Propensity Score Matching on Canadian Household Expenditure on Children's Education

Binqi Li, Id:1004871123

20 DEC 2020

Code and data supporting this analysis is available at Gitgub repo:

Abstract

Investment in children's education matters due to its relations to the development of the next generation,the quantity and the quality of children will have an impact on the future labor force. The household expenditure on children's education as a form of investment shows parent's awareness and evaluations of the importance of education. However, this awareness might be affected by the parents' education level, as parents who have higher education might have a higher incentive to invest in children's education. This paper will follow this assumption and further testify to this social-economic issue using the propensity score matching method.

Keywords

Household expenditure on children's education, Household education level, propensity matching score

Introduction

Macroeconomics has often defined education as the subsequent component in involving the nation's economic growth,expenditure on education is a form of investment in ensuring the quality of education,as the increasing return of knowledge would generate positive externality which is beneficial to both socially and privately. Under the social perspective, Human capital as the outcome of education investment would further be allocated to contribute to the nation's productivity. On the other hand, a high-skilled worker with a post-secondary education diploma will generally receive higher return of income. Investment in education is often regarded with two main factors: institution and individual contribution. Regardless of the education devotion made by institutional and social domains, assuming individuals are receiving similar institutional education resources across Canada, researching factors that may derive household expenditure on education as the portal to navigate whether family education background would change the awareness of the importance of education. This is where statistical investigation took place to analyze a potential correlation in the field of household expenditure on parents education level.As the correlation is referred from observational study by given survey data provided with randomization sampling, where rigid design of control bias is possibly avoided in turns of researching social issues(MarianiI, 2013)..

This investigation is inspired by "The more the merrier"(Black,2005), a paper studied the effect of family size and birth control on children's education, focusing household evaluation on kids' quantity and quality based on the allocation of their household income. The result of the children's education is insignificant to household composition as well as birth control within households.This conclusion has driven the attention of other factor that may potentially correlated to the spending on education.According to the concept based on the family economics, family educations determines the income of the family, parents who received higher

education will have supported knowledge base in convincing and expecting children in attending higher education institution.

The perceived concept of household income level has determined the level of expenditure on normal and inferior goods. Education,similarly, can be taken as consumption of Education services and investment on human capital. This study will take account of this factor but mainly evaluate the education spending proportion with respect to household income,focus on household recognition on the importance of investing in children's education. Considering the weighting of household expenditure using opportunity cost,the outcome of spending income on children's education has the opportunity cost of the amount of expense consumed in other sources.

Further concluding household's expectation on children's future development, and whether it is derived by the household education background. As spending income on children's education has the opportunity cost of the amount of expense being consumed in other sources. Therefore, the outcome is a household willingness to give up other consumption choices for children's education.

Following the aim of this study, Propensity score matching will be used to target whether parents' education background is an effective treatment for expenditure on children's education. It does not only reflect parents' evaluation on the importance of education but also conveys their expectation on children, as investment on any assets has unpreventable risks.This study will be using two datasets from canada Statistics on Canadian household spending 2009 to match household with the same number of children, province and similar age group to investigate potential causal link on household with higher education background as treatment group leads to a higher spending on children's education. Further conclusions will be depicted in based on the presentation of model and data analysis, flaws found through this entire procedure of developing the results will also be pointed in the weakness section.

Data

The dataset for this study is based on the Survey of Canadian household spending(SHS),the dataset is downloaded from the CHASS database website which is originally provided by Statistic Canada.The survey was recorded in the year 2009 indicating Canadian household fundamental expenditure in reflecting Canada's GDP and CPI for Economics investigation purpose, many questions were asked related to basic information regarding potential expenses made within each household, for example, the dwelling characteristics, information of house equipment, and personal information. The target population for this survey is to cover each household among 10 territories of Canada, however, exceptions were people in the military and residents of institutions which make up 2% of the population is excluded from the target population. The Survey of household spending conducted the total number of frame population where households were able to reach across Canada with the proper response of 17,792 sample population.

