My title*

My subtitle if needed

Rahma Binth Mohammad

August 7, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

2 Data

2.1 Sources

A good indicator of inflation is the Consumer Price Index (CPI), as it measures the change in price using a fixed basket of goods and services. Statistics Canada is a government agency that collects information on Canada's economy (cite). Thus, it has the data that provides the CPI for numerous years. Statistics Canada also provides us with other datasets such as the average household income and the average household expenditures, in a year for all the provinces. We have chosen to look at Ontario solely, as we hold firmly Ontario behaves very differently due to its high population in relation to the other provinces. The data for Canada as a whole is complicated as there are many missing values for some provinces. Our analysis and the data provided by Statistics Canada makes it possible to reproduce our model for different provinces using the github repository. While there are other datasets available, these three datasets are provided by the same source and has data for provinces across Canada. The expenditure and cpi data also provides the information for three additional specific categories, food, shelter, and recreation, making it easier to isolate and answer our research question. Section 2.3 will illustrate the visualised data.

This paper's objective is to address the following questions using data collected from a government census: (1) Does changes in prices affect expenditure? (2) Does changed in prices affect expenditure as income changes?

---

*Code and data are available at: https://github.com/rahmabinth/consumer-behaviour

2.1.1 Spending Data

`spending.csv`

This dataset contains the average total expenditures of the Ontario resident for the years 2010 to 2021, excluding years 2018 and 2020. It also contains the breakdown of the expenditures, food purchased (from restaurant and stores), shelter (rented and owned), utilities such as water, fuel, and electricity for the principal accommodation, clothing, transportation (private and public), healthcare, recreational goods and services.

2.1.2 Inflation Data

`cpi.csv`

The dataset provides CPI for every year. It is not seasonally adjusted for changes as it is an annual average. Thus, it is actual price paid by consumers and not adjusted for seasonal fluctuations. The base year for the CPI calculation is 2002 where CPI is 100. Thus, if our CPI is 120 for a year, it indicates a 20% increase since 2002. This value is useful as an indicator for inflation, as it provides the increase and comparison between years, rather than a absolute price values.

2.1.3 Income Data

`income.csv`

Throughout the province of Ontario, there exits people in many different financial situations. Statistics Canada provides us with this dataset that contains the average number of people in an income category. For example, "persons with income under \$5000", "persons with income of \$5000 and over", "persons with income of \$10 000 and over." Please note as for 2020, COVID-19 government income support and benefits are included in income estimates.

2.2 Methodology

The language and environment used for this analysis is R (R Core Team 2020), alongside the tidyverse (cite), rstanarm (cite), cansim (cite), knitr (cite), ggplot2 (cite), scales (cite), modelsummary (cite), and marginaleffects (cite) packages.

We downloaded the data directly from Statistics Canada using the cansim package. We cleaned it, identified missing values and wrote csv files to make it easily accessible using R studio.

