**Attaining Higher Degree Will Bring You Higher Income?**

Canberk Kandemir

54411160

**Introduction**

This paper aims to explore the possible relationship between the level of education and income, specifically what economist may call *returns to schooling*. This delineates that schooling is an investment in human capital, with a monetary payoff similar to that of a financial investment(Angrist & Pischke, 2015). Linear regression reduces these possible multiple comparisons to a single coefficient that captures the weighted effect of one additional year of education on incomes. In theory, holding other factors constant, schooling has a positive effect on the wage of an individual, since higher education is associated with a higher amount of human capital, and thus it yields higher productivity as well as higher income. We will explore other variables that can describe the ceteris paribus effects of schooling on wages. Those variables that have collinearity with schooling could be associated with higher income, such as fulltime workweeks ethnicity, education level, gender, and immigration status. For a complete study of the ceteris paribus of schooling on wages, some of these were used in multiple linear regression. Unravelling the impact of education on income has significance in economic policies, specifically for governments spending on education. If we conclude higher levels of education can be associated with a higher level of income for individuals, then increasing funding for education in the long-term can be lucrative for total capital gain. Such assumptions are compatible with macroeconomic theories relating to human capital and how it yields higher productivity per worker. However, there are certain limitations like the quality of education which cannot be easily obtained from the data.

**Literature Review**

One of the research we reviewed was "Effect of Education on Wage Earning". The paper used

data of Georgia residents selected from a survey regarding income, education, and other factors.

The study found that an additional year of education increases individual income by almost

$8000 per year. They obtained estimates of the multiple linear regressions coefficient of

education, experience, sex, marital status, and type of employment to measure the impact of

educational level on wage income of individuals. With an individual's income acting as the

outcome variable, there were other factors they manipulated such as years of education, years of

experience, age, weeks worked in the past year, language, and disability. Furthermore, they

provided dummy variables for marital status, gender. In both the cases of married and unmarried

women, they tend to have a lower income than both married and unmarried men; however,

whether this is due to sexism or sector they work, cannot be highlighted through regression.


The final paper evaluated was"Does the Attained Level of Education Affect the Income Situation

of Households?" The research used the data Czech people from a survey conducted between

2005-2009. The level of education was defined by the level obtained by the household member

with the highest salary. These levels were defined as no education, primary education,

secondary education, and higher education. They found that the most vulnerable group to

poverty consists of individuals with primary or no education. Surprisingly, the study concludes

that a higher level of schooling does not guarantee a lower risk of poverty. This is due to a

person with the highest level of education had the lowest disposable income which can be unique

to the Czech Republic (Turčínková&Stávková, 2015). The study concludes that a higher level of

education generally provides a better situation in terms of risk of poverty. Unless they lose their job or a stable source of income. In this case, they are more affected than those with a lower level of education, due to higher spending habits as well as the difficulty in adjusting to a new job.

**Data**

In order to execute this analysis, multiple variables that we believed would affect a person's income were included. With a person's income acting as the dependent variable, there are many other factors which could possibly impact employment income, thus not all of them were included.  In order to capture the fundamental causes of high income, we decided to explore the foundational requirements of employment. Some of the more significant variables we analyzed were the level of education each individual received and whether they work fulltime. Since the data on education was categorical by degree and qualification, some liberties were executed and the degrees higher than or equal to bachelor level were converted into a dummy variable 1, and 0 for anything less than the bachelor level education. We also defined 1 for people who work fulltime and 0 for otherwise. The predicted variable is transformed based on certain specifications to create a dummy variable. People who have a higher income than the average, $55,806.40, received 1 and people who do not have a higher income received 0.  We are testing whether obtaining a high-level degree can be associated with high income as well as the amount of fulltime working individuals. Therefore regression models are employed to predict the outcome. Our independent variables are endogenous, thus they are prone to change within the model.

**What is the source of data?**

We analyzed 427,805 individuals who are both immigrants and non-immigrants from Canada in the ages between 20 and 59. All data for both independent and dependent variables were taken from the 2016 Census Public Use Microdata File(PUMF). The PUMF data is a survey conducted by Statistics Canada that takes data from households across the provinces and looks at various variables including age, weight, marital status, employment status, etc.

