Are grocery prices attainable?*

Analysing grocery prices through an Affordability Index

Cristina Burca

December 3, 2024

Abstract

We analyse the probability of Kamala Harris winning in the 7 swing states of USA using baysian modeling;the states we look at are Arizona, Michigan, Pennsylvania, Nevada, Georgia, Wisconsin and North Carolina. USA is a large country with its national economy affecting the global economic conditions, which is why predicting the future President of USA will help understand the future economic condiotion of the world. Through baysian modelling we found that North Carolina (47.26% support for Harris), Nevada(46.43% support for Harris), Wisconsin(48.38% support for Harris), Michigan(47.30% support for Harris), Pennsylvania(48.02% support for Harris) will vote Kamala Harris, while the rest of the swing states might vote Donald Trump, meaning that majority of the swing states will vote Kamala Harris. Provided this we predict that Kamala Harris will win the election. By creating a baysian model based on 'polls of polls' where we compare the results from different polls, we are able to make this prediction.

Table of contents

1 Introduction **2**

2 Data Overiview **3**

2.1 Citations . . . 3

2.2 Variables . . . 3

2.3 Data Analysis . . . 4

2.4 Methedology and Measurement . . . 6

2.5 Outcome Variables . . . 6

2.6 Predictor Variables . . . 7

3 Model **7**

3.1 Model Setup . . . 7

3.1.1 Model Justification . . . 8

3.2 Results . . . 8

*Code and data are available at: Are Groxeries Attainable.

4 Discussion **9**

4.1 Model Discussion 9

4.2 Projected Affordibility Index 12

4.3 Limitations 12

4.4 Next Steps 13

5 Appendix **14**

5.1 Data Cleaning 14

5.2 Methodological Exploration of Data Collection 14

5.2.1 Grocery Data Collection And CPI Calculation 14

5.2.2 Wage Data Collection 16

5.3 Linear Model Summary Statistics 18

5.4 Postrior Predictive Check 18

1 Introduction

Inflation in Canada has increased significantly across all sectors, raising concerns about affordability for households. Among these, the rapid rise in grocery prices has emerged as a critical issue. Since May 2020, grocery prices have increased by 22.5% and continue to climb each month, straining the budgets of Canadians and creating widespread economic anxiety ([CTV News]). In this context, the intersection of rising inflation, stagnant wage growth, and sector-specific price surges has made essential groceries increasingly unattainable for many households. This paper explores the relationship between grocery prices, general inflation (CPI), and wages in Canada from 2017 to 2023. Through the development of an Affordability Index, we assess whether wage growth has kept pace with rising grocery prices, and we predict the future affordability of groceries under current economic trends.

Using historical data, this paper analyzez trends in general inflation (CPI) and sector-specific grocery inflation, alongside average wage growth, to evaluate their combined effect on affordability. The Affordability Index, calculated as the ratio of average wages to grocery prices, is used to track affordability trends over time and forecast future affordability challenges. This analysis sheds light on whether wage growth has been sufficient to counteract inflation and maintain stable grocery affordability for Canadians.

The primary aim of this paper is to predict the affordability of groceries in the coming years based on historical trends in grocery prices, inflation, and wages. Specifically, we seek to determine whether steady increases in inflation and wages will be sufficient to sustain grocery affordability for the average Canadian household. The analysis assumes linear growth patterns in wages and inflation, while acknowledging the unpredictability of economic disruptions.

The analysis explains that grocery-specific inflation has consistently outpaced both general inflation and wage growth since 2017. While wages have increased steadily, they have not been sufficient to offset the sharp rise in grocery prices, resulting in a gradual decline in grocery affordability. This finding underscores the disproportionate impact of sector-specific inflation on Canadian households. Projections based on current trends suggest that, without significant policy or economic intervention, affordability will continue to decline in the coming years.

The affordability of groceries is a critical issue for Canadian households, as it directly impacts food security, quality of life, and financial well-being. Rising grocery prices amidst slow wage growth exacerbate economic inequality and place undue pressure on lower-income households. Understanding the drivers of affordability and projecting future trends provides essential insights for policymakers to design targeted interventions that address inflationary pressures and wage stagnation.

The remainder of this paper is structured as follows. Section 2 provides an overview of the data sources, variables, and key trends observed in grocery prices, wages, and inflation. Methodology, measurements and outcome/predictor variables are also described and determined. Section 3 introduces the model and its justification. Section 3.2 presents the model findings, followed by a Section 4 that interprets the results, arriving to a conclusion, as well as discusses implications, limitations, and future research directions. Finally, Section Section 5 describes the data cleaning process, summary tables and checks for the models, and an in depth discussion of the sampling and data collection process of the datasets used in this paper.

2 Data Overiview

2.1 Citations

The datasets used in this analysis are retrieved from Statistics Canada (2024c). The main dataset is acquired from Statistics Canada (2024b), containing grocery prices of set units for each month of the year. This dataset is recorded beginning January 2017, and thus analysis with this data set will only be from 2017. The second dataset from Bank of Canada (2024) contains monthly records of Consumer Price Index (CPI), all-items excluding the effect of indirect taxes. CPI is used as an indicator of change for consumer prices, or in other words, rate of inflation. It is measuring the prices, and thus price difference, of a set number of common groceries and household goods at a set quantity(unit) representing the average Canadian household. The third dataset contains records of yearly wages of Canada from Statistics Canada (2024a). The three datasets were merged together for analysis.

This paper was complied using the prgramming language R (R Core Team (2023)). Data was complied, cleaned, and vizualised using packages Tidyverse (Wickham et al. (2019)), Patchwork (Pedersen (2023)), janitor ((**janitor?**)) and Knitr (Xie (2023)). Models were ran and summarized using rstanarm (Gelman et al. (2023)) and modelsummary (Arel-Bundock (2023)).

2.2 Variables

The important variables that are focused on throughout this paper:

- `Date`: Contains the date of the recording of the observations, formatted as YYYY-MM-DD. Days are set to 01, since daily data is not recorded.
- `CPI`: Contains monthly values of recorded CPI. CPI is a unitless measure, where the average price level in the abse year is set to 100.
- `Wage`: Yearly average wage of all employees from all industries.
- `CPI_Normalized` and `Wage_Normalized`: calculated columns to measure.
- `Average_price`: calculates the average price based off selected items that make up a Canadians' "essential grocery basket".
- `Affordability` : A variable created as an Affordability Index, calculated by Average price divided by Wage.

The Affordability Index is a measure that reflects the ease in which households or individuals can afford a list of essential goods, relative to income– in this context we consider average wage. Higher vales of the index indicate better affordability, while lower values suggest affordability struggles.

Items included in Average_Price include… The total average is a very rough estimate of popular items purchased by Canadians. The average is not particularly inportant, but the impact and correlation between the two variables is.

More about data cleaning in the appendix section appendix Section 5.1.

