Estimating the Financial Return to Education Between Fields of Study

Jeffrey McRae
Department of Management Economics and Finance, Business and Economics, University of Guelph, Guelph, ON Canada. Faculty supervisor: Louise Grogan. For correspondence, please email: jeffreymc@bell.net

Abstract
The Mincer regression equation was utilized to compute the expected financial return to education from additional years of education across the 2016 Canadian population. Data was taken from the 2016 Canadian Census of Population to create the populations of interest. Three subpopulations were then derived from the collected data to represent Canadians with different major areas of study, namely business, humanities, and engineering. Mincerian regressions were run using these subsections to determine how the financial return to education differs between distinct majors. Additional multiple regressions included an interaction term between sex and years of schooling in an attempt to determine whether an individual’s sex affects their expected return to education given a specific area of study. The regression results indicate that business majors boast the largest average expected return to education, while engineering majors boast the lowest. Subsequently, in relation to business majors, sex is not found to have an impact on expected financial returns. Future research may add to the findings in this paper by expanding the scope to include all areas of study in addition to deciphering whether the expected return to education for a given major is consistent throughout all major Canadian universities.

Keywords: Canada, Mincer earnings function, return to education, expected salary

Introduction
The Mincer earnings function is arguably the most widely used equation in empirical economic work. It is a multi-variable linear regression that may be used to help explain and estimate an individual’s salary as a function of their education and labour market experience (Parinos, 2016). This creates a useful tool for individuals and policy makers who wish to analyze the average financial return of additional investments in education and compare how this return varies throughout different demographics. Much of the preceding literature concludes there is a positive correlation between education and lifetime earnings (Webber, 2014). This conclusion aligns with previous research done using Mincerian regressions, which establishes an expected percentage increase in salary for each additional year of schooling an individual undergoes. This is also known as the return to education. This estimated return has been observed to change across different countries and sample populations (Montenegro & Patrinos, 2014), but will the expected return differ when comparing major areas of study within a single country?

The following research paper seeks to answer this question as it relates to the Canadian population—specifically, how the estimated average return to education differs between the distinct majors of business, engineering, and the humanities. In conjunction with this question, a subsequent analysis will gauge whether an individual’s gender impacts this expected return. It is hypothesized that areas of study that are traditionally understood to be academically rigorous and highly specialized, such as engineering, will boast a high-mean value for years of education across its population. This notion derives from the idea that to break into a highly specialized or technical field, individuals must undergo many years of education and training. It is assumed that those who enrol in extensively demanding programs intend to work in the field in which they are studying. This should result in many of the participants of these fields of study completing equal or similar years of schooling. With these presumptions in mind, it is reasonable to expect the assumed homogeneity of education to result in years of schooling being a poor predictor of income for technical fields. To clarify why this may be the case, imagine a population of engineers who graduate with the same degree; the perceived value of each individual (the determinate of how much they get paid) will not be dependent on education because it is not acting as a differentiating variable. Rather, variances in pay will be increasingly weighted on other variables, such as experience or factors not accounted for within the Mincerian equation. Logically following this...
Estimating the Financial Return to Education Between Fields of Study (McRae)

hypothesis, it is expected that the coefficient will likely be quite imprecisely measured, due to little variation in X. Therefore, years of schooling regressed on income will boast a smaller coefficient for the engineering subsection compared to business or the humanities. Additionally, contingent on how significant experience is at predicting income, the engineering major is expected to have a smaller R² value, indicating the model accounts for a small percentage of the variation in wage.

