

## 大学教育回报：基于大学扩招的自然实验

刘生龙 胡鞍钢\*

---

**内容提要** 本文基于 1999 年中国高等教育扩招这个自然事件，利用断点回归设计方法检验高等教育与就业、收入和生活满意度之间的因果关系。实证研究发现，第一，高等教育显著地提高了城镇居民的就业率。第二，高等教育与城镇居民和农村居民的收入之间都存在着显著的因果关系，高等教育对城镇居民收入的年度回报率为 17.1%，对农村居民收入的年度回报率为 15.9%。第三，高等教育对全部样本的生活满意度的影响不显著，但是在农村样本中，本文发现了高等教育提高生活满意度的证据。尽管高等教育扩招之后出现了一些问题，如大学生就业困难等，但是本文仍然发现了高等教育扩招之后，高等教育导致人力资本加速积累，以及农村居民生活满意度有所增加等积极意义。

**关键词** 高等教育扩招 大学教育回报 断点回归设计

---

### 一 引言

过去 30 多年时间里，中国经历了经济的高速增长，一些学者认为中国之所以保持数十年的经济高速增长，其中一个重要的因素就是人力资本的迅速增加（Fleisher et al., 2010; Fang et al., 2012）。在文化大革命结束之后，1977 年中国恢复高考。1999 年中国开启了高等教育扩招的序幕，此后，大学生入学人数迅速增加。从 1977 年到 1998

---

\* 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；胡鞍钢，清华大学公共管理学院、清华大学国情研究院，电子邮箱：anganghu@mail.tsinghua.edu.cn。作者感谢全国哲学社会科学规划办公室高端智库项目（20155010298）的资助，感谢北京大学中国社会科学调查中心（ISSS）提供中国家庭追踪调查（CFPS）数据，感谢审稿人意见，文责自负。

年，中国的高考录取人数从 27 万增加到 108 万，年均增长率为 6.8%<sup>①</sup>。1999 年高校开始大规模扩招，大学录取人数为 160 万人，1999 年之后大学录取人数仍然保持着高速增长的态势，到 2014 年达到 698 万人，1998 - 2014 年大学录取人数年均增长率为 12.4%。图 1 刻画的是中国 1977 年以来大学录取率和录取人数的发展趋势，从中可以看到，1999 年中国的大学录取率和录取人数在 1998 年的基础之上有一个大的跳跃。

高校扩招使得更多的人接受了高等教育，中国的人力素质迅速提高。但是另一方面，高等教育扩招也导致了大学生就业困难问题，由于大量的大学生毕业之后涌入劳动力市场造成大学生的就业压力，很多大学毕业生面临着结构性失业问题。大学生在劳动力市场上竞争加剧，初次就业的大学生平均工资上涨幅度缓慢。而随着人口老龄化的趋势日益明显，中国的农民工短缺问题也逐渐显现，初次就业的农民工的工资快速上涨。大学生就业难与农民工工资上涨之间形成了鲜明的对比，考虑到读大学的成本，于是引出了这样一个问题：读大学值得吗？

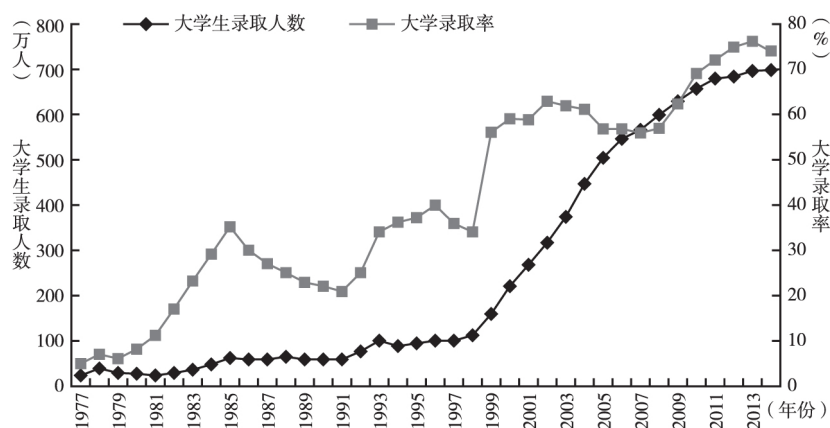


图 1 1977 - 2014 年中国大学录取率和录取人数变化趋势

资料来源：根据《中国统计摘要》和教育部公布数据计算得到。

本文研究的目的是回答接受大学教育是否值得这一问题。为了对这个问题进行回答，这里面很明显的一个理论着手点就是大学教育的成本和收益问题，如果收

① 1978 年高考录取人数在 1977 年基础之上有一个跃升，这是因为“文革”十年高考中断，积累了许多想要上大学而没能上大学的学生，所以集中在 1977 年和 1978 年报考，导致 1978 年高考录取人数跃升。

益大于成本，那么接受大学教育就是值得的。大学教育的收益主要体现在两个层面上：第一个层面是物质方面的，主要体现在对就业和收入的促进作用上。第二个层面主要体现在精神层面，比如说对生活满意度的影响，当然生活满意度也会受到物质层面的影响，但是它还包含人的认同感、身份感等因素。接受大学教育是有成本的，比如说最直接的成本就是大学学费，还有因为上大学所花费的时间成本，还有一些间接的成本如天赋能力一般的人为了考取大学而不得不十分努力所付出的“痛苦”成本等等。因为成本无法观测，所以我们不能够通过大学教育的净收益（即收益减去成本）来回答大学教育是否值得的问题，而只能通过一种间接的方法来比较在其他条件都控制的情况下，上大学的人和没有上大学的人在就业、收入和生活满意度之间的差异。如果其他条件相同，上大学的人相对不上大学的人来说，就业率、收入和生活满意度明显更高一些，那就足以说明接受大学教育是值得的。本文的研究正是基于这种逻辑框架来估计大学教育对就业、收入和生活满意度的因果影响。

基于 1999 年高等学校扩招这样一个自然事件，本文将主要借助断点回归设计（regression discontinuity design，以下简称 RDD）方法，论证高等教育对收入、就业和生活满意度的影响。由于大规模的高校扩招发生在 1999 年，中国的小学入学年龄是 6 岁，到高中毕业参加高考时的年龄是 18 岁，这就意味着 1999 年高等教育扩招主要影响的是 1981 年第 3 季度以后出生的人群。这样，我们可以以出生日期作为排序变量（rating variable），1981 年第 3 季度及以后出生的人群作为处理组人群（treatment cohort），1981 年第 3 季度之前出生的人群作为控制组人群（control cohort），检验高校扩招对两类人群大学入学率的影响，进而识别大学教育对就业、收入和生活满意度的影响。

基于 RDD 的估计结果，本文研究发现，高校扩招对城镇就业有着显著正向的影响，接受大学教育使得城镇样本就业率增加 31.9%。大学教育提高了城镇居民和农村居民的收入，城镇居民大学教育的年度回报率约为 17.1%，农村居民大学教育的年度回报率约为 15.9%。在农村样本中，存在着大学教育与生活满意度之间的因果证据。总体而言，虽然大学扩招使得中国大学生面临着很大的就业压力，但大学教育对提高中国劳动生产率具有一定的积极意义。

本文下面的结构安排为：第二部分提出大学教育回报因果关系识别策略；第三部分是本文使用的数据说明及其描述性统计分析；第四部分是基于 RDD 估计结果的实证分析；第五部分是对估计结果的有效性检验及分析；第六部分是结论。

## 二 大学教育回报的因果识别策略

自经济学家明瑟 (Mincer) 提出教育与收入决定的人力资本理论框架以来,教育的回报率问题就成为经济学界长期以来关注的一个话题。基于明瑟收入方程,一些经济学家利用微观数据对中国的教育回报进行了测算,许多研究者发现中国 20 世纪 80 年代的教育回报率非常低 (Fleisher & Wang, 2005)。一些经济学家认为,随着从计划经济向市场经济转变,中国的教育回报率会逐渐上升。后来的许多研究者通过实证研究发现,1990 年以后中国的教育回报率的确明显上升 (Zhang et al., 2005)。2000 年以后,越来越多的研究者运用工具变量法 (IV) 对中国的教育回报进行估计 (Heckman & Li, 2004; Chen & Hamori, 2009)。

