Rising Tides of Debt: The Evolving Landscape of Student Debt in Canada*

An Analysis of Changes in Student Debt Levels and Sources from 2000 to 2020

Ravnit Lotay

April 19, 2024

The dynamics of student debt in Canada have changed notably over the last two decades, shaped by shifts in government policy and broader socioeconomic factors, including increased tuition fees for international students. This study examines the evolution of student debt from 2000 to 2020, focusing on the amounts owed, sources of debt, and repayment status across different educational levels. By analyzing data from a survey of graduates, this research uncovers trends in student debt, emphasizing how factors like higher tuition for international students and changes in loan forgiveness policies have influenced borrowing and repayment habits. The results aim to shed light on the financial challenges students face and support policy recommendations to reduce student debt burdens.

Table of contents

---

*Code and data are available at: https://github.com/Ravnit202/Decadal-Shifts-in-Canadian-Student-Debt-2000-to-2020-Analysis.

1 Introduction

The landscape of student debt in Canada has experienced significant changes over the past decades, influenced by a myriad of economic and policy shifts. Particularly noteworthy is the period following the election of the Liberal government in 2015, led by Prime Minister Justin Trudeau, which marked a shift in federal policies affecting higher education funding and student loans. This paper explores the evolution of student debt across different education levels and

provinces in Canada, including comprehensive data from all regions, from Quebec to British Columbia, and the territories such as Yukon, the Northwest Territories, and Nunavut.

Our analysis is rooted in the study of data collected between the years 2000 and 2020, sourced from government and educational institutions, which provides insights into the amounts owed by graduates, the percentage of those who manage to clear their debts, and the lingering debts post-graduation.

These models help to quantify the impact of changes in policy and economic conditions on student debt levels, offering a detailed breakdown by province and level of study. The findings from this study not only reflect the direct outcomes of specific governmental policies but also align with broader economic conditions that influence students' financial decisions and capabilities.

This introduction sets the stage for a deeper investigation into how governmental policy shifts and economic cycles affect student debt, guiding readers through a factual and analytical journey that reveals the nuances of financial burdens borne by Canadian students. The subsequent sections will further detail the methodologies employed, present the results, and discuss the implications of our findings in the context of current and future educational and economic policies.

2 Data

2.1 Methodology

The dataset employed for this study provides an extensive overview of student debt in Canada, covering various provinces and education levels. This dataset includes both raw and processed data regarding the amounts of debt held by graduates, categorized by province, level of study, and the source of the debt (government or non-government).

Data processing and analysis were performed using the R programming language, leveraging a suite of packages designed to facilitate data manipulation, visualization, and statistical modeling. The primary tools included:

- **tidyverse** (Wickham et al. 2023): Utilized for its comprehensive data manipulation capabilities.
- **ggplot2** (Wickham 2016): Employed for creating advanced visualizations.
- **dplyr** (Wickham, François, et al. 2023): Used extensively for data transformation tasks.
- **readr** (Wickham, Hester, et al. 2023): Enabled efficient reading and handling of CSV data formats.
- **loo** (Vehtari et al. 2023): Used to compare models.
- **arrow** (Richardson et al. 2024): Applied for reading and writing data in Parquet format, enhancing data storage efficiency.

- **janitor** (Firke 2023): Helped clean the dataset by formatting variable names and pruning missing values.
- **testthat** (Wickham 2023): Ensured the reliability of our data transformations and model functions through systematic testing.
- **rstanarm** (Goodrich et al. 2024): A Bayesian modeling framework used for statistical inference, enhancing the analysis of complex data relationships.

Additionally, a Datasheet has been included to accompany the paper. This Datasheet includes metadata about the variables used, the methods of data collection, and any important notes for understanding the data. This resource aims to make the research transparent and reproducible. You can find the Datasheet in the appendix of this paper.

2.2 Data Source and Measurements

The primary dataset utilized in this research comprises student debt data across Canada, including regions such as Yukon, the Northwest Territories, and Nunavut, along with all other provinces. This dataset was obtained from the Government of Canada's open data portal and is available here. The metadata associated with this dataset, although available, was not necessary for the analyses conducted in this study.

Note, in the dataset, "Canada" refers to all of Yukon, the Northwest Territories, and Nunavut.

For the analyses, two main aspects of student debt were focused on: percentage-based assessments and dollar-value assessments of student debt, as categorized under different educational levels and sources of debt. Sample data can be seen below in Table 1.

