Original Research Article

The Causal Effect of High Education on Individuals' Income in Rural and Urban China

Tao Zou

Communication University of China School of Data Science and Media Intelligence ,Beijing 100020, China

Abstract: Children's education has been a hot topic in China in recent years. To explore the causal impact of education on earnings, this paper uses the PSM method in the counterfactual framework. The paper uses CGSS observational data to artificially construct control and experimental groups, with high education as the treatment variable and the annual personal income as the outcome variable. The study shows that the net effect of education on income is the same for both urban and rural individuals. This shows that China's labour market recognizes education fairly.
Keywords: PSM ; ATT ; Logistic model ; Stratification matching

1. Introduction

Education is a way to enhance the value of human capital, and it is also seen as the best way for underprivileged children to make the leap to higher social class. Income levels reflect to some extent the social class of individuals and are the most direct indicator of the role of education in improving human capital. As China's economy continues to develop, the demand for highly qualified labour for social jobs is rising.

2. Theory and method

Propensity Score Matching (PSM) is the method used to simulate the experimental group versus the control group[1]. PSM is a two-step process. The first step is to calculate a propensity score for each sample. The propensity score represents the probability of the sample receiving the experimental treatment, which is predicted by a logistic model. The second step is to match individuals with similar propensity scores. If two sample can be matched, it means that they have the same probability of receiving the experimental treatment. The sample that actually receives the treatment belongs to the experimental group and the other sample that actually do not receive the treatment belongs to the control group. Such differences in the outcome variable in both groups is precisely the causal effect in the counterfactual framework.

Specifically, the expression for calculating the Average Treatment effect for the Treated (ATT) is:

$$ATT = \pi[E(Y_1|w=1) - E(Y_0|w=1)] + (1-\pi)[E(Y_1|w=0) - E(Y_0|w=0)] \#(1)$$

Where $w=1$ denotes individuals in the experimental group and $w=0$ denotes individuals in the control group. Y_1 denotes the outcome variable for the treated individuals, while Y_0 denotes the outcome variable for individuals who didn't receive the treatment. π denotes the proportion of individuals in the experimental group to the overall sample. If the experimental individuals are assigned to the experimental and control groups in a random way, there are:

$$E(Y_1|w=1) = E(Y_1|w=0) \#(2)$$

$$E(Y_0|w=1) = E(Y_0|w=0) \#(3)$$

Thus, the calculated expression for ATT is reduced to:

$$ATT = E(Y_1|w=1) - E(Y_0|w=0) \#(4)$$

If P is used to denote an individual's propensity score, then the formula for calculating ATT is:

$$ATT = E[E_P(Y_1|w=1, P) - E_P(Y_0|w=0, P)] \#(5)$$

3. Data and variables

This paper uses China General Social Survey (CGSS) Data from 2013, 2015 and 2017. This is a large-scale, China-wide sampling project initiated by Renmin University of China.

In this paper, relatively high educational attainment is used as the treatment variable. Due to the existence of Credential Inflation over the last few decades, the same level of education has had a different value at different times[2]. In order to distinguish individuals who have acquired a high level of education, this paper distinguishes the top 25% of individuals with the highest level of education in

Copyright © 2021 Tao Zou
doi: 10.18282/l-e.v10i9.3200
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

each 5-year period as relatively highly educated.

The factors that influence an individual's attainment of a high level of education are divided into two parts, one being personal qualities and the other being family background. Age, gender and household registration are the main components of personal qualities. In rural and western areas of China, the inequality of access to education is significantly greater within women than men[3]. In addition, household registration can affect children's income by limiting access to educational opportunities[4]. This paper selects individuals with birth years between 1931 and 1990 and classifies the hukou variable into 0-1 variables according to whether or not these people were ever rural household registration.

Income is the outcome variable in this paper, which is measured by the individual's total annual income for the last year in the CGSS.