近年来,一些研究者开始利用制度变量作为教育的工具变量估计中国城镇的教育回报率。比如,利用中国健康与营养调查 (CHNS) 1997 年、2000 年、2004 年和 2006 年的数据, Fang et al. (2012) 以 1986 年中国的义务教育法实施所导致的干预组与参照组之间的教育差异作为工具变量对中国城镇的教育回报进行测算。研究结果表明,义务教育法的实施使得中国总受教育年限增加了约 0.8 年,而工具变量回归结果表明中国的教育回报率高达 20%。而运用同样的数据, La (2014) 的研究结果表明,当用义务教育法作为工具变量进行参数估计时,中国的教育回报率仅为 5.85%,而且这一估计结果并没有通过显著性检验; 而用最低劳动年龄法 (provisions on the prohibition of using child labor) 作为工具变量时,中国城镇居民的教育回报率达到了 8.98%。

随着伪随机实验在应用经济学的兴起,现在越来越多的学者将这些伪实验的方法也应用到教育回报的研究上 (Clark & Martorell, 2014)。虽然已有的文献利用中国的数据就中国的教育回报率问题已经做了许多有价值的研究,然而,前述文献普遍采用传统的方法来估计中国教育回报率,用伪实验的方法来验证中国教育回报率的文献并不多见。与以往文献不同的是,本文通过 1999 年高校扩招所导致的大学入学率迅速提高这样一个自然实验,利用断点回归设计 (RDD) 的方法估计中国高等教育的回报率。

本文中,我们定义大学教育 (college) 为一个二元变量,即一个人的最终学历为大专及以上文凭者, college 取值为 1; 一个人的最终学历为大专以下者,取值为 0。令  $D$  为个人出生日期距离受高等教育扩招影响初始年份的距离,当  $D \geq 0$  时,

个人受到大学扩招的影响，此时扩招（expansion）取值为 1；当  $D < 0$  时，扩招（expansion）取值为 0。

为了进行因果推断，我们假定  $(Y_0, Y_1)$  是一个人上大学和不上大学潜在的结果变量，其中  $Y_0$  是没有上大学的结果变量， $Y_1$  是上大学的结果变量。本文的结果变量包含三个，分别为是否就业、年收入水平和生活满意度。上大学的因果效应为  $\alpha = Y_1 - Y_0$ ，也就是一个人上大学和不上大学结果变量的差异。然而，由于一个人要么上大学，要么不上大学，因此  $\alpha$  是不可观测的。尽管  $\alpha$  不可观测，但其反映大学教育的因果回报，因此是本文感兴趣的参数。

本文中个体受到大学教育扩招影响的人并不一定上大学，也就是说即使  $D$  大于 0 时，个体也不是说都能够上大学，总有相当一部分人不论大学是否扩招都不会上大学（never takers），也有相当一部分人不论大学是否扩招都会上大学（always takers），因此个人上大学的概率在  $D$  大于 0 时并不是 1，而在  $D$  小于 0 时，也不是 0。这就意味着在大学扩招影响的临界点（cut point）附近，即  $D$  等于 0 的附近，个体上大学的概率有一个断点，但是断点的距离小于 1。当个体不能够随意选择  $D$  到底大于等于 0，还是小于 0 时，即个人不能完全对本文的排序变量进行操纵时，本文的断点可以帮助我们解决大学教育的内生性问题，从而对大学的教育回报进行因果识别。于是，在  $D$  等于 0 的附近，大学教育的因果回报为：

$$E[\alpha | college] = \frac{E[Y | D = 0^+] - E[Y | D = 0^-]}{E[college | D = 0^+] - E[college | D = 0^-]} \quad (1)$$

### 三 数据及其描述性统计

#### （一）数据来源

本文研究所使用的数据来自中国家庭追踪调查（CFPS），该数据是北京大学中国社会科学调查中心实施的一项旨在通过对全国样本家庭及其成员的调查获取个体、家庭和村/居委会层次的多时点信息，反映中国经济、社会、人口、教育和健康的变迁，为学术研究和政策决策提供可靠的数据来源。为了保证调查的效率和质量，CFPS 数据使用了计算机辅助面访调查技术（CAPI），考虑到中国巨大的地区差异，CFPS 抽样采用了内隐分层（implicit stratification）、多阶段、多层次、与人口规模成比例的概率抽样方式（PPS）。

本文的研究数据包含两轮的 CFPS 调查数据，分别是 2010 年的基线调查和 2012 年

第一轮的追踪调查。CFPS 样本覆盖 25 个省、市、自治区，代表了中国 95% 的人口。2010 年的基线调查共采访约 14960 户家庭 42590 名个人，其中成人 33600 名，少儿 8990 名。2012 年的追踪调查包含的成人样本为 35727 名，少儿为 8625 名<sup>①</sup>。

我们对上述 CFPS 两年的样本进行了处理，具体来说就是仅保留 1981 年第 3 季度出生人群前后各 9 年（即 1972 – 1989 年，前后各 36 个季度）的样本，在校生因为最终所获得的学历和收入不清楚，因而也从样本中进行了排除。由于本文将采用 RDD 方法进行实证检验，在 RDD 应用中，箱宽（bin width）的选择非常重要。与 Card & Shore-Sheppard（2004）和 Banks & Mazzonna（2012）一样，本文选择的箱宽为一个季度。在 2012 年的 CFPS 中，一些个体的出生月份缺失，因此在样本中进行了排除。最终我们剩下的样本数量为 14822 个，其中，1981 年第 3 季度以后出生（处理组，即  $D > 0$ ）的样本为 6479 个，之前出生（控制组，即  $D < 0$ ）的样本为 8343 个，1981 年第 3 季度出生的人群为本文的出生临界点（cut point，即  $D = 0$ ）<sup>②</sup>。

## （二）描述性统计

表 1 报告了样本的描述性统计。第 1 列是所有样本的描述性统计，可以看到，样本中拥有大学及以上学历的人群占比为 14.3%。第 2 列和第 3 列报告的分别是控制组人群和处理组人群的描述性统计，当本文将样本控制在临界点前后 9 年以内时，控制组和处理组样本较为均衡，处理组人群占总样本达到 43.7%。控制组中获得大学及以上学历的比例为 12.0%，处理组中获得大学及以上学历的比例为 17.3%，从第 4 列可以看得出，处理组人群获得大学学历的比例显著高于控制组。控制组人群的平均收入水平也明显低于处理组人群。此外，控制组的就业率明显高于处理组，但就生活满意度而言，处理组人群明显要高一些。第 5 列和第 6 列分别给出了城镇和农村样本的描述性统计。可以看到，城镇的大学及以上人口比重、收入水平和就业水平明显高于农村，但是从生活满意度来说，农村人口明显更高一些。

① 没有包含在样本中的省份为内蒙古、海南、西藏、宁夏、青海和新疆。

② 2012 年的 CFPS 调查没有个人的出生月份信息，但由于是追踪调查数据，2012 年的调查样本多数与 2010 年是相同的，因此，2012 年大部分调查个体的出生季度可以从 2010 年的调查中获得，2012 年调查数据中未知出生季度的个体没有包含在本文的回归样本中。

