Degrees of Difference: Analyzing the Impact of Education on Earnings in Canada*

Yuanyi (Leo) Liu

April 18, 2024

In this paper, we explore the connection between education and hourly wages in Canada, focusing on data from the year 2000. Our findings indicate a clear trend: higher educational levels correlate with increased average hourly wages for individuals aged 25 to 54. This research highlights the importance of education in determining earning potential and suggests that investment in education could have long-term economic benefits. The study provides evidence for policy implications regarding educational incentives and workforce development strategies in Canada.

Table of contents

*Code and data are available at: https://github.com/leoyliu/Analyzing-the-Impact-of-Education-on-Earnings-in-Canada

1 Introduction

The relationship between education and earnings is a well-established topic of interest within the field of labor economics, providing insights into the broader socio-economic fabric of a country. As Canada's economy continues to diversify and specialize in various sectors, the value of education in this landscape remains a critical question for policymakers and the public. While previous studies have explored this link, there is an ongoing need to update and deepen our understanding of how this dynamic plays out in the modern economy.

This paper narrows the focus to the Canadian labor market, where the interplay between educational achievement and wages within the core working age group, those aged 25-54, offers a mirror to the efficacy and value of educational advancements in the workplace. This demographic is selected for its relevance to the labor force and policy implications, excluding younger individuals still likely in the education system and older individuals, who may present outlier educational experiences that do not align with the central workforce.

This investigation seeks to address a gap in the current literature by providing a targeted analysis of the 25-54 age group in Canada since the year 2000, a period marked by rapid technological change and economic evolution. By applying a linear model to wage data categorized by education level, this study uncovers a positive correlation between educational attainment and hourly wages, suggesting that higher education can be linked to improved wage outcomes in this key demographic.

In this study, our primary focus is to estimate the effect of education on hourly wages in Canada. Our estimand is the incremental average hourly wage rate for individuals within the specified education levels, compared to the next lower education level. By quantifying this effect, we aim to capture the economic value of educational attainment and how it translates into wage premiums across various stages of a typical Canadian worker's career span.

The structure of the paper is organized as follows: Following Section 1, Section 2 presents the data, detailing the data sources, analytical techniques, and the rationale behind the chosen methods. Section 3 then delves into the specifics of the linear model analysis, laying out the statistical underpinnings that support our investigation. After that, Section 4 discusses the results, elaborating on the observed trends and patterns in wage rate data. Section 5 interprets these findings in light of the current economic and educational context in Canada,

exploring potential factors influencing these trends, drawing connections to broader socio-economic issues, and providing suggestions for future research in this area.

2 Data

This section aims to offer an insightful understanding of the dataset utilized in our analysis, which serves as the foundation for our examination of the relationship between education and average hourly wage rates in Canada, specifically within the 25-54 age demographic.

2.1 Source and Methodology

This study utilizes a dataset sourced from the Open Government Portal of Canada [], specifically designed to track the correlation between educational attainment and hourly wages across various demographic segments of the Canadian workforce. The dataset covers the period from the year 2000 onwards, providing a longitudinal view of wage trends in relation to educational background.

Alternative datasets, such as those from Statistics Canada's Labour Force Survey, were considered but were not selected due to their less detailed categorization of education levels and their broader focus on employment without specific wage breakdowns by education.

The data was processed and cleaned using R(R Core Team 2020), a powerful statistical programming language. The selection of this dataset was motivated by its direct relevance to the research question and its regular updates, ensuring that the data remains relevant for observing current trends. Initial data processing involved filtering the dataset to focus solely on entries categorized under 'Canada' in geography, eliminating data irrelevant to the national focus of this research. This step was crucial to maintaining the clarity and relevance of the analysis. For key operations, please refer to the Section A.

2.2 Variables

To better understand data, key variables extracted for this study include the year of data collection, education level, age group, and the average hourly wage rate. Notably, we have streamlined the 'Education level' from a broad range of categories into a numeric variable that aligns with ascending educational attainment, facilitating a quantitative analysis of its impact on wages. The age group has been limited to the 25-54 years range to focus on the most economically active segment of the population, avoiding potential outliers from younger individuals with less work experience and older individuals whose advanced education, such as doctoral degrees, could distort the analysis. Wage rates are presented in Canadian dollars per hour, reflecting pre-tax earnings, inclusive of tips, commissions, and bonuses.

