都在 5% 的显著性水平下通过检验。这再一次说明不同人群之间的大学教育回报率存在根本的差异，在这种情况下，LATE 和 ATE 不能完全反映不同人群的大学教育回报率。

精炼样本下半参数 LIV 估计主要是对全样本下 LIV 估计的稳健性检验。可以看到，ATE 估计和 ATUT 估计在精炼样本和全样本之间的差异并不明显，而且在精炼样本下 ATT 也没能通过显著性检验。总体来说，精炼样本的估计结果与全样本估计结果并没有很大的不同，在数值上略微不同主要是因为非参数估计的窗宽发生变化所导致。由于精炼样本半参数 LIV 估计与全样本半参数 LIV 估计非常接近，证实半参数 LIV 估计具有较高的稳健性，这一结论与 Heckman & Vytlacil (2005) 和 Carneiro et al. (2011) 是一致的。

表 6 的估计结果表明 ATUT 的估计值明显高于 ATT 的估计值，说明那些没有接受高等教育的个体高等教育回报率更高，再一次证明家庭背景越差的人群大学教育回报率更高。这是因为平均来说，没有接受大学教育的人群家庭背景更差，这一点可以从表 1 的描述性统计看到，比如说没有接受大学教育的个体母亲在其 14 岁时参加工作的概率更低，而且在农村居住的比率要远远高于接受大学教育的个体。

表 6 大学教育的处理效应

模型	MLE	全样本 LIV	精炼样本 LIV
ATE = $E(\beta)$	0.987 *** (0.176)	0.789 ** (0.386)	0.894 ** (0.374)
ATT = $E(\beta D=1)$	0.880 ** (0.384)	0.665 (0.549)	0.813 (0.561)
ATUT = $E(\beta D=0)$	1.003 *** (0.331)	1.571 ** (0.632)	1.429 *** (0.462)
LATE		1.243 *** (0.204)	1.269 *** (0.212)
OLS		0.307 *** (0.023)	0.300 *** (0.024)

注：括号中的数值是标准误；*、** 和 *** 分别表示在 10%、5% 和 1% 的显著性水平下通过检验；MLE 与半参数 LIV 所使用的数据完全相同，因此所对应的 LATE 和 OLS 估计是相同的；LATE 估计值实际上等于 IV 估计值；所有的估计结果都同时控制了表 1 中的解释变量以及省份和观测年份虚拟变量；精炼样本的 LIV 估计窗宽设定为 0.16。

资料来源：根据中国综合社会调查（CGSS）2010-2013 年数据计算得到。

六 结论及政策启示

教育回报是个人以及社会最优教育决定的重要组成部分，过去的半个多世纪里，无数文献研究了大学教育回报率问题。尽管人们普遍注意到教育的内生性选择问题，各种实证手段被用来解决教育的内生性，出现了许多代表性文献，使得教育的回报问题成为经济学界和社会学界经久不衰的话题。然而，传统的方法如工具变量法、双胞胎设计和断点回归设计等等往往都假设教育回报不存在异质性，或者仅仅存在可观测的异质性，这些可观测的异质性包括城乡之间、男女之间等等的教育回报差异。然而，当教育回报存在根本的异质性，即教育回报与教育选择本身相关时，传统的估计方法无法刻画这种不可观测的异质性。Heckman & Vytlačil（2005，2007）通过扩展 Willis & Rosen（1979）和 Bjorklund & Moffit（1987）的模型，解决了这种无法观测的教育异质性回报问题。

利用 Heckman & Vytlačil（2005，2007）和 Carneiro et al.（2011）所提出的研究框架，本文基于广义罗伊模型估计大学教育的边际处理效应。利用 CGSS（2010-2013 年）数据，本文分别使用 MLE 和半参数 LIV 方法对中国大学教育的异质性回报进行估计。本文的研究证实中国大学教育的回报率与大学教育选择之间存在明显的相关关系，也就是说大学教育回报率对于大学选择倾向不同的人群来说具有根本上的差异。本文还比较了 MLE 估计和半参数 LIV 估计，发现虽然两种估计方法都发现随着个人选择大学教