表 1 变量的描述性统计

| 变量名      | 全部                | 控制组               | 处理组              | P 值   | 城镇                | 农村                | P 值   |
|----------|-------------------|-------------------|------------------|-------|-------------------|-------------------|-------|
|          | (1)               | (2)               | (3)              | (4)   | (5)               | (6)               | (7)   |
| 样本量      | 14822             | 8343              | 6479             |       | 6924              | 7898              |       |
| 大学教育     | 0.143<br>(0.350)  | 0.120<br>(0.325)  | 0.173<br>(0.379) | <0.01 | 0.252<br>(0.434)  | 0.048<br>(0.214)  | <0.01 |
| 对数年收入    | 6.582<br>(4.335)  | 6.494<br>(4.393)  | 6.697<br>(4.256) | <0.05 | 7.391<br>(4.187)  | 5.871<br>(4.338)  | <0.01 |
| 是否就业     | 0.681<br>(0.466)  | 0.725<br>(0.446)  | 0.625<br>(0.484) | <0.01 | 0.702<br>(0.458)  | 0.664<br>(0.472)  | <0.01 |
| 生活满意度    | 3.351<br>(1.026)  | 3.329<br>(1.031)  | 3.381<br>(1.019) | <0.01 | 3.323<br>(1.002)  | 3.377<br>(1.047)  | <0.01 |
| 季度年龄     | 119.37<br>(21.45) | 135.50<br>(11.28) | 98.59<br>(11.04) | <0.01 | 120.29<br>(20.68) | 118.56<br>(22.07) | <0.01 |
| 性别(男性=1) | 0.459<br>(0.540)  | 0.466<br>(0.499)  | 0.450<br>(0.590) |       | 0.464<br>(0.499)  | 0.455<br>(0.574)  |       |
| 城乡(城镇=1) | 0.467<br>(0.500)  | 0.484<br>(0.500)  | 0.445<br>(0.497) |       |                   |                   |       |
| 民族(汉族=1) | 0.896<br>(0.305)  | 0.891<br>(0.311)  | 0.902<br>(0.297) |       | 0.935<br>(0.246)  | 0.862<br>(0.345)  |       |
| 已婚(在婚=1) | 0.806<br>(0.395)  | 0.932<br>(0.252)  | 0.645<br>(0.479) |       | 0.790<br>(0.408)  | 0.821<br>(0.383)  |       |
| 父亲受教育年限  | 5.423<br>(3.012)  | 5.225<br>(3.142)  | 5.550<br>(2.973) | <0.05 | 7.884<br>(4.096)  | 3.265<br>(1.621)  | <0.01 |
| 母亲受教育年限  | 4.071<br>(2.293)  | 3.874<br>(2.306)  | 4.229<br>(2.514) |       | 6.103<br>(2.517)  | 2.290<br>(1.375)  | <0.01 |

注：括号中的数值是标准差；控制组是没有受到 1999 年高校扩招影响的人群，处理组是受到 1999 年高校扩招影响的人群；第 4 列是控制组和处理组中连续变量差异的 P 值；第 7 列是城镇和农村样本连续变量差异的 P 值。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

#### 四 实证结果及分析

很多研究教育回报的文献都已经指出最小二乘 (OLS) 估计存在内生性问题，其内生性主要来自于三个方面：首先，天生的能力和家庭背景会同时影响教育和被解释变量，这会使得教育回报的估计被高估；其次，教育具有信号效应 (signal effect)，它往往可以作为某些高端职业的入门前提，忽视这种效应会低估教育的回报率；最后，教育的测量误差问题，在进行入户调查时，教育程度往往容易高报，而收入回报往往容

易低报，这会导致教育的回报率被低估。OLS 估计无法反映高等教育与收入、就业和生活满意度之间的因果关系，而经济学者更关注的是教育的因果回报，因此有必要解决教育的内生性问题。正如引言部分已经说到的，与先前主要通过 OLS 或传统的两阶段最小二乘（2SLS）估计中国教育回报的文献不同，本文将采用 RDD 方法来检验高等教育对就业、收入和生活满意度的因果影响。

在前面我们已经明确指出，本文中大学扩招政策存在从不接受者（never-takers），也存在总是接受者（always-takers），因此这是一种典型的模糊断点模型（fuzzy RDD），遵循标准的程序，对于这种模糊断点模型，我们采用两阶段最小二乘（2SLS）进行参数估计（Angrist & Lavy, 1999）。

由于方程式（1）在附近的区域与工具变量的估计原理是一致的，这里的 College 变量可以用是否受到大学扩招（expansion）作为工具变量进行 2SLS 估计。具体而言，本文中大学教育对结果变量影响的方程形式如下：

$$y_i = \alpha_0 + \alpha_1 college_i + h(D) + \alpha_2 X + \varepsilon_i \quad (2)$$

这里的 X 是表 1 中其他一些控制变量向量，包括性别（男性 = 1）、是否城镇人口（城镇 = 1）、是否汉族（汉族 = 1）和是否已婚（已婚 = 1）。所有的估计结果都控制了观测年份和省域虚拟变量， $\varepsilon_i$  是误差项。为了进行因果识别，我们还通过如下方程对内生变量进行第一阶段的估计，第一阶段回归方程形式如下：

$$college_i = \beta_0 + \beta_1 expansion_i + g(D) + \beta_2 X + \mu_i \quad (3)$$

这里  $\mu_i$  是误差项。方程（2）中的  $h(D)$  反映临界点（ $D=0$ ）前后  $y_i$  和  $D$  之间的拟合关系式，方程（3）式中的  $g(D)$  反映临界点（ $D=0$ ）前后  $College_i$  与  $D$  之间的拟合关系式。根据对  $h(D)$  和  $g(D)$  拟合关系式的设定不同，可以分为非参数拟合和多项式拟合，因而产生  $\alpha_i$  的非参数估计和参数估计。本文将先对  $\alpha_i$  进行非参数估计，作为稳健性检验，本文也给出了  $\alpha_i$  多项式估计结果。

#### （一）描述性证据

本文首先通过图形直观地判断高等教育扩招前后大学教育和结果变量的变化趋势。受大学扩招的影响，1981 年第 3 季度以后出生的人群上大学的概率应该明显有一个上升的过程，图 2 证实了这一点。图 2 刻画了 1981 年第 3 季度前后各 9 年之内出生人群的出生时间与大学教育获得率之间的关系<sup>①</sup>。图中的点是每一个出生季度人群获得大学

<sup>①</sup> 由于 1990 年以后出生且在上大学的人群在本文的样本中大部分还没有进入劳动力市场，因此本文选择 1981 年前后 9 年出生的人群作为观测样本。



教育的比率 (cell mean), 图中的线是大学教育与出生季度之间的二次多项式拟合曲线。可以看到, 在通过临界点的地方, 大学教育有一个明显的跳跃。当我们将样本分成城镇样本和农村样本之后, 我们发现不论是在城镇地区还是在农村地区, 高等教育扩招都导致了大学教育明显的跳跃。从图 2 中的所有样本可以看出, 在临界点附近 1981 年第 3 季度之前出生人群的大学教育获得率大约为 15%, 而之后出生人群的大学教育获得率大约为 21%, 也就是说高等教育扩招使得大学获得率大约发生了 6 个百分点的跳跃。

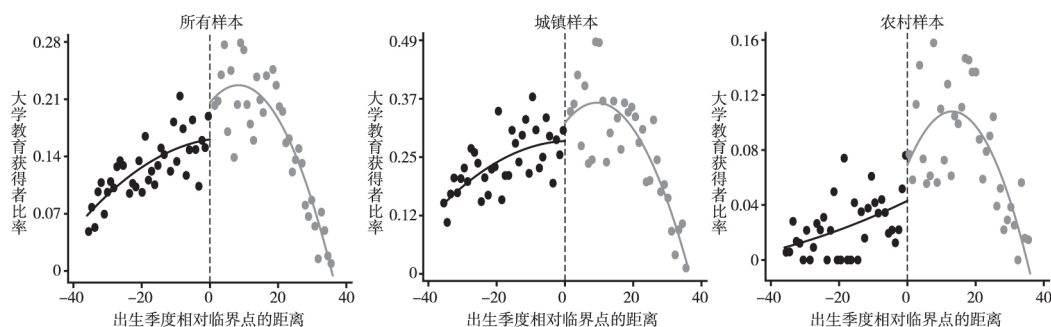


图 2 受高校扩招影响临界点前后出生人群的高等教育获得率断点

资料来源: 根据 CFPS 2010 年和 2012 年数据计算得到。

从图 2 可以看到, 由于大学扩招, 临界点 ( $D = 0$ ) 前后人群的大学教育获得率出现了明显的跳跃。如果大学教育对那些结果变量具有因果影响, 那么结果变量在临界点前后也应该出现明显的跳跃。

图 3 刻画的是 1981 年第 3 季度前后各 9 年之内出生人群的出生时间与就业之间的关系, 与前述一样, 这里的点代表的是不同出生年份和季度人群的就业比率 (cell mean), 图中的线是就业率与出生年份和季度的二次多项式拟合曲线。从图 3 所有样本可以看到, 在出生临界点前后, 就业率似乎存在断点。当我们将样本分为城镇样本和农村样本之后, 我们发现城镇样本临界点前后出生人群的就业率存在明显的跳跃, 而农村样本的处理组人群的就业率在临界点处的跳跃则不是很明显。