2.3 Variables

For our analysis, we have to use all three dataset values and thus, we chose to use data for the years 2010 to 2021, excluding 2018 and 2020, given the data was only available in all given datasets for these years. However our model does predict the behaviour in 2018 and 2020, using the other years' data. We also created numerous csv files that isolated the specific data. **?@sec-appendix** thoroughly explains the structure of the csv files.
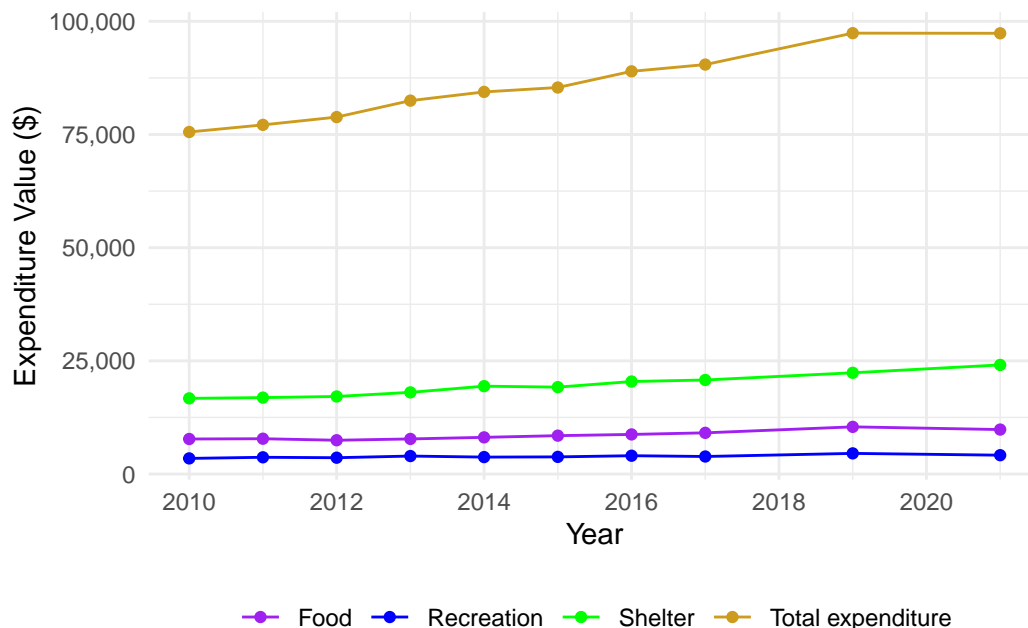

Figure 1: Household spending per year

From the spending dataset, we chose to select the values for all food expenditures, all shelter expenditures, recreation expenditures, and total expenditures. This choice was made because these are principal expenditures for everyone. Recreation was added to this mix, to identify how consumers behave when inflation is high. Figure 4 indicates this relationship. In Figure 1, we observe a steady increase in recreation expenditures, while food, shelter, and overall expenditures show a more rapid growth. Note for simplicity, we did not include the other categories used for total expenditures. For all four of the categories, the mean and median values were close to each other, indicating the data is not skewed and did not have many outliers.

From the CPI dataset, we chose to look at the annual average to keep it consistent with the annual average of expenditures. The variables

