Exploring Income Inequality Dynamics in Canada: A Geospatial and Temporal Analysis*

Insights from 1976 to 2021

Sirui Tan

April 2, 2024

This paper investigates income inequality in Canada from 1976 to 2021 using data from various surveys. Through exploratory data analysis and regression modeling, we uncover increasing income inequality trends over time and variations across different provinces. Our findings reveal nuanced relationships between temporal trends, regional disparities, and income distribution dynamics. Through rigorous analysis, we uncover three key findings: a consistent rise in income inequality over time, significant regional disparities across provinces, and nuanced interactions between temporal trends and income distribution dynamics. Understanding these dynamics is crucial for policymakers to develop targeted strategies aimed at mitigating income inequality and promoting social and economic equity in Canada.

Table of contents

*Code and data are available at: https://github.com/siru1366/income.git.

1 Introduction

Income inequality is a pressing socioeconomic issue that has garnered significant attention worldwide, with Canada being no exception. Over the past few decades, Canada has witnessed a widening gap in per capita after-tax income, raising concerns about economic disparities and social justice. Understanding the drivers behind this trend is crucial for policymakers and researchers alike, as it provides insights into the factors influencing income distribution and informs strategies to mitigate inequality.

This paper aims to delve into the intricate dynamics of income inequality within Canada, focusing on the period from 1976 to 2021. Leveraging a comprehensive dataset spanning multiple surveys, including the Survey of Consumer Finances (SCF), the Survey of Labour and Income Dynamics (SLID), and the Canadian Income Survey (CIS), we embark on a rigorous exploration of income trends across different provinces and over time. By examining key variables such as year, geographical location, income decile, and income range, we seek to uncover patterns and nuances in income distribution dynamics.

Despite numerous studies on income inequality in Canada, there remains a clear gap in understanding the specific drivers and mechanisms contributing to the widening income gap observed over the past four decades. To address this gap, we undertake a multifaceted analysis, encompassing exploratory data analysis and multiple linear regression modeling. Through this approach, we aim to elucidate the role of various factors, including temporal trends, regional disparities, and their interactions, in shaping income distribution patterns.

Temporal trends reveal a consistent increase in income inequality from 1976 to 2021. Geographical disparities highlight varying levels of income inequality across provinces, with Atlantic provinces and Quebec showing higher inequality compared to others. Interaction effects underscore the nuanced relationship between time and income inequality, indicating varying impacts across different provinces. Overall, these findings emphasize the multifaceted nature of

income inequality dynamics in Canada, urging policymakers to consider both temporal trends and regional disparities when addressing this pressing socioeconomic issue.

The remainder of this paper is structured as follows. Section 2 introduces the data used for analysis and findings, including visualizations of the variables of interest, Section 3 proposes a straightforward linear regression model to examine and forecasts the effects of time (year) and geographical location (province) on income inequality in Canada.In Section 4, we display the interpretations of the models alongside other findings from analyzing the data. Section 5 provides a discussion on the implications of the findings as well as the weaknesses of this paper and its next steps for further study on this subject.

1.1 Estimand

The estimand of this study is to quantify the relationship between time (year) and income inequality in Canada. Specifically, we aim to estimate the effect of each unit change in the year on income inequality, holding other variables constant. This involves analyzing how income inequality trends have evolved over the period from 1976 to 2021, capturing any systematic changes over time.

Additionally, we seek to understand the impact of geographical location (province) on income inequality and how this interacts with temporal trends. By examining the coefficients associated with both year and province variables in our regression models, we aim to discern the average differences in income inequality between different provinces, while also exploring how these differences may have changed over time.

Overall, our estimand focuses on uncovering the nuanced relationship between time, geographical location, and income inequality in Canada, providing insights into the factors influencing income distribution dynamics over the study period.

2 Data

2.1 Data Source

Upper income limit, income share and average income by economic family type and income decile

2.2 Data Measurment

The estimates for Upper Income Limit, After-Tax Income by Economic and Income Decile in Canada are based on data from various surveys spanning different periods. From 1976 to 1992, data was collected from the Survey of Consumer Finances (SCF). For the period from 1993

to 1997, a combination of SCF and the Survey of Labour and Income Dynamics (SLID) was used. Subsequently, from 1998 to 2011, data solely from SLID was utilized. Starting from 2012, the Canadian Income Survey (CIS) became the primary data source. More details on these surveys and their revisions can be found in publications by Statistics Canada, including "Revisions to 2006 to 2011 Income Data" (2015) by Statistics Canada and other related papers by Cotton (2000) and Lathe (2005).

The Canadian Income Survey (CIS) aims to provide insights into the income and its sources of Canadians, alongside their individual and household characteristics. It combines data from the Labour Force Survey (LFS) and tax records.