The quantitative independent variables that are selected as regressors for the model are education and fulltime work weeks. The table below presents the summary statistics for the selected variables and the histograms show the distribution of income based on the degree received.

| Variable | Mean | Std.Dev. | Min | Max |
|---|---|---|---|---|
| Employment Income | 53103.32 | 63848.87 | -50000 | 1200829 |
| Levels of Education | 6.623478 | 9.936491 | 0 | 88 |
| Full-time or part-time weeks worked in 2015 | 1.179722 | .3839562 | 1 | 2 |

for labels, *See Appendix.*

The linear parameter assumption is met through employing STATA as the analytical tool, and the random sample assumption is met by the methodology of Statistics Canada. No perfect collinearity assumption can be validated by regressing the quantitative independent variables

with each other. $R^2$ value is smaller than .05 (.0001 for the level of education and the fulltime work weeks).

**Results**

To run the regression model we applied the specification we mentioned earlier to create dummy variables.

Simple Regression

*Income=.365426 +.2020771\*Level of Education*

| Source | SS | df | MS | | Number of obs | = | 427,805 |
|---|---|---|---|---|---|---|---|
| | | | | | F(1, 427803) | = | 16148.92 |
| Model | 3815.31604 | 1 | 3815.31604 | | Prob > F | = | 0.0000 |
| Residual | 101072.023 | 427,803 | .236258331 | | R-squared | = | 0.0364 |
| | | | | | Adj R-squared | = | 0.0364 |
| Total | 104887.339 | 427,804 | .245176153 | | Root MSE | = | .48606 |

| HINC | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| highdegree | .2020771 | .0015902 | 127.08 | 0.000 | .1989604 | .2051938 |
| _cons | .365426 | .0009027 | 404.83 | 0.000 | .3636568 | .3671952 |

Multiple Regression

*Income= -.105218 + -.4185485 \*Level of Education + .1430733\* Weeks Worked + .9708082*

*\*Weeks_Education*

| Source | SS | df | MS | | Number of obs | = | 427,805 |
|---|---|---|---|---|---|---|---|
| | | | | | F(3, 427801) | > | 99999.00 |
| Model | 72283.9841 | 3 | 24094.6614 | | Prob > F | = | 0.0000 |
| Residual | 32603.3547 | 427,801 | .076211497 | | R-squared | = | 0.6892 |
| | | | | | Adj R-squared | = | 0.6892 |
| Total | 104887.339 | 427,804 | .245176153 | | Root MSE | = | .27606 |

| HINC | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| highdegree | -.4185485 | .0011332 | -369.35 | 0.000 | -.4207695 | -.4163274 |
| wwks | .1430733 | .0011254 | 127.13 | 0.000 | .1408676 | .1452791 |
| wwksdgree | .9708082 | .0010891 | 891.36 | 0.000 | .9686735 | .9729429 |
| _cons | -.105218 | .0010484 | -100.36 | 0.000 | -.1072728 | -.1031631 |

Analysis

The simple regression model shows a positive coefficient which means having a high level of

education will result in an increase in income by the education coefficient. From the $R^2$ value of

the simple regression, we see that the level of high education alone can explain only %3 of

high-income variation in the sample and this is due to exogenous variables.

On the other hand, Multiple regression can explain %68 of high-income variation. By

comparing both models we see that the coefficient of education becomes negative. Specifically

having a high level of education result in a decrease in income. However, *Weeks Worked*

variable has a positive coefficient which means switching to fulltime can result in an increase in the income. Furthermore, our interaction term, *Weeks_Education,* is our estimator which is also positive and thus result in an increase in income if an individual achieves a high level of education and works fulltime: $\beta_3 > 0$

Statistical Inference

The high t-score in the simple regression tells us the model is statistically significant in determining high income. On the other hand, in the multiple regression model, low t-score for the level of education indicate lower significance for the outcome. In Stata term, regression model tells us that the level of education has a relatively negligible effect on the dependent in the multiple regression. However, the full-time workweek is a significant indicator of income, and the interaction term(*Weeks_Education)* statistically solidifies the effect of education due to its high t value.