Table 1: Preview of unique geographic regions in the dataset on student debt across Canada

geo	level_of_study	type_of_debt_source	$	$%
Alberta	College	Graduates who owed money for their education to any source (government or non-government)	NA	54
BritishColumbia	College	Graduates who owed money for their education to any source (government or non-government)	NA	45
Canada	College	Graduates who owed money for their education to any source (government or non-government)	NA	49
Manitoba	College	Graduates who owed money for their education to any source (government or non-government)	NA	36
NewBrunswick	College	Graduates who owed money for their education to any source (government or non-government)	NA	66
NewfoundlandandLabrador	College	Graduates who owed money for their education to any source (government or non-government)	NA	59

Table 1: Preview of unique geographic regions in the dataset on student debt across Canada

geo	level_of_study	type_of_debt_source	$	$%
NovaScotia	College	Graduates who owed money for their education to any source (government or non-government)	NA	56
Ontario	College	Graduates who owed money for their education to any source (government or non-government)	NA	47
PrinceEdwardIsland	College	Graduates who owed money for their education to any source (government or non-government)	NA	68
Quebec	College	Graduates who owed money for their education to any source (government or non-government)	NA	52
Saskatchewan	College	Graduates who owed money for their education to any source (government or non-government)	NA	50

Each entry in the dataset corresponds to a specific record of debt, detailing the geographic region, level of study, source of the debt, and debt metrics (either as a percentage or a dollar value).

2.3 Data Preparation and Cleaning Process

The dataset used in our analysis was prepared to ensure consistency and relevance for examining student debt trends across Canada. The initial data, sourced from the Open Government Portal, underwent a series of cleaning and normalization steps to ready it for in-depth analysis.

2.3.1 Data Loading and Initial Cleaning

The raw dataset was loaded using the `read_csv` function from the `readr` package, which is part of the `tidyverse` collection of R packages. Initial cleaning involved standardizing column names with the `clean_names` function from the `janitor` package to ensure consistency in syntax and readability. Several columns that were not relevant to our analysis were removed, such as metadata and identifiers not used in the model creation or analysis.

2.3.2 Data Transformation

Further transformations included: - Converting the reference year to an integer format for easier time-series analysis.
- Standardizing the geographic labels to remove inconsistencies in spacing.
- Adjusting the education levels to broader categories for a more generalized analysis, particularly aggregating all postgraduate studies into a single category.

- Separating the debt values into dollars and percentage terms based on the unit of measure, facilitating separate analyses for each type.

2.3.3 Data Normalization

To aid in comparative analysis and enhance model performance, the dollar and percentage values of student debt were normalized. Normalization was performed by scaling each value to a 0-1 range based on the minimum and maximum values observed in the dataset for each measure.

2.3.4 Data Storage

After cleaning and normalization, the dataset was saved in a `.parquet` format to maintain data integrity and ensure efficient loading for subsequent analyses. The `write_parquet` function was used from the `arrow` package, allowing for seamless integration with our data analysis pipeline.

This preparation pipeline was crucial for ensuring that the data was in an optimal format for conducting the detailed statistical analyses and model training described in the following sections.

2.4 Variables of Interest

There are several variables of interest, however, th two most important are the value of debt in dollars and the value of students in debt (percent). Below, all of the key variables will be listed out.

2.4.1 Main Variables

- `ref_date`: Tracks the year for each data record, crucial for spotting how student debt has changed over time.
- `geo`: Indicates the geographical area within Canada, encompassing all provinces and territories. This variable helps pinpoint regional differences in student debt.
- `level_of_study`: Classifies the educational level of graduates (College, Undergraduate, Postgraduate), key to exploring variations in debt by educational attainment.
- `type_of_debt_source`: Identifies whether the debt comes from government loans or private sources, which informs our understanding of how students finance their education.
- `statistics`: Details the type of data provided, such as the average debt at graduation or the percentage of graduates with debt.

2.4.2 Financial Measures

- `value_dollars`: The actual amount of debt in Canadian dollars, used for straightforward financial analysis.
- `value_percent`: Reflects the percentage of graduates impacted by specific debt statistics.
- `normalized_value_dollars`: A normalized version of `value_dollars`, scaled between 0 and 1 based on the observed range.
- `normalized_value_percent`: A normalized version of `value_percent`, scaled between 0 and 1 based on the observed range.

2.5 Model Setup

To analyze the factors influencing student debt levels in Canada, two Bayesian regression models wer utilized to predict normalized values of student debt, both in percentage and dollar terms. These models help us understand the relative importance of various predictors like level of study, geographical location, type of debt source, reference date, and statistics types on student debt.

2.5.1 Model Description

Two models were created — one for percentage-based and dollar-based debt values. For each model, predictors include `level_of_study`, `type_of_debt_source`, `geo` (geographical location), `ref_date` (reference year), and `statistics` (type of statistical measure reported). The outcome variable for the percentage model is `normalized_value_percent`, and for the dollar model, it is `normalized_value_dollars`.