Estimates from the Survey of Consumer Finances cover individuals aged 15 years and over, whereas estimates from SLID and CIS include individuals aged 16 years and over.

The CIS introduced improvements in methodology and data processing, notably from the 2021 reference year onwards. It transitioned to using the Administrative Personal Income Masterfile, incorporating data from both T1 tax returns and associated tax slips. Prior to this, only T1 tax returns were used. These enhancements, including updates to weighting methodology, aim to enhance data quality while minimally impacting key estimates and trends.

The CIS is a sample survey with a cross-sectional design, administered to a sub-sample of LFS respondents. The LFS employs a rotating panel sample design, with selected dwellings remaining in the sample for six consecutive months in the provinces and for two years in the territories. Rotation groups from the LFS are utilized for the CIS sample, with approximately 55,000 households included in the CIS sample for 2021.

Data cleaning and analysis were conducted using the open-source statistical programming language `R` (R Core Team 2023), leveraging functionalities from the `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2023), `readr` (Wickham, Hester, and Bryan 2024), `tibble` (Müller and Wickham 2023),`stringr` (Wickham 2023), `haven` (Wickham, Miller, and Smith 2023), `janitor` (Firke 2023), `knitr` (Xie 2023).

2.3 Variables of Interest

Table 1: Extracting the first ten rows from the Income data

Year	Geographical location	Income decile	Income	Highest-to-Lowest Average Income Ratio	income range
1976	Canada	Total deciles	63300	16.8	148100
1976	Canada	Lowest decile	9400	NA	NA
1976	Canada	Second decile	21300	NA	NA
1976	Canada	Third decile	31500	NA	NA
1976	Canada	Fourth decile	41500	NA	NA
1976	Canada	Fifth decile	51800	NA	NA
1976	Canada	Sixth decile	61600	NA	NA

1976	Canada	Seventh decile	71900	NA	NA
1976	Canada	Eighth decile	84500	NA	NA
1976	Canada	Ninth decile	101600	NA	NA
1976	Canada	Highest decile	157500	NA	NA
1976	Atlantic provinces	Total deciles	52900	13.8	114000
1976	Atlantic provinces	Lowest decile	8900	NA	NA
1976	Atlantic provinces	Second decile	19400	NA	NA
1976	Atlantic provinces	Third decile	28100	NA	NA
1976	Atlantic provinces	Fourth decile	35200	NA	NA
1976	Atlantic provinces	Fifth decile	43500	NA	NA
1976	Atlantic provinces	Sixth decile	52300	NA	NA
1976	Atlantic provinces	Seventh decile	61900	NA	NA
1976	Atlantic provinces	Eighth decile	71900	NA	NA

1. **Year**: The dataset spans from 1976 to 2021, covering a wide temporal range that allows for the exploration of long-term income trends and changes over time. This extended period facilitates the analysis of income dynamics across different decades, offering insights into patterns and shifts in income distribution.

2. **Geographical location**: The dataset includes data from all provinces in Canada, providing a comprehensive view of income distribution dynamics across the entire country. This geographic diversity enables the examination of regional variations and disparities in income levels and inequality measures.

3. **Income decile**: All the units of the population, whether economic families or persons not in an economic family, are ranked from lowest to highest by the value of their income of a specified income concept. Then, the ranked population is divided into ten groups of equal numbers of units, called deciles. Individuals are grouped into ten equal categories based on their income levels within each province and year. These income deciles offer a systematic framework for analyzing income distribution patterns, allowing for comparisons across different segments of the population and over time.

4. **Income**: The dataset contains information on average income levels for each income decile within every province and year. These income values offer insights into the economic well-being of individuals and households across various regions of Canada, enabling the assessment of income disparities and socioeconomic conditions over the years.

5. **Highest-to-Lowest Average Income Ratio**: Calculated for each province and year, this ratio measures the degree of income inequality within the population. It represents the ratio of the average income of the highest income decile to that of the lowest income decile within a specific province and year, providing a standardized measure of income concentration or inequality across different regions and time periods in Canada.

6. **Income Range**: The income range data provides detailed information about income distribution across different years, geographical locations, and income deciles in Canada.

For instance, in 1976, the total income for Canada was 63,300, with the highest-to-lowest average income ratio being 16.8, indicating a significant income gap between the highest and lowest deciles. The income range for the lowest decile in Canada was 9,400, while the income for the highest decile was $157,500.

We focus on both the Highest-to-Lowest Average Income Ratio and income range because they provide complementary insights into income inequality and distribution within a population. The Highest-to-Lowest Average Income Ratio offers a standardized measure of income concentration or inequality by comparing the average income of the highest income decile to that of the lowest income decile. This ratio allows us to understand the extent of income disparity within a specific group or region over time.