Table 1: First Ten Rows of Wage Rates Spanning from 2000 to 2019 For Cleaned Data

Year	Education Level	Age	Hourly Wage
2000	0 - 8 years	25-54 years	13.1
2000	High school graduate	25-54 years	15.8
2000	Post-secondary certificate or diploma	25-54 years	18.0
2000	Trade certificate or diploma	25-54 years	17.4
2000	Community college, CEGEP	25-54 years	18.1
2000	University certificate below bachelors degree	25-54 years	20.2
2000	University degree	25-54 years	23.4
2000	Bachelor's degree	25-54 years	22.5
2000	Above bachelor's degree	25-54 years	25.6
2001	0 - 8 years	25-54 years	13.3

Table 1, created with `kableExtra` (Zhu 2021), outlines the first ten records of our dataset, featuring average hourly wages for Canadians aged 25-54 across varying levels of education from the years 2000 to 2019. This snapshot reveals the variable structure of our analysis, which includes the `Year` of wage data, the defined `Education level` ranging from "0 - 8 years" to "Above bachelor's degree," and the consistent `Age group` focus. It provides an early indication of the ascending trend in wages with higher educational attainment, setting the stage for our deeper investigation into the economic value of education within Canada's workforce.

To transition from a tabular overview to a more graphical representation, we will visualize the data, which will allow us to spot patterns and trends that are not immediately apparent in numerical form.

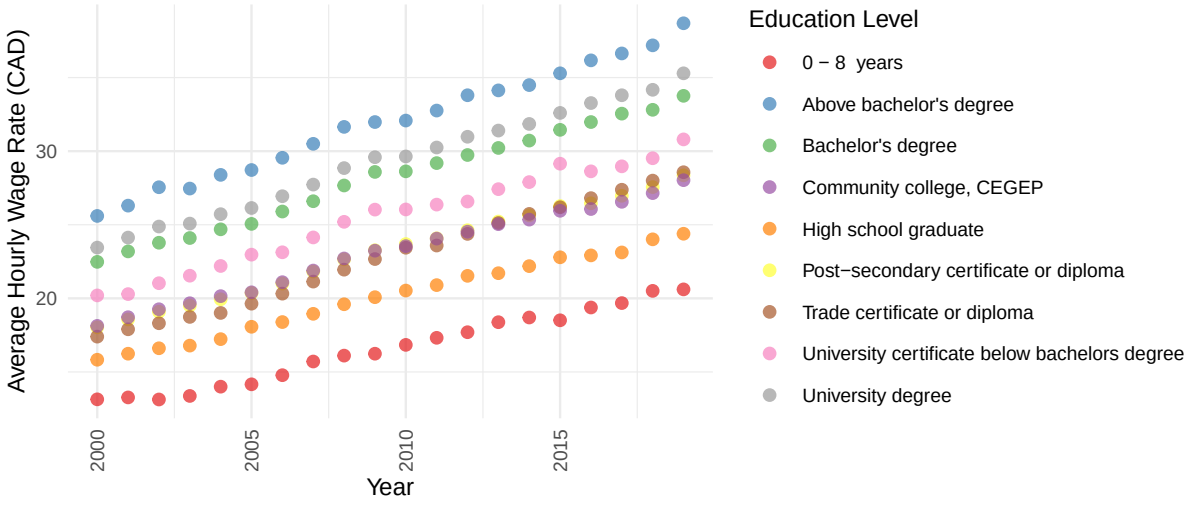

Figure 1: Visualization of Wage Rate from 2000 to 2019

Figure 1 illustrates the distribution of average hourly wage rates across different educational levels from 2000 to 2019 for Canadian workers aged 25-54. Each dot represents the wage rate for a specific education level in a given year, with color coding to distinguish between educational categories, ranging from "0 – 8 years" to "University degree." The spread and upward trend of the points suggest that individuals with higher educational qualifications tend to have higher average hourly wages, and this trend persists over the 20-year period.

2.3 Measurements

The measurement of variables within this dataset was handled to ensure accuracy and relevance. The 'Average Hourly Wage Rate' is calculated based on actual wages reported by employers, providing an objective measure of income. This data is then vetted and standardized by statistical agencies to ensure it reflects accurate and fair representations of wages across different demographics.

For educational attainment, data entry occurs as individuals enter the workforce or update their qualifications, with categorizations reflecting the highest level of formal education completed. This data is collected through surveys and employment records, often as part of broader demographic data collection efforts by governmental agencies.