图 4 刻画的是 1981 年第 3 季度前后各 9 年之内出生人群的出生时间与对数收入之间的关系。与图 2 和图 3 一样, 这里的点表示的是不同出生年份和季度的人平均的对数年收入 (cell mean), 图中的线是对数收入与出生季度的二次多项式拟合曲线。可以看到, 在所有样本中临界点前后有一个收入断点。当我们将样本分成城镇样本和农村样本后, 发现收入断点不论是在城镇样本中还是在农村样本中都比较明显。

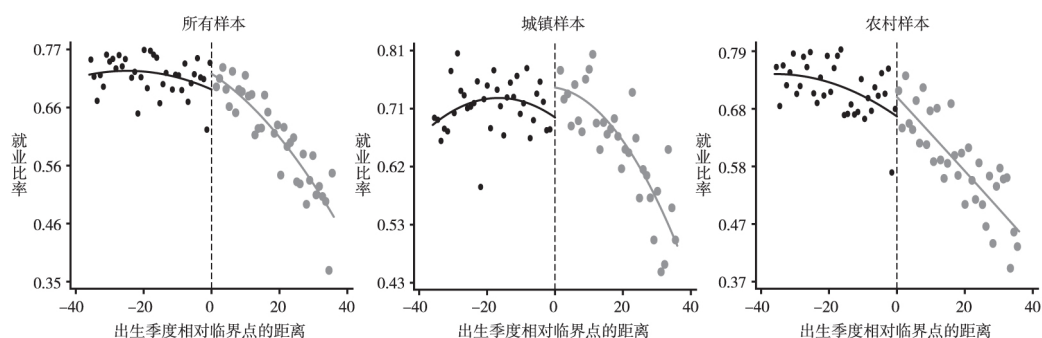


图3 受高校扩招影响临界点前后出生人群的就业断点

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

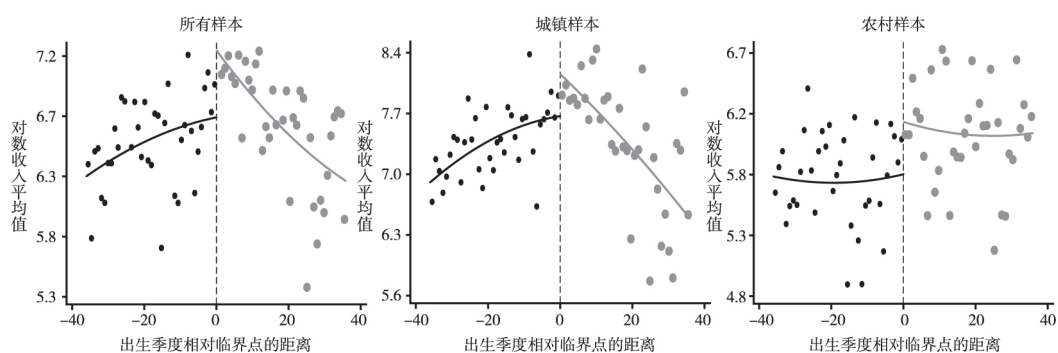


图4 受高校扩招影响临界点前后出生人群的收入断点

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

图5刻画的是1981年第3季度前后各9年之内出生人群的出生时间与生活满意度之间的关系。这里的点不同出生年份和季度人群的平均生活满意度 (cell mean)，图中的线是生活满意度与出生季度的二次多项式拟合曲线。可以看到，在临界点附近，个人生活满意度出现了一定的跳跃。将样本分成城镇样本和农村样本后，城镇样本临界点前后的生活满意度并没有明显的跳跃，而在农村样本中可以看到明显的跳跃。由于临界点两边的点都比较分散，临界点附近的断点是不是显著还得根据参数估计结果来进行判断。

## (二) 一阶段估计结果

在进行RDD的2SLS估计时，首先第一阶段的估计结果需显著。虽然图2的描述性证据表明大学扩招使得临界点附近前后人群出现了大学入学率的断点，但是尚不知道

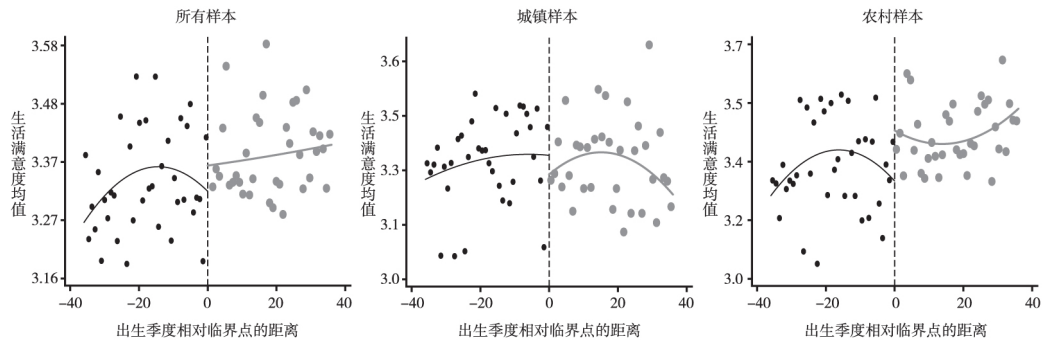


图 5 受高校扩招影响临界点前后出生人群的生活满意度断点

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

这些断点是否由更大的噪声造成，因此有必要直接给出参数的估计结果。本文首先确定断点回归模型的最优窗宽，然后在所有样本和最优窗宽条件下进行参数估计。从目前来看，在 RDD 估计过程中，有 3 种主要的确定最优窗宽的办法，分别是 Ludwig & Miller (2007) 提出的 cross-validation (即 CV 方法)，Imbens & Kalyanaraman (2012) 提出的 IK 方法和 Calonico et al. (2014) 提出的 CCT 方法，由于前两种只适合于尖锐断点 (sharp RDD) 的情形，本文的断点设计由于是模糊断点设计 (fuzzy RDD)，因此采用 Calonico et al. (2014) 的方法进行最优窗宽设定。本文首先对方程 (3) 的最优窗宽进行计算，结果为 20 个季度，即临界点前后 5 年时间。基于 Austin (2007) 的方法，本文选择三角核密度函数 (triangle) 的参数估计在不同设定下，全样本、城镇样本和农村样本的估计结果如表 2 所示<sup>①</sup>。

表 2 的第 1 列和第 2 列给出了所有样本大学扩招对大学教育的影响。可以看到，不论是所有样本还是最优窗宽，不论加不加入基准控制变量，大学扩招对大学教育都有着显著的影响，而且连影响的数值都没有发生很大变化，也就是说高等教育扩招之后，大学生的入学率提高了约 6.6 个百分点。在加入所有控制变量后，最优窗宽条件下，大学教育扩招对整体来说，大学教育获得率增加了 5.6 个百分点。后面，我们的分析都将集中在加入所有控制变量之后，最优窗宽条件下的参数估计结果。

① 由于 RDD 估计结果是一个局部平均处理效应 (LATE)，其估计具有典型的局部属性，也就是说它在临界点附近的估计是最应该受到关注的。相对于参数估计来说，非参数估计的优势就在于更加注重局部的分布情况。此外，RDD 非参数估计还较少受到其他控制变量的影响。因此本文以非参数估计结果作为主要分析的结果，以参数估计结果作为稳健性检验。

