Simon Fraser University
Summer 2024

What accounts for the difference in average earnings of immigrant and native-born workers in Canada? Could it be Education levels?

Matthew Luong
301339822
Jane Friesen
August 9, 2024

1. Introduction

As a student pursuing the education path, I have an interest in understanding how education influences economic success. I know that education is widely recognized as a key driver of individual economic achievement and is important in shaping economic growth. In Canada, just like many other countries, the relation between education and earnings is a fundamental aspect of the labor market. This paper explores the returns of education, specifically examining how obtaining different levels of education impact individual earnings.

The main research question for this case for a diverse country like Canada, is understanding why earnings differ between immigrant and native-born workers. The main question for this research is 'What accounts for the difference in average earnings of immigrant and native-born workers in Canada? Could it be Education levels?'

While there are more immigrants immigrating to Canada, I thought it would be interesting to find out their earnings compared to native-born workers, especially once I graduate and enter the workforce. Despite having comparable levels of formal education, immigrants often face wage gaps relative to their native born counterparts. To investigate this, I used the 2021 Canadian Census to assess the role of educational attainment in explaining the raw wage gap. By using the human capital model, I can examine whether the difference in the highest level of education obtained affects the gap between immigrants and native-born individuals. While higher levels of education generally lead to increased earnings for both immigrant and native-born workers, the impact of education on earnings may differ between these groups. This paper will explore whether immigrants with similar educational credentials as native-born workers still experience lower earnings and, if so, how education levels contribute to this.

This topic really resonates with me as I approach the end of my studies at SFU. I'm excited to explore how education impacts earnings, especially in a diverse workforce like Canada's. This research aligns with what I'm passionate about and helps me understand more about how education influences economic success.

2. Data and Methodology

A. Empirical model:

$$ln(w_i) = \beta_0 + \beta_1 Immstat_i + \beta_2 Female_i + \beta_3 Female\_Immig_i + \beta_4 Secondary_i + \beta_5 Some\_Secondary_i + \beta_6 Bachelors_i + \beta_7 Masters_i + \beta_8 HigherEd_i + \varepsilon_i$$

Where:

- $ln(w_i)$ represents the natural log of annual earnings of individual i.

- $Immstat$ represent immigrant status and Immigrant was coded as 2 in the code book but I re-coded to Immstat ==1 if they are an immigrant, if not immigrant then Immstat == 0 otherwise for non-immigrant status.

-$Female_i$ indicator variable for gender whether the individual is female or not

- $Secondary_i$ indicator variable for an individual whose highest level of completed education is secondary education

- $SomeSecondary_i$ indicator variable for individuals whose highest level of completed education is some post secondary education

- $Bachelors_i$ indicator variable for individuals whose highest level of completed education is a bachelor's degree

- $Masters_i$ indicator variable for an individuals whose highest level of completed education is a master's degree

- $HigherEd_i$ indicator variable for individuals whose highest level of completed education is defined as higher education such as a doctoral degree.

- ϵ_i is the random error term.

B. Describe my estimation sample:

To describe my estimation sample, I implemented several data restrictions to ensure a clean and relevant dataset, focusing on full-time, working-age individuals with clear employment and educational status. I began by removing observations with unrealistic or invalid wage values, specifically dropping cases where wages were equal to 99999999, 88888888, and 1. But then as for my extension that I will later touch on, I have decided to also drop wages equal to 1000, 2000, 3000, 4000, and 5000 to see how many observations I would lose and to see how much of a difference it would make to my results. These extreme values likely represent any irrelevant data or outliers that could possibly skew the results.

I want to restrict my attention to full-time workers by doing so by filtering the data to include only those who worked 49-52 weeks, as indicated by `wkswork == 6` in the codebook, which corresponds to a full year of employment. This ensures that my sample reflects individuals with consistent employment, which is relevant for studies on labor market behavior. Additionally, I restricted the dataset to individuals who were currently employed which was called ‘class of work’ (`cow == 1`) and I have restricted mainly full-time (`fptwk == 1`), to capture those fully participating in the labor market.

I further restricted my dataset to include only individuals aged 18 to 64 by including the age group coded as ‘agegrp = 6 and 18’, focusing on the working-age population and excluding minors and retirees whose labor market behavior differs and may impact the accuracy of my results. In terms of immigrant status, I excluded non-permanent residents and unavailable statuses by dropping observations where `immstat` equal to 88 (not available) or 3 (non-permanent residents), to focus on relevant information.

