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

Yuanyi (Leo) Liu

April 2, 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 across all age categories. 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 across different age groups in the modern economy.

This paper aims to fill the knowledge gap by analyzing the impact of education levels on the average hourly wage rate in Canada, with a specific focus on three age cohorts: young entrants (15-24 years), established workers (25-54 years), and senior professionals (55 years and over). The analysis begins in the year 2000, a period of significant technological advancement and economic change, offering a contemporary perspective on the education-wage relationship. By dissecting the data across these age groups, the study provides nuanced insights into the economic returns of education at various career stages.

Our research methodology involved a detailed statistical analysis of wage data, stratified by education level and age group, to uncover patterns and variations in earnings. The findings confirm the hypothesis that higher education correlates with increased wages, with the effect becoming more significant as workers age. This not only reinforces the case for educational investment but also suggests that such investment pays dividends throughout an individual's career, not just in the early years.

In this study, our primary focus is to estimate the effect of education on hourly wages across different age groups 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, while controlling for age. 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. The dataset captures the birth rates across various demographics in the United States from 1980 to 2020. The data provide a broader context, allowing an analysis of trends over time and across different economic cycles, including the period of the Great Recession.

2.1 Source and Methodology

Data on birth rates for women aged 15-44 were sourced from National Vital Statistics Reports covering 2015, 2019, and 2020 (Martin and Mathews 2017; Martin and Driscoll 2021; Hamilton and Osterman 2021), compiled from birth certificates across the US and the District of Columbia. Kearney, Levine, and Pardue (2022) provided aggregated birth data for six cohorts over five-year spans, derived from the NBER Natality Database (National Bureau of Economic Research 2021) and NCHS microdata (National Center for Health Statistics 2020) for 1980-2019. In this study, we simply used the replication data package provided by Kearney, Levine, and Pardue (2022) for analysis.

While there were alternative datasets available from other public and private sources, this dataset was chosen due to its comprehensive coverage, reliability, and the level of detail it offers. It includes data from the National Center for Health Statistics (National Center for Health Statistics 2020) and the U.S. Census Bureau, which are both reputable sources of demographic data.

The data was processed and cleaned using R(R Core Team 2020), a powerful statistical programming language. For key operations, please refer to the Section A.

2.2 Variables

To better understand the data, a summary table was developed to provide a detailed description of each variable, explaining its relevance and how it contributes to our understanding of the topic. Our focus is not only on the direct measures of birth rates but also on a range of demographic and economic variables that offer a comprehensive view of the factors influencing fertility decisions during the recessionary period.

2.3 Measurements

The measurement of birth rates in our study involves a detailed analysis of the annual number of live births per 1,000 women in various age groups, ranging from 15 to 44 years. These rates were extracted from datasets provided by the National Center for Health Statistics (NCHS) (National Center for Health Statistics 2020), ensuring accuracy and reliability in capturing demographic trends across the United States from 1980 to 2020.

For each age group, birth rates were calculated by dividing the total number of live births by the population of women in that age range, then multiplying by 1,000 to standardize the measure. This approach allows for comparison across different demographics and time periods, providing a clear picture of how birth rates have shifted, particularly in response to economic conditions such as the Great Recession.

In addition to age-specific birth rates, state-level data were also analyzed to identify geographic variations in birth rate trends. This involved mapping birth rates against state populations, taking into consideration the exclusion of data from states like Alaska and Hawaii where mapping constraints exist. Further details on this aspect of our study are discussed in Section 4.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

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 as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

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

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

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.

References

Hamilton, Martin, Brady E., and Michelle J. K. Osterman. 2021. "Births: Provisional Data for 2020." *Vital Statistics Rapid Release*, no. 12: 6. https://doi.org/https://doi.org/10.15620/cdc:104993.

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.

Martin, Hamilton, Joyce A., and Anne K. Driscoll. 2021. "Births: Final Data for 2019." *National Vital Statistics Report* 70 (2): 13. https://doi.org/https://dx.doi.org/10.15620/cdc:100472.

Martin, Hamilton, Joyce A., and T. J. Mathews. 2017. "Births: Final Data for 2015." *National Vital Statistics Report* 66 (1): 35. https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_01.pdf.

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

National Bureau of Economic Research. 2021. *NCHS Vital Statistics Natality Birth Data*. https://data.nber.org/data/vital-statistics-natality-data.html.

National Center for Health Statistics. 2020. *Restricted-Use Vital Statistics Data*. https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm.

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.