

are grocery prices attainable*

Analysing grocery prices with inflation and wages

Cristina Burca

November 30, 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

1	Introduction	2
2	Data Overview	2
2.1	Citations	2
2.2	Variables	2
3	Model	5
3.1	Model Justification	5
3.2	Discussion	5
3.3	Projected Affordability Index	6
4	Discussion	6
4.1	Limitations	6
4.2	Next Steps	7

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

5 Appendix	8
5.1 Data Cleaning	8
5.2 Methodological Exploration of CPI Data Collection	8
5.3 Posterior Predictive Check	8

1 Introduction

Inflation in Canada 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.

“How have grocery prices in Canada changed over time, and how do these changes, relative to inflation and wages, affect the accessibility of groceries across socioeconomic classes?”

Focus on a subset of comparable products (e.g., fruits vs. snacks, vegetables vs. processed foods). Analyze how prices for healthy and unhealthy groceries have changed over time.

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_Percentage:** A column created to capture the CPI increase as a percentage. The CPI recorded on January 2017 was 129, and is set to 0 in this variable. Continuous data is based off of January 2000.
- **Wage:** Yearly average wage of all employees from all industries.
- **CPI_Normalized** and **Wage_Normalized:** calculated columns to measure.
- **Average_price:** calculates the average prices of a “basket” summing all products in the data and dividing by the number of groceries.
- **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 affordibilty, while lower values suggest affordability struggles.

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

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.

CPI doesn't particularly affect wage increases.

May 2022/2023 - which correlations to focus on?

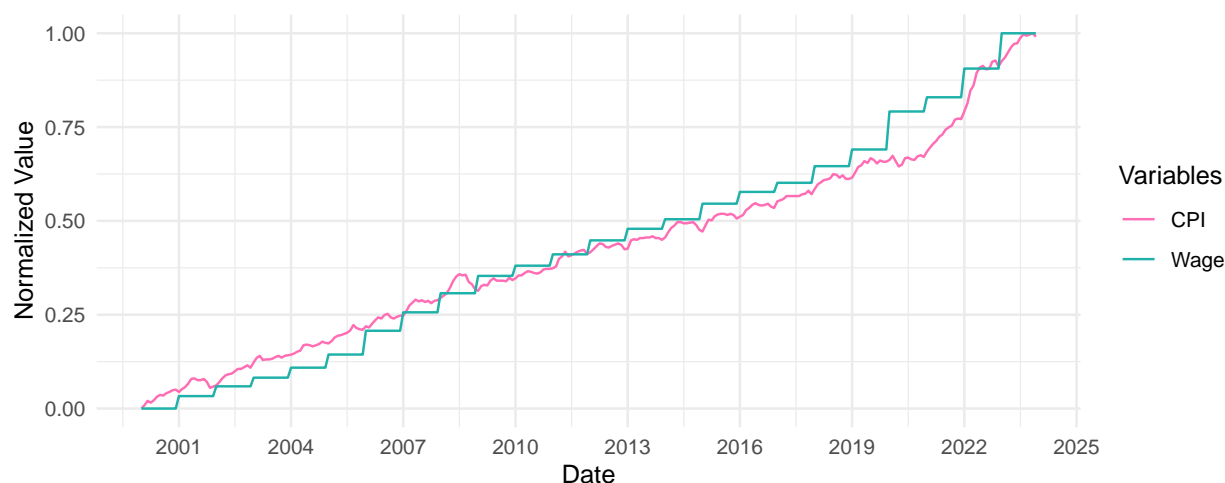


Figure 1: Normalized CPI and Wages Over Time from 2000 to 2023

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.

In Figure 3 we analyze the Affordability Index created. Graphed over the years beginning in 2017, we notice that ...