To test this hypothesis and analyze the return to education across majors of study, data will be taken from the 2016 Canadian Census of Population. This is the most recent census, which ensures our results are the most relevant to Canada’s current population and labour market. The entire data set contains information on 930,421 citizens and represents 2.7% of the population. This was the largest publicly available data set which contained the variables needed to carry out the necessary regressions. This census seeks to represent the entirety of the Canadian population residing within private dwellings throughout all provinces and territories. Non-permanent residents such as refugee claimants, holders of work or study permits, and members of their families living with them are also included in the survey. The response rate for the 2016 survey was 96.9% (up significantly from 2011, where the response rate was 68.6%). While all representative samples are subject to both sampling and non-sampling errors, the methodology employed by the government of Canada works to reduce the size of this error to an acceptable level (Information outlining the methodology used can be found under chapter 3 of the 2016 Census Individuals PUMF User Guide, beginning on page 126). The size and accuracy of this sample provides strong reason to believe it will act as a suitable data set for answering our research question.

Literature Review

Utilizing a Mincerian regression to estimate the percentage increase in salary as a result of additional years of schooling is hardly a groundbreaking concept. Many previous papers have confirmed the validity of this regression technique and answered a wide variety of salary-based questions.

Douglas Webber dissects the return to education in his 2014 paper “Is the Return to Education the same for everyone?” Webber retains that education is the most certain path to financial security, but he argues the return to education is not equally realized throughout all disciplines. Webber cites wage data which illustrates that the most technical fields, such as engineering and finance, produce significantly higher average earnings compared to less technical fields, such as the humanities. The higher salaries for individuals who complete degrees in technical fields result in larger premiums being realized per year of schooling. This juxtaposition of areas of study does not however alter this paper’s stated hypothesis. One may intuitively assume that the higher average salaries and increased years of schooling for technical fields will cause education to act as a strong predictor for income, and that this is true over the entire population. However, the research conducted for this paper divides the data into subsections that are based on area of study. As mentioned in the introduction, this is expected to lead to a low variance in education among engineers, resulting in years of schooling insignificantly predicting income.

The 2014 research paper authored by Claudio E. Montenegro and Harry Anthony Patrinos “Comparable Estimates of Returns to Schooling Around the World” uses an almost identical Mincer equation to that of this paper. The authors used their equation to estimate the return of education across 139 separate economies. Data was retrieved with the goal of compiling a harmonized sample that could be used in computing comparable results between all countries. Eight hundred and nineteen household surveys were used spanning 1980-2013. Running a Mincerian regression on the entire data set resulted in an estimated average return to education of 10.1% per year, with males averaging 9.6% and females averaging 11.7%. It could be expected that a similar result will be found throughout subsections of the Canadian data used for this study.

Afghanistan produced the lowest return to education at 1.6% for their entire population, while Rwanda was observed to produce the highest estimated return at 22.4%. Reviewing three years (1981, 1991, and 2001) of Canadian data, the reports correspond to returns of 9.2%, 11.2%, and 12.1%, respectively. This increasing trend provides reason to believe the 2016 census may result in a return to education higher than 12.1%.

Explanantion of Data

The 2016 Canadian Census of population originally included 930,421 observations with 123 different variables. To analyze the difference in estimated predicted earnings across separate areas of study, only four variables were needed from this data set. These are: CIP2011, whose values correspond to major areas of study; AGE; HDGREE, which signifies the individual’s highest degree earned; and TOTINC, which is the reported amount of total income before tax. Analyzing the variance between males and females with regard to expected earnings within a single area of study required an additional SEX variable taken from the 2016 Census data. A recode was necessary to transform select variables into values usable within the Mincerian regression model. The AGE variable originally assigned a coded number between 1 and 21 to each observation. Each code corresponded to an age range, the smallest being 0-1 years of age and the highest being 85 years and older. These codes were transformed by taking the mean of the age range and inputting it as a single value. For example, code 8, which corresponded to the age range 20-24 years old, was recoded to the value 22. HDGREE originally took on values of one to 13 based on the highest degree earned by the individual. This was recoded to YRSCHOOL and signified the number of years required to complete each degree, diploma, or certificate. Using the newly formed age and years of schooling variables,
an experience variable was generated. This was conducted by subtracting years of schooling from age and then subtracting an additional six years for the period before school begins. The generated experience variable assumes all individuals immediately begin work after school and never undergo a hiatus. In actuality, this is unrealistic, but for the sake of analysis this will be the best estimate of labour force experience.