From the income dataset, we chose to isolate the values for the category "all persons with income". This indicate all tax fillers and dependants that earn income from all sources. This was done to keep our analysis consistent.
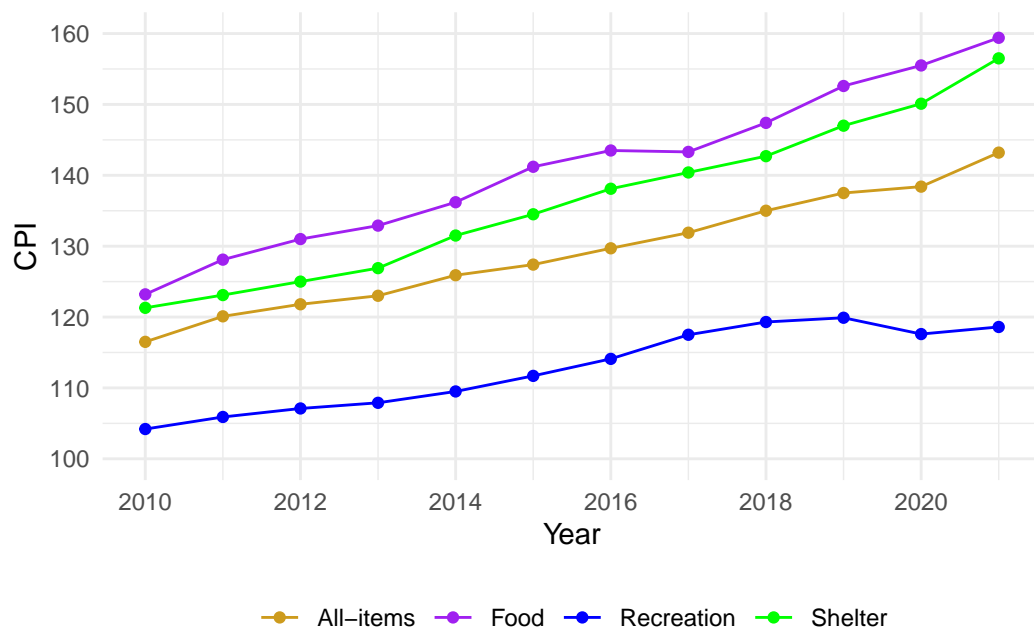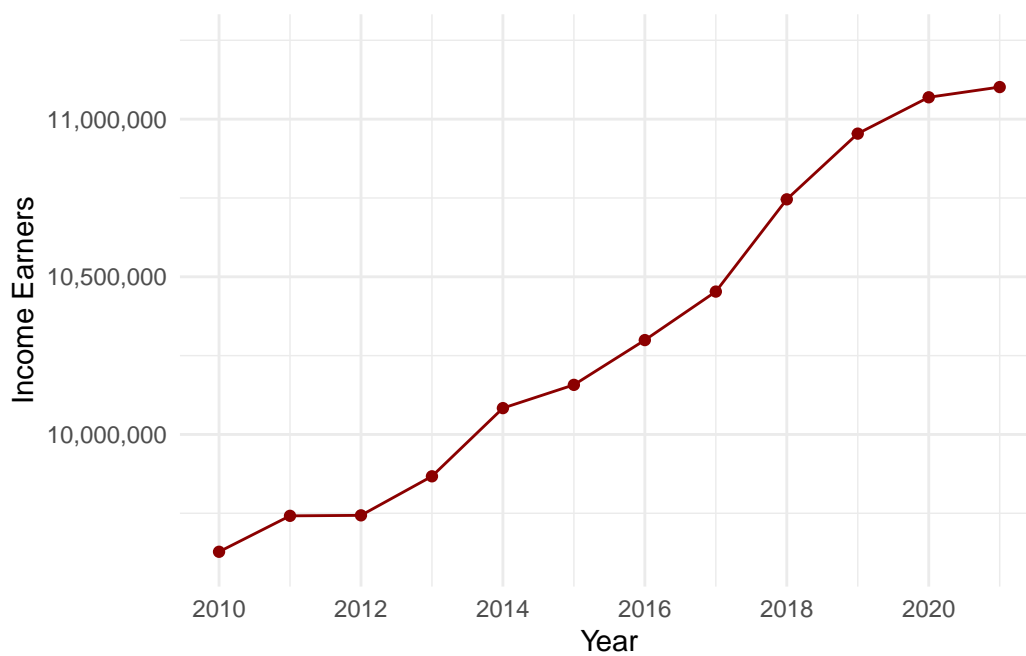
3