For education levels, I excluded data for individuals where the highest degree (`hdgree`) was 88 (Not Applicable) or 99 (Not Available), as these categories do not provide meaningful information for analyzing educational attainment. I then created a new variable to categorize the highest level of education that an individual successfully completed. I grouped the education level by "secondary," "some_secondary," "bachelors," "masters," and "highered" (which includes doctorate degrees and other higher education qualifications).

These restrictions ensure that the dataset is focused on full-time, working-age, employed individuals with clear and relevant wage and educational information, providing a more accurate sample for analyzing data and the factors influencing wage disparities. By cleaning the data and applying these filters, the analysis can yield more meaningful insights into the behavior and outcomes for immigrants versus native born individuals.

3) Results

A) Summary Statistics

The summary statistics table provides a clear picture of the key variables in the 2021 census dataset after we applied our restrictions, resulting in a total of 228,400 observations. One notable aspect is the distribution of immigrant status: about 25% of the sample are immigrants. This gives us a solid base to study how being an immigrant compared to native born individuals.

The gender distribution is well-balanced, with 46% of the sample being female. This almost equal split between male and female workers allows us to thoroughly analyze gender wage gaps. Additionally, 11.6% of the sample are female immigrants, which lets us explore the combined effects of gender and immigrant status on the raw wage gap.

Education levels vary widely in our dataset. We found that 20% of the sample have completed secondary education, 35% have some post-secondary education, 25% hold a bachelor’s degree, 85% have attained a master’s degree, and 1.6% have higher education, including doctorates. These varied level of education enable us to analyze how education impacts earnings for immigrants and native born individuals.

Overall, the summary statistics highlight important characteristics of my dataset, such as the significant representation of immigrants, a balanced gender distribution, and a diverse range of education levels. These factors are essential for understanding the labor market and will help us conduct a meaningful analysis for my research.

B) Regression results

Model 1: Raw Wage Gap (Column 1)

The first model presents the raw wage gap, showing that immigrants earn approximately 3.4% less in logarithmic wages compared to non-immigrants, with a coefficient of -0.0342. This suggests a wage disadvantage for immigrants in this model.

Model 2: Interaction with Gender (Column 2)

The second model introduces an interaction term for gender and immigrant status. Here, the immigrant coefficient (-0.026) indicates that male immigrants earn 2.6% less than native-born males. The interaction term (female_immig) has a coefficient of -0.020, indicating that female immigrants earn an additional 2% less than female non-immigrants. This results in female immigrants earning a total of (-0.026 + (-0.020) = -0.046) or 4.6% less than female non-immigrants. The inclusion of the interaction term highlights that the wage gap is larger for female immigrants compared to male immigrants, suggesting a disadvantage for female immigrants when interpreting model 2.

Model 3: Controlling for Education (Column 3)

The third model includes controls for various levels of education, providing a more detailed view of the wage differences when being compared to high school dropouts which is the reference group in this case. Model 3 displays the coefficient for immigrants of -0.128, indicating that male immigrants earn 12.8% less than native-born males with the same level of education. This larger wage gap, compared to the second model, suggests that when controlling for education, immigrants earn substantially less. The contrast between models 2 and 3 implies that the smaller gap observed in model 2 (2.57%) could partly be because immigrants in the sample generally have higher levels of education. Thus, when comparing immigrants and non-immigrants with similar educational attainment, the wage discrepancies become more noticeable.

Continuation of Model 3: Education and Wage Outcomes

The completion of the highest level of education significantly impacts earnings, as reflected in the coefficients for the different levels of education variables in the third regression model. For example, secondary education has a coefficient of -0.043, indicating a slight decrease in earnings compared to highschool dropouts. This could mean that possibly those who dropped out decided to pursue a different type of education to earn more such as completing a certificate or labour work Those who completed the highest level is some post-secondary education (some_post) see an increase in earnings by 14.8% (coefficient of 0.148) when compared to the reference group of highschool dropouts. A bachelor's degree follows with a 39.8% wage increase (coefficient of 0.398), while a master's degree further boosts earnings by 57% (coefficient of 0.570). This model summarizes that higher education (HigherEd) yields the highest return, increasing wages by 70.4% (coefficient of 0.704).

I also found an interesting fact about the difference between individuals with secondary education and those with a bachelor's degree. By adding the coefficients for secondary education and bachelor's degree, I found that the difference = 0.398 - (-0.043) = 0.441. This indicates that individuals with a bachelor's degree earn approximately 44.1% more compared to those with only secondary education. This impact of obtaining higher education supports the importance of promoting bachelor’s degrees.