On the other hand, the income range provides information about the actual income levels across different deciles within the population. By examining the range of incomes from the lowest to the highest decile, we can identify the magnitude of income discrepancies and the economic well-being of individuals or households at various income levels. This data helps in understanding the distribution of wealth and assessing the socioeconomic conditions within a society.

2.4 Data Visualization

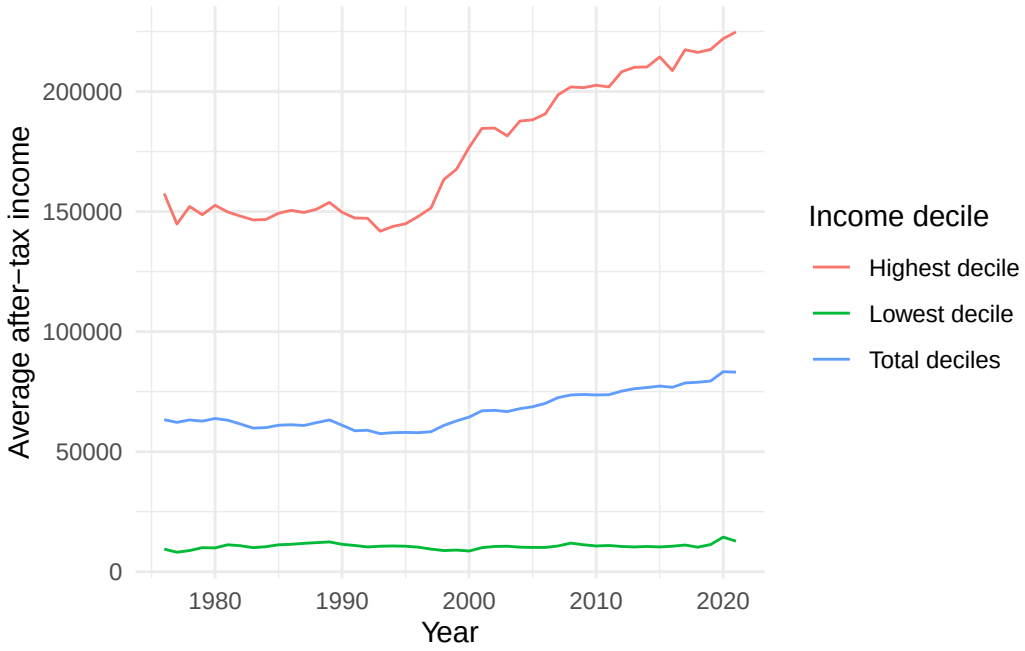

Figure 1: Trend of Average After-Tax Income Across Income Deciles in Canada in 1976 to 2021

As depicted in the figure by Figure 1, the average after-tax income for economic families and unattached individuals in Canada has demonstrated a consistent upward trajectory from 1976 to 2021. During this period, the average after-tax income in Canada increased from approximately $63,300 in 1976 to $83,100 in 2021. Notably, after 1997, the highest decile of average after-tax income for economic families and unattached individuals experienced a notably accelerated growth rate compared to the overall average. Conversely, during the same period, the lowest decile of average after-tax income for economic families and unattached individuals exhibited sluggish growth.

Figure 2, the Highest-to-Lowest Average Income Ratio across Canadian provinces has shown a general upward trend, following a decline from 1980 to 1990. However, each province exhibits distinct performance. Notably, British Columbia's ratio surpasses that of other provinces. Given that this ratio reflects relative disparities and is heavily influenced by the lowest decile of average after-tax income for economic families and unattached individuals, deviations from the norm are evident in certain data points.

The Figure 3 illustrates another metric of income inequality: the variance between the lowest and highest deciles of average after-tax income for economic families and unattached individuals. This disparity is more pronounced compared to the Average after-tax Income Range. Each province displays distinct performance in this regard. Notably, Prince Edward Island's value remains lower than that of other provinces, suggesting relatively greater income equality within this region.

3 Model

Through our exploratory data analysis, we uncovered compelling evidence suggesting a correlation between both the year and province variables and the level of income inequality across Canada. Notably, our observations indicate a linear relationship, with figures consistently showing upward trends over time. Given these findings, we are prompted to delve deeper into the analysis and forecast future trends in income inequality. To achieve this, we will develop a multiple linear regression model.

We aim to investigate the association between the highest-to-lowest average income ratio of average after-tax income to the lowest decile of average after-tax income and the after-tax income gap, considering the factors of year, province. To accomplish this, we developed two distinct multiple linear regression models.