Below is the general setup for both Bayesian models:

$$
\begin{aligned}
y_i | \mu_i &\sim \mathrm{N}(\mu_i, \sigma^2) \\
\mathrm{link}(\mu_i) &= \beta_0 + \beta_1 \times \mathrm{level\_of\_study}_i + \beta_2 \times \mathrm{type\_of\_debt\_source}_i \\
&\quad + \beta_3 \times \mathrm{geo}_i + \beta_4 \times \mathrm{ref\_date}_i + \beta_5 \times \mathrm{statistics}_i \\
\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 &\sim \mathrm{Normal}(0, 2.5) \\
\sigma &\sim \mathrm{Normal}^+(0, 2.5)
\end{aligned}
$$

Each β coefficient represents the effect of its corresponding predictor on the outcome variable, under a Gaussian family assumption with a link function appropriate for continuous data. These models are run using the `rstanarm` package, allowing us to incorporate prior knowledge and uncertainty directly into our inference process.

2.5.2 Model Justification

The use of two models is due to the different insights each type of data provides within the dataset:

- **Percentage-based Model**: Focuses on relative debt measures, like the percentage of graduates with debt. This helps compare debt levels over time and across different demographics without the influence of inflation or changes in dollar amounts.

- **Dollar-based Model**: This model assesses the actual amounts of debt, illuminating the financial challenges faced by graduates. It's crucial for evaluating the direct financial impact of student debt and guiding policies on loan amounts, repayment plans, and financial aid.

Having separate models allows for a more tailored analysis under the most suitable statistical methods, enhancing the accuracy and usefulness of the results. Choosing Bayesian regression supports the integration of existing knowledge and manages the detailed nature of economic data effectively.

3 Results

In this section, the student debt data will be visualized through various graphs and present the results from our statistical models.

3.1 Debt Distribution and Trends

The line graphs labeled Figure 1 and Figure 2 reveal an overall trend of increasing average debt in absolute dollar terms, while the percentage of students in debt shows a decline post 2010.

The boxplots under Figure 3 display the variability of debt across different levels of study. Notably, the postgraduates (Master's and PhDs) have the most widespread debt values.

The histograms at Figure 4 and Figure 5 provide a detailed look at the distribution of debt values across geographic regions, demonstrating how location and level of study jointly influence the financial experiences of graduates. Notably, there are several missing rows mentioned by R, indicating some sort of issue with these two charts. However, the general trend should be acceptable as it aligns with the rest of the results.
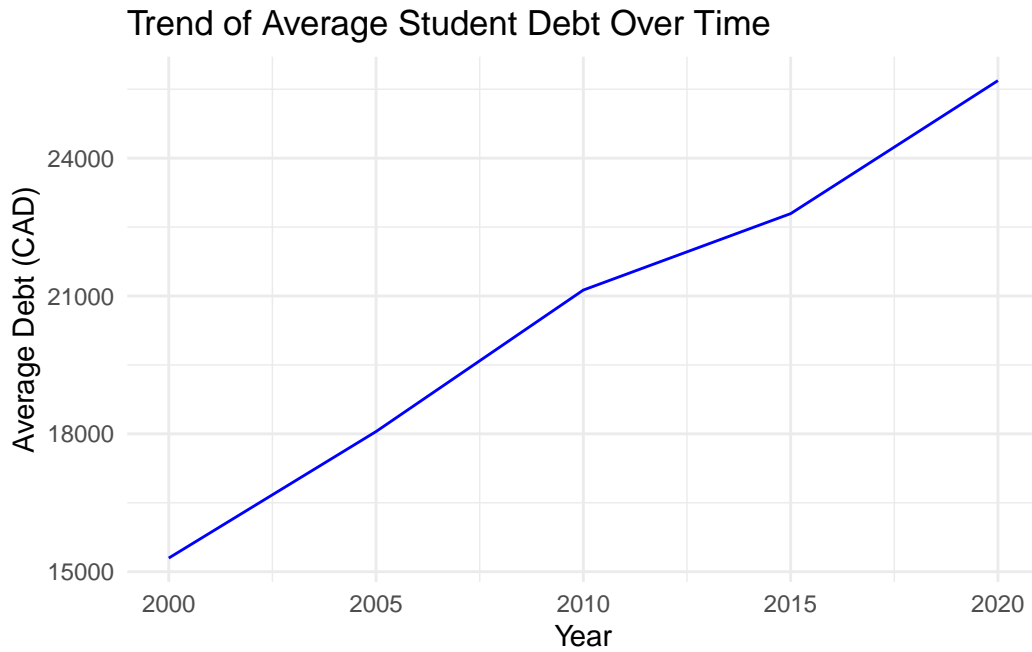
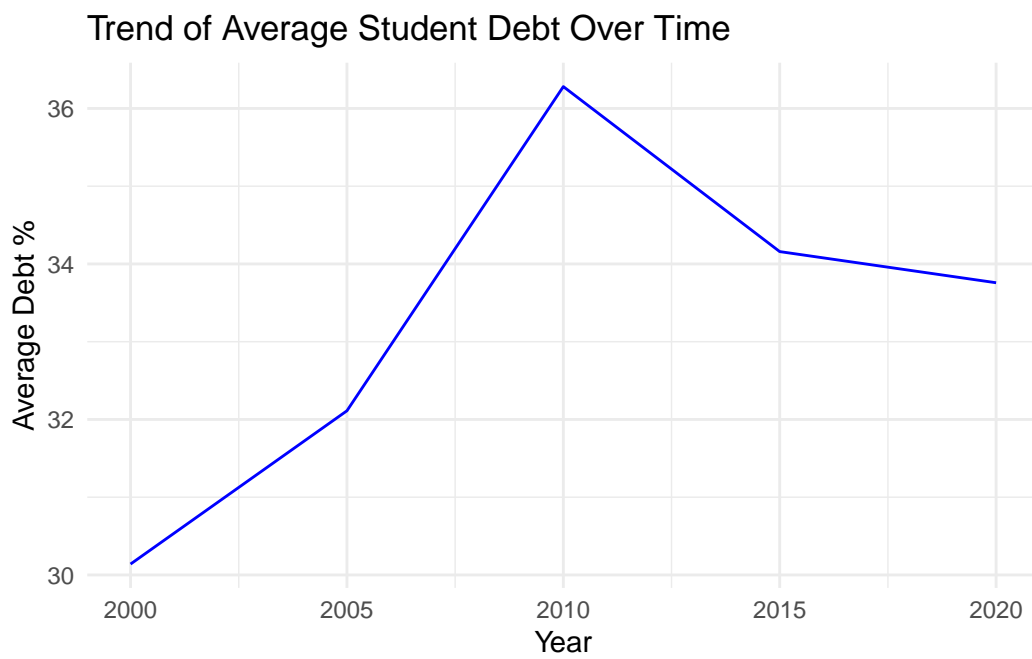Figure 1: Trend of Average Student Debt Over Time