Secondly I found that the wage difference between native-born and immigrant women with higher education is significantly interesting. I can calculate the wage difference between native-born and immigrant women who have attained higher education (e.g., a master's degree) by taking the value of native-born women with higher education = 0.704 and immigrant women with higher education = -0.128 + 0.704 + 0.0253 = 0.6013. This concludes that native-born women with higher education earn more compared to immigrant women with the same level of education. This calculation highlights the persistent wage gap even among highly educated women, indicating that immigrant women are at a disadvantage despite their educational qualifications.

C) Extension

For my extension, I focused on refining the sample by excluding wage observations between $1000 and $5000, as I found these values to be less relevant to my analysis and that if I am setting my restrictions for full-time workers and those who work 49-52 weeks in a year shouldn't be earning $1000 to $5000 in a year. In particular, I applied the following restrictions which resulted in dropping observations where wages equal $1000 (18,443 observations deleted), $2000 (10,273 observations deleted), $3000 (9,727 observations deleted), $4000 (9,126 observations deleted), and $5000 (8,695 observations deleted). In total, 56,264 observations were removed from the dataset.

Upon imposing these restrictions, I observed that these restrictions of these wage observations did not lead to any meaningful changes in the overall results of my regression analysis. The coefficients and their significance levels remained relatively the same compared to the original findings. This suggests that the presence of wages in the $1000 to $5000 range had a minor impact on the results, and their exclusion did not affect the conclusions drawn from the data. However, I suspect that further restrictions will similarly have little to no effect on the overall results.

4. Conclusion

In this study, I analyzed wage differences using the 2021 census dataset, focusing on immigrant status, gender, and levels of education. The dataset consisted of 228,400 observations, with about 25% of immigrants in the sample, and a balanced gender distribution consisting of 46% female. Notably, 11.6% of the sample were female immigrants.

The analysis began with the regression in Model 1, which presented the raw wage gap, showing that immigrants earn approximately 3.4% less compared to non-immigrants. Model 2 introduced an interaction term for female and immigrant status, revealing that male immigrants earn 2.6% less than native-born males, while female immigrants earn 4.6% less than female non-immigrants, highlighting a larger wage gap for female immigrants.

Model 3 consisted of the controls for education, demonstrating that male immigrants with the same level of completed education as native-born males earn 12.8% less, indicating a substantial wage disadvantage even after accounting for education. This model resulted in higher educational completion and increased earnings as bachelor's degree led to 39.8% increase in earnings, a master's degree led to a 57.0% increase, and higher education led to a 70.4% increase in earnings.

Additionally, the wage difference between individuals with secondary education and those with a bachelor's degree was found to be 44.1%, emphasizing the positive impact of higher education in relation to earnings. However, a persistent wage gap remains between native-born and immigrant women with higher education, suggesting that immigrant women face disadvantages despite their highest level of education obtained.

For the extension of this study, I excluded wage observations between $1000 and $5000. This concluded to have little to no effect on the main results, confirming the consistency of the findings as there was not much of a change in the values of the coefficients. Overall, the results highlight the difference range in earnings based on immigrant status, gender, and highest level of education obtained. These findings provided me with a better understanding of what accounts for the differences in average earnings for immigrants and native-born individuals.

5. References

Statistics Canada. (2023). *Census Profile. 2021 Census of Population* (Catalogue number 98-316-X2021001). Ottawa. Released November 15, 2023 (accessed July 10, 2024)

Summary results: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
lnwage	228400	11.09	.61	8.7	13.78
immigrant	228400	.25	.43	0	1
female	228400	.46	.50	0	1
female_immig	228400	.12	.32	0	1
secondary	228400	.21	.41	0	1
some_post	228400	.35	.49	0	1
bachelors	228400	.25	.44	0	1
masters	228400	.08	.28	0	1

Main Results

	(1)	(2)	(3)
VARIABLES	lnwage	lnwage	lnwage
immigrant	-0.034***	-0.026***	-0.128***
Std. Error	(0.002)	(0.003)	(0.003)
female		-0.191***	-0.246***
Std. Error		(0.002)	(0.002)
female_immig		-0.020***	0.025***
Std. Error		(0.005)	(0.005)
secondary			-0.043***
Std. Error			(0.004)
some_post			0.148***
Std. Error			(0.004)
bachelors			0.398***
Std. Error			(0.005)
masters			0.570***
Std. Error			(0.006)
highered			0.704***
Std. Error			(0.01)
Constant	11.10***	11.19***	11.03***
Std. Error	(0.001)	(0.002)	(0.004)
Observations	228,400	228,400	228,400
R-squared	0.001	0.028	0.139

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1