After the variables were properly generated, the set of data was filtered to only include viable observations. This resulted in observations being dropped if they had missing values due to non-responses or errors in entering the data. The strategic decision was then made to only include observations for individuals between the ages of 21 and 60. The lower age limit was implemented to ensure all individuals were old enough to have had the chance to complete a four-year degree. The upper limit was implemented to ensure data was not skewed by retirees who might report negative income, while having high levels of experience. The age band was also felt to generate a more representative picture of the current full-time labour force.

After the filters were properly applied, the data set was reduced to 474,323 observations, which made up our complete sample population, representing Canadian citizens aged 21 to 60 years old who reported positive pre-tax income in 2015. The summary statistics of this sample are provided in Figure 1. The average number of years of schooling across this population was observed to be 13.8 years, with a standard deviation of 2.35. This tells us that 68.27% of all individuals present in the sample have completed between 11.45 and 16.15 years of schooling.

When the data sample is divided into subsets corresponding to the three majors of interest, varying summary statistics are observed (See Figures 2, 3, & 4). These results communicate a large difference between the average salary of engineering majors ($73,094.93) and those who studied the humanities ($48,873.81). This was to be expected after the review of previous literature. It is also worth noting that the average years of schooling are the highest among humanities majors (15.67) and lowest among engineering majors (14.13).

This finding partially contradicts our hypothesis by rejecting the idea that engineers need high-level degrees to enter the industry. In an attempt to resolve this contradiction, it is hypothesized that the completion of a technical bachelor’s degree or comparable training sufficiently prepares individuals for high paying jobs throughout the work force, negating the need to pursue higher level degrees. This alteration to the hypothesis additionally helps justify the notion that the level of education is a poor indicator of salary increases within the population of engineering majors.

Regression Results

The regressions run for this paper took the form:

\[
\ln(TOTINC) = B_0 + B_1 \text{YRSCHOOL} + B_2 \text{EXP} + B_3 \text{EXP}^2 + B_4 \text{iSEXRC} + U_i
\]

Total income was inputted as our dependent variable to analyze how an individual’s earnings changed as a result of variations in the independent (explanatory) variables. The decision was made to measure the change in the log of income rather than the dollar amount, so the co-efficient would be expressed as a percentage change.

The first independent variable, \(B_0\), represents the intercept value of income when the explanatory variables equal zero. \(B_1\) represents the percentage change in income as a result of each additional year of schooling. This is the most important variable concerning the analysis of return on education. Two terms were inputted to account for the variation in income as a result of experience: \(\text{EXP}\) and \(\text{EXP}^2\). The rationale behind including a squared experience term was to account for the decreasing function of return to experience. Put simply, it is expected that each additional year of experience does not account for as large an increase in income as the previous year. The fifth variable on the right side of the equation, \(B_4 \text{iSEXRC}\), is a dummy variable used to account for the variance in pay between males and females. Lastly, \(U_i\) is a catch-all variable, which represents all the factors that influence salary but are not explicitly inputted into our model. All regression formulas used account for heteroskedasticity in the results by using the robust standard error function of our statistics program.

Running this regression model on our entire population resulted in the findings in Figure 5. The first number to take note of is the \(R^2\) value of 0.1383. This number signifies that the entire model accounts for 13.8% of the variation in the dependent variable. The percentage increase on salary, as a result of an increase in years of schooling, was calculated at 10.2%. This is almost identical to the 10.1% return to education found among all wealthy countries (Montenegro & Patrinos, 2014). The \(T\) scores of each of the variables also provide encouraging results, as their values are all greater than 1.96, implying our coefficient results are statistically significant and the null hypothesis may be rejected.

