



国际教育学院

School of International Education

工具变量的应用

教育对收入的影响

公选课《可计量的社会》

讲课人：何濯羽

参考论文：Angrist & Krueger (1991)



引言



研究方法



研究发现



学者简介

Part **1**

引言

1-1 研究问题

1-2 研究难点

课前讨论：教育的价值是什么？



梁植 (1987年~)

- 清华大学2007年十佳歌手
- 清华大学2010届法律学士
- 清华大学2012届金融硕士
- 清华大学2012级新闻传播博士
- 2013年第4届北京大学生戏剧节最佳男演员奖
- 2014年第21届北京大学生电影节全国主持人选拔大赛全国总冠军

“我现在有很大的纠结，因为我面临明年毕业嘛，然后我本科学了法律，硕士学了金融，博士学了新闻传播，我也在想，应该做什么样的工作，然后会让这三个专业能发挥作用。”

—— 2014年参加《奇葩说》海选时的发言

课前讨论：教育的价值是什么？



**名校是镇国重器，
不是用来找工作用的！**

课前讨论：贷款上大学，值得吗？

观看视频： *Student debt: Is a degree worth the financial strain?* (CBS Sunday Morning, 2022 May 15)



研究问题

教育对收入有什么影响？

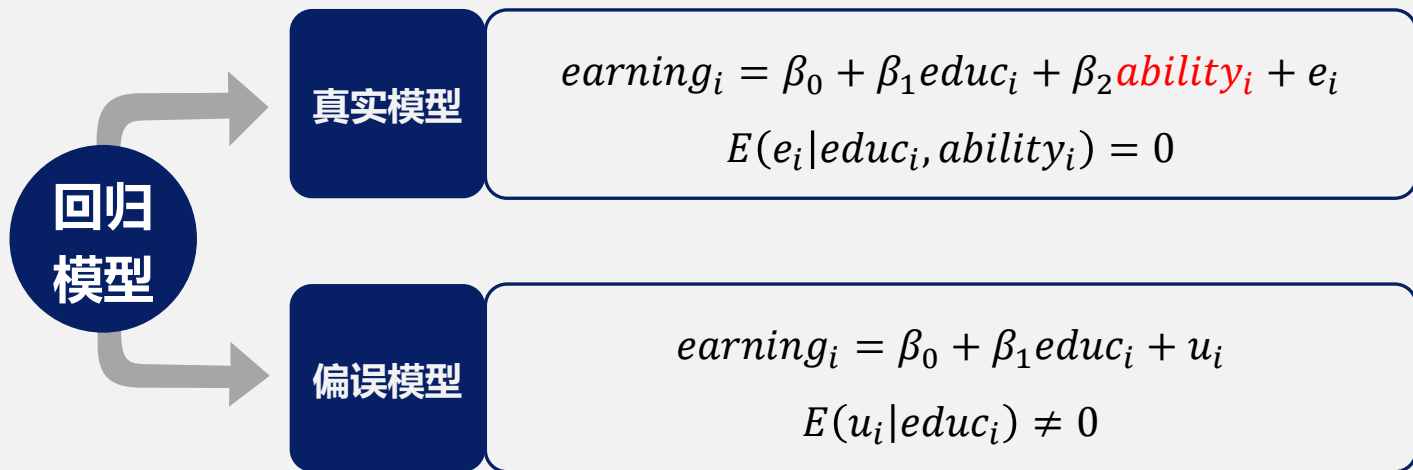
令 $Y_i(1)$ 代表我接受高等教育的未来收入， $Y_i(0)$ 代表我不接受高等教育的未来收入。