2.3 Data Analysis

This section provides an overview of the datasets used, the variables of interest, and key trends observed in the data. The analysis focuses on understanding grocery affordability in Canada by evaluating changes in the Consumer Price Index (CPI), grocery prices, and average wages over time.

The first variables examined are `CPI` and `Wage` from 2000 to 2023. In Figure 1, we analyse normalized values of yearly average wages and monthly CPI, as they are measured on different scales. This plot examines the long term relationship between general inflaiton and wages.

The period of 2000 to 2020 exhibits a steady correlation between the two predictors, with predictable inflation rates, suggesting a balanced economic trend. Post-2020, disruptions caused by the COVID-19 pandemic result in sharp CPI growth, outpacing wage increases. This resulted in a decline in demand of goods and services, which is reflected in the CPI. By 2022, CPI rapidly inflates due to increase consumer demand as economies reopened, but wages remain on a somewhat steady increase.

Figure 1: Normalized CPI and Wages Over Time from 2000 to 2023

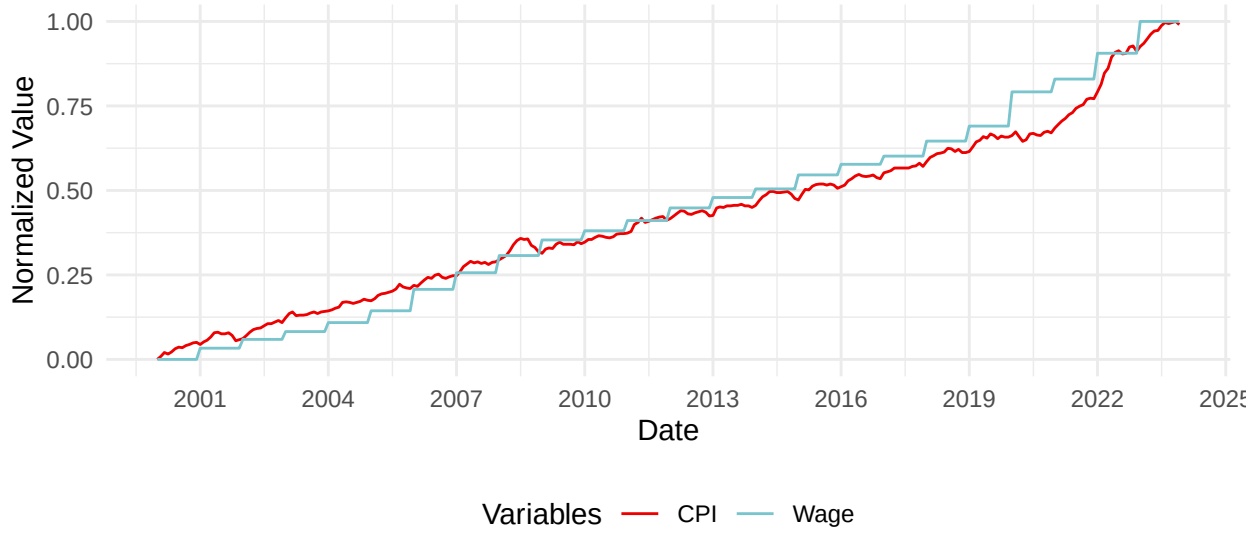


Figure 2 compares grocery-specific inflation with general inflation through normalized values. Both CPI and grocery prices have increased significantly since 2017, but grocery prices exhibit steeper growth (~32% increase compared to ~20% for CPI).

Figure 3 traks changes in the grocery Affordability Index over between 2017 to 2023. The Affordability Index has shown a gradual decline (~4.6% decrease), reflecting rising grocery prices that are not fully offset by wage growth. Temporary improvements in affordability during 2019 and 2021 correspond to slower grocery price growth. This number brings concern about the affordability of groceries in Canada, in addition to other inflating expenses, such as rent and utilities.

The data demonstrates the increasing gap between grocery-specific inflation and general inflation, revealing that grocery prices have risen at a much steeper rate than other consumer goods, particularly since 2020. Wages, while increasing steadily, have not fully kept pace with the rise in grocery prices, contributing to a gradual decline in the Affordability Index—a measure designed to reflect the ease with which households can afford essential groceries relative to average wages.

Table 1 further supports this analysis by quantifying the year-over-year percentage changes in CPI, grocery prices, and wages. Specifically: In 2022, grocery prices increased by 10.7%, significantly outpacing both CPI (6.7%) and wage growth (4.2%), marking a concerning affordability gap. While wage growth greatly increased in 2020 (6.0%), it did not sustain pace in subsequent years, particularly as grocery-specific inflation

Figure 2: Normalized CPI and Wages Over Time from 2017 to 2023

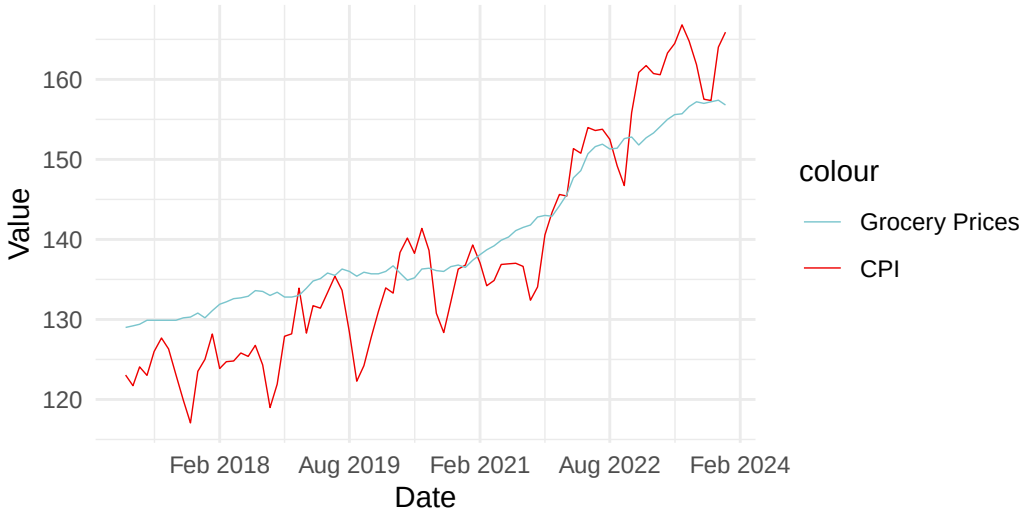
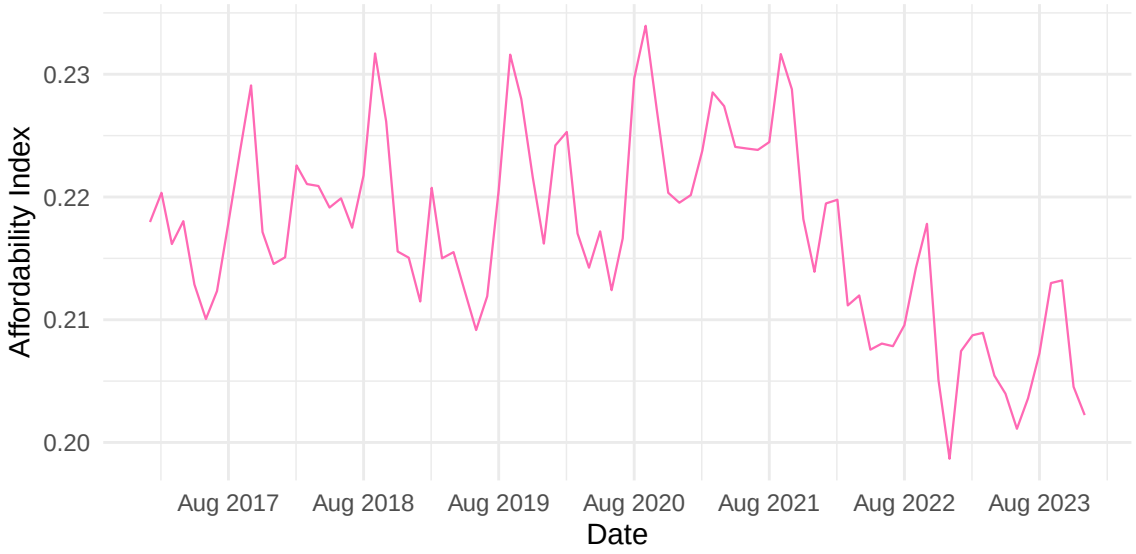