The next step in our research is to run an identical regression on the three sub populations (See Figures 6, 7, & 8). The most prominent and revealing comparison that can be made between these generated statistics is the difference between the expected return to education for business majors (7.07%) and engineering majors (3.81%). The expected return to education for the humanities (4.68%) is also relatively low. This finding yields the conclusion that students studying business or management should more readily pursue additional degrees than their peers in humanities or engineering. However, this conclusion assumes that the goal of each individual is to increase their level of income relative to others within their field of study, which in reality is not always the case.

The comparison of \(R^2\) values also returns an interesting conclusion. The model accounts for much less of the variation in wage for engineers (6.53%) compared to that of business students (10.43%) and humanities majors (12.17%). These results align with the hypothesis that income is not as readily predicted by the Mincerian regression for engineers.
The last note to be made on the results of these regressions pertains to the similar YRSCHOOL co-efficient for engineering and the humanities. While these results imply a similar expected percentage increase in income per year of schooling for both majors, it is hypothesized this finding has occurred for different reasons. As shown in the summary statistics, engineering majors have much higher incomes and lower years of schooling, on average, than those studying the humanities. This finding indicates that humanities majors are on average undergoing more years of schooling; however, this additional education is not translating to increases in average salary. One explanation for this result may be a lack of jobs specific to graduates of humanities. This would result in large numbers of graduates taking jobs that do not require their degree qualifications. For these cases, the level of education may not be a significant indicator of expected job performance. Effectively, other factors such as experience and on-job training would be expected to play a larger role in determining income, which is exactly what we see from our regression on the humanities sub sample.

Focusing on the population of business majors, an additional regression was run to measure if the return to schooling differs between males and females. Previous literature indicates that around the world, on average, women receive a larger return to education than their male counterparts (Montenegro & Patrinos, 2014). Seeing as business majors enjoy the largest expected return to education of the three subgroups, it is fitting to analyze how this return differs between genders. A Mincerian regression was run using the original formula with the addition of an interaction term between the dummy variable SEXRC and YRSCHOOL. The results are found in Figure 9.

The co-efficient of our interaction term (ISEXXYRSCH_1) was calculated at -.0005428. This is interpreted to mean male business majors earn a negative return of 0.05% per year of schooling when compared to women; however, the T score for this variable indicates there is no difference in returns to schooling between male and female business majors. Therefore, a conflicting result has been found between the population of Canadian business students and the world average.

Conclusion

The purpose of this paper was to answer the following questions: How does the estimated expected return to education differ between groupings of individuals within a major of study? And subsequently, are returns to education for business majors affected by an individual’s gender? Data was taken from the 2016 Canadian Census of Population to create the sample populations of interest. Mincerian multilinear regressions were utilized to produce the statistics necessary to answer the stated research questions.

Results yielded that Canada, as a whole, is on par with the return to education present in other countries with comparable economies. Breaking up the data into subsections of majors—specifically humanities, engineering, and business—resulted in the following findings:

Business majors benefit from the largest return to education of our three groups. Humanities majors, on average, undergo the most years of schooling but have the lowest level of income. The variables present in the Mincerian regression account for the lowest amount of variation in income for engineers. And finally, gender of individuals does not significantly impact the return to education of business majors. These concluding results confirm the stated hypothesis: education is not a strong predictor of income for engineering majors when compared to a population of those within the same area of study.

Additional research should increase the scope of this paper by preforming similar estimations for all major areas of study and comparing the returns to education. It would be valuable for upcoming Canadian students to have a better understanding of how accurately they can estimate their future income based on area of study and desired degree. Follow-up research should additionally analyze which Canadian universities and colleges boast the strongest predictors of increases in salary within groupings of majors.

Acknowledgements

I would like to acknowledge Lucia Costnzo for her assistance in retrieving the necessary data for conducting the research present in this paper. I would also like to acknowledge Professor Louise Grogan for her continued support and feedback throughout the process of writing this paper.

References

Montenegro, C., Anthony Patrinos, H. (2014). Comparable Estimates of Returns to Schooling Around the World. World Bank and Department of Economics, University of Chile. WPS7020.