The dataset is directly provided by respondents, each respondent has accomplished the survey voluntarily through the online survey and telephone interview. However, many of the missing values were respondents who refused to complete the survey or those who failed to complete the survey due to technical difficulties. For those non-respondents, statistics Canada weights and analyzes characteristics regarding non-responders and allocated them to respondents following similar traits to minimize the bias. (Canada, 2008)

The survey of household spending (SHS) as a nationwide survey, having access to detailed and comprehensive information along with the support from the government to help present the societal situation. Therefore, the survey performed under government institutions conveyed a researchful source to future investigating of true living conditions and household's habits regarding the allocation of resources. It is a strong database to support this study in revealing economic behaviors with the information provided around Canada.

The treatment for this study is the average family education level, this variable is generated with the average of respondent education level rated from 1-10 with the corresponding degree of study and spouse education level following the same methodology. To make treatment binary, households having an education level of 1 is where they achieved a high-level education background with the average above 5 in the survey, which both response and spouse have finished post-secondary and higher education. Those who had a low level of

education are pointed to as 0, meaning a family having average education with only a secondary degree or lower.

The outcome for this study is the average spending on each child in the household, due to households raising different numbers of children, the best way to evaluate the average spend on each child is to understand the expenditure decisions made affect by the family educational background. So to get the average education expenditure on each child is to first add up the number of children to the age of 25 and using the total education expenditure divided the number of children to have the value with each education investment within the household.

Variable selection for multiple linear regression model with the initial purpose of avoiding multicollinearity(multicollinearity,2020), where only the correlation between the explanatory variable and response variable can be found other than correlation found between each predictor. So to prepare propensity score matching techniques, looking for the potential connection between household education level and household expenditure on each child, variables like age, province, the number of children are the basic personal information that considered to facilitated propensity score matching due to its non-correlate.

Household income is not considered in this study, because of its strong correlation with the treatment group, as a higher education level has a large possibility of receiving higher income, this direct causal link may affect the result of the regression model's accuracy, so it would not be considered in further investigation.

Model/Methodology

The logistic regression model is firstly utilized in this study to indicate the treatment of levels of family education background as the response variables with predictors. As a logistic model is used to calculator the propensity score of family education level and preparing next step for matching.In this case the level of family education background is the outcome of the interest based on the explanatory variables like age, number of kids and province predicting the proportion of household achieving a higher education level.The logistic linear regression model is:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{Agegroup} + \beta_2 x_{province} + \beta_3 x_{numberofchildren} + \epsilon$$

β_0 represents the probability of individuals in the 0 age group (age under 25), from a masked recorded province with number of children in the household receiving a high level of education. As β_1 represents estimated coefficient of age group.β_2, on the other hand, indicating the estimated coefficient of province while β_3 represents the estimated coefficient of number of children in the household.

Age groups and number of children are the numeric variables,age groups are recorded from the level of 0 to 14, covering ages up till 84 years old. The reason for choosing age group instead of age is due to people within the same age group will have a similar life experience, so that expenditure decisions were made between the factors of age would not have much difference. Provinces on the other hand also appeared as numeric variables as each province is labeled into a specific number, the outcome of