Figure 3: Normalized CPI and Wages Over Time from 2017 to 2023



accelerated in 2022 and 2023. Even in years with relatively low CPI growth, such as 2021 (3.3%), grocery prices exhibited slower growth (0.9%), yet wage increases (2.1%) were insufficient to offset other cumulative inflationary pressures.

The primary objective of the analysis is to assess the relationship between grocery prices, general inflation (CPI), and wages through a predictive linear model. This model serves to quantify the degree to which wage growth offsets inflation in the context of grocery affordability

Table 1: Inflation and Wage Growth Comparison

Year	% Change in CPI	% Change in Grocery Prices	% Change in Wages
2018	2.2%	1.4%	2.8%
2019	1.9%	4.0%	2.7%
2020	0.6%	4.3%	6.0%
2021	3.3%	0.9%	2.1%
2022	6.7%	10.7%	4.2%
2023	3.8%	7.1%	5.0%

2.4 Methedology and Measurement

This study relies on data collected from three primary sources—CPI data, grocery price data, and wage data—each with distinct methodologies and inherent limitations. While the full details of the sampling and data collection processes are provided in Section 5.2, the following key points summarize the methodology relevant to this study.

Grocery price and CPI:

The Consumer Price Index (CPI) is a broad indicator of inflation that tracks price changes across a fixed basket of goods and services, mainly groceries and household goods (Bank of Canada (2024)). Data for CPI is collected monthly, targeting a representative sample of goods, geographic regions, and retail outlets. CPI data prioritizes households in urban and rural areas but excludes specific populations. Representative products are chosen based on popularity, price stability, and their share of total consumer expenditures. Prices are collected from high-revenue retail outlets, flyers, and online sources during the first two weeks of every month. Quality changes in products are accounted for through implicit and explicit methods to ensure the index reflects "pure" price changes over time.

Wage data:

Wage data was obtained from Statistics Canada's Labour Force Survey (LFS), a monthly survey covering approximately 100,000 individuals across Canada ((**wage?**)). Data is collected through interviews and electronic surveys, with one household member often acting as a proxy for others. Weighting techniques correct for non-responses and ensure the final estimates align with census projections for key demographic and regional groups. Imputation methods address missing or inconsistent responses.

2.5 Outcome Variables

The Affordability Index serves as the outcome variable in the proceeding linear model. This index measures the ease with which households can afford a set of essential grocery list described. It is defined as:

$$\text{Affordability Index} \frac{\text{Wage}}{\text{Average Price}},$$

providing a direct measure of purchasing power for groceries, capturing how well wages can offset grocery price inflation. Higher values of the index indicate better affordability (wages adequately covering grocery prices), while lower values signal affordability challenges.

2.6 Predictor Variables

The model includes the following predictor variables to explain variations in the Affordability Index:

- **CPI (Consumer Price Index):** The CPI is a unitless measure that tracks the overall inflation rate of goods and services relative to a base year. CPI is included to capture general inflation trends and their impact on grocery affordability. Although CPI is not specific to groceries, it provides context on broader economic conditions that may influence wages and purchasing power. A negative relationship is expected because since CPI is expected to increase grocery prices, and thus becoming less affordable.
- **Average Price:** The Average Price variable represents the mean cost of the predefined basket of grocery items essential to Canadian households. This variable isolates grocery-specific inflation, enabling an analysis of how rising grocery prices impact affordability. A negative relationship is expected because higher grocery prices reduce purchasing power and affordability.
- **Time:** This is included as a cumulative measure of the number of months since the start of the dataset in January 2017.

These predictors were selected to assess whether wage growth has kept pace with rising grocery prices and how affordability has changed over time.

3 Model

The motivation of the linear model employed is to evaluate the relationship between grocery affordability and its predictors– CPI, Time and Average Price–, to quantify the elasticity of affordability with respect to these predictors. Elasticity reflects the proportional change in affordability in affordability for a 1% change in a given predictor, providing a practical interpretation of how affordability responds to economic changes.

A linear model was chosen for its simplicity and straightforward interpretability of the relationships between variables. Specficially, a log-log linear regression model is chosen as the relationship between affordability and the predictors is expected to be approximately linear in logarithmic terms. The log transformation ensures linearity in log space, stabilize variance, and reduce the impact of outliers. Moreover, elasticity coefficients derived from log-log models simplify the assessment of proportional changes in affordability relative to changes in predictors.

3.1 Model Setup

The model specification is as follows:

$$\log(y_i)|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \beta_0 + \beta_1 \times \log(\text{CPI})_i + \beta_2 \times \log(\text{Average Price})_i + \beta_3 \times \log(\text{Time})_i \tag{2}$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{6}$$

$$\sigma \sim \text{Exponential}(1) \tag{7}$$

Where:

- y_i is the dependent variable, representing the Affordability Index(Wage/ Average Price),

Table 2: Summary statistics of model

	Elasticity Model
(Intercept)	−0.284
	(0.270)
log(CPI)	0.397
	(0.105)
log(Average_Price)	−0.685
	(0.064)
log(Time)	0.024
	(0.003)

- μ_i: is the mean response, a function of the predictors,
- β_0 is the intercept, representing baseline affordability when all predictors are set to zero,
- β_1 the elasticity of affordability with respect to CPI, indicating the proportional change in affordability for a 1% change in the CPI,
- β_2 is the elasticity of affordability with respect to Average Price, representing impact of the changes in average grocery prices on affordability,
- β_3 is a numerical variable of Date turned into a cumulative column of days for each month of the year. Captures the change in μ_i over time.