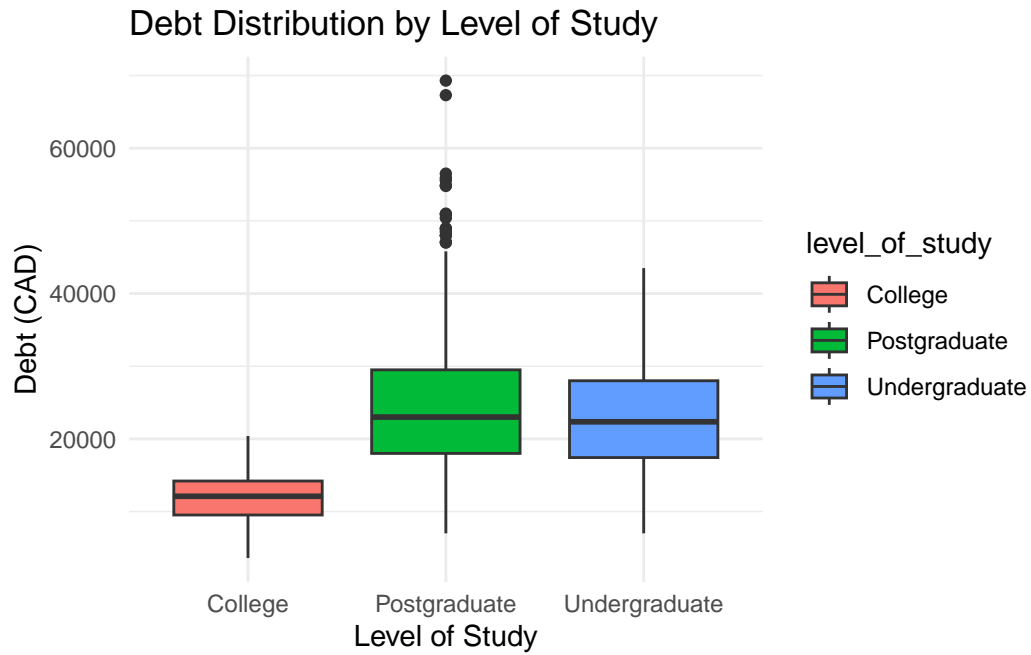Figure 2: Trend of Average Student Debt Over Time
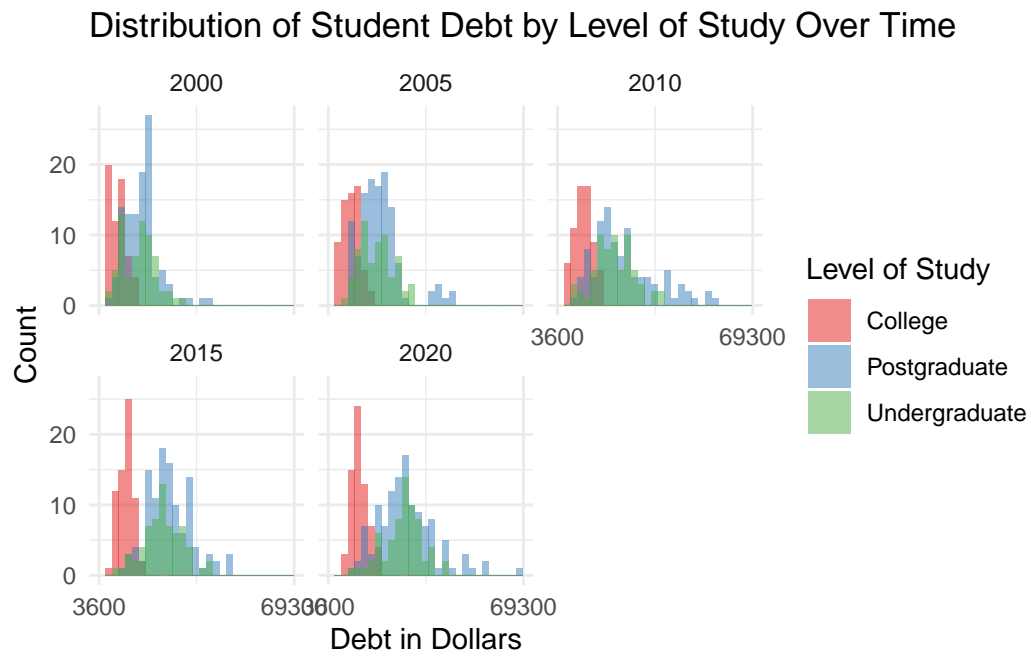
Figure 3: Debt Distribution by Level of Study