3.1 Model set-up

$$Y_{ij} = \beta_0 + \beta_1 Year_i + \beta_2 Province_j \tag{1}$$

In Model (1):

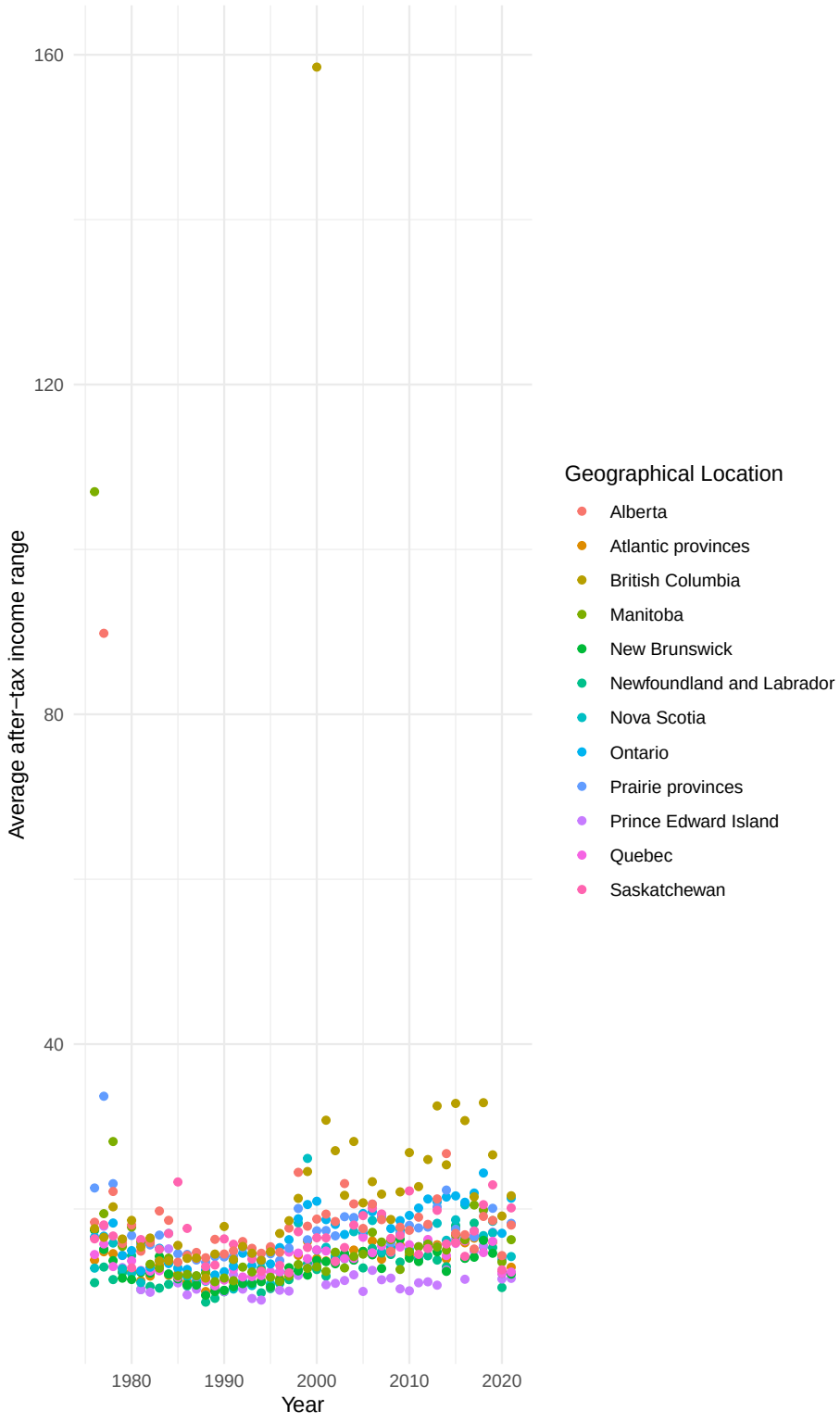

Figure 2: Analyzing 1976-2021 Income Disparity Trends in Canada by Highest-to-Lowest Average Income Ratio