高等教育对我的收入的影响是 $Y_i(1) - Y_i(0)$

高等教育对总体收入的影响是 $E[Y(1) - Y(0)]$

难点： 我们永远无法同时观察到 $Y_i(1)$ 和 $Y_i(0)$ 。

研究难点：天赋难以测量



Part **2**

研究方法

2-1 工具变量

2-2 背景介绍

2-3 假设检验

工具变量

$$earning_i = \beta_0 + \beta_1 educ_i + u_i \quad \text{结构方程}$$

$$educ_i = \alpha_0 + \alpha_1 Z_i + \varepsilon_i \quad \text{第一阶段方程}$$

排他性 条件

$$E(Z_i u_i) = 0$$

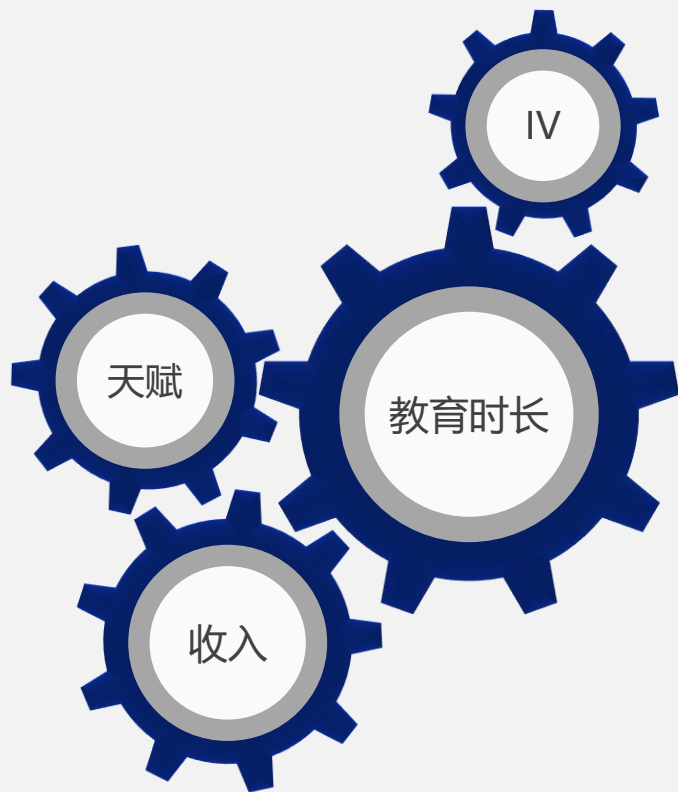
工具变量与结构方程中的误差项不相关；换言之，工具变量只能通过 $educ_i$ 影响结果变量（即第一阶段方程描述的影响渠道）。

相关性 条件

$$Cov(Z_i, educ_i) \neq 0$$

工具变量与 $educ_i$ 有较强的线性相关性。

有效的工具变量：出生年份×出生季度



$$\ln W_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_c Y_{ic} \xi_c + \rho E_i + \mu_i$$

$$E_i = \mathbf{X}_i \boldsymbol{\pi} + \sum_c Y_{ic} \delta_c + \sum_c \sum_j Y_{ic} Q_{ij} \theta_{jc} + \epsilon_i$$