Patrinos, H. (2016). Estimating the return to schooling using the Mincer equation. IZA World of Labor 2016, 278 doi: 10.15185/izawol.278

Webber, D. (2014) Is the return to education the same for everybody?. IZA World of Labor 2014, 92 doi: 10.15185/izawol.92
### Tables and Figures

| Variable       | Obs    | Mean       | Std. Dev. | Min | Max   |
|----------------|--------|------------|-----------|-----|-------|
| YRSCHOOL       | 474,323| 13.81312   | 2.353309  | 0   | 19    |
| SEXRC          | 474,323| 0.4956137  | 0.4999813 | 0   | 1     |
| EXP            | 474,323| 20.32923   | 11.75968  | 0   | 51    |
| TOTINC         | 474,323| 54167.28   | 74803.74  | 1000| 1586814|

Figure 1. Summary statistics for the entire population

```
. summarize YRSCHOOL SEXRC EXP TOTINC if(CIP2011==3)
```

| Variable       | Obs    | Mean       | Std. Dev. | Min | Max   |
|----------------|--------|------------|-----------|-----|-------|
| YRSCHOOL       | 15,242 | 15.6692    | 2.077087  | 0   | 19    |
| SEXRC          | 15,242 | 0.3696365  | 0.4827221 | 0   | 1     |
| EXP            | 15,242 | 17.5559    | 11.16834  | 0   | 51    |
| TOTINC         | 15,242 | 48873.81   | 67330.32  | 1000| 1586814|

Figure 2. Summary statistics for humanities majors

```
. summarize YRSCHOOL SEXRC EXP TOTINC if(CIP2011==5)
```

| Variable       | Obs    | Mean       | Std. Dev. | Min | Max   |
|----------------|--------|------------|-----------|-----|-------|
| YRSCHOOL       | 66,140 | 15.03877   | 2.198026  | 0   | 19    |
| SEXRC          | 66,140 | 0.3731781  | 0.4836525 | 0   | 1     |
| EXP            | 66,140 | 20.10434   | 11.03769  | 0   | 51    |
| TOTINC         | 66,140 | 69095.11   | 105370.3  | 1000| 1586814|

Figure 3. Summary statistics for business and management majors

```
. summarize YRSCHOOL SEXRC EXP TOTINC if(CIP2011==8)
```

| Variable       | Obs    | Mean       | Std. Dev. | Min | Max   |
|----------------|--------|------------|-----------|-----|-------|
| YRSCHOOL       | 62,357 | 14.1308    | 2.201516  | 0   | 19    |
| SEXRC          | 62,357 | 0.904726   | 0.2935954 | 0   | 1     |
| EXP            | 62,357 | 21.21274   | 11.33123  | 0   | 51    |
| TOTINC         | 62,357 | 73094.93   | 86703.77  | 1000| 1586814|

Figure 4. Summary statistics for architecture, engineering, and related technologies majors
Figure 5. Regression results for the entire population

| lnTOTINC | Coef.   | Std. Err. | t       | P>|t|  | [95% Conf. Interval] |
|----------|---------|-----------|---------|-------|---------------------|
| YRSCHOOL | 0.101855 | 0.008151  | 124.96  | 0.000 | 0.100258 - 0.1034531 |
| EXP      | 0.0786066| 0.005391  | 145.82  | 0.000 | 0.07755 - 0.0796631  |
| EXPSQ    | -0.001428| 0.000132  | -108.54 | 0.000 | -0.0014538 - -0.0014022 |
| 1.SEXRC  | 0.3166398| 0.002903  | 109.07  | 0.000 | 0.3109501 - 0.3223295 |
| _cons    | 8.045126 | 0.0126017 | 638.41  | 0.000 | 8.020427 - 8.069825  |