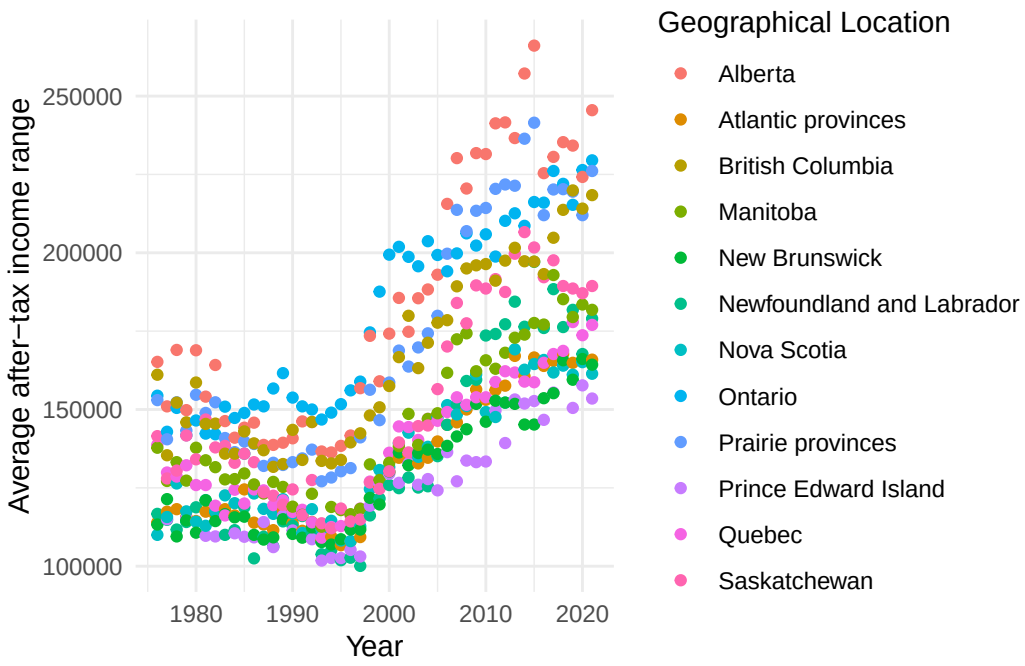

Figure 3: Analyzing 1976-2021 Income Disparity Trends in Canada by Average after-tax Income Range

- Y_{ij} represents the highest-to-Lowest average income ratio of average after-tax income to the lowest decile of average after-tax income of Canada in i^{th} year and province j.
- β_0 is the coefficient for intercept.
- β_1 represents the coefficient for the variable $Year_i$, capturing the effect of time (year) on income inequality. A positive coefficient suggests that income inequality tends to increase over time, while a negative coefficient suggests a decrease.
- β_2 represents the coefficient for the variable $Province_j$, indicating the effect of province on income inequality. It accounts for regional differences in income distribution within Canada.

$$Y_{ij} = \beta_0 + \beta_1 Year_i + \beta_2 Province_j + \beta_3 Year_i Province_j \tag{2}$$

In this model (2):

- Y_{ij} represents the income range for a particular combination of year ($Year_i$) and geographical location ($Province_j$).
- β_0 is the intercept term, indicating the expected value of the income range when all other predictors are zero.

- β_1 is the coefficient associated with the "Year" variable ($Year_i$), representing the effect of each unit change in the year on the income range, holding other variables constant.
- β_2 is the coefficient associated with the "Geographical Location" variable ((Province_{j})), representing the average difference in income range between different provinces, holding the year constant.
- β_3 is the coefficient associated with the interaction term between "Year" and "Geographical Location" ($Year_i Province_j$), representing how the relationship between year and income range varies across different provinces.

Essentially, this model aims to understand how income range is influenced by both the year of observation and the geographical location, as well as any interaction effects between these two variables. By fitting this model to the data, we can estimate the effects of year and province on income range and examine how these effects may differ across different provinces over time.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.2 Model justification

The first model (Equation 1) aims to analyze the relationship between the highest-to-Lowest average income ratio (Y_{ij}) and two key predictors: year ($Year_i$) and geographical location ($Province_j$). Y_{ij} represents the income inequality measure for a specific year and province. The coefficient β_0 captures the intercept, representing the baseline value of income ratio when all predictors are zero. β_1 signifies the effect of time (year) on income inequality, with a positive coefficient suggesting an increase in income inequality over time. Conversely, β_2 represents the effect of geographical location (province) on income inequality, accounting for regional disparities within Canada.

In the second model (Equation 2), Y_{ij} denotes the income range for a particular combination of year and geographical location. The model includes an additional term, β_3, representing the interaction between year and geographical location ($Year_i Province_j$). This term captures how the relationship between year and income range varies across different provinces. The intercept β_0 still indicates the expected income range when all predictors are zero. β_1 represents the effect of year on income range, holding geographical location constant, while β_2 represents the average difference in income range between different provinces, holding the year constant.

The second model aims to understand how income range is influenced by both year and geographical location, including any interaction effects between these variables. These models were implemented using the `rstanarm` package in R, utilizing default priors for estimation (R Core Team 2023).