Each coefficient β_i is assigned a Normal prior $\mathcal{N}(0, 2.5)$ to regularize estimates and prevent overfitting. The residual variance σ is modeled using an Exponential prior to ensure positive values. Each predictor captures an aspect of the Affordability Index:

- **Consumer Price Index (CPI)**: Captures inflation and changes in general cost of goods and services, and the primary determinant of grocery prices, and evidently, the Affordability Index. As CPI increases, affordability is expected to decline.
- **Average Price**: Represent average monthly price of the defined basket in (**datasec?**). Measures the cost of groceries comparatively.
- **Time**: Represented as a cumulative numeric variable corresponding to the number of days since the start of the dataset. Captures temporal change and long term shifts.

The decision to exclude Wage as a predictor stems from the construction of the Affordability Index, which is already defined as the ratio of Wage to Average Price. Including Wage would introduce redundancy and lead to multicollinearity, and produce almost exact estimates.

3.1.1 Model Justification

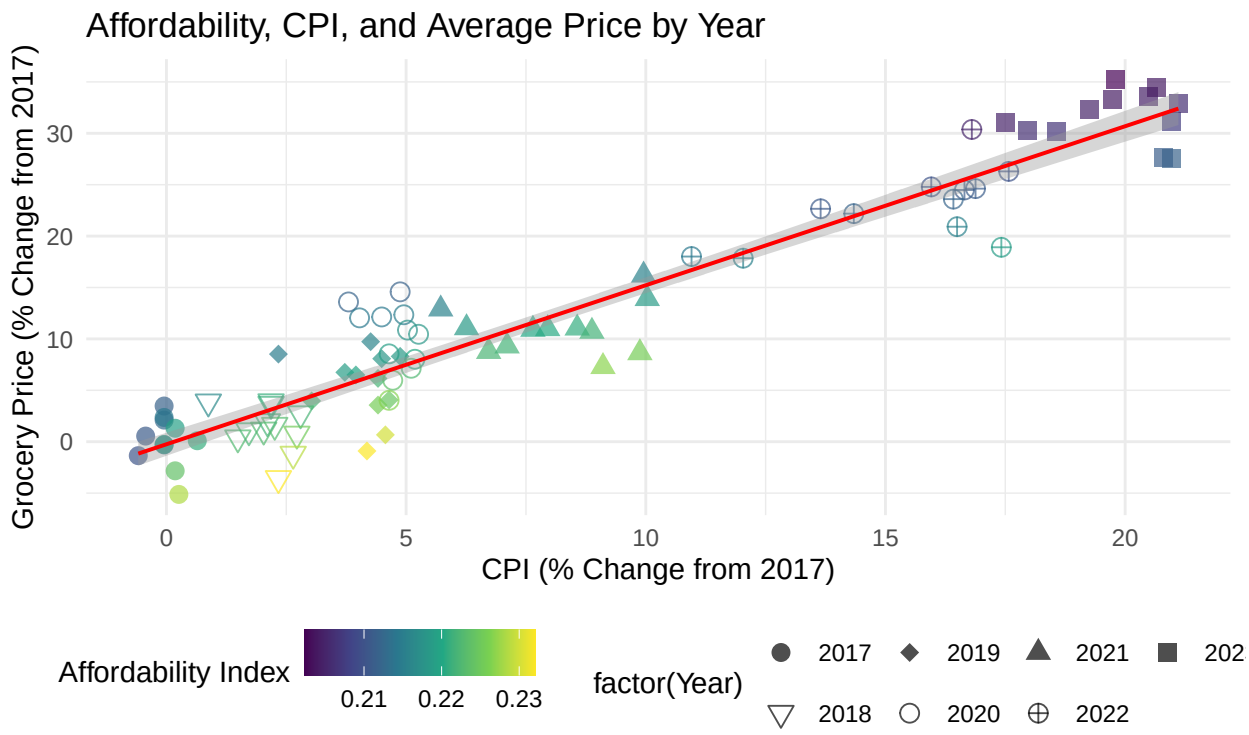
Given these trends, the model is expected to demonstrate a strong negative relationship between grocery prices and affordability, a positive but smaller effect of wage growth on mitigating affordability challenges, and a nuanced role of CPI that reflects its correlation with broader wage adjustments and economic trends.

3.2 Results

Table 2 summarizes the coefficients of the model.

Table 3 provides a summary of yearly averages and log-transformed values for key variables, offering an intuitive connection between the regression model and the observed data. These values will be interpreted further in the discussion section:

Figure 4: Affordability, CPI, and Average Price by Year



4 Discussion

4.1 Model Discussion

Inflation is often a reliable indicator of affordability for goods, including groceries. However, because inflation reflects changes in the prices of a wide variety of goods and services, it does not directly isolate grocery-specfic affordability. By examining the predicted Affordability Index from the linear model, this study assesses whether wage growth has kept pace with rising grocery prices through elastic relationships with CPI, average price, and time.

The model some contrasting relationships between the Affordability Index and its predictors. Firstly, the intercept values of the elasticity model in Table 2 are intereprted as:

- **The intercept (-0.284)** represents the expected log(Affordability) when all predictors are zero, serving as the baseline for the model.
- **log(CPI) coefficient (0.397)**: The positive coefficient states that a 1% increase in CPI is associated with a 0.397% increase in the Affordability Index, holding other predictors constant. While this may seem counter-intuitive —given that inflation typically reduces affordability— it likely reflects a correlation between CPI and wage growth.
- **log(Average_Price) coefficient (-0.685)**: The negative coefficient states that a 1% increase in Average Price is associated with a 0.685% decrease in Affordability, holding other predictors constant. This aligns with expectations, as higher grocery prices directly reduce purchasing power.
- **log(Time) coefficient (0.024)**: The positive coefficient states that a 1% increase in time is associated with a 0.024% increase in the Affordability Index, holding other predictors constant. While small, this positive trends suggests gradual, incremental improvements in affordability over the study period.

The intercept coefficients of the model shows that general inflation (CPI) has a positive, although seemingly contradictory, relationship with grocery affordability, while grocery-specific inflation strongly reduces affordability. Over time, affordability shows a slight but insufficient improvement, highlighting that rising grocery prices are the primary driver of declining affordability.

