

You have more money in your wallet. This does not mean you will spend extra on food*

Examining the relationship between disposable income and food expenditure for American consumers

Ahnaf Alam

April 14, 2024

Abstract

We examined the relationship between changes in food spending and increases in disposable income. Utilizing quarterly country-level data from the Federal Reserve Economic Data (FRED) spanning a 15-year period from 2007 to 2022, we employed both a simple linear regression model and a multiple regression model. Our analysis revealed a economically insignificant increase of \$20 million in food spending for every billion-dollar rise in disposable income. From a policy perspective, our findings indicate that solely targeting disposable income is unlikely to substantially impact food spending behaviors. Therefore, policies designed to boost consumer spending should explore alternative approaches.

Table of contents

1. Introduction	2
2. Data	3
2.1. Software and R-packages	3
2.2. Incorporating FRED data	3
2.3. Dataset characteristics	4
2.4. Measurement	4
3. Model	5
3.1. Model set-up	5
3.1.1. Simple linear regression	6

*Code and data are available at https://github.com/AhnafAlam1/income_and_expenditure

3.1.2. Multiple linear regression	6
3.2. Model justification	10
4. Results	11
5. Discussion	12
5.1. Accounting for savings	12
5.2. Interpreting consumer behavior	12
5.3. Factoring in elasticities	13
5.4. Weaknesses and next steps	13
Appendix	15
A. Additional data details	15
A.1. Datasheet	15
B. Model details	22
B.1. Posterior predictive check	22
B.2. Diagnostics	23
References	24

1. Introduction

Over the past thirty years in the US, wages have consistently risen, while tax rates have remained relatively stable. This has resulted in people having more disposable income, which is the money left after taxes are deducted from one's income. Disposable income is crucial to consider as it reflects the actual amount of money available for spending. Studies indicate that when individuals have higher disposable income, they generally spend more (Friend 1946). We are particularly interested in examining this relationship in the context of food.

Food is something everyone needs to live. People might change what kinds of food they like over time, but everyone still needs to eat. In this project, we're looking at how closely linked disposable income is to how much people spend on food. This matters because if there's a strong connection, it suggests that when people make more money, they tend to spend more on basic goods and services like food. Further, this could help predict how much people will spend on luxury items in the future. For instance, if we find a strong connection, it might mean that when people earn more, they spend a bigger chunk of their money on essentials like food, leaving less for luxury items.

Numerous studies, notably summarized by Ernst Engel's law, suggest that as income and family size increase, spending on food goes up. However, the proportion of income spent on food decreases

(Lewbel 2008). This means wealthier families tend to spend more on food, but it is a smaller part of their overall income. Taking inspiration from Engel’s law, we want to see how food spending changes when we account for other spending categories instead of just income differences across families. Additionally, most past studies used income as the predictor variable, but we’re using disposable income because we think it gives a better idea of overall wealth. The estimand of this paper is the measure of how much food spending changes (in billions of USD) when disposable income increases by a billion USD.

In Section 2, we discuss the data we used for this project, including information of the software and packages used. This is followed by a explanation of the models that we used, which can be found in Section 3. In Section 4, we present the results of the models and Section 5, we interpret these results, and consider how savings, consumer preference and elasticities play into this discussion on food expenditure and income. We finish the project with weaknesses and potential areas of improvement in the future.

2. Data

2.1. Software and R-packages

We create this project using statistical software, R (R Core Team 2023). For cleaning and re-purposing the data, we used `tidyverse` (Wickham et al. 2019) package and graphs, we relied on `ggplot2` (Wickham 2016). The data used in this paper comes from `fredr` (Boysel and Vaughan 2021) package. We further used `rstanarm` (Goodrich et al. 2022) for modelling. Lastly, we used `kableExtra` (Zhu 2021) and `viridis` (Garnier et al. 2024) for aesthetics purposes.

2.2. Incorporating FRED data

We gathered our data from Federal Reserve Economic Data (FRED), an extensive online database housing a multitude of economic time series, both at the US national and international levels. Utilizing the `fredr` package, we integrated six distinct datasets into one dataframe. These datasets were:

- Real personal consumption expenditure: Food
- Real disposable personal income
- Real personal consumption expenditure: Durable Goods
- Real personal consumption expenditure: Nondurable Goods
- Real personal consumption expenditure: Services
- Real personal consumption expenditure: Healthcare

For a detailed breakdown of each dataset, refer to Table 1. Across all datasets, a few consistent features are notable. Firstly, we confined our data selection to the quarterly period spanning from 2007 to 2022. We anchored our starting point in 2007 due to the unavailability of durable/nondurable goods data prior to 2007. We were interested in a comprehensive 15-year modeling period, which culminated in 2022.

All datasets underwent seasonal adjustment and were indexed to 2017 dollars. Seasonal adjustments mitigate the influence of periodic fluctuations, such as strikes or unusual weather events, which can obscure genuine economic trends. By referencing all data to the 2017 price level, we adjusted for inflation over time, facilitating accurate comparisons across various periods. Additionally, utilizing real economic data rather than nominal figures allowed for valid comparisons between different categories across time frames.