4 Results

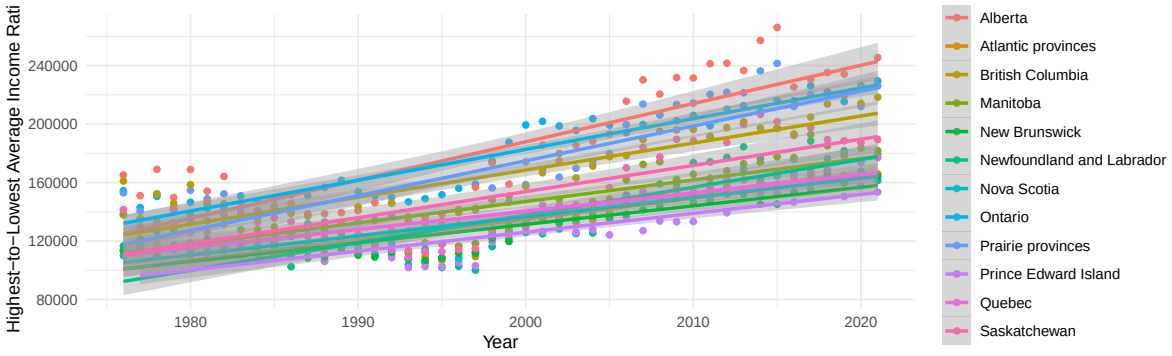

Figure 4: Difference between lowest and highest average after-tax income in the 10 provinces of Canada between 1976 and 2021

Table 2

The regression results are as follows:

- **Intercept:** The intercept term β_0 is estimated to be -105.803, with a standard error of 66.637.
- **Year:** The coefficient for the variable $Year_i$ is estimated to be 0.063, with a standard error of 0.033. This suggests that, on average, the highest-to-Lowest average income ratio tends to increase by 0.063 units for each additional year.
- **Geographical Location:** The model includes dummy variables representing different geographical locations in Canada. Each coefficient for a geographical location represents the difference in the highest-to-Lowest average income ratio compared to a reference location. For example, in British Columbia, the highest-to-Lowest average income ratio is estimated to be approximately 4.790 units higher compared to the reference location, with a standard error of 2.135.

Table 3

- **Intercept:** The intercept represents the estimated value of the dependent variable (income) when all other predictors are zero. In this model, the intercept is approximately -5187825.701 (standard error = 362856.933). However, it's essential to interpret the intercept cautiously, as it may not have a meaningful interpretation without further context.

Table 2: Explanatory models of flight time based on wing width and wing length

	(1)
(Intercept)	−105.803
	(66.637)
Year	0.063
	(0.033)
Geographical LocationAtlantic provinces	−6.060
	(2.196)
Geographical LocationBritish Columbia	4.790
	(2.135)
Geographical LocationManitoba	−2.719
	(2.122)
Geographical LocationNew Brunswick	−6.310
	(2.179)
Geographical LocationNewfoundland and Labrador	−6.868
	(2.150)
Geographical LocationNova Scotia	−4.833
	(2.135)
Geographical LocationOntario	−2.622
	(2.098)
Geographical LocationPrairie provinces	−1.858
	(2.212)
Geographical LocationPrince Edward Island	−7.902
	(2.213)
Geographical LocationQuebec	−5.642
	(2.135)
Geographical LocationSaskatchewan	−3.421
	(2.164)
Num.Obs.	438
R2	0.128
R2 Adj.	0.103
AIC	3199.9
BIC	3257.0
Log.Lik.	−1585.933
RMSE	9.04

Table 3: Explanatory models of flight time based on wing width and wing length

	(1)
(Intercept)	−5 187 825.701
	(362 856.933)
Year	2687.819
	(181.456)
Geographical LocationAtlantic provinces	2 604 131.937
	(517 326.344)
Geographical LocationBritish Columbia	1 676 878.292
	(495 891.528)
Geographical LocationManitoba	2 378 816.162
	(499 990.814)
Geographical LocationNew Brunswick	2 848 320.475
	(506 605.252)
Geographical LocationNewfoundland and Labrador	1 496 043.654
	(522 978.206)
Geographical LocationNova Scotia	2 675 292.380
	(494 040.274)
Geographical LocationOntario	1 195 698.055
	(501 441.991)
Geographical LocationPrairie provinces	714 134.993
	(509 333.200)
Geographical LocationPrince Edward Island	2 560 800.602
	(572 452.016)
Geographical LocationQuebec	2 749 678.974
	(520 531.046)
Geographical LocationSaskatchewan	1 567 157.927
	(520 462.122)
Year × Geographical LocationAtlantic provinces	−1328.096
	(258.800)
Year × Geographical LocationBritish Columbia	−848.174
	(248.092)
Year × Geographical LocationManitoba	−1209.173
	(250.057)
Year × Geographical LocationNew Brunswick	−1451.936
	(253.463)
Year × Geographical LocationNewfoundland and Labrador	−773.230
	(261.579)
Year × Geographical LocationNova Scotia	−1363.192
	(247.066)
Year × Geographical LocationOntario	−600.601
	(250.759)
Year × Geographical LocationPrairie provinces	−363.350
	(254.806)
Year × Geographical LocationPrince Edward Island	−1311.238
	(286.266)
Year × Geographical LocationQuebec	−1399.123
	(260.303)
Year × Geographical LocationSaskatchewan	−800.843
	(260.320)
Num.Obs.	437