For an in-depth technical breakdown of the statistical analyses and additional visualizations, readers are directed to the appendices, where these aspects are discussed in detail. This ensures that the main text remains accessible while still providing resources for more technically inclined readers.

Correlations between the variables will be examined at Section 3, providing insights into how education levels are associated with hourly wages. By understanding these relationships, we can better infer the potential impact of education on earnings within the Canadian economy.

3 Model

The objective of our modeling approach is to quantify the relationship between educational attainment and hourly wages. Our analysis employs a Bayesian framework to assess how changes in the level of education correlate with variations in wage rates among Canadian workers aged 25-54.

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in **?@sec-model-details**.

3.1 Model set-up

Define y_i represent the average hourly wage rate for the i^{th} individual. The predictor variable, x_i, corresponds to the numeric value assigned to each education level. The model can be described by the following equations:

$$\begin{aligned} y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \beta_0 + \beta_1 x_i \\ \beta_0 &\sim \text{Normal}(0, 2.5) \\ \beta_1 &\sim \text{Normal}(0, 2.5) \\ \sigma &\sim \text{Exponential}(1) \end{aligned}$$

We run the model in R (R Core Team 2020) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

The model is executed in R(R Core Team 2020) using the rstanarm package (**rstanarm?**). Default priors from rstanarm are utilized, set with a mean of zero and a conservative standard deviation, reflecting no strong prior beliefs about the magnitude of the education coefficients.

3.2 Model justification

Given the economic theory and previous empirical findings, we anticipate that higher educational qualifications would correspond to higher hourly wages. This is encapsulated in the assumption that β_i should have a positive effect on y_i. Through Bayesian analysis, we aim to capture the degree to which educational attainment can predict wage rates, thus providing empirical evidence to support or refine this hypothesis.

Mathematical notation is incorporated using LaTeX syntax, enabling precise formulation of our statistical model. For example, the average effect of education on hourly wage rate is denoted by β.

4 Results

Section 4 presents the core findings from our analysis of US birth rates, specifically focusing on the changes observed over time, across different age groups of young women, and by state.

In conclusion, our findings provide a complete perspective of dropping birth rates in the United States, demonstrating both temporal and demographic tendencies. The upcoming Section 5 will look at the various variables and biases that may be driving these trends.

5 Discussion

The findings of this study offer a layered perspective on the decline in birth rates among young people in the United States, particularly in the context of the Great Recession. Our analysis provides an understanding of the multifaceted influences on demographic trends.

5.1 Findings

In our replication of Kearney, Levine, and Pardue (2022), we reaffirmed the central finding that the Great Recession has had a lasting impact on birth rates among young people in the United States. Our analysis of Figures 1 and 2 showed a pronounced decline in birth rates across all age groups, with the steepest decrease observed among those aged 15-29. This trend extended beyond the economic recovery, suggesting that the Great Recession may have fundamentally altered the family planning trajectory of a generation.

5.2 Economic Impact Insights

The dramatic drop in birth rates among young Americans during and after the Great Recession demonstrates the profound impact of economic hardship on reproductive decisions. This decline in birth rates is not simply coincidental with the date of the recession, but rather indicates a deeper, more systemic influence of economic insecurity on personal life choices, notably the decision to establish or expand a family. The persistence of this pattern, even in the years after economic recovery, suggests that the effects of the recession went beyond immediate financial hardship, influencing long-term views of financial security and stability. This is further evidenced by the lack of rebound in birth rates post-recession, which might have been anticipated if the decline were solely due to immediate economic pressures.

The economic model of fertility provides a framework for understanding these patterns by taking into account the cost of pregnancy and raising as well as the opportunity costs associated with parental time and resources. The recession likely increased these expenditures and opportunity costs, making the decision to have children more difficult. Unemployment and job uncertainty may have caused a reevaluation of the feasibility of having children, resulting in the postponement or avoidance of childbirth. Furthermore, the economic downturn may have shifted young adults' expectations and desires for financial security, impacting their family planning decisions.

5.3 Societal and Technological Influences

Apart from economic factors, the decline in birth rates among young people also reflects broader societal and technological shifts. The period following the Great Recession coincided with significant changes in social norms, increased educational and career opportunities for

women, and advancements in reproductive technology. These factors collectively have empowered individuals, especially women, to make more independent decisions regarding their reproductive health and family planning.