表2 大学扩招对大学教育的影响

|        | 全部样本                 |                      | 城镇样本                 |                      | 农村样本                 |                      |
|--------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|        | (1)                  | (2)                  | (3)                  | (4)                  | (5)                  | (6)                  |
| 所有样本   | 0.066 ***<br>(0.021) | 0.065 ***<br>(0.023) | 0.068 **<br>(0.029)  | 0.077 **<br>(0.033)  | 0.058 ***<br>(0.014) | 0.059 ***<br>(0.016) |
| 观察值数   | 14822                | 14818                | 6924                 | 6923                 | 7898                 | 7895                 |
| 最优窗宽   | 0.058 ***<br>(0.015) | 0.056 ***<br>(0.013) | 0.067 ***<br>(0.019) | 0.070 ***<br>(0.019) | 0.051 ***<br>(0.010) | 0.051 ***<br>(0.011) |
| 观察值数   | 7231                 | 7229                 | 3782                 | 3587                 | 3649                 | 3642                 |
| 基准控制变量 | 否                    | 是                    | 否                    | 是                    | 否                    | 是                    |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和\*\*\*分别表示在10%、5%和1%的显著性水平下通过检验；所有回归方程都包含了省份和年份虚拟变量；基准控制变量包括父亲的受教育年限、母亲的受教育年限、性别（男性=1）、民族（汉族=1）、城乡（城镇=1）和婚姻（在婚=1）。

资料来源：根据CFPS 2010年和2012年数据计算得到。

第3列和第4列给出了在不同设定下大学扩招对城镇样本大学教育获得率的影响，可以看到，大学扩招使得城镇样本大学教育获得率增加了7.0个百分点。第5列和第6列给出了大学扩招对农村样本大学教育获得率的影响，可以看到，大学扩招使得农村样本大学教育获得率增加了5.1个百分点。

### （三）大学教育的回报率估计

表3和表4是大学教育的经济和社会回报的2SLS估计结果。我们将首先分析大学教育对就业的影响，然后分析大学教育对收入和生活满意度的影响。

引言部分已经提到，自1999年高等院校大规模扩招以后，从2003年开始，中国大学生毕业人数开始加速上升。随着大量的大学生进入劳动力市场，中国劳动力市场的人力资源素质应该说有了很大的提升。但是，由于每年劳动力市场就业岗位有限，而大学生毕业人数又逐年增加，因而大学生就业的问题也日渐凸显。越来越多的大学生面临着结构性失业的压力，其平均的期望工资水平也就越来越低。那么大学教育究竟对就业会产生什么影响呢？

表3报告了大学教育获得对就业影响的2SLS估计结果。一般情况下，接受到的教育程度越高，劳动技能应当越高，在就业市场上应该占据更多的优势，表3的结果支持了这一点。在全样本的情况下，大学教育显著地促进了就业。这就意味着，相对于未接受大学教育的人群而言，已经接受大学教育的人群在中国就业市场上表现出明显的优势。当我们将样本分成城镇样本和农村样本之后，我们发现接受高等教育与城镇

样本的就业有着显著的因果关系，而跟农村样本的就业之间不存在因果关系。一个可能的解释是，由于城镇地区存在大量的知识和技术密集型的产业，这些行业往往需要更高的技能和复杂的知识，在这些行业里，大学教育获得者往往更能胜任，因而在就业市场上占据优势。而在农村地区，农业是主要的产业，农业并不是一个知识密集型的行业，大学教育获得者在农村劳动力市场上并不占优，因此大学教育对农村就业影响不显著。

表 3 大学教育对就业影响的 2SLS 估计

|      | 全部样本                |                      | 城镇样本               |                      | 农村样本             |                  |
|------|---------------------|----------------------|--------------------|----------------------|------------------|------------------|
|      | (1)                 | (2)                  | (3)                | (4)                  | (5)              | (6)              |
| 所有样本 | 0.202 **<br>(0.085) | 0.250 ***<br>(0.075) | 0.414 *<br>(0.246) | 0.429 ***<br>(0.144) | 0.065<br>(0.137) | 0.151<br>(0.138) |
| 观察值数 | 14822               | 14818                | 6924               | 6923                 | 7898             | 7895             |
| 最优窗宽 | 0.166<br>(0.152)    | 0.151 **<br>(0.072)  | 0.350 *<br>(0.207) | 0.319 ***<br>(0.155) | 0.097<br>(0.178) | 0.125<br>(0.104) |
| 观察值数 | 7231                | 7229                 | 3782               | 3781                 | 3449             | 3448             |
| 控制变量 | 否                   | 是                    | 否                  | 是                    | 否                | 是                |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和\*\*\*分别表示在 10%、5% 和 1% 的显著性水平下通过检验；所有模型均控制观察年份和区域虚拟变量；基准控制变量包括父亲的受教育年限、母亲的受教育年限、性别（男性=1）、民族（汉族=1）、城乡（城镇=1）和婚姻（已婚=1）。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

自 2000 年以后，中国加速进入老龄化社会。中国 65 岁及以上人口的比重 1990 年为 5.6%，2000 年达到 7.0%，2010 年达到 8.9%。1990—2000 年增加了 1.4 个百分点，2000—2010 年增加了 1.9 个百分点。人口老龄化加速的同时，中国 20 世纪 80 年代劳动力无限供给的状况也已发生转变。近些年来，中国一些沿海地区相继出现了“民工荒”的问题，普通劳动力的短缺使得未获得大学教育的劳动力比过去劳动力无限供给时更容易获得就业的机会。而另一方面，在连续多年实行大学扩招之后，中国的大学毕业生达到了前所未有的规模，2014 年有将近 700 万大学生毕业。大量的大学生进入劳动力市场必然会加剧就业市场上的竞争，很多大学生由于初次就业不理想从而选择继续寻求更体面的就业机会，相对于过去而言，就业市场上大学生之间的竞争变得更加激烈，由此出现了大学生就业难的问题。但是，相对于普通劳动力而言，大学教育获得者由于拥有知识和技能方面的优势，因而在城镇劳动力市场上仍然具有明显的优势。总体而言，在城镇样本中，相对于未接受大学教育的人而言，接受大学教育

的个体就业概率要高出 31.9%。

前面已经提到过，由于普通劳动力趋于短缺，工资水平不断增加，而大学生工资增幅缓慢，这就引出这样一个问题，中国大学教育的工资或者收入回报是否还存在呢？表 4 给出了大学教育影响收入的 2SLS 估计结果。可以看到，在最优窗宽下，大学教育对收入有着显著正向的影响。在加入基准控制变量之后（第 2 列），大学教育显著地增加了个人的收入，大学教育对收入对数的影响为 3.434 个点，相当于收入对数的 0.79 个标准偏移。由于本文中，受到大学及以上教育人群的平均受教育年限比没有受大学教育的人多 8.2 年，因此，年度回报率为 16.2%<sup>①</sup>。

表 4 大学教育对收入影响的 2SLS 估计

|        | 全部样本                |                     | 城镇样本                 |                     | 农村样本                |                     |
|--------|---------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
|        | (1)                 | (2)                 | (3)                  | (4)                 | (5)                 | (6)                 |
| 所有样本   | 6.233 **<br>(2.993) | 6.190 **<br>(3.041) | 5.669 ***<br>(2.510) | 5.335 **<br>(2.357) | 6.988 **<br>(3.444) | 7.442 **<br>(3.465) |
| 观察值数   | 14605               | 14601               | 6835                 | 6834                | 7770                | 7767                |
| 最优窗宽   | 3.412<br>(2.363)    | 3.434 *<br>(2.246)  | 3.379 *<br>(1.992)   | 3.271 *<br>(2.005)  | 3.323 *<br>(1.867)  | 3.653 **<br>(1.738) |
| 观察值数   | 7130                | 7128                | 3542                 | 3541                | 3588                | 3587                |
| 基准控制变量 | 否                   | 是                   | 否                    | 是                   | 否                   | 是                   |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和\*\*\*分别表示在 10%、5% 和 1% 的显著性水平下通过检验；所有模型均控制观察年份和区域虚拟变量；基准控制变量包括父亲的受教育年限、母亲的受教育年限、性别（男性=1）、民族（汉族=1）、城乡（城镇=1）和婚姻（在婚=1）。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

当我们将样本分成城镇样本和农村样本之后，发现大学教育对城镇样本和农村样本收入的影响都显著为正。从表 4 第 4 列可以看到，大学教育对城镇居民对数收入的影响为 3.271 个点，相当于城镇居民对数收入的 0.78 个标准偏移。由于城镇居民中受过大学教育的人群受教育年限比未受大学教育的人群高出 7.5 年，因此城镇大学教育的年度回报率为 17.1%。而从表 4 第 6 列可以看到，大学教育对农村居民对数收入的影响为 3.653 个点，相当于农村居民收入对数的 0.84 个标准偏移。由于农村居民中受过大学教育的人群的受教育