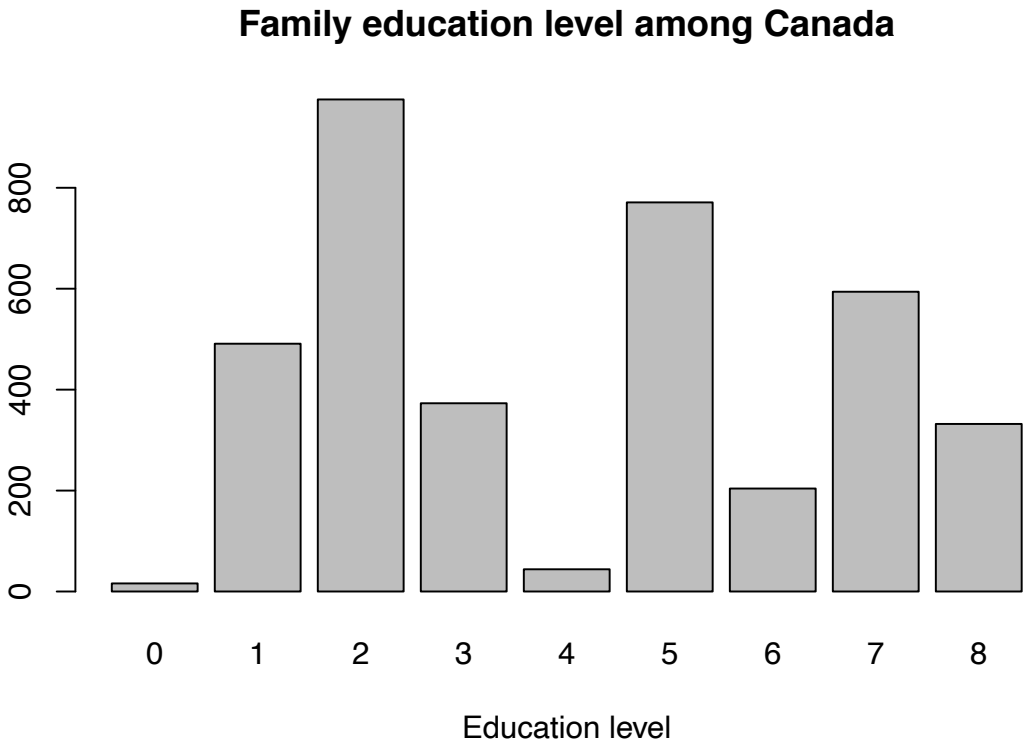
$$log(\frac{p}{1-p})$$

is the proportion of people receiving education level based on these predictors.

All the model is implemented in the r studio using r code.Furthermore,taking the propensity score of households having a higher education background, creating a column indicating propensity score of fitted value with each observation, the goal is to match observations which had similar propensity score with or without the treatment of high family education using arm package.After matching the data, the next step is to see how the level of family education would affect the amount of money spend on children's education.So the propensity score regression is constructed to see the correlation with each predictors including the treatment variable. Here is the propensity regression formula:

$$EducationExpentidure = \beta_0 + \beta_1 x_{familyeducation} + \beta_2 x_{Agegroup} + \beta_2 x_{province} + \beta_4 x_{numberofchildren} + \epsilon$$

Result



Based on the histogram drawn above with the number of cases in the y axis respected to the counts of education level in each range for x-axis. The plot indicates that the number of education levels of 2, 5 and 7 appeared to be the most where 5 is community college, 2 is secondary high school, 7 as Bachelor's degree.

Here is the summary table indicating the estimated coefficient and p-value of each predictors regarding the logistic model:

Variable	correlation coefficient	p-value
intercept	-0.1398	0.2398
province	-0.0028	0.1139
age group	0.0553	0.00149
number of Children	-0.0313	0.4219

The regression estimation of the logistic regression indicates that the beta1-province and number of children has coefficient correlation of -0.002754 and -0.03128, with the p-value is significant which has no evidence to reject the null hypothesis.The age group of respondents showed to have the correlation coefficient of 0.0565 and significant p-value in having strong evidence to reject null hypothesis. Meaning holding another predictor fixed, a unit of increase in age would result in 0.056294 addition to the

$$log(\frac{p}{1-p})$$

The overall result using a method of propensity matching score matching according to the output of the

	(1)
(Intercept)	1685.574 ***
	(347.830)
ed_level	1433.215 ***
	(184.972)
province	-2.361
	(4.940)
age	249.512 ***
	(42.128)
number_of_children	-671.359 ***
	(110.425)
N	3800
R2	0.037
logLik	-38243.628
AIC	76499.255

*** p < 0.001; ** p < 0.01; * p < 0.05.

multiple linear regression with estimated coefficient and p-value indicated a conflict result as the study initially proposed. The output table presented the predictors of education level, province,age and number of children with the likelihood of 1516.875,-6.323,273.152,-640.553 respectively where only the predictor of province shown to be significant in this case in failing to reject the null hypothesis, would be excluded in the final model.The contradiction with the investigation propose has suggested potential issue and problem in the study.