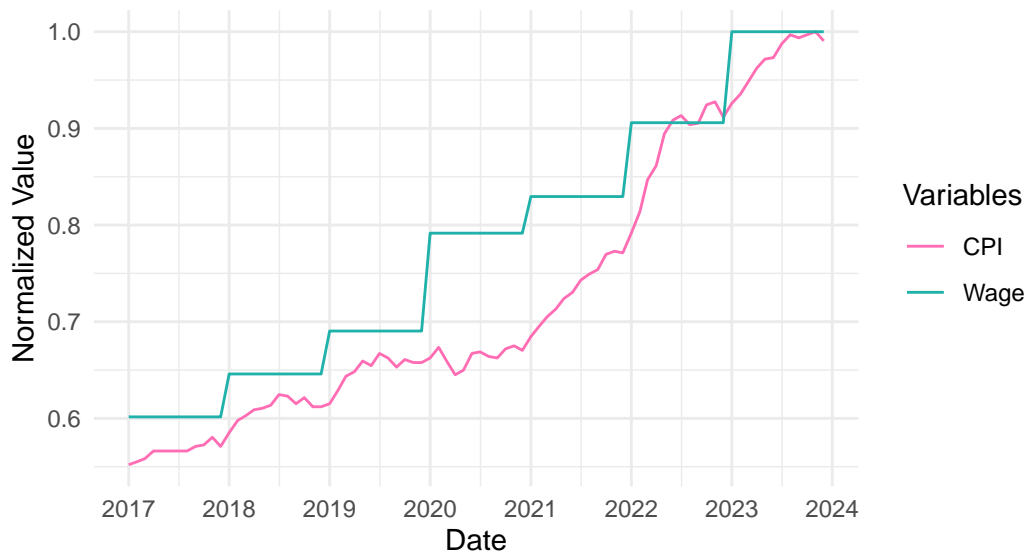


Figure 2: Normalized CPI and Wages Over Time from 2017 to 2023

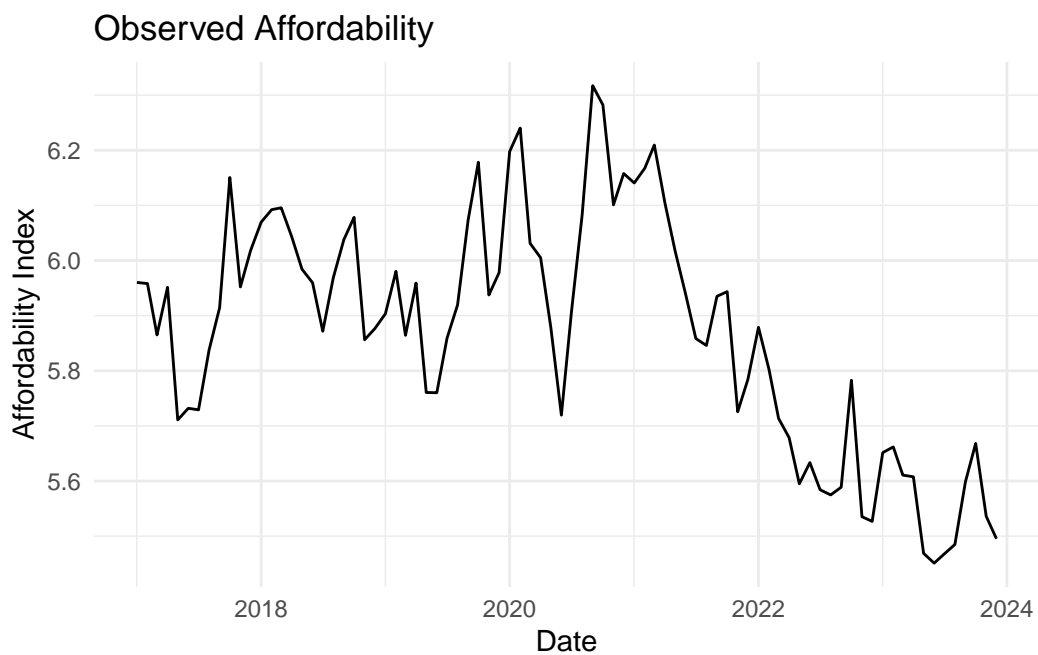


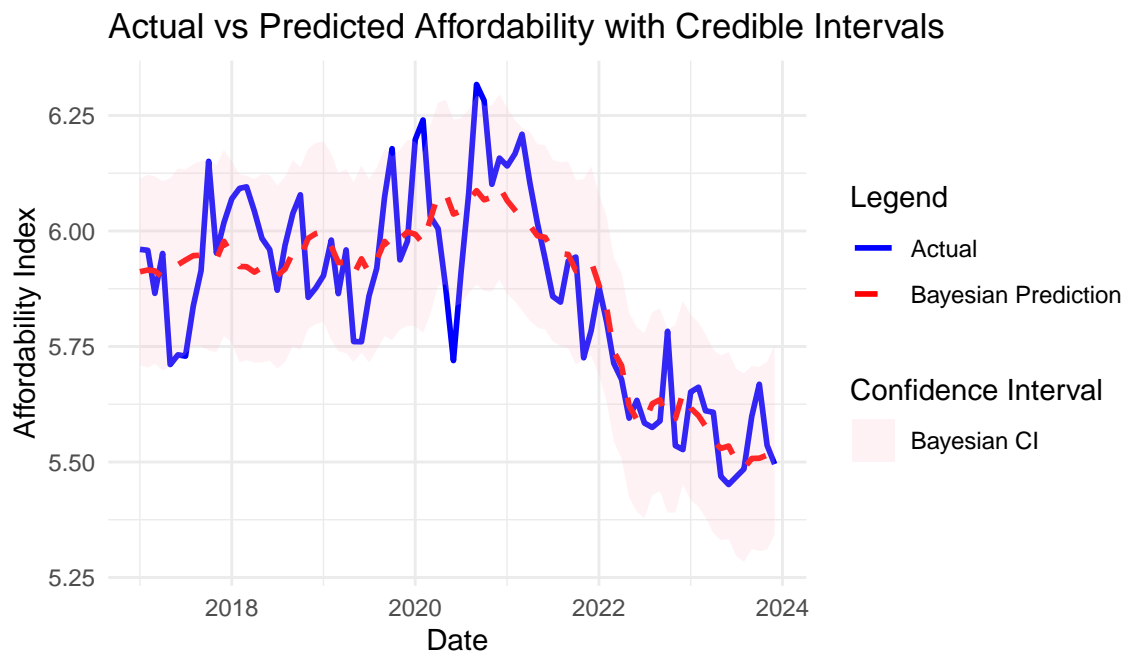
Figure 3: Normalized CPI and Wages Over Time from 2017 to 2023