Figure 2: Consumer Price Index (CPI) per Year

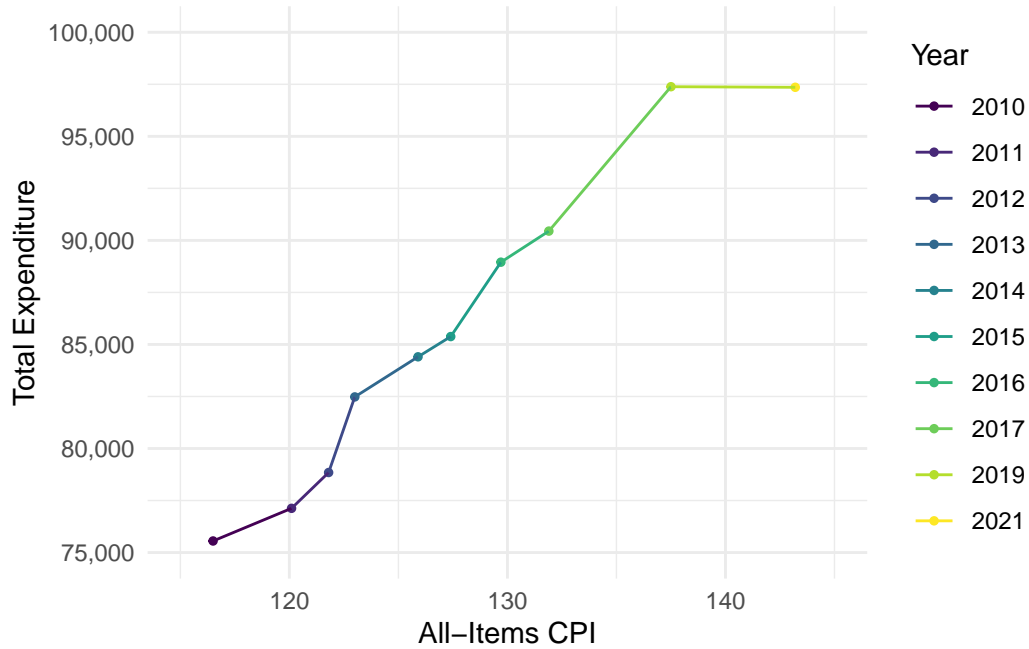Figure 3: Number of Income Earners

4

Figure 4: CPI and total expenditure over the year

3 Model

The goal of our model is to forecasting the total expenditure in a household in a year, based solely on a the Consumer Price Index, and the number of people in Ontario earning an income. We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

3.1.1 Simple linear regression

The model we employ is:

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \beta_0 + \beta_1 \times \text{cpi}_i + \epsilon \tag{2}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\sigma \sim \text{Exponential}(1) \tag{5}$$

Where:

- y_i as the total expenditure in Canadian dollars in year i.
- cpi_i is the annual Consumer Price Index (CPI) for year i

3.1.2 Multiple linear regression

The model we employ is:

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{6}$$
$$\mu_i = \beta_0 + \beta_1 \times \text{cpi}_i + \beta_2 \text{income}_i + \epsilon \tag{7}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{8}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{9}$$
$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{10}$$
$$\sigma \sim \text{Exponential}(1) \tag{11}$$

Where: * y_i as the total expenditure in Canadian dollars in year i. * cpi_i is the annual Consumer Price Index (CPI) for year i * $income_i$ is the number of people in Ontario earning an income