方程释义

W_i 代表 i 每周的薪资收入；

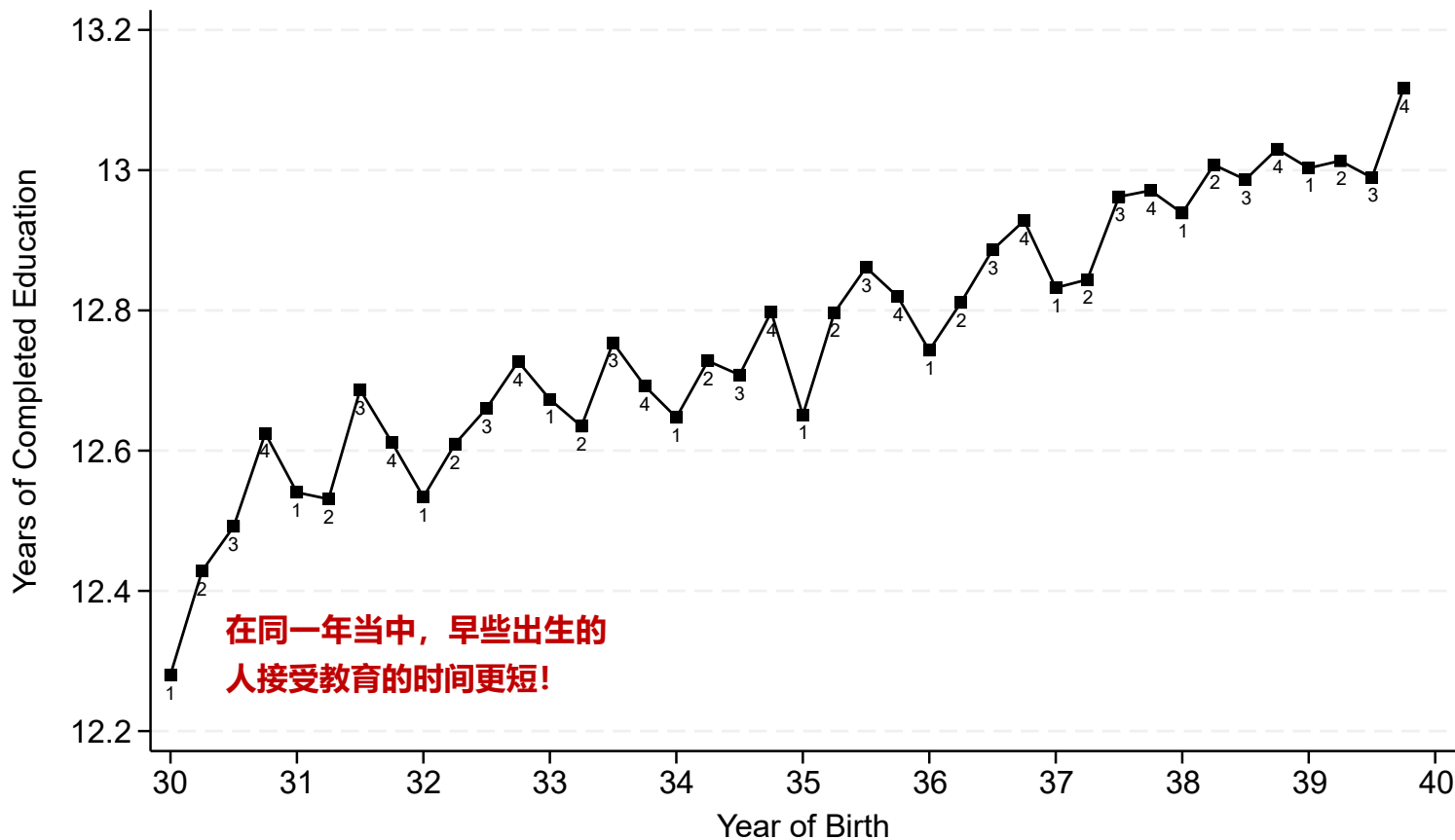
\mathbf{X}_i 代表控制变量；

E_i 代表 i 接受教育的时间长度（单位：年）

$Y_{ic} = 1$ if i 出生于 c 年；

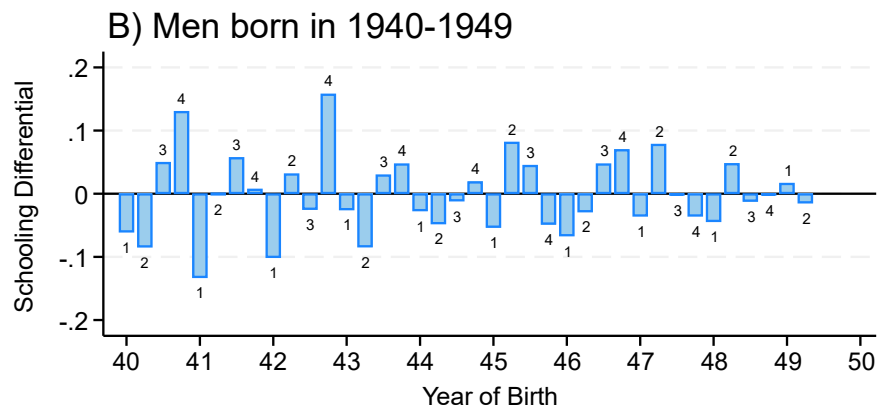
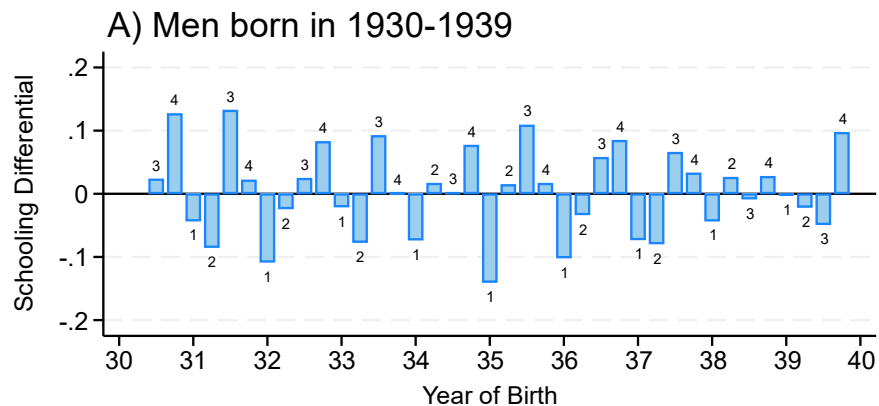
$Q_{ij} = 1$ if i 出生于第 j 季度。