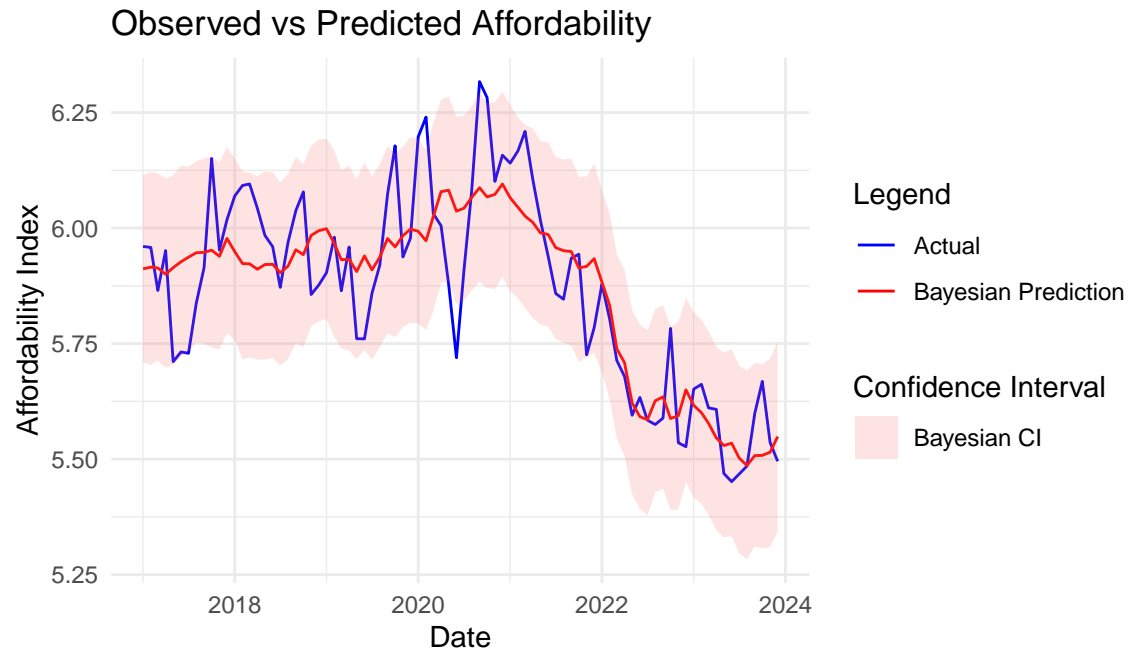
3 Model

3.1 Model Justification

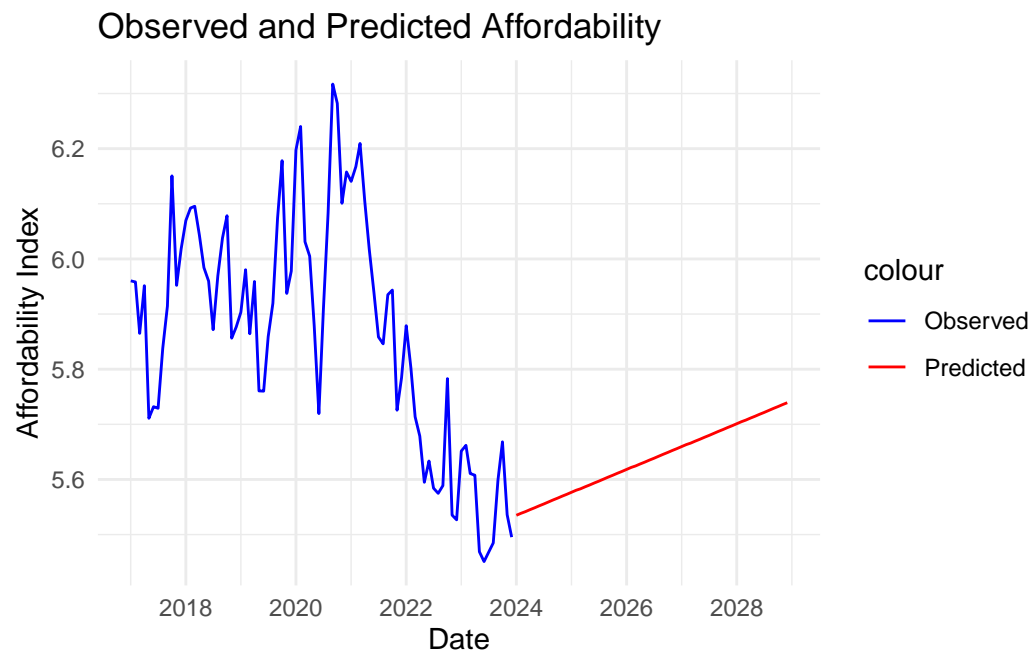
3.2 Discussion

- 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.





3.3 Projected Affordability Index



4 Discussion

4.1 Limitations

- limited produce

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