3.1.3 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results
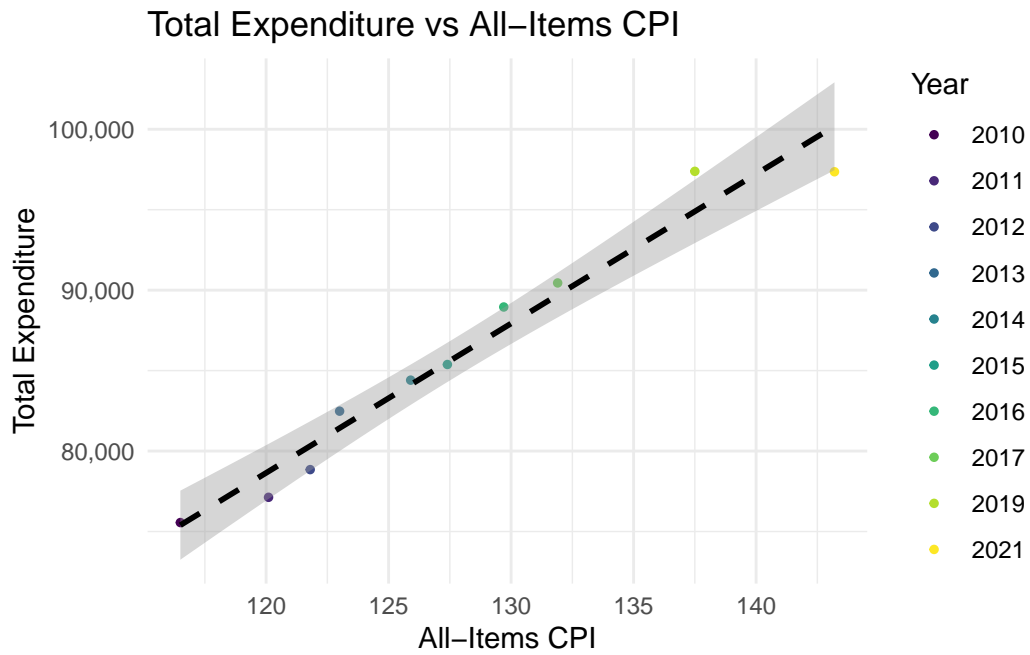
Our results are summarized in Table 1.
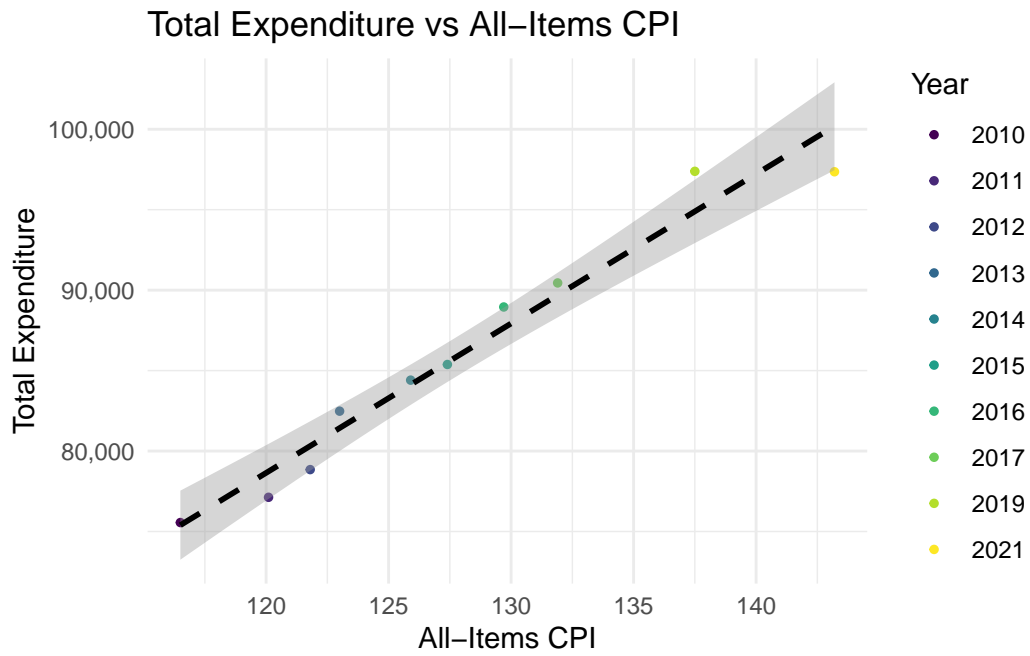
Figure 5: Linear model

Figure 6: Linear model

Table 1: Explanatory models of

	First model	Second model
(Intercept)	$-32\,899.89$	$-59\,979.55$
	(9842.07)	$(18\,660.19)$
cpi	928.77	225.52
	(76.80)	(437.28)
income		0.01
		(0.01)
Num.Obs.	10	10
R2	0.944	0.955
R2 Adj.	0.920	0.932
Log.Lik.	-88.550	-87.273
ELPD	-92.0	-91.1
ELPD s.e.	3.1	2.3
LOOIC	184.1	182.1
LOOIC s.e.	6.1	4.6
WAIC	182.4	180.5
RMSE	1501.29	1326.89

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

Posterior predictive check

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

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