Table 3: Yearly Averages, Affordability Index, and Percentage Changes with Wage

Year	Avg CPI	Avg Grocery Price	Affordability Index (%)	Percent CPI (%)	Percent Grocery Price (%)	Percent Wage (%)
2017	129.96	123.39	21.75	0.00	0.00	0.00
2018	132.71	125.07	22.05	2.11	1.36	2.80
2019	135.26	130.12	21.78	4.07	5.45	5.59
2020	136.11	135.70	22.15	4.73	9.98	11.97
2021	140.56	136.94	22.41	8.15	10.98	14.35
2022	150.02	151.64	21.09	15.43	22.89	19.16
2023	155.72	162.42	20.66	19.82	31.63	25.09

Considering CPI and Grocery Prices in Table 3, it is apparent that these averages are rising. However, their increase is disproportionate. Between 2017 to 2023, CPI increase by ~20% (from 129.96 to 155.72), while groceries rose by ~32% (from $123.40 to $162.42). The faster rise in grocery prices indicates that grocery-specific inflation is growing faster than general inflation trends of Canada. Even if wages or broader economic adjustments help offset general inflation, they may not adequately address the sharp rise in grocery-specific costs.

The disproportionate growth is reflected in the Affordability Index, which declined from 0.217 in 2017 to 0.207 in 2023, representing a ~4.6% drop in grocery affordability over the six years. While the model shows small improvements in affordability during certain years (e.g., 2019 and 2021), these appear to be temporary and may align with periods of wage growth or slower grocery price inflation. For instance, during the COVID-19 pandemic, restrictions between 2020-2022 limited the sale of goods and services, temporarily lowering Canada's CPI and stabilizing grocery price growth. However, affordability declined sharply in 2022 and 2023 as grocery prices increased, highlighting that wage growth has not kept up with the rising cost of essential goods.

The positive correlation between CPI and affordability suggests that wages or economic adjustments have kept pace with general inflation, preventing broader grocery affordability declines. This indicates that broad inflation is not inherently detrimental to grocery affordability when accompanied by proportional increases in income or government support. In contrast, **the strong negative elasticity of Average Price highlights that grocery-specific inflation is the primary driver of affordability challenges**. Even if wages rise to match general inflation, rising grocery prices can counteract that increase.

The contrasting effects of CPI and Average Price highlight the importance of separating general inflation trends from grocery-specific price trends. The positive correlation between CPI and grocery affordability suggests that wages or other economic compensations may have kept pace with general inflation. This indicates that broad inflation is not inherently detrimental to grocery affordability if accompanied by proportional increases in income or government support. The strong negative impact of grocery prices highlights that grocery-specific inflation is a more significant concern. Even if wages rise to match general inflation, higher grocery prices can still disproportionately reduce affordability.

The faster rise in grocery prices compared to CPI reveals that grocery-specific inflation is growing too rapidly for wages to keep up. As grocery prices increased by 32%, affordability declined, indicating that rising wages or other economic adjustments have been insufficient to maintain purchasing power for essential goods like groceries. This dynamic poses significant challenges for households and policymakers alike.

Figure 5 compares all predictors Affordability Index, Grocery Inflation, Wage Growth, and an additional metric Wage-Price Gap, to measure the gap that interprets if Wage is keeping up with price.

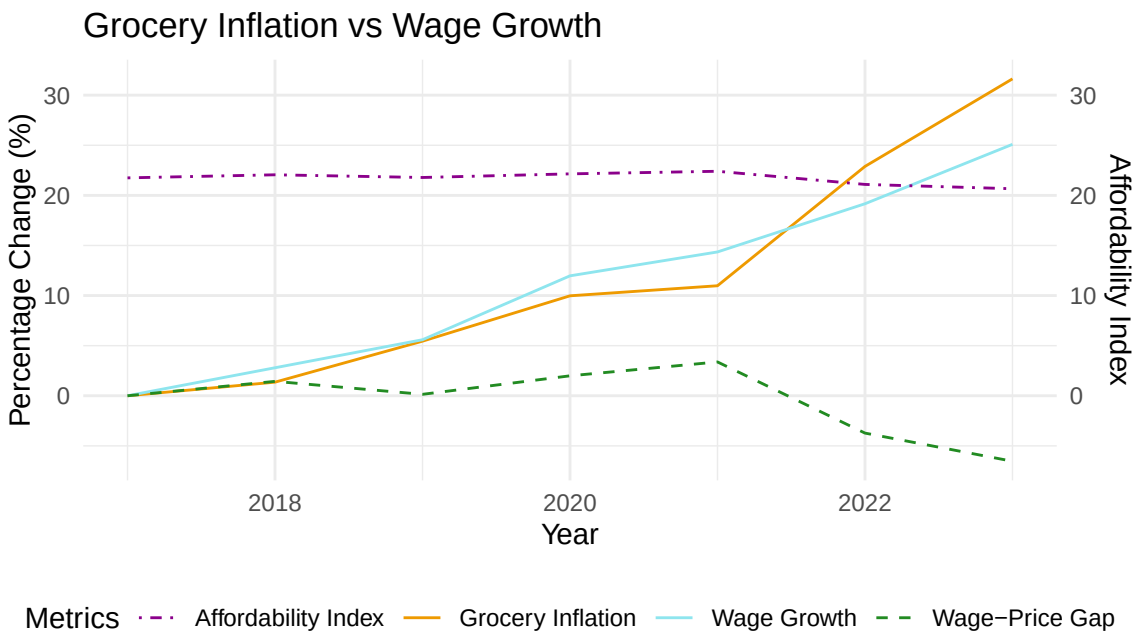


Figure 5: Yap

The sharp upward trend in grocery inflation after 2020 indicates that grocery-specific prices are growing faster than general inflation, particularly during and after the COVID-19 Pandemic. While wages grew steadily over the years, it did not match the rapid rise in grocery prices, particularly after mid-2021. The consistent gap between grocery inflation and wage growth highlights that wages are failing to keep pace with grocery-specific inflation, directly impacting affordability.

The wage-price gap represents the difference between wage growth and grocery inflation. A negative gap indicates that grocery prices are rising faster than wages, reducing affordability. The gap is relatively stable before 2020 but starts to decline sharply after 2021, turning increasingly negative by 2023, aligning with grocery inflation and slower wage growth during this period. A persistently negative wage-price gap confirms that grocery prices are rising too quickly for wages to offset, leading to worsening affordability.

The Affordability Index declines slightly from 2017 to 2023, reflecting a cumulative drop of ~4.6%. The overall trend shows a decline in affordability, particularly after 2021, when the wage-price gap becomes significantly negative. The flat trajectory of the Affordability Index relative to the sharper decline in the wage-price gap suggests that broader economic factors may have mitigated some of the impact of rising grocery prices. However, these adjustments are insufficient to fully counteract the affordability challenges. The declining Affordability Index aligns with the negative wage-price gap, confirming that affordability has worsened as grocery prices outpace wage growth.

During 2017 to 2020, there was stable affordability: With grocery inflation and wage growth rose at similar rates, resulting in a relatively stable wage-price gap and Affordability Index. This reflects a period of equilibrium where wage adjustments kept pace with rising grocery prices.

But during 2021 to 2023, there was declining affordability: Where grocery inflation accelerated, while wage growth slowed, widening the wage-price gap and reducing the Affordability Index. This period highlights the disproportionate impact of grocery-specific inflation on affordability, as wages failed to keep up with rising grocery prices.