Extensions

    Instruments

    We can include instruments to our model to see whether there will be a significant effect of majors of study on education for determining high income. To see whether the coefficient of education would be negative in LeastSquares Regression model, we ran our regression model.

```
Instrumental variables (2SLS) regression        Number of obs   =     427,805
                                                 Wald chi2(2)    =    46139.82
                                                 Prob > chi2     =      0.0000
                                                 R-squared       =      0.1110
                                                 Root MSE        =     .46686
```

| HINC | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| highdegree | .2216354 | .002496 | 88.80 | 0.000 | .2167434 | .2265274 |
| wwks | .3531013 | .0018628 | 189.56 | 0.000 | .3494503 | .3567523 |
| _cons | .0694825 | .0018262 | 38.05 | 0.000 | .0659031 | .0730618 |

```
Instrumented:  highdegree
Instruments:   wwks 2.CIP2011 3.CIP2011 4.CIP2011 5.CIP2011 6.CIP2011
               7.CIP2011 8.CIP2011 9.CIP2011 10.CIP2011 11.CIP2011
               13.CIP2011 88.CIP2011
```

The data only explains %10 high-income variation in the sample. According to this

model, if an individual achieves a high level of education it will result in high income due

to positive coefficient of education.

Conclusion

Overall the results are consistent with the reviewed literature and they show the significance of a

high level of education on high income. We can see how only one extra year of education can

potentially increase a person's income highlights the significance of earning a good education.

However, as one would expect a person's education and weeks worked fulltime per week also

play a role in income earnings. This help in confirming common assumptions that are present

today. Although it is not surprising to see a positive correlation between weeks worked fulltime

and high income, the interaction effect highlights its effect with combined with education which

is statistically significant due to high t-score. However, in the multiple regression model, the effect of education remains as a negligible indicator of income by itself. One of the limitations in determining the high income was quality of education which was absent from the sample. To reduce biases in the model:

1. B1- wide age group( 20-59) no specification of ethnicity.
2. B2-Our control group is those who achieved less than bachelor level which makes 67.78 of population
3. B3- The treatment does not say anything about the quality of education.

The following can be used to reduce each section of the bias is as shown below:

1. B1: Specification can be done for the population
2. B2: Reduction of the population can be done.
3. B3: The grades can be considered to determine the quality of the treatment

Future Research

There are various exogenous variables we excluded in this research, which can be correlated with income such as immigration status, major of study, gender, etc. These can be included in future research to highlight the milestones of obtaining high income. To do that one can test for instrument relevance and overidentification. Subsequently, the relevant instrument can be employed such as gender, immigration status, etc. With the inclusion of other variables: gender, immigration status marital status as dummy variables. We can make explore the interactions of these variables and select the best estimators. This is will also serve to clarify gender bias related assumptions.

References

Angrist, J. D., & Pischke, J. S. (2015). *Mastering Metrics: the path from cause to effect*. Princeton: Princeton University Press.

Statistics Canada. (2018, April 24). Public Use Microdata File (PUMF) Collection. Retrieved from https://www150.statcan.gc.ca/n1/pub/11-625-x/11-625-x2010000-eng.htm.

Talley, Q., Wang, T., & Zaski, G. (n.d.). Effect of Education on Wage Earning.

Turčínková, J., & Stávková, J. (2012, November 10). Does the Attained Level of Education Affect the Income Situation of Households? Retrieved from https://www.sciencedirect.com/science/article/pii/S1877042812040578.

**Appendix**

Label define                                                                      *fig.1*

| Code | Description |
|---|---|
| No degree | No certificate, diploma or degree |
| Secondary | Secondary (high) school diploma or equivalency certificate |
| Trades | Trades certificate or diploma other than Certificate of Apprenticeship or Certificate of Qualification |
| Quali | Certificate of Apprenticeship or Certificate of Qualification |
| less | Program of 3 months to less than 1 year (College, CEGEP and other non-university certificates or diplomas) |
| 1 to 2 | Program of 1 to 2 years (College, CEGEP and other non-university certificates or diplomas) |
| More than 2 | Program of more than 2 years (College, CEGEP and other non-university certificates or diplomas) |
| certificate | University certificate or diploma below bachelor level |
| Bache | Bachelor's degree |
| Above bach | University certificate or diploma above bachelor level |
| med | Degree in medicine, dentistry, veterinary medicine or optometry |
| Master | Master's degree |
| doctorate | Earned doctorate |

| Not available | Not available |
|---|---|
| Not applicable | Not applicable |