As women's educational attainment and labor force involvement have increased, so has the opportunity cost of childbirth, potentially contributing to the drop in birth rates. Technological developments have given women greater control over their fertility, allowing for more intentional planning around childbirth. Social changes, such as delayed marriage and a growing acceptance of childlessness or smaller family sizes as realistic lifestyle options, further compound these tendencies.

The interaction of these economic, social, and technical elements has altered the landscape of family planning for young Americans. While the economic model of fertility provides a framework for understanding these developments, including sociological and technical effects provides a more complete view of the dynamics affecting current birth rate patterns.

5.4 Weaknesses and Future Research Directions

One limitation of our study is the potential for unobserved variables that could affect birth rates, such as cultural shifts and changes in social norms, which were not fully captured in the data. Additionally, the original study did not account for the influence of the Affordable Care Act and its potential impact on family planning decisions, an area our study also does not explore.

Future research should aim to disentangle these complex relationships further, perhaps through longitudinal studies or by incorporating more detailed data on individual socioeconomic status. Understanding these dynamics is crucial for developing policies that support young people in their family planning decisions during and after economic downturns.

Appendix

A Data Manipulation and Cleaning

Most of the data in our dataset was previously cleaned for the project we are replicating. Thus, **?@fig-trends-in-birth-rate** directly used data from `outputs/data/fig_1.csv` without any further cleaning necessary, as it simply displays the birth rate from 1980 to 2020. Thus, the cleaned data is directly stored to `outputs/data/fig_1.csv` from `inputs/data/fig_1.csv`.

In **?@fig-trends-in-birth-rate-in-young-people**, the dataset was imported from `inputs/data/fig_2.csv` and used `dpylr` (Wickham et al. 2023) to select the birth rates of the six age groups (from 15 to 44). Furthermore, the selected data was renamed using `tidyverse` (Wickham et al. 2019) and changed into names that represent the content of the data better. Moreover, these data is then pivoted using `tidyverse` (Wickham et al. 2019) in order to be graphed properly using `ggplot2` (Wickham 2016).

In **?@fig-birth-rate-comparison-old-and-young**, the first dataset was imported from `inputs/data/fig_3.csv` and the dataset containing American states information was from `mapdata` (Richard A. Becker and Ray Brownrigg. 2022). The first dataset was first sliced to remove birth rate information regarding the states of Alaska and Hawaii, as these two states can't be properly shown in the dataset from `mapdata` (Richard A. Becker and Ray Brownrigg. 2022). Then, a temporary data frame is created with tibble from `tidyverse` (Wickham et al. 2019) with the abbreviated state names from `inputs/data/fig_3.csv` and the state full names. Moreover, the abbreviated state name in the first dataset is switched with the state full name using left_join from `dpylr` (Wickham et al. 2023), which is then selected and renamed using `tidyverse` (Wickham et al. 2019). After this, the adjusted dataset with the state full names and birth rate is merged with the second dataset that contains the American states information for further graphing with `haven` (Wickham, Miller, and Smith 2023).

In all three figures, the library `here` (Müller 2020) was used to ensure that the file path should be accessible in all directories.

The transformation from raw to analysis-ready data involved substantial cleaning, including filtering by geography (Canada), type of work (both full- and part-time), and specific wage type (average hourly rate). Moreover, we excluded broad or ambiguous categories like 'Some high school' and 'Total, all education levels' to ensure clarity in our analysis. The resulting 'Education_numeric' variable was constructed using a sorted factorization process, mapping each education level to a corresponding numeric value.

In presenting our findings, we will incorporate graphs that visualize the relationship between education and wage rates, providing an intuitive understanding of the data's characteristics. Tables summarizing the observations, along with discussions on the graphs, will be included to elucidate key trends and observations. Summary statistics and any notable relationships between the variables will also be featured, offering a detailed overview of the dataset. Should

the volume of this information warrant it additional detail, these elements may be relegated to appendices to maintain the accessibility of the data section.

References

Kearney, Melissa S., Phillip B. Levine, and Luke Pardue. 2022. "The Puzzle of Falling US Birth Rates Since the Great Recession." *Journal of Economic Perspectives* 36 (1): 151–76. https://doi.org/10.1257/jep.36.1.151.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richard A. Becker, Original S code by, and Allan R. Wilks. R version by Ray Brownrigg. 2022. *Mapdata: Extra Map Databases*. https://CRAN.R-project.org/package=mapdata.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://haven.tidyverse.org.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/, https://github.com/haozhu233/kableExtra.