检验相关性条件：受教育的时间长度与出生季度相关？！



数据来源：美国1970年和1980年人口普查数据

检验相关性条件：受教育的时间长度与出生季度相关？！



该现象出现的原因：**美国义务教育法的年龄规定！**

- 1) 美国小学通常要求学生在入学当年的1月1日前年满6周岁。因此，早出生的学生入学时比其他学生稍大一些。例如：Amy出生于2000年1月15日（第1季度），Bob出生于2000年11月15日（第4季度），则Amy和Bob都可以在2007年入学。但是，在2007年1月1日，Amy是6.96岁，Bob是6.13岁。
- 2) 美国许多州设有义务教育法，要求学生在达到特定年龄（通常是16、17或18岁）之前必须留在学校。因此，出生较早的学生将更早地达到特定年龄，更早地被允许退学，因此，受教育的时间更短。

检验相关性条件：受教育的时间长度与出生季度相关？！

| Outcome variable | Birth cohort | Mean | Quarter-of-birth effect ^a | | | <i>F</i> -test ^b [<i>P</i> -value] |
|--|--------------|-------|--------------------------------------|-------------------|-------------------|---|
| | | | I | II | III | |
| Total years of education | 1930–1939 | 12.79 | –0.124 (0.017) | –0.086 (0.017) | –0.015 (0.016) | 24.9 [0.0001] |
| | 1940–1949 | 13.56 | –0.085 (0.012) | –0.035 (0.012) | –0.017 (0.011) | 18.6 [0.0001] |
| High school graduate | 1930–1939 | 0.77 | –0.019 (0.002) | –0.020 (0.002) | –0.004 (0.002) | 46.4 [0.0001] |
| | 1940–1949 | 0.86 | –0.015 (0.001) | –0.012 (0.001) | –0.002 (0.001) | 54.4 [0.0001] |
| Years of educ. for high school graduates | 1930–1939 | 13.99 | –0.004 (0.014) | 0.051 (0.014) | 0.012 (0.014) | 5.9 [0.0006] |
| | 1940–1949 | 14.28 | 0.005 (0.011) | 0.043 (0.011) | –0.003 (0.010) | 7.8 [0.0017] |
| College graduate | 1930–1939 | 0.24 | –0.005 (0.002) | 0.003 (0.002) | 0.002 (0.002) | 5.0 [0.0021] |
| | 1940–1949 | 0.30 | –0.003 (0.002) | 0.004 (0.002) | 0.000 (0.002) | 5.0 [0.0018] |
| Completed master's degree | 1930–1939 | 0.09 | –0.001 (0.001) | 0.002 (0.001) | –0.001 (0.001) | 1.7 [0.1599] |
| | 1940–1949 | 0.11 | 0.000 (0.001) | 0.004 (0.001) | 0.001 (0.001) | 3.9 [0.0091] |
| Completed doctoral degree | 1930–1939 | 0.03 | 0.002 (0.001) | 0.003 (0.001) | 0.000 (0.001) | 2.9 [0.0332] |
| | 1940–1949 | 0.04 | –0.002 (0.001) | 0.001 (0.001) | –0.001 (0.001) | 4.3 [0.0050] |

a. Standard errors are in parentheses. An $MA(+2, -2)$ trend term was subtracted from each dependent variable. The data set contains men from the 1980 Census, 5 percent Public Use Sample. Sample size is 312,718 for 1930–1939 cohort and is 457,181 for 1940–1949 cohort.