Figure 6. Regression results for humanities majors

| lnTOTINC | Coef.   | Std. Err. | t       | P>|t|  | [95% Conf. Interval] |
|----------|---------|-----------|---------|-------|---------------------|
| YRSCHOOL | 0.0468354| 0.0042747 | 10.96   | 0.000 | 0.0384565 - 0.0552143 |
| EXP      | 0.0850745| 0.0027335 | 31.12   | 0.000 | 0.0797166 - 0.0904325 |
| EXPSQ    | -0.0016527| 0.000766  | -21.59  | 0.000 | -0.0018027 - -0.0015026 |
| 1.SEXRC  | 0.2308946| 0.01707   | 13.53   | 0.000 | 0.1974353 - 0.2643539 |
| _cons    | 8.705953 | 0.0682792 | 127.51  | 0.000 | 8.572117 - 8.839788  |
Estimating the Financial Return to Education Between Fields of Study (McRae)

### Linear regression

| lnTOTINC | Robust |       |       | [95% Conf. Interval] |
|----------|--------|-------|-------|----------------------|
|          | Coef.  | Std. Err. | t  | P>|t| |         |
| YRSCHOOL | .0707322 | .0023139 | 30.57 | 0.000 | .0661969 | .0752675 |
| EXP      | .0729666 | .0014799 | 49.30 | 0.000 | .0700659 | .0758673 |
| EXPSQ    | -.0013207 | .0000372 | -35.54 | 0.000 | -.0013935 | -.0012479 |
| 1.SEXRC  | .3274383 | .0085379 | 38.35 | 0.000 | .310704 | .3441725 |
| _cons    | 8.675138 | .0377537 | 229.78 | 0.000 | 8.60114 | 8.749135 |

**Figure 7.** Regression results for business and management majors

### Linear regression

| lnTOTINC | Robust |       |       | [95% Conf. Interval] |
|----------|--------|-------|-------|----------------------|
|          | Coef.  | Std. Err. | t  | P>|t| |         |
| YRSCHOOL | .0381226 | .0021188 | 17.99 | 0.000 | .0339698 | .0422754 |
| EXP      | .0665595 | .0014826 | 44.89 | 0.000 | .0636537 | .0694654 |
| EXPSQ    | -.0012962 | .0000348 | -37.21 | 0.000 | -.0013644 | -.0012279 |
| 1.SEXRC  | .363811 | .014923 | 24.38 | 0.000 | .3345619 | .39306 |
| _cons    | 9.283528 | .0363832 | 255.16 | 0.000 | 9.212217 | 9.354839 |

**Figure 8.** Regression results for architecture, engineering, and related technologies majors
Figure 9. Regression results for business and management majors including an interaction term between sex and years of schooling

| lnTOTINC       | Robust Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------------|--------------|-----------|-------|------|----------------------|
| EXP            | 0.0729693    | 0.0014798 | 49.31 | 0.00 | 0.0700689 - 0.0758697 |
| _ISEXRC_1      | 0.3356868    | 0.0725469 | 4.63  | 0.00 | 0.3194948 - 0.3518787 |
| YRSCHOOL       | 0.0709516    | 0.0029724 | 23.87 | 0.00 | 0.0651286 - 0.0767776 |
| _ISEXYRSCH_1   | -0.005428    | 0.0047466 | -0.11 | 0.90 | -0.0098461 - 0.0007604 |
| EXPSQ          | -0.0013207   | 0.000372  | -35.54| 0.00 | -0.0013935 - -0.0012478 |
| YRSCHOOL       | 8.671833     | 0.0469432 | 184.73| 0.00 | 8.579824 - 8.763841   |
| _cons          |              |           |       |      |                      |

Table 1. Summary of results

| Population                           | # of Observations | Mean YRSCHOOL | R²   | Coef  |
|--------------------------------------|-------------------|---------------|------|-------|
| Entire Population                    | 474,323           | 13.813        | .138 | .1019 |
| Humanities                           | 15,242            | 15.669        | .122 | .0468 |
| Business and management              | 66,140            | 15.039        | .104 | .0707 |
| Architecture, engineering, and related technologies | 62,357            | 14.131        | .065 | .0381 |