① 这里需要说明的是，本文中受过大专及以上教育的人群的平均受教育年限为 15.9 年，没有受过大专教育的人群的平均受教育年限为 7.3 年，因此大学教育的年度回报率为  $\sqrt[8.2]{3.434 - 1} = 0.162$ 。

年限比未受大学教育的人群高出 8.8 年，年度回报率为 15.9%。大学教育对城镇居民收入的影响明显高于对农村居民收入的影响。一个可能的解释是，中国长期以来城乡经济发展不平衡，形成城乡二元结构。城镇地区的行业类型往往更加需要高技术人才，对于一些高技术行业来说，只有接受高等教育的高技术人才能够胜任，而这些行业往往收入也比较高。相对于城镇的就业行业来说，农村地区主要行业是农业，相对于城镇地区而言，农村地区行业准入的教育壁垒没有那么高，因此高等教育的回报率也就更低一些。

由于人口结构转型，中国的劳动力短缺开始显现，对于普通的劳动力而言，找到一份工作变得越来越容易。然而，由于普通劳动力劳动技能提升有限，往往无法适应更加复杂的工种，因而长期来看，收入上升空间有限。受过高等教育的人虽然由于数量越来越庞大，就业竞争压力自然增大。但是，相较于普通劳动力而言，高等教育接受者能够从事更加复杂的劳动和工作任务，因此在总体的就业市场上相对于普通劳动者更具优势。而且，高等教育接受者一旦就业，由于他们具有较好的教育基础，往往能够通过自学适应更加复杂的工作岗位，因此，其收入的增长潜力也就更大，最终接受高等教育的那部分人的平均收入会显著高于未接受高等教育的普通劳动者。

高等教育与收入之间存在显著的因果关系，这就说明个人要想获得更高的收入水平，或者获得更高的收入增长潜力，接受高等教育是有必要的。由于本文研究使用的数据中，收入是当前观察值，教育水平是过去已经完成的，因此，高等教育与收入之间的因果关系其实在一定程度上也就反应了高等教育与生产率之间的因果关系。尽管中国的高等教育扩招受到了许多争议，比如说大学生失业、大学教育质量下降等，但是由于高等教育显著地提高了中国劳动者的劳动生产率，从这一点来看，高等院校扩招仍然具有一定的积极意义。

生活满意度是衡量某一社会中人们生活质量的的重要参数，往往可以用来评价一个人的主观幸福感。个人生活满意度与收入、健康、社会关系等都是有关联的，在一定程度上反映了个人的整体福利水平。前面探讨高等教育对就业和收入的影响，下面主要分析高等教育与个人整体福利水平，即生活满意度之间的因果关系。

表 5 列出了大学教育对所有样本、城镇样本和农村样本生活满意度影响的 2SLS 估计。第 1 列和第 2 列是所有样本情况下的回归结果，可以看到，在全样本下，大学教育对生活满意度的影响没有通过显著性检验，在最优窗宽下引入控制变量后也没有通过显著性检验。当我们将所有样本分成城镇样本和农村样本之后，我们发现，高等教育对城镇居民生活满意度的影响没有通过显著性检验，但大学教育对农村居民生活满意度的影响在最优窗宽下，不论是否引入基准控制变量，都显著为正。这就意味着，

虽然在所有样本条件下，大学教育对中国居民生活满意度的影响不显著，但是存在证据表明大学教育对农村居民的生活满意度有着正向的影响。

表 5 大学教育对生活满意度影响的 2SLS 估计

|        | 全部样本              |                  | 城镇样本             |                   | 农村样本               |                     |
|--------|-------------------|------------------|------------------|-------------------|--------------------|---------------------|
|        | (1)               | (2)              | (3)              | (4)               | (5)                | (6)                 |
| 所有样本   | 0.691<br>(0.751)  | 0.591<br>(0.797) | 0.115<br>(0.880) | -0.007<br>(0.767) | 1.620<br>(1.136)   | 1.537<br>(1.087)    |
| 观察值数   | 14236             | 14232            | 6776             | 6775              | 7460               | 7457                |
| 最优窗宽   | 1.504*<br>(0.879) | 1.487<br>(0.877) | 0.418<br>(1.026) | 0.430<br>(0.987)  | 2.822**<br>(1.282) | 3.343***<br>(1.093) |
| 观察值数   | 6968              | 6966             | 3521             | 3520              | 3447               | 3446                |
| 基准控制变量 | 否                 | 是                | 否                | 是                 | 否                  | 是                   |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和\*\*\*分别表示在 10%、5% 和 1% 的显著性水平下通过检验；所有模型均控制观察年份和区域虚拟变量；基准控制变量包括父亲的受教育年限、母亲的受教育年限、性别（男性=1）、民族（汉族=1）、城乡（城镇=1）和婚姻（在婚=1）。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

#### （四）稳健性检验

前面部分给出了方程（2）和（3）中的  $h(D)$  和  $g(D)$  非参数拟合条件下  $\beta_1$  和  $\alpha_1$  的估计值。通常人们在进行 RDD 估计时，还有一种方式就是将  $h(D)$  和  $g(D)$  在临界点前后设定为低阶多项式进行拟合。本文先进行 4 阶以内的多项式拟合，然后根据拟合优度跟 AIC 准则发现，当多项式阶数选择为 2 阶时是最优的多项式阶数。我们给出了在最优窗宽条件下方程（2）和（3）的参数估计结果，如表 6 所示。

表 6 的板块一给出了大学扩招影响大学教育获得率的参数估计结果，可以看到，大学扩招显著地增加了处理组大学教育获得的概率，估计结果与非参数估计值较为接近。板块二、三、四分别列出了大学教育对就业、收入和生活满意度影响的 2SLS 估计值，结果表明大学教育显著地增加了城镇样本的就业率，对农村样本的就业率没有产生影响；不论城镇还是农村，大学教育显著地增加了个人的收入；大学教育对城镇居民生活满意度没有影响，但是显著地增加了农村居民的生活满意度。表 6 的结果再次表明，大学教育与城镇居民就业率、所有居民的收入和农村居民的生活满意度之间存在因果关系。这一结论与非参数估计值是相一致的。



表 6 大学教育回报的参数估计结果

|                           | 全部样本                |                     | 城镇样本                |                     | 农村样本                 |                      |
|---------------------------|---------------------|---------------------|---------------------|---------------------|----------------------|----------------------|
|                           | (1)                 | (2)                 | (3)                 | (4)                 | (5)                  | (6)                  |
| 板块一：大学扩招对大学教育获得率的影响       |                     |                     |                     |                     |                      |                      |
| 最优窗宽                      | 0.057 **<br>(0.025) | 0.056 **<br>(0.025) | 0.068 **<br>(0.033) | 0.062 **<br>(0.032) | 0.055 ***<br>(0.022) | 0.056 ***<br>(0.022) |
| 观察值数                      | 7231                | 7229                | 3782                | 3587                | 3649                 | 3642                 |
| 板块二：大学教育对就业影响的 2SLS 估计    |                     |                     |                     |                     |                      |                      |
| 最优窗宽                      | 0.204<br>(0.171)    | 0.342 *<br>(0.191)  | 0.426 *<br>(0.288)  | 0.369 **<br>(0.180) | 0.112<br>(0.811)     | 0.314<br>(0.452)     |
| 观察值数                      | 7231                | 7229                | 3782                | 3781                | 3449                 | 3448                 |
| 板块三：大学教育对收入影响的 2SLS 估计    |                     |                     |                     |                     |                      |                      |
| 最优窗宽                      | 4.111 *<br>(2.474)  | 3.857 **<br>(1.749) | 4.002 *<br>(2.312)  | 3.155 *<br>(1.862)  | 3.251 *<br>(1.816)   | 4.388 *<br>(2.329)   |
| 观察值数                      | 7130                | 7128                | 3542                | 3541                | 3588                 | 3587                 |
| 板块四：大学教育对生活满意度影响的 2SLS 估计 |                     |                     |                     |                     |                      |                      |
| 最优窗宽                      | 2.203<br>(1.672)    | 2.608<br>(1.629)    | 2.884<br>(2.011)    | 2.441<br>(1.697)    | 2.997 **<br>(1.257)  | 3.121 **<br>(1.474)  |
| 观察值数                      | 6968                | 6966                | 3521                | 3520                | 3447                 | 3446                 |
| 基准控制变量                    | 否                   | 是                   | 否                   | 是                   | 否                    | 是                    |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和 \*\*\* 分别表示在 10%、5% 和 1% 的显著性水平下通过检验；所有模型均控制观察年份和区域虚拟变量；基准控制变量包括父亲的受教育年限、母亲的受教育年限、性别（男性=1）、民族（汉族=1）、城乡（城镇=1）和婚姻（在婚=1）。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