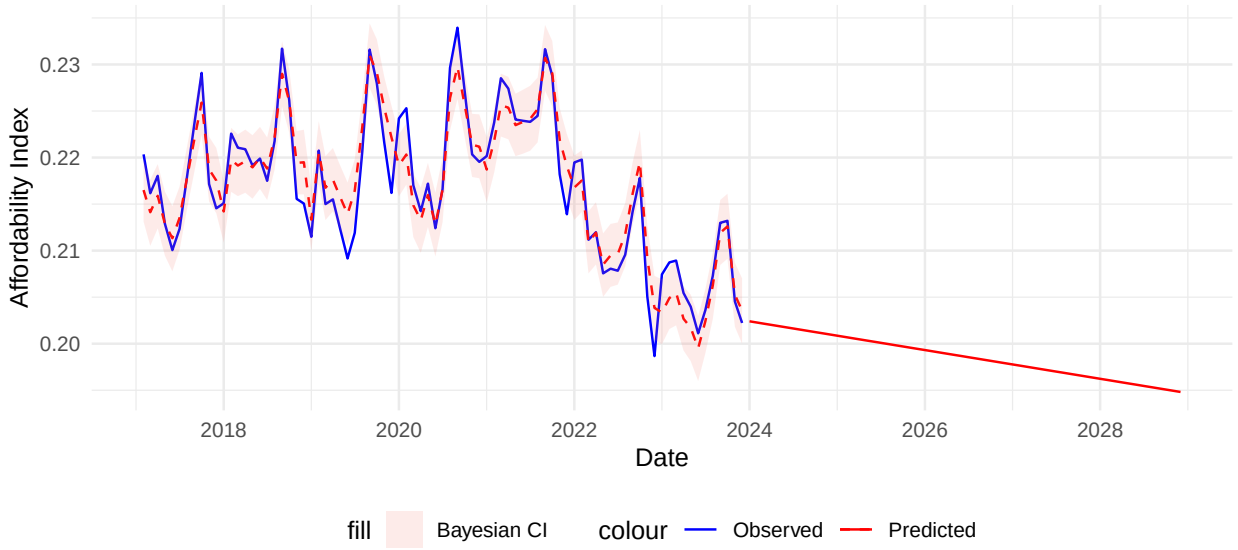
Ultimately, Grocery affordability has declined, as shown by the 4.6% drop in the **Affordability Index** and the increasingly negative wage-price gap since 2020. Wage growth has partially mitigated the impact of general inflation (CPI), as reflected in the stable Affordability Index prior to 2020. However, wage growth has

not kept pace with the rapid rise in grocery prices since 2021, leading to significant affordability challenges.

4.2 Projected Affordibility Index

Figure 6 provides the observed and predicted Affordability Index over time and suggests trends for future affordability under the current conditions. This model is ran using Gelman et al. (2023) and summarized using Arel-Bundock (2023).

Figure 6: Yap



The model aligns closely with the observed Affordability Index for the period between 2017 and 2023, capturing the overall trend in affordability while smoothing short-term fluctuations. Short-term fluctuations, such as the spikes during 2018–2019 and 2021, are likely caused by temporary market conditions, that is wage adjustments, temporary stabilization of grocery prices, or external shocks like COVID-19). These fluctuations are minimized by the Bayesian prediction. The predicted Affordability Index exhibits a gradual but clear decline, reflecting the underlying trend of affordability worsening over time. This aligns with the observed data and highlights that grocery-specific inflation continues to outpace wage growth under current conditions. The model projects a further decline in the Affordability Index beyond 2023, with affordability expected to fall below 20% by 2028. This decline is not immediately visible in raw data due to short-term fluctuations but becomes evident in the long-term projection.

The model assumes current policies and economic conditions remain unchanged, making it difficult to predict the impact of unpredictable events such as strikes, government interventions, or global economic disruptions. While it provides a general sense of where affordability is headed, that is all it is, a highly indefinite prediction. External factors could significantly alter the predicted trend.

Predicting affordability is inherently uncertain due to the potential influence of short- and long-term external factors (e.g., pandemics, economic shocks, or policy changes). However, the model is valuable for understanding the underlying trends in affordability and providing a general sense of direction for future planning.

4.3 Limitations

The study focuses exclusively on grocery prices and their relationship to general inflation (CPI) and wages, without considering affordability in other sectors (e.g., housing, transportation, utilities, etc.). This narrow

scope may underestimate the broader economic pressures faced by households.

The analysis uses average wages across all workers, which does not account for variations across income classes, industries, or geographic regions. This aggregation may obscure disparities in how affordability challenges are distributed across different groups.

The "average grocery basket" used in the study is limited in scope and may not reflect the diverse consumption patterns of Canadian households. Certain produce or specialty items could experience different inflation trends, impacting affordability unevenly.

4.4 Next Steps

Expand the research to compare affordability across different sectors, such as housing, transportation, and utilities, to better understand how grocery affordability fits into overall household expenses.

Investigate wage growth across different income classes, industries, provinces and geographic regions to understand how affordability challenges vary among groups. This would provide a more nuanced picture of affordability disparities.

Evaluate how factors such as unemployment rates, household savings, or government subsidies influence affordability trends. This would provide a more comprehensive understanding of the economic context.

By addressing these limitations and exploring the proposed next steps, future research could provide a more actionable analysis of grocery affordability in Canada.

5 Appendix

5.1 Data Cleaning

This section will be explaining the data cleaning process documented in `/scripts/03-clean_data`.

I used a 3 different datasets, split between 5 files. `old_inflation_data` and `inflation_data` are from the same source, but `old_inflation_data` contains years 2000 to 2023. The same was done for `old_wage_data`. This decision was done to make a broader analysis over many years easier, since data for `grocery_data` is only available from 2017 and on- variables `avg_wage_data` and `inflation_data` only contain data from 2017 and on as well. Similar cleaning processes were used for each variable. The data is loaded in and the first couple of rows are skipped since they contains subheadings?. One of the rows is kept for the column names using `colnames`, which then are renamed. `gsub` is used to rename the column names to remove spaces and characters that would conflict with the code. mutate was used to extract year and month columns for `grocery_data` and `inflation_data`,and `old_inflation_data`, since these variables are recorded monthly. `left_join` Is used to merge `inflation_data` to `grocery_data` by variables `year` and `month`.

`ref_cpi` is created to calculate the percentage change in CPI , `CPI_Percentage`. Then, I took the values of wages for each year and manually added the value for each year in new variables `inflation_wage_data` and 'old_inflation_wage_data', both variables having columns of cpi, wage and date, where wage has the same value throughout the year since the record is yearly. Lastly, csv files are written in the folder `/02-Analysis_data` for variables `grocery_data`, `inflation_data`, `avg_wage_data`, `inflation_wage_data` and `old_inflation_wage_data`.