4.2 Next Steps

- Perhaps a better question would be how does affordability compare to all sectors, not just grocery prices ?

5 Appendix

5.1 Data Cleaning

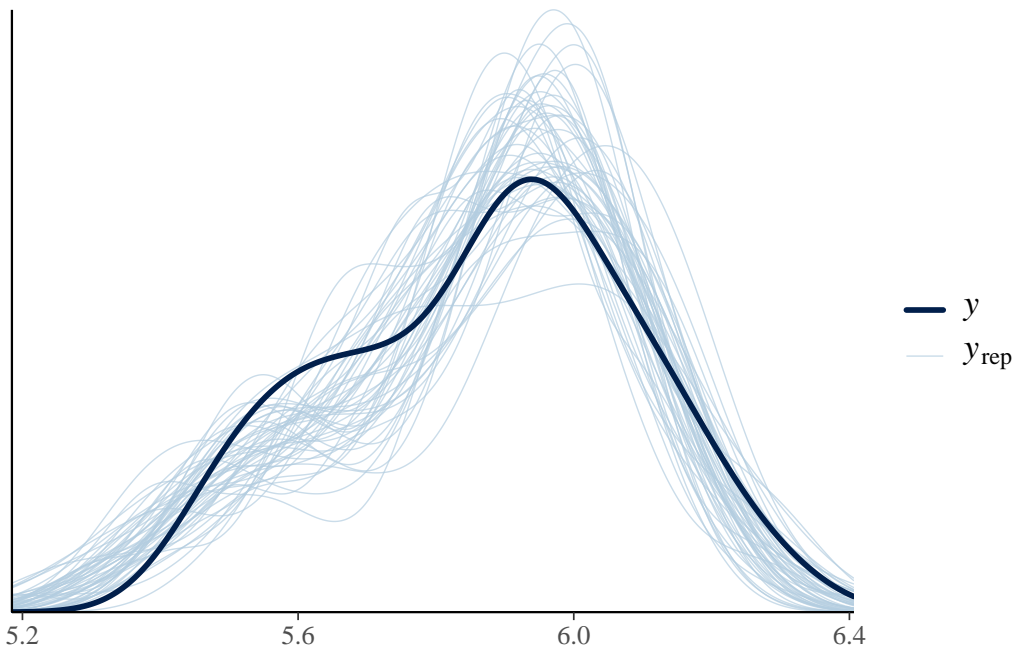
This section will be explaining the data cleaning process documented in `/scripts/03-clean_data`.