b. *F*-statistic is for a test of the hypothesis that the quarter-of-birth dummies jointly have no effect.

OLS回归检验：

$$E_{icj} - MA_{cj} = \alpha + \sum_j^3 \beta_j Q_{icj} + \varepsilon_{icj}$$

我们发现：第一季度出生的男性平均接受教育的年数比同年最后一个季度出生的男性低约1/10。

检验排他性条件：筛查出生季度影响个人收入的其他渠道

相比于相关性条件的检验，我们很难完美地检验排他性条件的成立。

Angrist & Krueger (1991) 考虑了众多工具变量影响个人收入的渠道，然后——讨论它们的不可能性。

3. 大学毕业生的收入

Angrist & Krueger (1991) 对大学毕业生（他们的教育没有因义务教育而延长）的样本数据进行了 TSLS 回归分析（因变量是收入，自变量是出生季度），发现没有显著结果。

1. 教育心理学

相比于同级的其他学生，出生在第一季度的学生更年长、更成熟，因此可能更熟练地掌握知识。这一渠道的存在意味着该研究的发现存在 downward bias——即：研究结果本应呈现更大的数值。

2. 父母的社会地位

一些学者认为：孩子的出生季度可能与父母的社会地位相关，而父母的社会地位会影响孩子工作后的收入。可是，Lam & Miron (1987) 提供了许多证据证明这两者不相关。

Part 3

研究发现

研究发现：使用1970年人口普查数据得出的结论

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1920–1929: 1970 CENSUS^a

| Independent variable | (1) OLS | (2) TSLS | (3) OLS | (4) TSLS | (5) OLS | (6) TSLS | (7) OLS | (8) TSLS |
|-------------------------------|--------------------|--------------------|---------------------|---------------------|--------------------|---------------------|---------------------|---------------------|
| Years of education | 0.0802 (0.0004) | 0.0769 (0.0150) | 0.0802 (0.0004) | 0.1310 (0.0334) | 0.0701 (0.0004) | 0.0669 (0.0151) | 0.0701 (0.0004) | 0.1007 (0.0334) |
| Race (1 = black) | — | — | — | — | 0.2980 (0.0043) | -0.3055 (0.0353) | -0.2980 (0.0043) | -0.2271 (0.0776) |
| SMSA (1 = center city) | — | — | — | — | 0.1343 (0.0026) | 0.1362 (0.0092) | 0.1343 (0.0026) | 0.1163 (0.0198) |
| Married (1 = married) | — | — | — | — | 0.2928 (0.0037) | 0.2941 (0.0072) | 0.2928 (0.0037) | 0.2804 (0.0141) |
| 9 Year-of-birth dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 8 Region of residence dummies | No | No | No | No | Yes | Yes | Yes | Yes |
| Age | — | — | 0.1446 (0.0676) | 0.1409 (0.0704) | — | — | 0.1162 (0.0652) | 0.1170 (0.0662) |
| Age-squared | — | — | -0.0015 (0.0007) | -0.0014 (0.0008) | — | — | -0.0013 (0.0007) | -0.0012 (0.0007) |
| χ^2 [dof] | — | 36.0 [29] | — | 25.6 [27] | — | 34.2 [29] | — | 28.8 [27] |

a. Standard errors are in parentheses. Sample size is 247,199. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the State, County, and Neighborhoods 1 percent samples of the 1970 Census (15 percent form). The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