Figure 4: Debt Distributions Across Geographic Regions and Study Levels
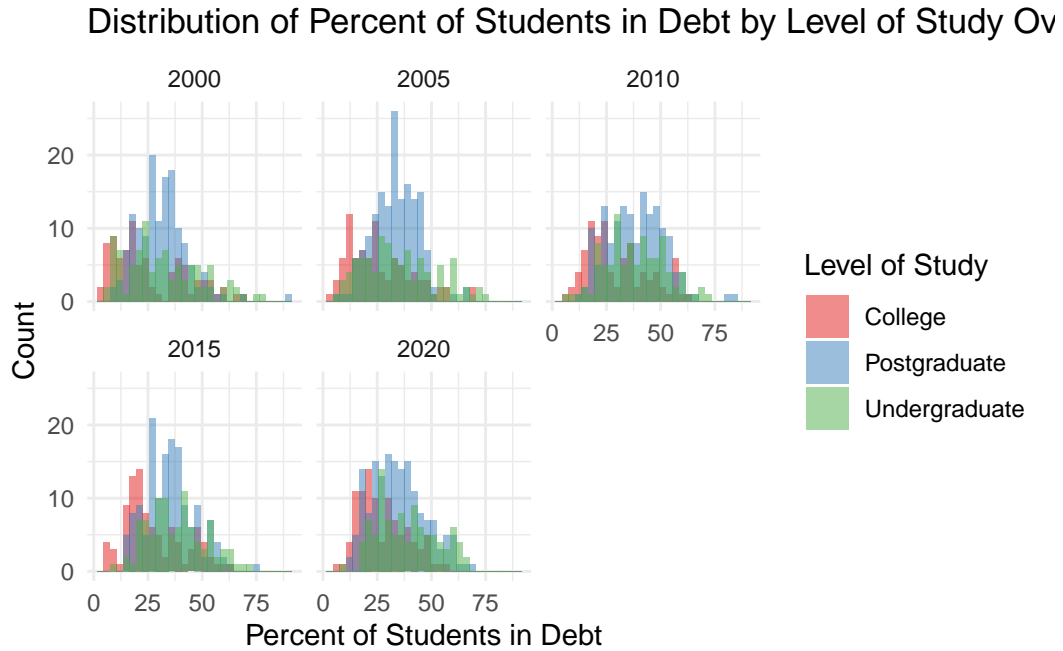
Figure 5: Debt Percent Distributions Across Geographic Regions and Study Levels

3.2 Model Results

This section discusses the findings from our statistical models analyzing student debt in Canada, presented in two formats: dollar amounts and percentage terms. The models help to identify which factors significantly affect student debt, aiding policymakers and educational institutions in decision-making.

3.3 Model Coefficients Overview

3.3.1 Percent-Based Model Coefficients

Coefficient	Estimate	Description
(Intercept)	-4.09	Represents the baseline normalized percentage of student debt when all predictor variables are zero. Indicates a baseline adjustment downward, suggesting lower debt percentages without specific influencing factors.
Level of Study - Postgraduate	0.08	Higher education levels slightly increase the percentage of debt.
Level of Study - Undergraduate	0.11	Undergraduates experience a slightly higher increase in debt percentage than postgraduates.
Type of Debt Source - Government	-0.07	Associated with a decrease in the percentage of debt, suggesting more favorable borrowing terms.
Type of Debt Source - Non-government	-0.10	Leads to lower overall debt percentages.
Geographic Regions - Manitoba	-0.03	Low regional influence on debt percentage with slightly lower percentages in Manitoba.
Geographic Regions - Quebec	-0.04	Low regional influence on debt percentage with slightly lower percentages in Quebec.
Year (ref_date)	0.002	A small annual increase in the percentage of debt, suggesting incremental growth in debt burden over time.
Statistics - Paid off debt	-0.03	Indicates that graduates who have paid off their debt typically have a lower remaining percentage of debt.