Wages used are averages of all industries, of part-time and full-time employees, for people 15 and older.

5.2 Methodological Exploration of Data Collection

The reliability of affordability analysis is determined by the accuracy of the considered variables. This appendix will highlight the importance of how CPI data, grocery data, and wage data was collected and proceed, to ensure the credibility of the data.

5.2.1 Grocery Data Collection And CPI Calculation

The CPI is an indicator used to measure the rate of inflation. It is calculated by using price data collected for specific goods and services, such us groceries,household goods, gas, utilities, etc., and thus includes the grocery data used in this paper. The CPI has implementations in:

- contracted payments, used to preserve purchasing power of a value,
- obtaining constant dollar estimates of variables such as income of expenditure values,
- monitoring implementation of economic policy,
- research, such as cause and effect of inflation.

The target population includes households in urban and rural households in Canada. Citizens with distinctive circumstances, such as collective households, those living on Indian reserves, official representatives, chronic care patients, prison inmates, etc. are excluded from the target population.

The CPI price sample is derived from a basket of representative foods and goods across all geographical areas and stores. Only stores with high sales revenues are included. The timing of price collection is set for the first two weeks of every month to ensure consistency, but may vary depending on the nature of the good. Season items are only observed in their respective months. To ensure the compliance of international guidelines regarding price index methodology, manual and automated processes are implemented to ensure product characteristics over time.

5.2.1.1 Sampling Approach

The general sampling approach for the grocery data involves three main stages to ensure representative data.

- stage 1: Areas are divided into subdivisions of municipalities and neighborhoods. Population count is used in the selection process of these subdivisions to prioritize areas with larger populations.
- stage 2: Representative outlets (retail stores, small stores, or online retailers) are selected based on largest consumer purchases.
- stage 3: Small samples of products are chosen from each product class to represent the broader class, based on popularity and research. Products with volatile prices, significant basket weights, and lower cost are prioritized. Each product class is weighed based on its share of total consumer expenditure.

Price collection is conducted by many methods through retail stores, flyers or online. For some product classes that have national. Provincial pricing, prices may be collected centrally through retailer websites. Units of products are primary as possible, with goods usually measured per unit, per kilogram, or per liter. More detailed explanation of sampling apporach in chapter 5 of (Canada (2023)).

5.2.1.2 Calculations of the Consumer Price Index

After samples are collected, the CPI is calculated through two stages.

During the first stage, price changes are calculated for group called *elementary aggregates*, that are small categories of similar products in specific locations. Average price change, known as the *elementary index* is calculated using Jevons formula, defined by:

$$I_{J,a}^{t-1;t} = \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^{t-1}} \right)^{1/n},$$

where:

- $I_{J,a}^{t-1;t}$ is the weighed Jevons price index for an elementary aggregate a between periods $t-1$ and t,
- n is the number of product prices i in a, and
- $\left(\frac{p_i^t}{p_i^{t-1}} \right)^{1/n}$ is the price relative for i between periods $t-1$ and t. The Jevons formula is the primary method used by Statistics Canada since 1995. It calculates a geometric mean of price relatives, to depend the ratio on previous prices. This formula can be interpreted as an average of price changes, as well as a change in average prices between periods, allowing for easy data analysis.

In the second stage, calculated elementary indices are combined into *aggregate indices* that measure price changes for larger categories. Each aggregate index is assigned a weight that accurately represents consumer expenditure, and then combined into a single CPI value for the entire country. More detail in chapter 6 of (Canada (2023)).

Adjustments are made to product prices to compensate for quality change in products over time. Implicit methods assume there is no major quality differences, and thus rely on indirect calculations. Some of these methods are:

- Direct Price Comparison: For where there is no quality difference in old and new products, such as gas or electricity,
- Overlap Pricing: applied when old and new products are sold concurrently,
- Mean Imputation: imputes price changes of new products based on average price movements of similar products,
- Quantity adjustment: Adjusts for changes in produce quantity, for example package size, assuming the quality of the product remains constant. This common for packaged food items and household items, and impacts most of the grocery data used this paper.

Explicit methods directly account for quality differences by comparing old and new products. Hedonic Quality Adjustment is used for products rapidly changing features, such as software, or services like internet and cellular plans. This method replies of regressions models to estimate relationships between price and product characteristics. Regression models are used to predict prices based on features, like CPU and RAM of a laptop, and separate regressions are run for features like download speed and upload speed of internet services. Regression variables are weighed accordingly to their relevance to the product or service. More detail in chapter 7 of (Canada (2023)).

By implementing these quality adjustments ensure that CPI reflects pure price changes.

5.2.1.3 Limitations of Methodolgy and Survey Samples

The calculation of the CPI is subject to several limitations that an impact the accuracy and representativeness. One limitation is the focus on private households, which many capture the consumption patterns of all Canadians. Combined households are not included in the survey, which can account for and percentage of the overall population of Ontario, and Canada. Different households have varying purchasing behaviors influenced by income, cultural and personal preferences, and household composition, which may not reflect the broader Canadian population. Furthermore, there is significant regional variation in price sampling, and limited coverage of specific products in certain areas. The lack of geographic representation may skew results, especially for locally popular or region-specific goods.

Another limitation is the reliance on a "judgmental" selection of products within each category. While this approach prioritizes commonly purchased items, it introduces subjectivity. Although selection is based on market research, it also depends on the expertise of those making the selection. This method may overlook less popular but significant products. Additionally, the handling/ manipulation of sales, promotions, and out-of-stock items presents challenges to accuracy. When items are unavailable or missing price data, estimates of the prices are made based on similar products or locations, or simply substituted. While these methods aim to maintain continuity in the index, it may introduce inaccuracies that do not fully capture the original products price dynamics or consumer preference.

These limitation highlight trade-offs in CPI calculations, ensuring representativeness while managing constraints in data collection.

5.2.2 Wage Data Collection

Wage data was obtained from the Labour Force Survey (LFS). the LFS is monthly household survey conducted by Statistics Canada to gather data on the Canadian Labour Market. A survey, in the form of a questionnaire, is completely by eligible individuals in the samples.

5.2.2.1 Sampling Methodology

The LFS uses a stratified multi-staged probability approach for their samples, dividing each province into large geographic strata. In the first stage, smaller geographic areas, known as clusters, are selected within each stratum. In the second stage, dwellings are chosen from these clusters.

The LFS utilizes a Rotating Panel Design, where initially selected households participate for six consecutive months in the collection of data, with about one-sixth of the sample replaced each month. The sample covers approximately 68,000 households monthly, collecting data from around 100,000 individuals. This approach balances the need for up-to-date information with the benefits of tracking changes over time.

5.2.2.2 Data Collection

The target population of the survey include all people aged 15 and older who's main country of residence is Canada, including citizens, non-permanent residents and permanent residents. Excluded from the population are those who live on reserves, full-time members of Armed Forces, and people living in institutions. Surveys

are conducted via in-person interviews, telephone interviews, or self-completed electronic questionnaires. The LFS practices proxy reporting, where one individual per household provides the information for the entire household, which accounts for a significant portion of the responses, about 65%.

The questionnaire is periodically redesigned to reflect the changes in the Canadian Labour Market. These changes are rigorously tested through review committees, focus groups, and pilot tests to ensure reliability. The most recent full questionnaire can be found here (Canada n.d.), effective since January 1st, 2023.

5.2.2.3 Error Detection and Imputation

As a sample survey, the LFS is subject to sampling and non- sampling errors, that can rise from data collection and processing. To mitigate these issues, the LFS has several error detection and imputation strategies to ensure data reliability. employed comprehensive training, through questionnaire testing, and questions are researched to ensure clarity and ambiguity. Interviewers undergo comprehensive training,through questionnaire testing, and questions are researched to ensure clarity and ambiguity.Additionally, electronic questionnaires are programmed to flag out-of-range or inconsistent responses for immediate human verification. Post-collection, further imputation methods are executed to handle missing or incorrect data. These methods include:

- Deterministic Imputation: Missing/incomplete data is replace with logically consistent values,
- Carry-Forward Imputation: Data from the previous month is carried over to the current month's missing values, when appropriate, and
- Donor Imputation: Data from a similar individual (considered the 'donor') with comparable characteristics to fill in gaps when the aim is to maintain consistency and minimize bias.

5.2.2.4 Weighting Adjustments

For non-responding households, weights are adjusted to assume that these households have similar characteristics to responding ones, in attempt to minimize non-response bias. After non-response adjustments, weights are refined to correct for coverage errors, when populations are misrepresented, by adjusting weights to ensure final weighted estimates align with census projections for demographic groups, age, gender, region, etc. Moreover, sampling variance is addressed by compounding overlapping consecutive monthly samples by stabilizing estimates from past surveys, to increase efficiency and reduce variance.

These major weight adjustments improve estimate accuracy by ensuring that surveys reflect the true population distribution, and thus are consistent and reliable, even when drawing from different samples.

5.2.2.5 Limitations of the LFS

While the LFS has a comprehensive process for collection information, it has some limitations that should be considered. The LFS does exclude some populations, as described above, which may not fully represent the labour market conditions for these groups.

Non-response remains a challenge in many surveys, even with attempted follow up efforts. Adjustments are made to account for these reports, but the assumption of similarity of dispositions between households may not hold true for all instances. Proxy reporting can also lead to inaccuracies due to the proxy's incomplete knowledge or misreporting. Furthermore, sampling variability from different samples could produce varied estimates, particular for smaller population groups or regions. While weight adjustments are implemented to attenuate this, the estimates may still be impacted by these fluctuations.

Despite these limitations, the LFS employes effective methodologies to maintain data quality. For a more detailed explication of the LFS, refer to (Canada n.d.).

5.3 Linear Model Summary Statistics

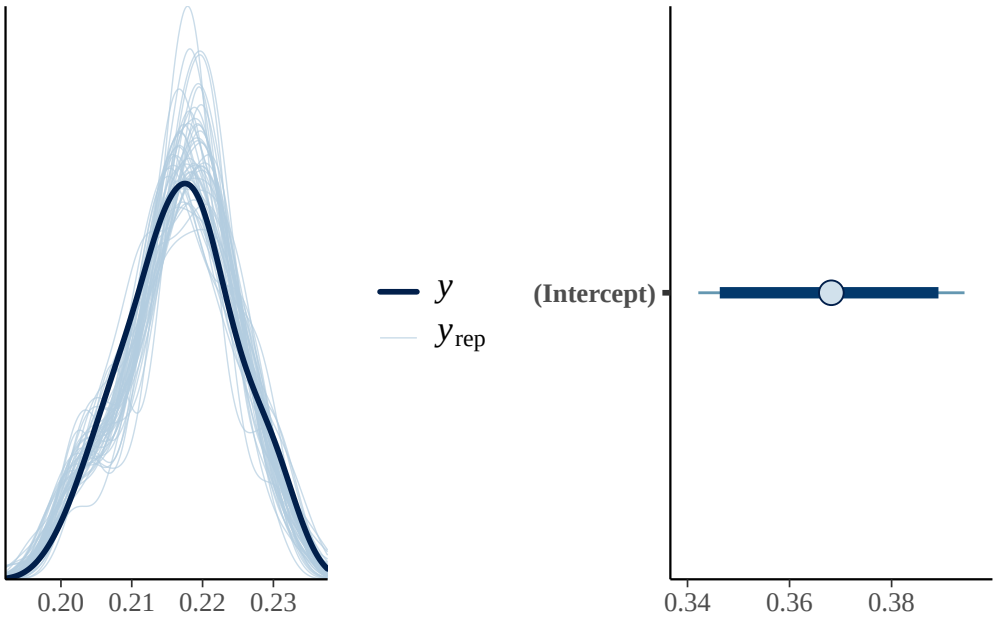
```
Call:
lm(formula = log(Affordability) ~ log(CPI) + log(Average_Price) +
    log(Time), data = model_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.031419 -0.012361 -0.001302  0.012941  0.046970

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.283988   0.270185  -1.051 0.296423
log(CPI)            0.397345   0.105289   3.774 0.000309 ***
log(Average_Price) -0.685090   0.063795 -10.739  < 2e-16 ***
log(Time)           0.024000   0.003373   7.115 4.49e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01746 on 79 degrees of freedom
Multiple R-squared:  0.7803,    Adjusted R-squared:  0.7719
F-statistic: 93.51 on 3 and 79 DF,  p-value: < 2.2e-16
```

5.4 Postrior Predictive Check



Arel-Bundock, Vincent. 2023. *Modelsummary: Beautiful and Customizable Model Summaries in r*. https://vincentarelbundock.github.io/modelsummary/.

Bank of Canada. 2024. "Consumer Price Index (CPI) statistics, measures of core inflation and other related statistics." https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1810025601.

Canada, Statistics. n.d. "Consumer Price Index (CPI) Survey." https://www.statcan.gc.ca/en/statistical-programs/instrument/3701_Q1_V9.

———. 2023. "The Consumer Price Index, 2023." https://www150.statcan.gc.ca/n1/pub/62-553-x/62-553-x2023001-eng.pdf.

Gelman, Andrew, Ben Goodrich, Jonah Gabry, and Aki Vehtari. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

Pedersen, Thomas Lin. 2023. *Patchwork: The Composer of Plots*. https://patchwork.data-imaginist.com/.

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

Statistics Canada. 2024a. "Average weekly earnings by industry, monthly, unadjusted for seasonality." https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410006401.

———. 2024b. "Monthly average retail prices for selected products." https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810024501.

———. 2024c. "Statistics Canada Homepage." https://www.statcan.gc.ca/en/start.

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.

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