提示：因变量是 $\ln W_i$ ，因此在解释估计值时应使用 $(e^{\beta\Delta} - 1) \times 100\%$ 公式。

研究发现：使用1980年人口普查数据得出的结论

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS^a

| Independent variable | (1) OLS | (2) TSLS | (3) OLS | (4) TSLS | (5) OLS | (6) TSLS | (7) OLS | (8) TSLS |
|-------------------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Years of education | 0.0711 (0.0003) | 0.0891 (0.0161) | 0.0711 (0.0003) | 0.0760 (0.0290) | 0.0632 (0.0003) | 0.0806 (0.0164) | 0.0632 (0.0003) | 0.0600 (0.0299) |
| Race (1 = black) | — | — | — | — | -0.2575 (0.0040) | -0.2302 (0.0261) | -0.2575 (0.0040) | -0.2626 (0.0458) |
| SMSA (1 = center city) | — | — | — | — | 0.1763 (0.0029) | 0.1581 (0.0174) | 0.1763 (0.0029) | 0.1797 (0.0305) |
| Married (1 = married) | — | — | — | — | 0.2479 (0.0032) | 0.2440 (0.0049) | 0.2479 (0.0032) | 0.2486 (0.0073) |
| 9 Year-of-birth dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 8 Region-of-residence dummies | No | No | No | No | Yes | Yes | Yes | Yes |
| Age | — | — | -0.0772 (0.0621) | -0.0801 (0.0645) | — | — | -0.0760 (0.0604) | -0.0741 (0.0626) |
| Age-squared | — | — | 0.0008 (0.0007) | 0.0008 (0.0007) | — | — | 0.0008 (0.0007) | 0.0007 (0.0007) |
| χ^2 [dof] | — | 25.4 [29] | — | 23.1 [27] | — | 22.5 [29] | — | 19.6 [27] |

a. Standard errors are in parentheses. Sample size is 329,509. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the 5 percent sample of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

提示：因变量是 $\ln W_i$ ，因此在解释估计值时应使用 $(e^{\beta\Delta} - 1) \times 100\%$ 公式。

研究总结

与出生季度相关的教育差异的出现是由于一些人出生日期的随机性。因为美国义务教育法的允许退学年龄限制，一些人“被迫”比其他人上学的时间更长。使用出生季节作为回归模型中教育变量的工具变量，我们发现 OLS 和 TSLS 对教育回报的估计值之间存在明显的相似性——均接近0.08，即接受教育的时长增加1年，会使得个人收入增长 $(e^{0.08} - 1) \times 100\% \approx 8.33\%$ 。



Part **4**

学者简介

4-1 Joshua D. Angrist

4-2 Alan B. Krueger

Joshua D. Angrist

Joshua David Angrist (1960年9月18日出生) 是一位以色列裔美国经济学家，也是**麻省理工学院经济学教授**。Angrist 与 Guido Imbens (荷兰裔美国经济学家、斯坦福大学经济学教授) 因其对因果关系分析方法论的贡献而被授予**2021年诺贝尔经济学奖**。

他在**计量经济学、教育经济学、劳动经济学和城市经济学**方面跻身世界顶尖经济学家之列，并以利用准实验研究设计(如工具变量)研究公共政策的影响而闻名。他是麻省理工学院Blueprint Labs的联合创始人兼联合主任，该实验室致力于研究美国人力资本与收入不平等之间的关系。



Alan B. Krueger

Alan Bennett Krueger (1960年9月17日~2019年3月16日) 是一位美国经济学家，曾任**普林斯顿大学政治经济学教授**和国家经济研究局研究员。2009年5月至2010年10月，他担任美国总统巴拉克·奥巴马提名的助理财政部长。2011年，他被奥巴马提名为**白宫经济顾问委员会主席**，并于2011年11月至2013年8月期间任职。

根据经济学研究论文的贡献度，他是世界排名最高的50位经济学家之一。他创新地运用经济学中的自然实验，挑战了最低工资对就业产生不利影响的经济学主流观点。他还对不平等和教育的经济影响的研究做出了突出贡献。





国际教育学院

School of International Education

讲课完毕 感谢聆听

复现论文所需数据与代码: https://github.com/IanHo2019/Return_to_Education_AK1991

讲课人: 何濯羽

参考论文: Angrist & Krueger (1991)