## 五 有效性检验

RDD 有效的前提条件主要有 3 个。首先，排序变量在临界点附近不能够完全受到人为的操纵，即个人不能够自由地选择在处理组里或者在控制组里。其次，除了核心变量（variable of interest，本文是大学教育获得率）之外，其他的控制变量在临界点附近不能够发生显著的跳跃，否则，我们无法判断到底是核心变量在临界点附近跳跃导致结果变量的跳跃，还是其他控制变量的跳跃导致结果变量的跳跃。因此本文中，除了大学获得率在临界点附近有断点之外，其他控制变量应该都是连续的。最后，断点是由政策变化引起的，而不能是由本身的发展趋势引起的，否则，断点不一定是由外生的政策变化引起，而是本身的发展趋势。本部分我们将对本文的 RDD 估计结果进行有效性检验。

### （一）排序变量的随机性检验

已有的研究表明，当排序变量（rating variable）不能够完全被个体操纵时，与传统的 IV 估计相比较而言，RDD 估计结果更具随机实验的特征（Lee，2008）。本文中的排序变量是出生年份和季度，个人不能够选择自己的出生日期，父母也不可能预测到 1999 年高考扩招从而使得自己的子女刚好出生在临界点之前或者之后。虽然不排除一些人为了某种目的在出生以后修改出生日期<sup>①</sup>，但是由于中国采用了户口登记制度，出生日期一旦确定要想修改是非常困难的。因此，本文中即使有人在临界点附近对出生日期进行了操纵，这种情况所占的比重也是极低的，因而不可能对总体结论造成影响。

图 6 给出了在临界点前后不同出生季度人群的频数散点图，可以看到，在临界点附近，不同出生日期的人群频数并没有发生明显的跳跃。该图说明本文的样本人群在临界点附近并不存在非随机排序的证据，即排序变量的随机性得以满足。

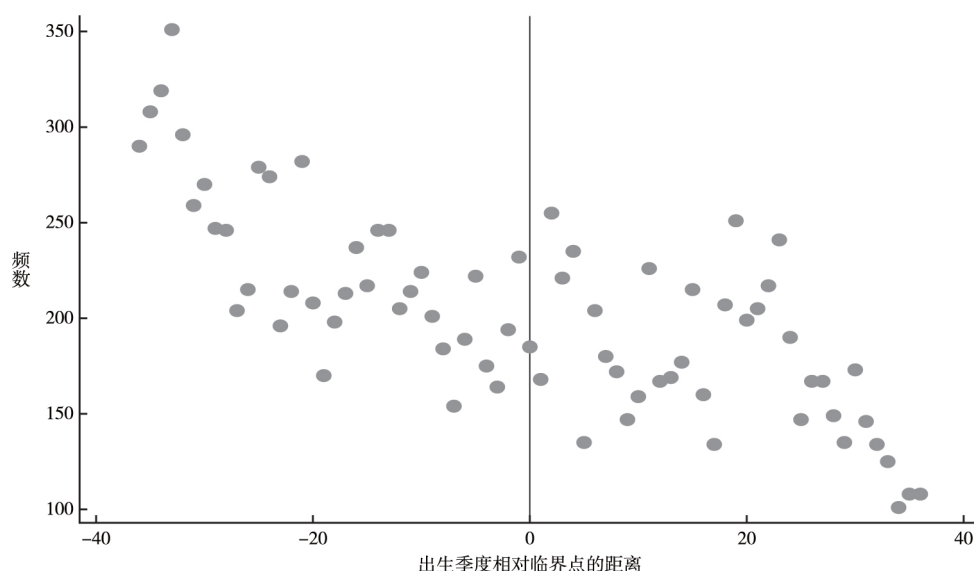


图 6 1981 年第 3 季度前后出生人口的频数分布

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

<sup>①</sup> 比如说为了升学或者升官降低或者增加年龄。

## （二）控制变量的连续性检验

RDD 估计有效要求除了内生性的核心变量在临界点处出现断点之外，其余的控制变量在临界点处应当是连续的。表 7 给出了本文中控制变量的连续性检验，从中可以看到，不论是在所有样本设定下，还是在最优窗宽设定下，本文中基准控制变量在临界点处都不能拒绝连续性假设，即控制变量的连续性条件是满足的。

表 7 控制变量的连续性检验

| 被解释变量 | 城镇                | 性别               | 民族                | 婚姻                | 父亲教育             | 母亲教育             |
|-------|-------------------|------------------|-------------------|-------------------|------------------|------------------|
| 所有样本  | -0.007<br>(0.017) | 0.003<br>(0.014) | -0.002<br>(0.005) | -0.006<br>(0.011) | 0.027<br>(0.038) | 0.046<br>(0.044) |
| 观察值数  | 14822             | 14822            | 14822             | 14822             | 14819            | 14818            |
| 最优窗宽  | -0.008<br>(0.018) | 0.025<br>(0.015) | -0.001<br>(0.011) | -0.004<br>(0.024) | 0.045<br>(0.045) | 0.045<br>(0.053) |
| 观察值数  | 7231              | 7231             | 7231              | 7231              | 7230             | 7229             |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和 \*\*\* 分别表示在 10%、5% 和 1% 的显著性水平下通过检验；所有模型都包含观察年份和省份虚拟变量。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

## （三）安慰剂（Placebo）检验

为了检验临界点前后的断点就是由 1999 年的高等院校扩招造成，而非由时间趋势造成，我们最后假定两个高考扩招的时间点，分别是 1997 年或者 2001 年。1997 年高校扩招年份对应的出生临界点分别是 1979 年第 3 季度前后，而 2001 年对应的则是 1983 年第 3 季度前后，通过构造这两个反事实的高考扩招政策变化的时间点，我们可以对前文高考扩招效应进行敏感性检验。如果反事实的结果不成立，就说明本文高等教育获得的跳跃就是由高校扩招造成的，而非时间趋势的结果。表 8 报告了高考扩招政策变动影响的反事实估计结果。

我们首先假定高等教育扩招提前了 2 年<sup>①</sup>，即 1997 年开始高等教育扩招。由于我们假定的扩招年份提前了两年，在进行断点检验时，主要观察 1979 年第 3 季度前后 2 年出生人群的高等教育获得率在假定的临界点，即 1979 年第 3 季度前后是否存在断点。从表 8 第（1）列和第（2）列可以看到，不论是否加入控制变量，不论在临界点前后

① 由于在做这种稳健性检验时不能够包含发生了显著跳跃的临界点前后出生人群，而为了保证一定的观察值数，我们选择临界点前后 2 年的时间作为假想的临界点做安慰剂检验。

2 年还是前后 1 年，高等教育获得率都没有出现明显的断点。之后我们假定高等教育扩招推迟了 2 年，即 2001 年开始高等教育扩招。这一次在进行断点检验时，主要观察 1983 年第 3 季度前后 2 年出生人群的高等教育获得率在假定的临界点，即 1983 年第 3 季度前后是否存在断点。从第（3）列和第（4）列可以看到，我们再次发现在所有设定下断点都是不存在的。表 8 的结果再次证实 1981 年第 3 季度前后出生人群的高等教育获得的断点是高考政策引起的变动，而非时间趋势的结果。