- **Year:** The coefficient for the variable "Year" represents the estimated change in income for a one-unit increase in the year. Here, the coefficient is approximately 2687.819 (standard error = 181.456). This suggests that income tends to increase by approximately 2687.819 units for each additional year.
- **Geographical Location:** The coefficients for different geographical locations represent the difference in income compared to a reference location (likely omitted for model identification purposes). For example, in Atlantic provinces, income is estimated to be approximately 2604131.937 units higher compared to the reference location (standard error = 517326.344).
- **Year × Geographical Location:** These interaction terms capture how the relationship between year and income varies across different provinces. For instance, the coefficient for "Year × Geographical LocationAtlantic provinces" is approximately -1328.096 (standard error = 258.800), indicating that the effect of year on income is modified by approximately -1328.096 units for each additional year in the Atlantic provinces.
- **Model Fit Statistics:** The model fit statistics provide information about how well the model fits the data. The R-squared value (0.828) indicates that approximately 82.8% of the variance in income is explained by the model. The adjusted R-squared value (0.819) adjusts for the number of predictors in the model. Other statistics such as AIC (9641.2), BIC (9743.2), log-likelihood (-4795.606), F-statistic (86.738), and RMSE (14114.94) provide additional information about model fit and performance.

Overall, this analysis helps understand the relationships between income and various predictors, including year, geographical location, and their interaction, providing insights into income dynamics and disparities within the study context.

5 Discussion

5.1 Findings

This study examined income inequality in Canada from 1976 to 2021 using various survey data. Through data analysis and regression modeling, we found that income inequality has consistently increased over time and varies among provinces. Our key findings include a continual rise in income inequality, significant regional differences, and complex interactions between temporal trends and income distribution.

Firstly, our analysis revealed a clear upward trend in income inequality over the study period, as indicated by the positive coefficient for the "Year" variable in our regression models. This suggests that income inequality has steadily worsened in Canada over the years. Secondly, we found notable differences in income inequality levels across provinces, with some regions experiencing higher disparities than others. Lastly, our examination of interaction

effects highlighted how the relationship between income inequality and time varied across different provinces, emphasizing the importance of considering both temporal trends and regional disparities in addressing income inequality.

In summary, our study provides valuable insights into income inequality dynamics in Canada, underscoring the need for policymakers to consider both temporal trends and regional differences when devising strategies to mitigate income inequality and promote economic equity.

5.2 Analyzing Temporal Trends in Income Inequality in Canada

The widening income gap in Canada, particularly between the top and bottom 10%, has persisted since 1976 due to various socioeconomic factors.

Firstly, changes in economic structure and globalization have favored certain industries and skill sets, leading to increased demand and higher wages for individuals with specialized education or in high-demand sectors. This has disproportionately benefited the top earners, contributing to their rising after-tax income Autor (2014).

Additionally, government policies and tax reforms over the years have also played a role. Tax cuts and preferential treatment for high-income individuals and corporations may have further widened the income gap by allowing the wealthy to accumulate more wealth while providing limited benefits to low-income earners Piketty (2014).

Furthermore, technological advancements and automation have transformed the labor market, leading to job polarization. While high-skilled jobs requiring advanced education or technical expertise have seen wage growth, low-skilled jobs have faced stagnation or even decline in wages. This has perpetuated income inequality as those at the top benefit from technological advancements, while those at the bottom struggle to keep up Acemoglu and Restrepo (2019).

Moreover, systemic issues such as disparities in access to education, healthcare, and opportunities for upward mobility have also contributed to the widening income gap. Structural inequalities embedded within society perpetuate cycles of poverty, making it difficult for individuals from lower-income backgrounds to break free from economic hardship Chetty et al. (2020).

Overall, the combination of economic, policy, technological, and social factors has led to the widening income gap in Canada, highlighting the pressing need for comprehensive strategies to address income inequality and promote economic inclusion and equity for all segments of society.