Coefficient	Estimate	Description
Statistics - Large initial debt	-0.02	Shows that graduates with larger debts at graduation typically have a lower remaining percentage of debt.

3.3.2 Dollar-Based Model Coefficients

Coefficient	Estimate	Description
(Intercept)	-15.07	A negative intercept suggesting an adjustment from a base level in dollar terms, indicating a starting point lower than zero when other variables are not considered.
Level of Study - Postgraduate	0.19	Higher education levels correlate with higher dollar amounts of debt, with postgraduates accruing slightly more than undergraduates.
Level of Study - Undergraduate	0.16	Correlates with higher dollar amounts of debt.
Type of Debt Source - Government	-0.06	Associated with lower dollar amounts of debt, suggesting more favorable terms.
Type of Debt Source - Non-government	-0.08	Also associated with lower dollar amounts of debt.
Geographic Regions - Quebec	-0.11	Demonstrates a significant negative effect on the dollar amounts of debt.
Year (ref_date)	0.007	Indicates a slight yearly increase in the dollar amount of student debt.

Coefficient	Estimate	Description
Statistics (Remaining Debt)	-0.04	Suggests that the average remaining dollar debt decreases over time, potentially due to repayment or forgiveness programs.

These coefficients offer insights into the various factors affecting both the amount and proportion of student debt across different demographics and years. Smaller coefficients suggest a minor influence, whereas larger absolute values indicate a stronger impact from those factors.

3.3.3 Leave-One-Out Cross-Validation (LOO)

The LOO method helps us understand the effectiveness of our models in generalizing beyond the observed data. It is particularly useful in identifying overfitting and underfitting, ensuring that the models are just right for our analytical needs.

Table 4: LOO results for the percentage-based and dollar-based models.

Model	ELPD_loo	p_loo	LOOIC
Percent Model	819.9974	19.34336	-1639.995
Dollar Model	1208.1980	21.51010	-2416.396

Figure 6: LOO Cross-Validation Results for Both Models

The LOO results shown at Figure 6 include ELPD, the effective number of parameters (p_loo), and LOO Information Criterion (LOOIC) for both the percent-based and dollar-based models. These values help us understand how well the models predict and their level of complexity.

Metric Percent Model Dollar Model ELPD_loo 820.0 1208.2 p_loo 19.3 21.5 LOOIC -1640.0 -2416.4

ELPD_loo: Higher numbers are better as they indicate more accurate predictions. p_loo: Fewer parameters reduce the risk of overfitting. LOOIC: Lower numbers are better, indicating a model that fits the data well.

These outcomes show that our models are well-suited for analyzing student debt trends in Canada, capable of handling different scenarios and data conditions effectively.

4 Discussion

4.1 Analysis of Results

The analysis shows a clear trend: while fewer students are taking on debt in later years, those who do face higher education costs. This situation suggests a need for policies aimed at reducing the financial burden of higher education. Additionally, the rise in debt levels may be linked to the increasing number of international students who typically pay higher tuition fees. Data also indicate that students pursuing Master's and PhD programs are most likely to accrue higher amounts of debt, whereas college students tend to have the lowest debt levels.

4.2 Real-World Application and Policy Impact

The Canadian government has actively been working on policies to reduce the financial burden of student loans. Noteably, under the leadership of the Liberal Party since their election in 2015, the Canadian government has attempted to reduce financial challenges of higher education through various student loan forgiveness programs. These efforts are visible in the annual increase in loan forgiveness amounts and are tailored to assist specific groups such as healthcare professionals working in remote or rural regions of Canada. Such targeted measures reflect the significant impact of regional and career-related factors on student debt levels, as identified through the analysis of data trends within this research. For a deeper dive into these initiatives, including the specifics of the loan forgiveness programs, please visit this CBC news article (Press 2017).

Additionally, government reports spanning 2014-2015 and 2015-2016 provide comprehensive insights into the evolving policies concerning student loans during these periods. These documents are essential for understanding the broader implications of policy decisions on student borrowing and repayment trends. Not only do they highlight shifts in policy but also align closely with the patterns and insights derived from the data utilized in our models. For more contextual understanding, these reports can be accessed online: 2014-2015 report (Canada 2017) and 2015-2016 report (Canada 2018).

These documents and policies play a pivotal role in shaping the financial landscapes for current and future students, aiming to mitigate the challenges associated with rising education costs. By examining these resources, stakeholders and policymakers can gain valuable insights into effective strategies for student debt management, ensuring that financial aid programs and loan forgiveness schemes are optimally designed to meet the needs of Canadian students.