表 8 利用假想的高考扩招年份检验大学教育的连续性

| 大学教育获得率    | (1)               | (2)               | (3)              | (4)              |
|------------|-------------------|-------------------|------------------|------------------|
|            | 1997 年为假想的临界点     |                   | 2001 年为假想的临界点    |                  |
| 临界点前后 8 季度 | -0.030<br>(0.024) | -0.037<br>(0.019) | 0.023<br>(0.021) | 0.025<br>(0.018) |
| 观察值数       | 1699              | 1698              | 1755             | 1755             |
| 临界点前后 4 季度 | -0.032<br>(0.025) | -0.031<br>(0.018) | 0.047<br>(0.032) | 0.080<br>(0.051) |
| 观察值数       | 950               | 950               | 1064             | 1064             |
| 基准控制变量     | 否                 | 是                 | 否                | 是                |

注：括号中的数值是稳健的出生年份聚类标准误；\*、\*\*和 \*\*\* 分别表示在 10%、5% 和 1% 的显著性水平下通过检验。

资料来源：根据 CFPS 2010 年和 2012 年数据计算得到。

## 六 结论

改革开放后，中国经历了高速经济增长。随着义务教育法的实施和高等院校扩招，中国人的平均受教育年限加速上涨。人力资本快速积累成为中国经济高速增长的主要动力之一。然而，1999 年开始的高校扩招一方面使得越来越多的中国人接受了高等教育，加速了人力资本的积累，同时也带来了大学生面临着越来越难的就业问题。而另一方面，自进入 21 世纪后，中国人口老龄化加速，劳动力短缺的问题开始显现。不少地区出现“民工荒”问题，普通劳动力成本迅速上涨。普通劳动力工资上涨与大学生就业难形成鲜明的对比，于是社会上出现了这样一种观点，即“读书无用论”。

由于教育的内生性使得 OLS 估计可能发生偏误，本文基于 1999 年高等院校扩招这

样一个事实，采用 RDD 方法检验中国的高等教育对个体就业、工资和生活满意度的因果影响。我们先是基于一个简化式模型检验高等教育扩招对大学入学率的影响，发现在受高校扩招影响临界点的前后出现了明显的高等教育获得率的断点，高校扩招使得处理组人群获得大学教育的概率提高了 5.6 个百分点。当我们将所有样本分成城镇样本和农村样本后，我们发现不论是城镇样本还是农村样本，临界点前后都存在大学获得率的断点。大学扩招使得城镇样本处理组人群获得大学教育的概率提高了 7.0%，使得农村样本处理组人群大学教育的概率提高了 5.1%。

接下来我们基于非参数结构式模型检验大学教育获得与个人就业、收入和生活满意度之间的因果关系。我们发现在城镇样本中，接受大学教育的人就业率显著地高于未接受大学教育的人，而在农村样本中，是否接受大学教育对就业的影响不显著。在城镇样本中，相较于未接受大学教育的人群而言，受过大学教育的人就业高出 31.9%。不论是城镇样本，还是农村样本，大学教育都显著地提高了个人收入。大学教育对城镇居民收入的年度回报率为 17.1%，对农村居民收入的年度回报率为 15.9%。虽然大学教育对个人生活满意度的总体影响不显著，但是我们也发现了大学教育提高农村居民生活满意度的证据。本文的最后还给出了 RDD 的参数估计检验结果和 RDD 估计有效的前提假设检验，所有的检验均支持前述结论的有效性。

当然，本文的研究也存在一些不足之处。首先，本文的样本人群平均年龄仅为 29.8 岁，样本年龄比较年轻，许多人可能刚刚从学校毕业进入劳动力市场，这部分人群由于正在找工作，或者工作刚刚起步，收入水平有可能并不是劳动生产率的体现，受到了教育文凭符号效应（signal effect）的影响可能性较大。因此，要想更加准确地反映大学教育的生产率效应，还得进一步寻找证据。其次，由于本文的样本人群相对来说比较年轻，一些本应该在处理组中的人群因为仍然在校而被排除掉了，这就使得大学扩招对大学教育的影响估计有所偏低。最后，RDD 估计结果本身是一种局部处理效应，估计的因果关系在临界点附近是有效的，该估计结果具有较高的内部有效性，但是缺乏外部有效性。

在忍受上述几点不足之后，本文的研究表明，虽然高等教育扩招之后，中国的大学毕业生面临很大的就业压力，但是我们发现大学教育对城镇居民就业和所有居民的收入有着显著的正向促进作用，此外，我们也发现大学教育提高了农村居民生活满意度的微弱证据。本文在一定程度上肯定了中国 1999 年高等院校扩招的积极意义，同时也证实了大学教育对提高劳动生产率存在显著的积极影响。

## 参考文献：

- Angrist , Joshua & Victor Lavy ( 1999) . Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* , 114 ( 2) , 533 – 575.
- Austin , Nichols ( 2007) . Causal Inference with Observational Data. *Stata Journal* , 7 ( 4) , 507 – 541.
- Banks , James & Fabrizio Mazzonn ( 2012) . The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design. *The Economic Journal* , 122 , 418 – 448.
- Calonico , Sebastian , Cattaneo Matias & Rocio Titiunik ( 2014) . Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* , 82 , 2295 – 2326.
- Card , David & Lara Shore-Sheppard ( 2004) . Using Discontinuity Eligibility to Identify the Effect of the Federal Medicaid Expansions on Low-Income Children. *Review of Economics and Statistics* , 86( 3) , 752 – 766.
- Chen , Guifu & Shigeyuki Hamori ( 2009) . Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach. *China Economic Review* , 20 , 143 – 152.
- Clark , Damon & Paco Martorell ( 2014) . The Signal Value of a High School Diploma. *Journal of Political Economy* , 122( 2) , 282 – 318.
- Fang , Hai , John Rizzo , Scott Rozelle & Richard Zeckhauser ( 2012) . The Returns to Education in China: Evidence from the 1986 Compulsory Education Law. *NBER Working Paper* , No. 18189.
- Fleisher , Belton , Haizheng Li & Mingqiang Zhao ( 2010) . Human Capital , Economic Growth , and Regional Inequality in China. *Journal of Development Economics* , 92( 2) , 215 – 231
- Fleisher , Belton & Xiaojun Wang ( 2005) . Returns to Schooling in China Under Planning and Reform. *Journal of Comparative Economics* , 33( 2) , 265 – 277.
- Heckman , James & Xuesong Li ( 2004) . Selection Bias , Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000. *Pacific Economic Review* , 9( 3) , 155 – 171.

- Imbens , Guido & Karthik Kalyanaraman ( 2012 ) . Optical Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies* ,79 ,933 – 959.
- Lee , David ( 2008 ) . Randomized Experiment from Non-Random Selection in U. S. House Elections. *Journal of Econometrics* ,142( 2) ,675 – 697.
- Ludwig ,Jens & Douglas Miller ( 2007) . Does Head Start Improve Children’s Life Chances? Evidence from Regression Discontinuity Designs. *Quarterly Journal of Economics* ,122 ,159 – 208.
- La , Vincent( 2014) . Does Schooling Pay? Evidence from China. *MPRA Paper* , No. 54578.
- Zhang ,Junsen ,Yaohui Zhao ,Albert Park & Xiaoqing Song ( 2005) . Economic Returns to Schooling in Urban China ,1988 to 2001. *Journal of Comparative Economics* ,33( 4) ,730 – 752.

## Returns to College Education: A Natural Experiment from Expansion of Higher Education in China

Liu Shenglong & Hu Angang

( School of Public Policy & Management ,Tsinghua University;  
Institute for Contemporary China Studies ,Tsinghua University)

**Abstract:** Based on the empirical evidences after China’s expansion of higher education in 1999 ,this paper uses regression discontinuity design to test the linkage from college education to employment , income , and life satisfaction. The results of this paper are as follows: firstly , the expansion of higher education significantly increases the employment rate in the urban sample. Secondly , there is a significant cause relationship between higher education and income in both urban sample and rural sample. The annual return to college education is respectively 15.8 percent in urban sample and 13.7 percent in rural sample. Thirdly , there is no significant association between college education and life satisfaction in the total sample; but college education has impacts on life satisfaction in the rural sample. Although there are a series of problems and challenges after China’s expansion of higher education , including declining education quality and low employment after graduation , this paper still identifies some positive returns on China’s expansion of higher education.

**Keywords:** expansion of higher education , return to college , regression discontinuity design

**JEL Classification:** I26 , I28 , J24

( 责任编辑: 一帆)