育的动机越弱，或者心理成本越高，大学教育的回报率越高，但是由于 MLE 估计的假设前提不能得到满足，与 Carneiro et al. (2011) 的研究结论一样，本文也发现半参数 LIV 估计相较于 MLE 而言要更优一些。基于半参数 LIV 估计，本文的研究结果表明随着个人上大学选择的意愿降低，或者心理成本增加（即 U_D 增加），大学教育的 MTE 从 -45.73%（不显著）增加到 336.47%（显著），大学教育的 MTE 在大多数情况下并不等于其平均的处理效应（ATE），其 ATE 估计值为 78.92%，即大学教育的年均回报率为 19.73%。

Heckman & Vytlačil (2005, 2007) 所提出的 MTE 估计除了可以刻画不可观测的教育异质性回报之外，另一个优势在于可以通过 MTE 对教育的其他处理效应进行估计。这是因为 ATE、ATT、ATUT、LATE 和 OLS 估计都可以看成是 MTE 的加权平均值。本文最后根据 MTE 估计对其他处理效应也进行了估计，并各自进行比较。估计的结果表明 OLS 估计严重低估了中国的大学教育回报率，LATE 估计尽管能够准确反映因为工具变量的影响而接受大学教育的那部分人群即顺从者（compliers）的大学教育回报率，但是由于 LATE 估计结果明显高于 ATE 估计，因此，LATE 估计结果不能够代表所有人群的大学教育回报率。ATUT 估计全部都显著为正，且明显高于 ATT，而半参数估计的 ATT 都没有通过显著性检验。

所有研究结果都表明没有上大学的个体如果上了大学，其大学教育回报率更高，随着上大学的意愿降低或者心理成本增加，大学教育回报率不断增加。这就意味着，大学教育对家庭背景更差的个体来说回报率反而更高，这是因为对于这部分个体来说，由于家庭背景和生长环境较差，一旦这部分人上了大学，那么他们就很可能脱离生长的环境，从而实现阶层的跃迁，因此大学教育回报率更高。大学教育扩招的一个最明显的结果就是家庭背景更差的人群有了更多的机会接受大学教育，因此更多地促进了这部分人群的收入增长，客观上实现了缩小收入差距的结果。

参考文献：

- 刘生龙、周绍杰、胡鞍钢 (2016)，《义务教育法与中国城镇教育回报率：基于断点回归设计》，《经济研究》第 51 期，154 - 167。
- Abadie, Alberto (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics*, 113 (2), 231 - 263.
- Angrist, Josh & Alan Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106 (4), 979 - 1014.

- Ashenfelter, Orley & Cecilia Rouse (1998). Income, Schooling, and Ability: Evidence from A New Sample of Identical Twins. *The Quarterly Journal of Economics*, 113 (1), 253 – 284.
- Basu, Anirban (2014). Estimating Person-Centered Treatment (Pet) Effects Using Instrumental Variables: An Application to Evaluating Prostate Cancer Treatments. *Journal of Applied Econometrics*, 29 (4), 671 – 691.
- Basu, Anirban, Heckman James, Salvador Navarro-Lozano & Sergio Urzua (2007). Use of Instrumental Variables in the Presence of Heterogeneity and Self-selection: An Application to Treatments of Breast Cancer Patients. *Health Economics*, 16 (11), 1133 – 1157.
- Björklund, Anders & Robert Moffitt (1987). The Estimation of Wage Gains and Welfare Gains in Self-selection Models. *The Review of Economics and Statistics*, 69 (1), 42 – 49.
- Card, David (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *NBER Working Paper*, No. w4483.
- Card, David (1999). Chapter 30 The Causal Effect of Education on Earnings. In Orley Ashenfelter & David Card (eds.), *Handbook of Labor Economics*, Vol. 3. Elsevier, pp. 1801 – 1863.
- Carneiro, Pedro, James Heckman & Edward Vytlačil (2011). Estimating Marginal Returns to Education. *The American Economic Review*, 101 (6), 2754 – 2781.
- Chen, Guifu & Shigeyuki Hamori (2009). Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach. *China Economic Review*, 20 (2), 143 – 152.
- Cornelissen, Thomas, Christian Dustmann, Anna Raute & Uta Schönberg (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126 (6), 2356 – 2409.
- Eisenhauer, Patricia, James Heckman & Edward Vytlačil (2015). The Generalized Roy Model and the Cost-benefit Analysis of Social Programs. *Journal of Political Economy*, 123 (2), 413 – 443.
- Harmon, Colm & Ian Walker (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *The American Economic Review*, 85 (5), 1278 – 1286.
- Heckman, James & Edward Vytlačil (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the national Academy of Sciences*, 96 (8), 4730 – 4734.
- Heckman, James & Edward Vytlačil (2005). Structural Equations, Treatment Effects and