4.3 Limitations and Weaknesses

The models explored in this study provide insights into student debt trends in Canada, yet they do not reflect significant policy shifts that affect international students, such as changes

to tuition fees and visa regulations. These factors critically influence debt trends but are omitted from our analysis. Furthermore, the models do not incorporate the legislative change that eliminated interest on Canada Student Loans effective April 1, 2023. This policy change is anticipated to modify borrowers' debt management strategies, likely influencing repayment behaviors and overall debt levels.

Moreover, the models fail to encompass the scope of the expanded student loan forgiveness programs intended to alleviate the financial burdens faced by particular groups, such as healthcare workers in rural areas. The inability of the models to predict the impacts of these programs indicates a gap in capturing forthcoming trends in student debt, which could skew our understanding of future debt patterns. As these new policies take effect and additional data become available, the conclusions drawn from the current data may need to be reassessed.

To enhance their predictive accuracy and relevance, the models should be updated to include these recent legislative changes. This update would improve the models' capability to reflect the actual debt landscape that Canadian students face. Furthermore, addressing the management of variables such as value_dollars and value_percent is crucial. The initial approach led to many rows with missing values, potentially undermining the robustness of our analyses. A more careful approach to data handling could prevent data loss and provide a clearer picture of student debt, ensuring that the models more accurately represent the diverse realities of student borrowers.

Enhancing the models will enable them to provide more precise and actionable insights, supporting more informed policy decisions aimed at reducing the financial challenges of higher education in Canada. As policies continue to evolve and new data become available, it will be essential to continuously update the models to maintain their relevance and to provide stakeholders with reliable data to support effective interventions.

4.4 Conclusion

The research uses models to analyze student debt in Canada, showing that higher degrees like Master's and PhDs usually lead to more debt, while college programs often result in less. The models also highlight how international students face higher tuition fees, increasing their debt burden. These findings suggest the need for targeted policies that address the varying financial challenges students face across different levels of education and backgrounds.

Additionally, the models suggest that as the educational landscape evolves, particularly with changes affecting international students and various debt forgiveness programs, policy strategies must also adapt. This ongoing adaptation is crucial to effectively support students in managing their debt. The study underlines the importance of continuous updates and refinements to these policies to keep them relevant and effective.

A Appendix

A.1 Datasheet for Student Debt Data in Canada

A.1.1 1. Data Identification

- **Title**: Student Debt Data in Canada
- **Source**: Open Government, Government of Canada
- **URL**: Open Data Canada
- **Scope**: Nationwide (Canada), covering various provinces and territories
- **Timeframe**: 2000 to 2020
- **Purpose**: To analyze trends in student debt across different levels of education and sources of funding in Canada.

A.1.2 2. Data Collection Method

- **Collection Process**: Likely collected from surveys or administrative records from educational and financial institutions.
- **Sampling Strategy**: Data spans multiple provinces, suggesting broad geographic coverage, possibly stratified to ensure representativeness across regions and levels of education.
- **Frequency**: Annual, as indicated by the presence of a year column (`ref_date`).

A.1.3 3. Data Structure

- **Format**: CSV
- **Size**: 2,945 entries, 10 columns
- **Variables**: Explained in the Attribute Information section.

A.1.4 4. Attribute Information

- **ref_date**: Year of data collection (range from 2000 to 2020).
- **geo**: Geographic regions covered, including Canada overall and individual provinces (e.g., Ontario, Quebec).
- **level_of_study**: Education levels including College, Undergraduate, and Postgraduate.
- **type_of_debt_source**: Categories include debts owed to any source, government-sponsored loans, and non-government sources.
- **statistics**: Describes the data point (e.g., average debt, percentage with debt).
- **uom**: Unit of measure (Percentage or Dollars).
- **value_dollars**: Debt value in dollars.

- **value_percent**: Debt value as a percentage.
- **normalized_value_dollars**: Normalized dollar values for comparisons.
- **normalized_value_percent**: Normalized percentage values for comparisons.

A.1.5 5. Data Limitations and Considerations

- **Accuracy**: Data is accurate as of 2020. Newer data is needed for more accurate analysis.