4. Implementation

4.1 Propensity score estimation

After removing missing values and invalid samples, the total remaining sample size is 11,304, of which 7,174 are from rural household registration and 4,130 from urban household registration. The distribution of propensity scores predicted by the model corresponding to the two different household registration groups is shown below. The difference between the two distributions shows that rural individuals have less access to high education than urban individuals overall.

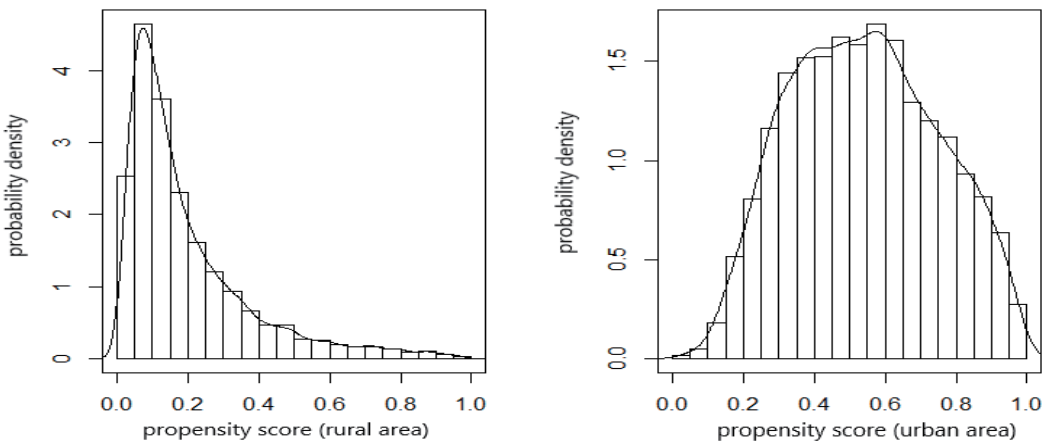

Figure 1 Distribution of propensity score

4.2 Stratification matching for propensity score

In this paper, the propensity scores are divided into 10 layers using 0.1 as the interval length. Individuals are classified into corresponding intervals based on their predicted propensity score. Individuals located in the same interval are divided into experimental and control groups according to whether they actually obtained a high level of education or not. Then the ATT values are obtained by differencing the incomes of control group and experimental group in the same propensity score interval, which are shown in detail in the Table 1. The weighted ATT is obtained by calculating the average of the ATT using the number of samples in the different propensity score layers as weights. In addition, groups with sample sizes less than 100 were removed before calculating the weighted ATT in order to obtain robust estimate. The ATT for rural area is 21998 yuan, while the ATT for urban area is 22915 yuan.

5. Conclusion

This paper is a study of the causal impact of high educational attainment on income in China. In order to artificially construct the experimental and control groups, this paper makes use of the propensity score matching method (PSM). PSM first calculates the probability of each individual obtaining a high level of education, and then matches individuals of similar probability to make a difference to calculate the causal effect.

In the process, this paper finds that urban individuals are overall more likely to obtain high levels of education than the rural. For both rural and urban individuals, the average income return from a high level of education is a little over RMB 20,000 per year. This suggests that China's labour market recognizes qualifications more equitably and that higher education is an effective pathway to break through social stratification.

References:

[1] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55.

[2] Fang C, Feng X. Social origins and educational attainment: A historical analysis based on 70 birth cohorts from CGSS datasets[J]. Sociological Studies, 2018, 33(2): 140-63.

[3] Xuemin Z, Zhenhua W. Multidimensional measurement and comparison of gender equality in education[J]. Education & Economy, 2019, 1: 16-24.

[4] Wu X, Zheng B. Household registration, urban status attainment, and social stratification in China[J]. Research in Social Stratification and Mobility, 2018, 53: 40-49.

[5] Chunling L. Social and Political Changes and Inequality in Edu-cational Opportunities: On the Impact of FamilyBackground and Institutional Factors on Educa-tional Attainment (1940-2001)[J]. 2003.