- Econometric Policy Evaluation. *Econometrica*, 73 (3), 669 – 738.
- Heckman, James & Edward Vytlacil (2007). Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Handbook of Econometrics*, Volume 6, Part B, 4875 – 5143.
- Heckman, James (2010). Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic literature*, 48 (2), 356 – 398.
- Heckman, James, Daniel Schmierer & Sergio Urzua (2010). Testing the Correlated Random Coefficient Model. *Journal of Econometrics*, 158 (2), 177 – 203.
- Heckman, James, Sergio Urzua & Edward Vytlacil (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88 (3), 389 – 432.
- Imbens, Guido (2010). Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic literature*, 48 (2), 399 – 423.
- Imbens, Guido & Josh Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62 (2), 467 – 475.
- Li, Hongbin, Pak Wai Liu & Junsen Zhang (2012). Estimating Returns to Education Using Twins in Urban China. *Journal of Development Economics*, 97 (2), 494 – 504.
- Nyblom, Martin (2014). The Distribution of Lifetime Earnings Returns to College. *Working Paper Series*, No. 2/2014, Swedish Institute for Social Research.
- Oreopoulos, Philip (2006). Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter. *The American Economic Review*, 96 (1), 152 – 175.
- Quandt, Richard (1958). The Estimation of the Parameters of A Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association*, 53 (284), 873 – 880.
- Robinson, Peter (1988). Root-N-consistent Semiparametric Regression. *Econometrica*, 56 (4), 931 – 954.
- Roy, Aditi Deb (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3 (2), 135 – 146.
- Willis, Robert & Sherwin Rosen (1979). Education and Self-selection. *Journal of political*

Economy, 87 (5), S7 – S36.

Zhong, Hai (2015). Does A College Education Cause Better Health and Health Behaviours?
Applied Economics, 47 (7), 639 – 653.

Who Benefits More from Higher Education Expansion? Evidence Based on General Roy Model

Liu Shenglong^{1,2} & Zheng Shilin³

(School of Public Policy & Management, Tsinghua University¹;

Institute for Contemporary China Studies, Tsinghua University²;

National School of Development, Peking University³)

Abstract: Estimating returns to education is an essential part of determining optimal schooling decisions, both at individual and social level. In the past, a lot of literatures have estimated returns to education. Traditional methods such as instrumental variable method, regression discontinuity design and twin design are mainly used to solve the endogeneity problem of education. However, there is essential heterogeneity in returns to education, that is, returns to education is correlated with selection of education, and traditional methods can't estimate such unobservable heterogeneity. Using research framework of Heckman & Vytlacil (2005, 2007), this paper estimates the marginal treatment of college education based on general Roy model. The result of this paper proves the correlation between returns to college education and college selection. According to semi-parametric local instrumental invariable (LIV) estimation, the results show that the MTE of college education increases monotonously with the increases of psychological cost of going to college. This paper finds that individuals with lower benefits are more likely to select college education. This pattern of reverse selection on gains is driven by unobservable family background. Individuals with disadvantaged backgrounds have higher psychological cost of receiving higher education, however, once they have, they benefit more from higher education.

Keywords: education, general Roy model, marginal treatment effect, propensity score, semi-parametric local instrumental variable

JEL Classification: I28, J31

(责任编辑:一帆)