A.2 Additional Model Information

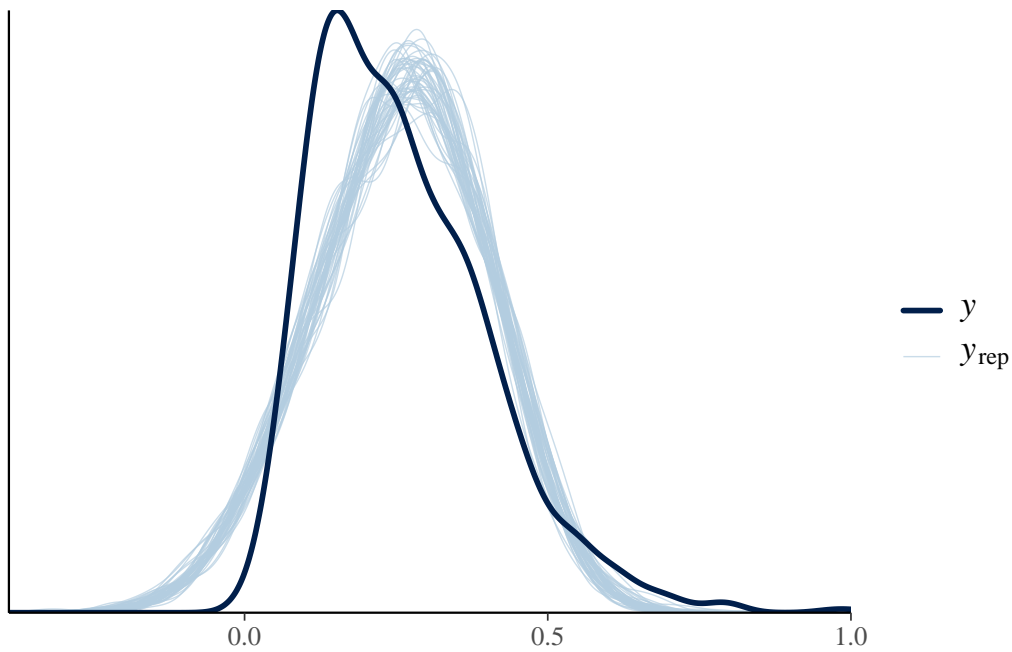### A.2.1 Dollar-Based Debt Model Posterior Predictive Checks

Figure 7: Dollar Model Diagnostics and Performance Metrics

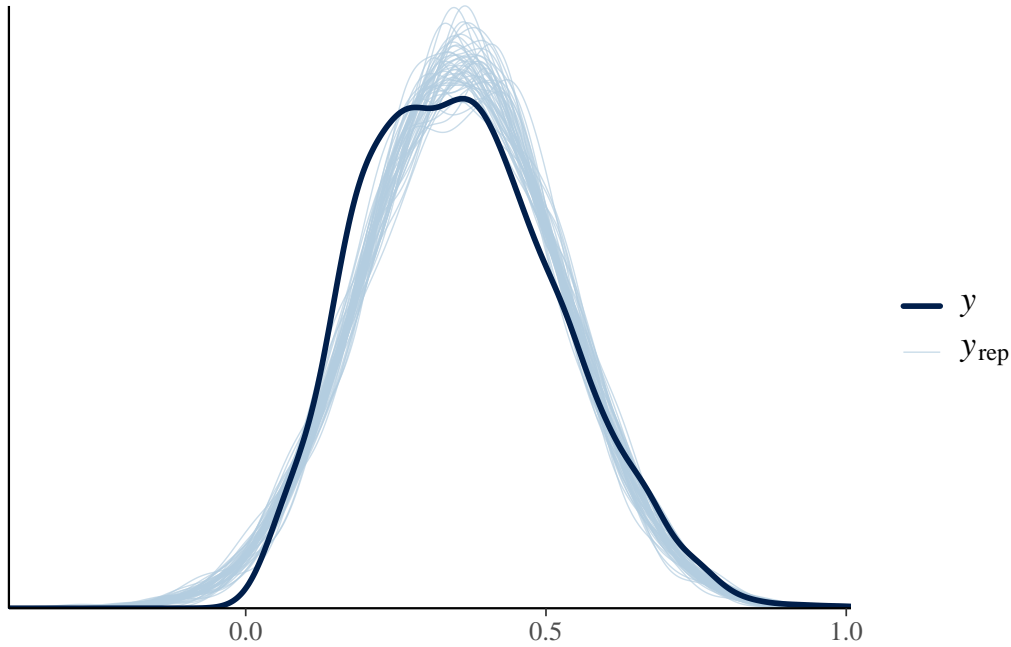### A.2.2 Percent-Based Debt Model Posterior Predictive Checks

Figure 8: Percent Model Diagnostics and Performance Metrics

References

Canada, Gov. 2017. "CLSP Annual Report 2014-2015." https://publications.gc.ca/collections/collection_2017/edsc-esdc/HS45-2015-eng.pdf.

———. 2018. "CLSP Annual Report 2015-2016." https://publications.gc.ca/collections/collection_2018/edsc-esdc/HS45-2016-eng.pdf.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Goodrich, Ben et al. 2024. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://CRAN.R-project.org/package=rstanarm.

Press, The Canadian. 2017. https://www.cbc.ca/news/politics/liberals-forgive-student-loans-again-1.3982728.

Richardson, Neal et al. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Vehtari, Aki et al. 2023. *Loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models*. https://CRAN.R-project.org/package=loo.

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

———. 2023. *Testthat: Unit Testing for r*. https://CRAN.R-project.org/package=testthat.

Wickham, Hadley et al. 2023. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https:

//CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, et al. 2023. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.