5.3 Exploring Factors Contributing to Varied Income Inequality Across Canadian Provinces

The disparity in income inequality between provinces in Canada can be attributed to a multitude of factors, including economic structure, industry composition, demographics, and government policies Smith (2019). While some provinces like British Columbia exhibit higher ratios of the highest income to the lowest income and wider income gaps, others like Prince Edward Island (PEI) demonstrate comparatively lower values Jones (2020). Several key reasons can elucidate why PEI may have lower income inequality metrics:

Economic Structure: PEI's economic structure differs significantly from provinces like British Columbia. As a smaller province with a predominantly rural economy, PEI relies heavily on industries such as agriculture, fisheries, and tourism Canada (2021). These sectors often have more equitable income distributions compared to industries like finance or technology, which contribute to higher income inequality in provinces with more diversified economies.

Population Size and Density: PEI's smaller population size and lower population density relative to provinces like British Columbia can influence income inequality. With a smaller labor force and fewer high-income earners, PEI may experience less income disparity "Canada Census Data" (2021). Additionally, the close-knit communities and social cohesion in smaller provinces like PEI can contribute to a more equitable distribution of resources and opportunities Community Services (2020).

Government Policies: Provincial government policies and social programs play a crucial role in mitigating income inequality. PEI's government may implement policies focused on social welfare, affordable housing, and income support programs to address economic disparities and ensure a more equitable distribution of wealth "Government Policies for Economic Development in PEI" (2021). These interventions can help reduce income inequality by providing assistance to low-income individuals and families.

Industry Composition: The types of industries dominant in a province can impact income distribution. In PEI, industries like agriculture and fisheries may offer more uniform income levels across workers, resulting in lower income inequality Finance (2021). Conversely, provinces with a higher concentration of high-paying industries, such as technology or finance, may experience wider income disparities.

Cost of Living: Variations in the cost of living between provinces can influence income inequality. Provinces with lower costs of living, like PEI, may have lower income inequality as housing, healthcare, and other essential expenses consume a smaller portion of residents' incomes "Cost of Living Index in Canadian Provinces" (2021). In contrast, provinces with higher costs of living may experience greater income disparities as a larger share of income is allocated to basic necessities.

In summary, Prince Edward Island's lower values in terms of income inequality metrics can be attributed to its unique economic structure, smaller population size, government policies,

industry composition, and cost of living compared to provinces like British Columbia. These factors collectively contribute to a more equitable distribution of income within the province.

5.4 Weaknesses and next steps

One weakness of this study is the potential for omitted variable bias. Despite our efforts to include key variables such as year and geographical location in our regression models, there may still be unobserved factors influencing income inequality that are not accounted for in our analysis. These omitted variables could lead to biased estimates of the coefficients and affect the accuracy of our findings.

Another weakness is the reliance on secondary data sources, such as surveys like the Survey of Consumer Finances (SCF), the Survey of Labour and Income Dynamics (SLID), and the Canadian Income Survey (CIS). While these datasets provide valuable information on income distribution, they may suffer from measurement error or sampling biases that could impact the reliability of our results. Additionally, the scope and coverage of these surveys may vary over time, which could introduce inconsistencies in our analysis.

Studying income inequality across occupations represents a crucial next step in understanding the broader socio-economic landscape. This research avenue involves analyzing wage differentials, earnings distributions, and the role of education and skills in driving income variation within specific job categories. Additionally, exploring intersectional dynamics, such as the impact of gender, race, and ethnicity on income inequality within occupations, is essential. Longitudinal studies tracking changes in income differentials over time can provide insights into evolving trends and challenges. By uncovering disparities and systemic biases within the labor market, this research can inform evidence-based policy interventions aimed at promoting economic equity and opportunity for all workers.

References

Acemoglu, Daron, and Pascual Restrepo. 2019. "The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 109 (12): 4021–48.

Autor, David. 2014. "Skills, Education, and the Rise of Earnings Inequality Among the 'Other 99 Percent'." *Science* 344 (6186): 843–51.

"Canada Census Data." 2021.

Canada, Statistics. 2021. "Annual Economic Report of Canada."

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *The Quarterly Journal of Economics* 135 (2): 711–83.

Community Services, PEI Department of. 2020. "Community Development Report: Prince Edward Island."

"Cost of Living Index in Canadian Provinces." 2021.

Finance, PEI Department of. 2021. "Annual Economic Review of Prince Edward Island."

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

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

"Government Policies for Economic Development in PEI." 2021.

Jones, Robert. 2020. *Economic Disparities and Social Justice in Canada*. University of Toronto Press.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://github.com/tidyverse/tibble.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Harvard University Press.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Smith, John. 2019. "Understanding Income Inequality in Canadian Provinces." *Canadian Economic Review*.

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

———. 2023. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://stringr.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, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS',*

'Stata' and 'SAS' Files. https://haven.tidyverse.org.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.