Disscussion

This study has indicated whether the family education level would affect the amount of household expenditure on children's education using propensity score of matching. The pairs were matched with similar straits of province, age group and number of children, the outcome of the propensity score matching illustrate how effective the treatment education background would affect household spending on education.The models were well presented in the data section.

The propensity score matching method used in the study to investigate whether the treatment would have the effect on the outcome of the average spending in each household. It is resulted with a significant p-value of proving the proposal for this study. However, other predators are shown to be significant in this case, showing correlation with the outcome, this indicated that the treatment is not effective for treating the outcome of the study, in other words, there is no direct and it is not the only causal link of family education level would affect the amount of investment spend on education for next generation, the family education background may combine other factors together have correlation on the outcome. Therefore, the assumption is violated with only the treatment that can influence the outcome of household expenditure on education.

So it is necessary to reconsider the factors that influenced the education expenditure within household, the family education is not the only factor that contribute to the outcome of the amount of spending, it has to do with the encounter and experience happened in each family, it may also depends the race of family background, the cultural influence on each family in leading them to make certain spending decisions.

weakness & next step

The result of this study may suggest many potential errors in the data cleaning process, the data is not prepared for the further investigation. So that the bias and inaccuracy of the dataset is generated to effect the outcome of the model. For example,the elimination of the single parent household due to some issues happening in code. With the minimization of the sample data, it may affect the overall outcome with less accuracy in terms of representing the target population.

The paper "The more the merrier"(Black,2005) investigated the education expense affected by the composition of the family, the survey has selected samples with family experienced a born of twins, as the unexpected treatment to the family and investigated the changes made to the expenditure due to the unexpected shock. Following the same method, this study should consider looking for a database that shows that treatment can be an unexpected shock of parents' education level, avoiding the cases when parents are saving and perparing their kids for education. With the effect of the unexpected treatment, a direct causal link will be more convictive in terms of social investigation.

To investigate the specifically the household education background have impact on children's education expenditure generates non constant variance residuals due to correlation between variables, residuals happened the case as both parents have raised a kid but still rely on kid's grandparents for living, so that the expenditure on kids will not rely on parents education level.One other potential residuals is that the amount of education fees is spend on them not for the children in the family.

The dataset chosen for this study more than 12 years ago, the outcome of social and family economic behavior might change throughout the time, it would not apply to the real situation nowadays, it is better to have a survey dataset that is in the recent decade to investigate the family choice of consumption.

The future step for this study is to have the survey that adapted the contemporary social trend and comparing the family economics behaviors with 2009, discovery the changes to analyze the changes of outcome effect by the throughout the time. This study can be utilized as a useful reference to make comparison with future studies.

Reference

(1)MarianiI, A. W. (2013, November 13). Observational studies: Why are they so important? Retrieved December 09, 2020, from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-31802014000100001#:~:text=Other%20advantages%20of%20observational%20studies,are%20easily%20and%20quickly%20performed.

(2)Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. "The more the merrier?The effect of family size and birth order on children's education." The QuarterlyJournal of Economics 120, no. 2 (2005): 669-700.

(3)Alexander, R. (2020, November 11). Running Through a Propensity Score Matching Example. Retrieved December 19, 2020, from https://q.utoronto.ca/courses/184060/files/10649324?module_item_id=1998838

(4)Multicollinearity. (2020, June 17). Retrieved December 22, 2020, from https://www.statisticssolutions.com/multicollinearity/

(5)Canada, S. (n.d.). Survey of Household Spending. Retrieved December 22, 2020, from https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey