

# 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 bayesian 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 condition of the world. Through bayesian 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 bayesian model based on ‘polls of polls’ where we compare the results from different polls, we are able to make this prediction.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data Overview</b>	<b>2</b>
2.1	Citations . . . . .	2
2.2	Variables . . . . .	3
<b>3</b>	<b>Data Analysis</b>	<b>3</b>
3.1	Measurement . . . . .	6
3.2	Outcome Variables . . . . .	6
3.3	Predictor Variables . . . . .	6
<b>4</b>	<b>Model</b>	<b>6</b>
4.1	Model Setup . . . . .	6
4.1.1	Model Justification . . . . .	7
4.2	Model Results . . . . .	7

---

\*Code and data are available at: [Are Groceries Attainable](#).

<b>5</b>	<b>Discussion</b>	<b>9</b>
5.1	Important discussion . . . . .	9
5.2	Projected Affordability Indexvery vague Preditived Model . . . . .	9
5.3	Limitations . . . . .	10
5.4	Next Steps . . . . .	10
<b>6</b>	<b>Appendix</b>	<b>11</b>
6.1	Data Cleaning . . . . .	11
6.2	Methodological Exploration of Data Collection . . . . .	11
6.2.1	Grocery Data Collection And CPI Calculation . . . . .	11
6.2.2	Wage Data Collection . . . . .	13
6.3	Linear Model Summary Statistics . . . . .	15
6.4	Posterior Predictive Check . . . . .	15

# 1 Introduction

Inflation in Cananda has increased rampantly in all sectors, growing concerns about affordability. Within this, the roles of inflation and stagnant wages have made groceries less attainable for Canadians. Grocery prices have risen 22.5% since May of 2020, and continue to rise each month, creating panic for households (News (2024)). In this paper, we explore the inflation rates of groceries in Canada from 2017 to 2023, relating it to wages. We then aim to predict the cost of groceries over the coming years, assessing if the increasing rate of wages is enough for groceries as an average Canadian.

**This paper examines the affordability of groceries in Canada between 2017 and 2023 by analyzing the relationship between grocery prices, inflation (CPI), and average wages. Using an affordability index, we assess whether wage growth has kept pace with rising grocery prices, with implications for future affordability under projected trends. How has the affordability of groceries in Canada changed over time, and to what extent has wage growth mitigated the impact of inflation on grocery prices?**

Overview paragraph

Estimand paragraph We want to predict grocery prices based on past grocery prices, inflation and wages? to determine if Canadians will be able to afford groceries in the coming years, assuming inflation and wages are increasing at a steady rate.

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. ?@sec-data....

# 2 Data Overview

## 2.1 Citations

The datasets used in this analysis are retrieved from Statistics Canada ((cite?)). The main dataset is acquired from (Statistics Canada 2024), 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 ((cite?)) 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. The three datasets were merged together for analysis.

## 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 base 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 values 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 important, but the impact and correlation between the two variables is.

More about data cleaning in the appendix section appendix (data?) cleaning.

## 3 Data Analysis

The first variables examined are CPI and Wage. In Figure 1, we analyse normalized values of yearly average wages and monthly CPI from 2000 to 2023, as they are measured on different scales. This graph illustrates the trend between CPI and wages. The period of 2000 to 2020 shows a predominantly economically steady correlation between the two variables, with predictable inflation rates. However, the Covid-19 pandemic began in 2020, causing restricted movement globally, subsequently disruption global economies and changing consumer spending behaviour, shown by the sharp growth after 2020. 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.

# A tibble: 576 x 9

	Date	CPI	Year	Month	Wage	CPI_Percentage	Wage_Percentage	Variable	Value
	<chr>	<dbl>	<int>	<int>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>
1	2000~	94	2000	1	16.7	NA	NA	CPI_nor~	0
2	2000~	94	2000	1	16.7	NA	NA	Wage_no~	0
3	2000~	94.6	2000	2	16.7	0.638	NA	CPI_nor~	0.00946
4	2000~	94.6	2000	2	16.7	0.638	NA	Wage_no~	0
5	2000~	95.3	2000	3	16.7	0.740	NA	CPI_nor~	0.0205
6	2000~	95.3	2000	3	16.7	0.740	NA	Wage_no~	0
7	2000~	95	2000	4	16.7	-0.315	NA	CPI_nor~	0.0158

8	2000~	95	2000	4	16.7	-0.315	NA Wage_no~ 0
9	2000~	95.4	2000	5	16.7	0.421	NA CPI_nor~ 0.0221
10	2000~	95.4	2000	5	16.7	0.421	NA Wage_no~ 0

# i 566 more rows

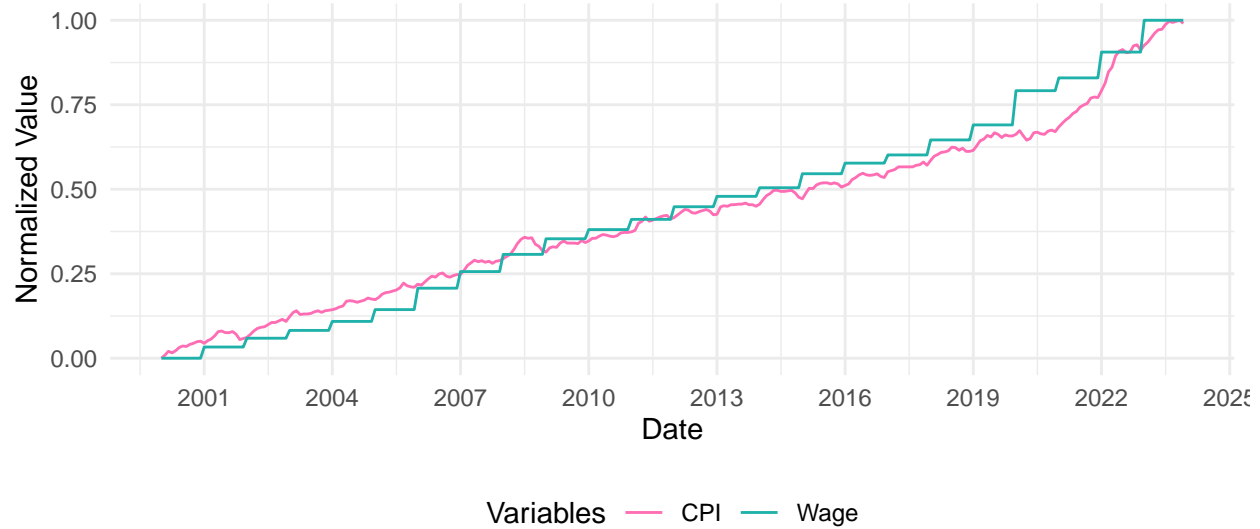


Figure 1: Normalized CPI and Wages Over Time from 2000 to 2023

in Figure 2 between 2020-2022, covid. CPI went down because people were buying things, and wages went up surprisingly, since many people lost their jobs during covid, suggesting that companies were saving more money, people were getting paid more perhaps through bonuses and government covid funding. It is clear that the two variables are correlated, but may not necessarily depend on each other.

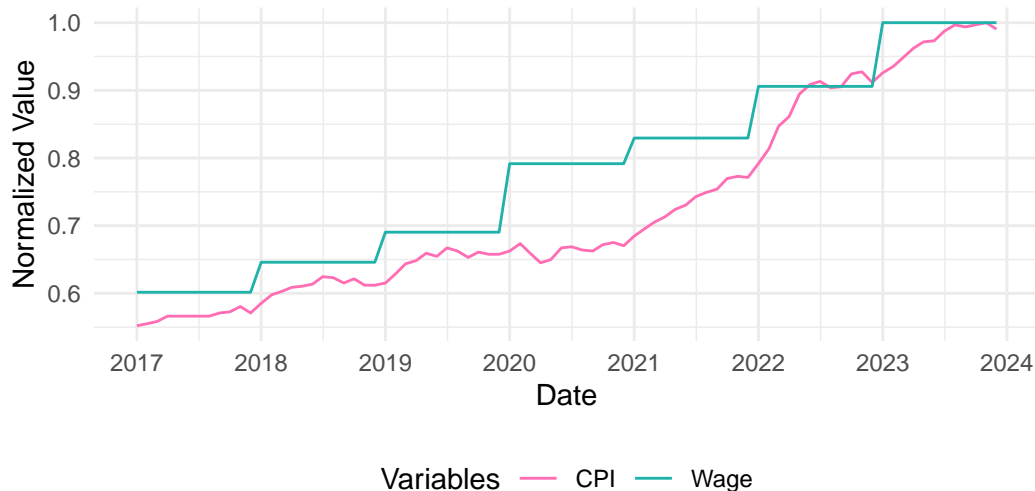


Figure 2: Normalized CPI and Wages Over Time from 2017 to 2023

Introducing `Average_Price`, Figure 3 depicts the trend of average grocery prices from 2017 onwards. It is obvious that grocery prices increase as inflation increases, but it is confounding how much inflation has risen in the past 7 years. A grocery cart that costed \$120 in 2017 can cost well above \$160 in 2023 – that is a 22%

inflation. This number brings concern about the affordability of groceries in Canada, in addition to other inflating expenses, such as rent and utilities.

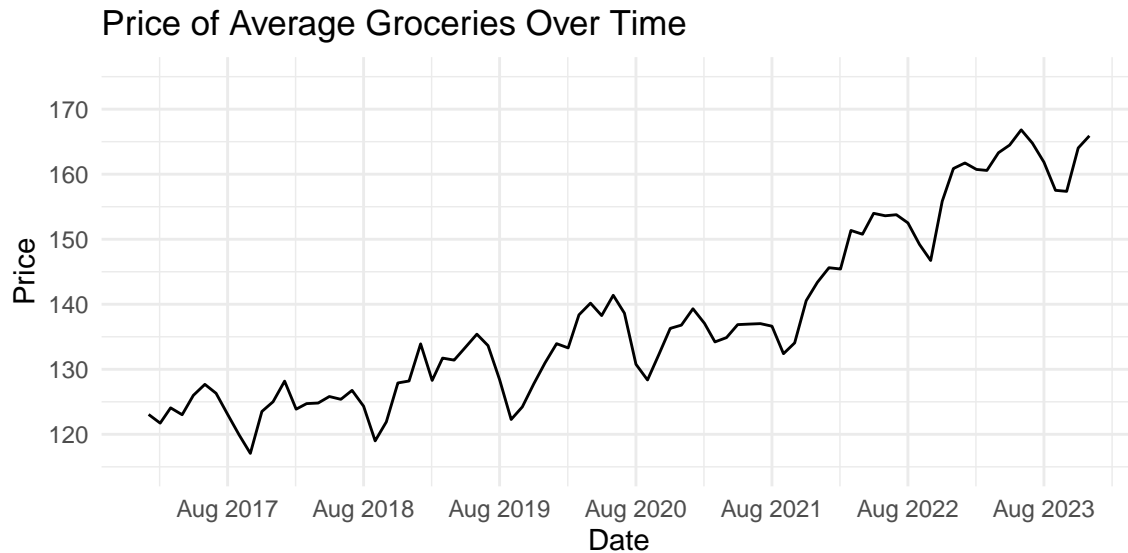


Figure 3: Normalized CPI and Wages Over Time from 2017 to 2023

To investigate this concern, we analyze the Affordability Index created in [?@fig-affordability](#). Graphed over the years beginning in 2017, we notice that the general trend of affordability is decreasing, implying affordability challenges.

even with increasing wages as CPI increases, there suggests that wage increasing is not enough to sustain the CPI increase.

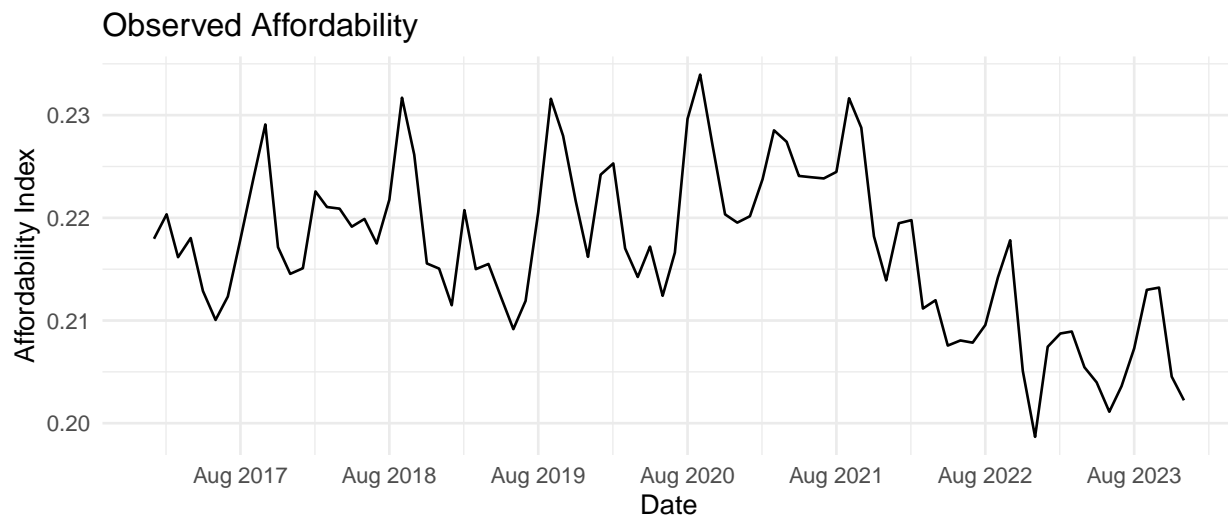
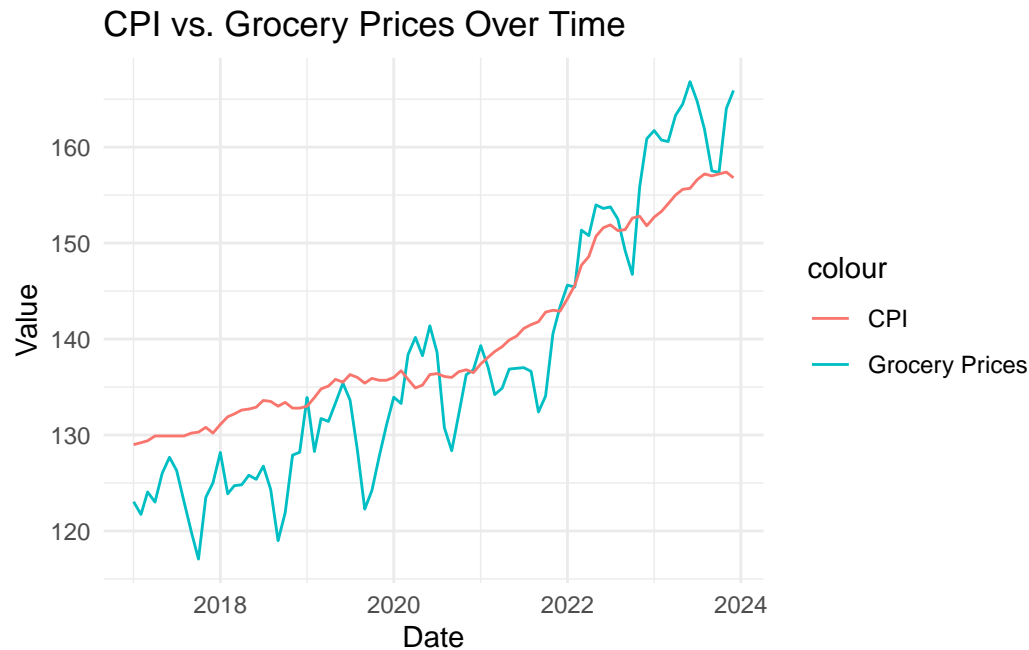


Figure 4: Normalized CPI and Wages Over Time from 2017 to 2023

We further explore this correlation through a model...

```
ggplot(grocery_data, aes(x = Date)) +  
  geom_line(aes(y = Average_Price, color = "Grocery Prices")) +
```

```
geom_line(aes(y = CPI, color = "CPI")) +
labs(title = "CPI vs. Grocery Prices Over Time", y = "Value", x = "Date") +
theme_minimal()
```



### 3.1 Measurement

### 3.2 Outcome Variables

### 3.3 Predictor Variables

## 4 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. Specifically, 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.

### 4.1 Model Setup

The model specification is as follows:

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

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

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

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

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

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

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

Where:

- $y_i$  is the dependent variable, representing the Affordability Index(Wage/ Average Price),
- $\mu_i$ : is the mean response, a function of the predictors,
- $\beta_0$  is the intercept, representing baseline affordability when all predictors are set to zero,
- $\beta_1$  the elasticity of affordability with respect to CPI, indicating the proportional change in affordability for a 1% change in the CPI,
- $\beta_2$  is the elasticity of affordability with respect to Average Price, representing impact of the changes in average grocery prices on affordability,
- $\beta_3$  is a numerical variable of Date turned into a cumulative column of days for each month of the year. Captures the change in  $\mu_i$  over time.

Each coefficient  $\beta_i$  is assigned a Normal prior  $\mathcal{N}(0, 2.5)$  to regularize estimates and prevent overfitting. The residual variance  $\sigma$  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 (**dataset?**). 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.

#### 4.1.1 Model Justification

## 4.2 Model Results

Table 1 summarizes the coefficients of the model.

Table 2 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:

Table 1: 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)

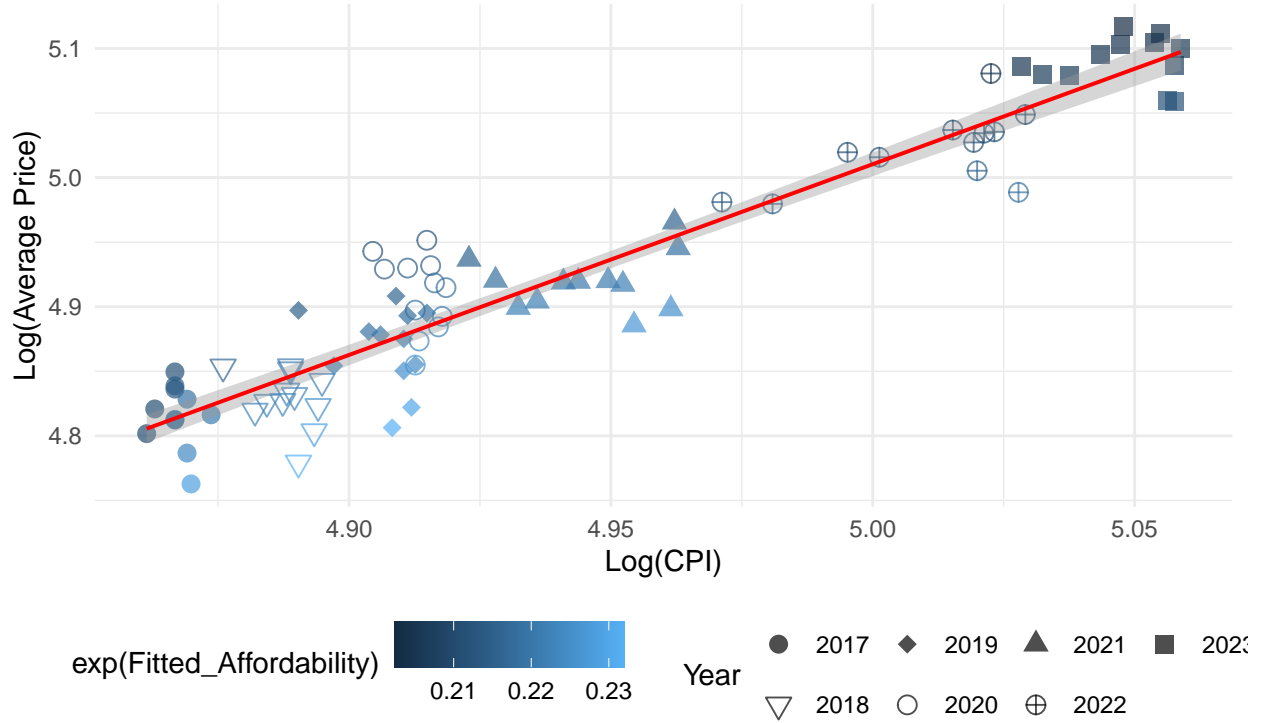


Figure 5: Affordability, CPI, and Average Price by Year



Table 2: Yearly Averages and Log-Transformed Values for Key Variables

Year	Avg CPI	Avg Grocery Price	Avg Affordability	Log CPI	Log Grocery Price	Affordability index
2017	129.964	123.395	0.217	4.867	4.815	0.217
2018	132.708	125.071	0.221	4.888	4.829	0.221
2019	135.258	130.118	0.218	4.907	4.868	0.218
2020	136.108	135.705	0.221	4.913	4.910	0.221
2021	140.558	136.943	0.224	4.946	4.920	0.224
2022	150.017	151.639	0.211	5.011	5.022	0.211
2023	155.717	162.423	0.207	5.048	5.090	0.207

## 5 Discussion

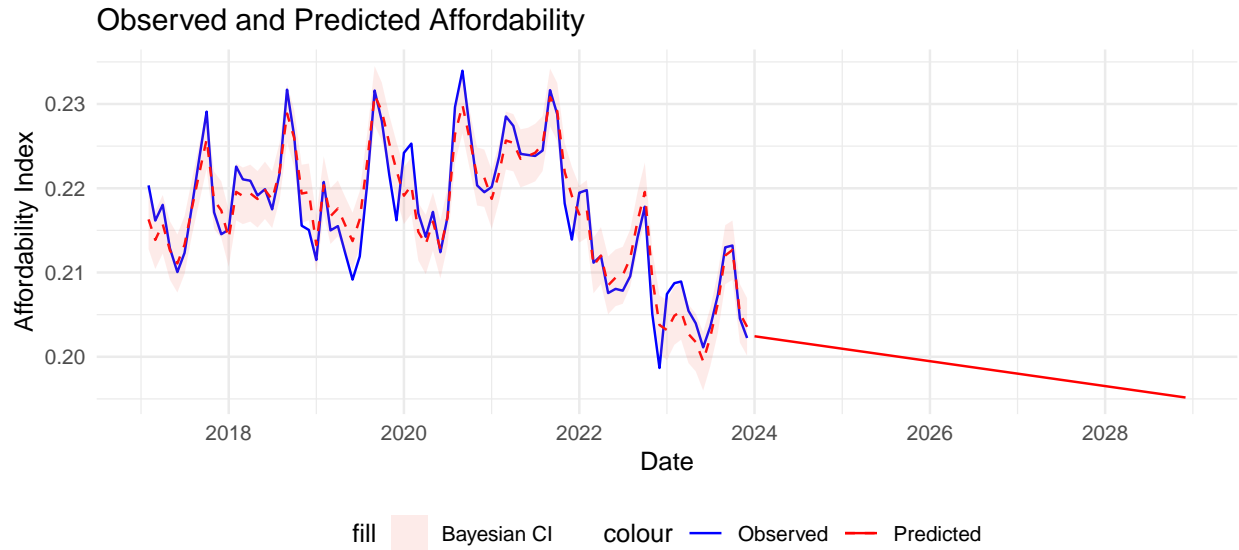
### 5.1 Important discussion

Inflation usually is a good indicator to the affordability of groceries, however, inflation takes into account many other characteristics. To focus on the affect of purely grocery prices, we use the results of the predicted Affordability Index from the linear model to asses whether wage growth has kept with rising grocery prices through elastic relationships.

Interpreting the intercept values of the elasticity model in Table 1: - **The intercept (-0.284)** is the expected  $\log(\text{Affordability})$  when all predictors are zero. -  **$\log(\text{CPI})$  coefficient (0.397)** states that a 1% increase in CPI is associated with a 0.397% increase in Affordability, holding other predictors constant. This may seem counter-intuitive, since we expect CPI to decrease affordability. However, the increase in CPI may correlate with wage growth, which could lead to higher affordability. -  **$\log(\text{Average\_Price})$  coefficient (-0.685)** states that a 1% increase in Average Price is associated with a 0.685% decrease in Affordability, which aligns with expectations. -  **$\log(\text{Time})$  coefficient (0.024)** states that a 1% increase in time is associated with a 0.024% increase in Affordability, suggesting a slight improvement in affordability over time.

### 5.2 Projected Affordability Indexvery vague Preditived Model

- hard to forecast short or long term events that affect CPI, such a strikes, government bills, diseases/sickness, etc, thus accurate model is very ahrd to predict. However, it is informative to have a general sense of where affordability will be in coming years.



- Model follows observed Affordability Index, capturing long-term trend
- The Bayesian prediction smooths out short-term fluctuations (e.g., the volatility in observed values during 2018 and 2019). This indicates that while real-world affordability may be influenced by sudden events (e.g., market shocks), the underlying trend is governed by systematic factors such as inflation and wage growth, which the model captures.
  - The smoothing effect helps reveal the underlying, systematic decline in affordability that may not be immediately visible in the raw data, which is influenced by noise or temporary fluctuations.
- The smoothing effect helps reveal the underlying, systematic decline in affordability that may not be immediately visible in the raw data, which is influenced by noise or temporary fluctuations.

### 5.3 Limitations

- limited produce
- limited data available (only from 2017)
  - more recent data, 2024 has high spikes in groceries

### 5.4 Next Steps

- Perhaps a better question would be how does affordability compare to all sectors, not just grocery prices ?
- not focus on wages as a whole, but different classes and wage ranges
- analyze data geographically, by province

## 6 Appendix

### 6.1 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.

### 6.2 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.

#### 6.2.1 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 as 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 or 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.

### 6.2.1.1 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 approach in chapter 5 of (Canada (2023)).

### 6.2.1.2 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 relies on 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.

### **6.2.1.3 Limitations of Methodology and Survey Samples**

The calculation of the CPI is subject to several limitations that 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 a 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 limitations highlight trade-offs in CPI calculations, ensuring representativeness while managing constraints in data collection.

## **6.2.2 Wage Data Collection**

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

### **6.2.2.1 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.

### **6.2.2.2 Data Collection**

The target population of the survey includes all people aged 15 and older whose 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.

### **6.2.2.3 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 replaced 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.

### **6.2.2.4 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.

### **6.2.2.5 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 employs effective methodologies to maintain data quality. For a more detailed explication of the LFS, refer to (Canada n.d.).

## 6.3 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 ***

---

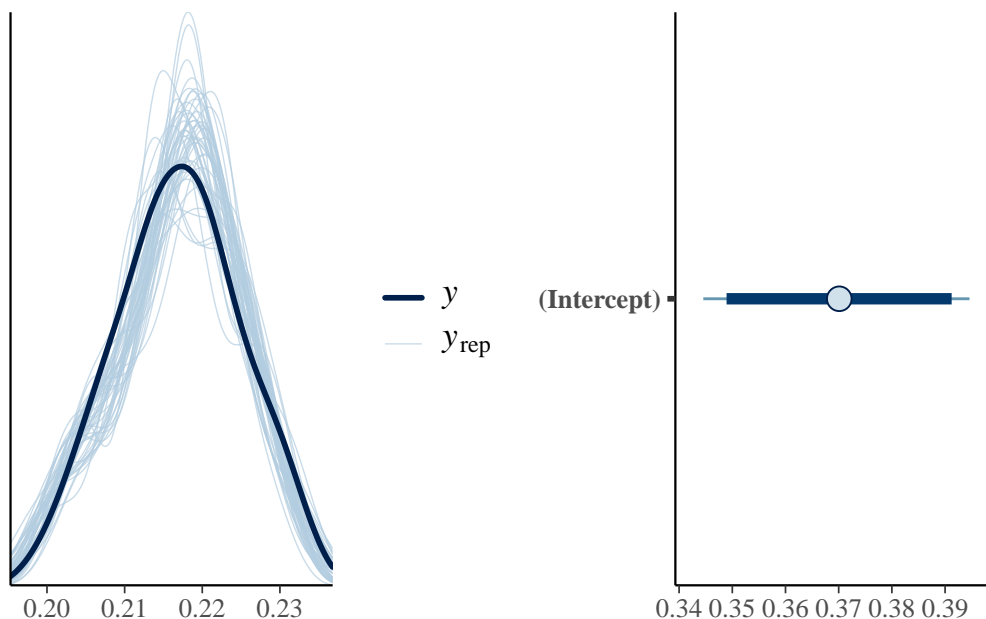
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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

## 6.4 Posterior Predictive Check



Canada, Statistics. n.d. “Consumer Price Index (CPI) Survey.” [https://www.statcan.gc.ca/en/statistical-programs/instrument/3701\\_Q1\\_V9](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>.

News, CTV. 2024. “Canadians Feel the Pinch at Grocery Store as Food Inflation Ticks Higher for First Time in Nearly a Year.” 2024. <https://www.ctvnews.ca/business/canadians-feel-the-pinch-at-grocery-store-as-food-inflation-ticks-higher-for-first-time-in-nearly-a-year-1.6940172>.

Statistics Canada. 2024. “Monthly average retail prices for selected products.” <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810024501>.