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`.

5.2 Methodological Exploration of CPI Data Collection

5.3 Postrior Predictive Check



Model Info:


```

function:      stan_glm
family:        gaussian [identity]
formula:       Affordability ~ CPI + Time
algorithm:     sampling
sample:        4000 (posterior sample size)
priors:        see help('prior_summary')
observations:  84
predictors:    3

```

Estimates:

	mean	sd	10%	50%	90%
(Intercept)	12.0	0.6	11.2	12.0	12.7
CPI	0.0	0.0	-0.1	0.0	0.0
Time	0.0	0.0	0.0	0.0	0.0
sigma	0.1	0.0	0.1	0.1	0.1

Fit Diagnostics:

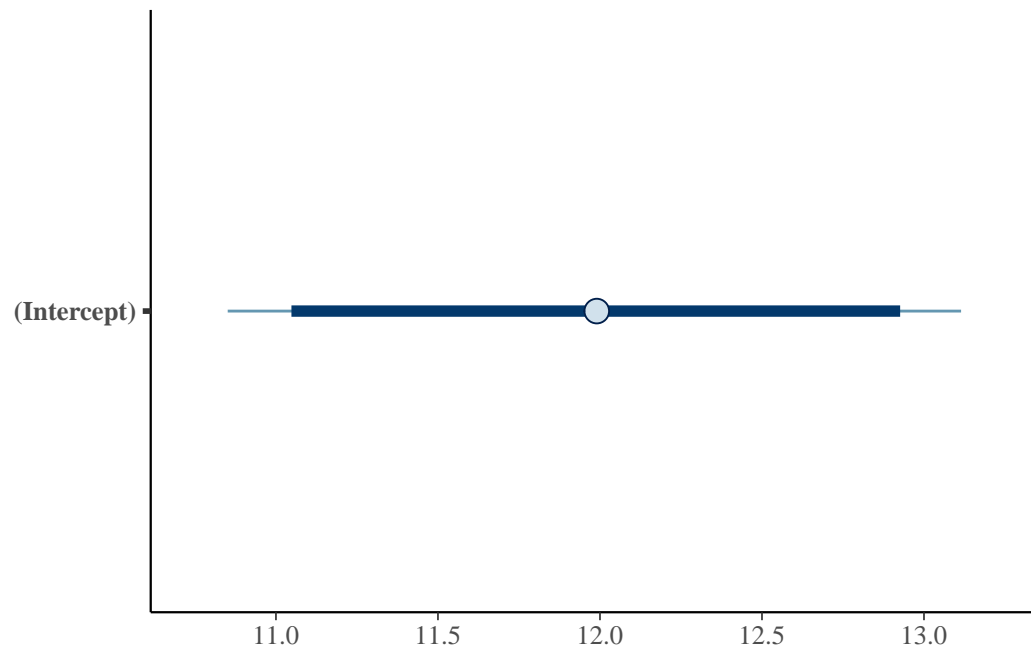
	mean	sd	10%	50%	90%
mean_PPD	5.9	0.0	5.8	5.9	5.9

The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for detail

MCMC diagnostics

	mcse	Rhat	n_eff
(Intercept)	0.0	1.0	1606
CPI	0.0	1.0	1598
Time	0.0	1.0	1619
sigma	0.0	1.0	2518
mean_PPD	0.0	1.0	3100
log-posterior	0.0	1.0	1440

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample si



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>.