2.3. Dataset characteristics

Each dataset is structured with five distinct columns, the most significant being “date” and “value.” The “date” column denotes the specific day of data collection. In the case of quarterly data, observations are restricted to the first day of January, April, July, and October. Notably, FRED does not provide data for August or December quarters. This omission is primarily due to practical considerations, as certain data may only become available after the quarter’s end, rendering updates before then unproductive.

Furthermore, the “value” column records expenditures in billions of US dollars. For a comparative analysis of each category, please refer to Figure 2. The cleaned data utilized for modeling and analysis is detailed in Table 2. The datasheet is accessible in Section A.1 for further reference.

2.4. Measurement

FRED datasets, essentially time series variants, are compiled by the US Federal Reserve and sourced from government agencies like the U.S. Census Bureau, Bureau of Labor Statistics, and U.S. Bureau of Economic Analysis (BEA), among others. These datasets are aggregated at the country level. We use national-level data because it provides a broader perspective on general economic trends in the U.S. All five datasets we have used originate from the BEA. The BEA gathers data from over 360 surveys and collections sponsored by various Federal and private agencies. Upon receiving these datasets, BEA makes adjustments to the raw survey data. For instance, they fill in missing gaps in the datasets. For example, when estimating changes in private inventories for Gross Domestic Product (GDP), the primary survey is the Census Bureau’s Monthly Wholesale Trade Survey does not include inventories from nonmerchant sellers. Therefore, BEA incorporates a separate survey to fill this gap, giving us a more comprehensive view of the overall economic conditions.

FRED collaborates with both public and private organizations, such as the BEA, to populate its comprehensive database. The goal of FRED is to serve as a centralized data bank where you can find information on all facets of the U.S. economy, eliminating the need to search various agencies individually for data. However, FRED does make further adjustments to BEA datasets before releasing them. For example, when aggregating data through averaging, summing, or end-of-period calculations, FRED typically overlooks missing values, except for the first and last observations of each month (“Getting to Know FRED” 2024).

Missing values commonly occur during statutory holidays, when federal offices are closed. Consequently, during such weeks, FRED reports data for only six days, excluding the holiday. End-of-period calculations are adjusted accordingly, based on the remaining available days within the month, minus one.

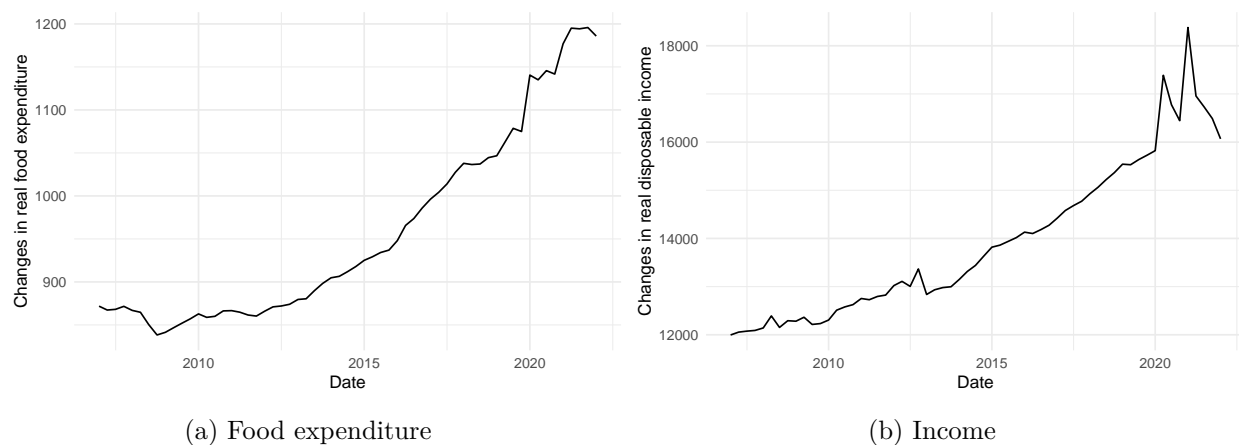


Figure 1: Levels of real consumption expenditure and income expenditure, in billions of dollars, chained to 2017 prices

3. Model

In this section, we briefly discuss Bayesian models that are being used in this analysis. Background details and model diagnostics can be found under [Appendix B](#).

3.1. Model set-up

Using `rstanarm` library, we evaluated two Bayesian model, with one being simple linear regression, and another being multiple linear regression. The simple linear regression explores whether an

Table 1: Description of the data variables

Expenditure variable	Description
Date	Date of data collected
Durable goods	Durable goods are goods that are more for future consumption than immediate consumption. These types of goods provides utility over a length of period. Examples include machinery, tools, appliances among others
Non-durable goods	Durable goods are anything that are generally consumed within a short period of time. Examples include food, clothing, cosmetics etc
Food	Expenditure in food by all Americans in a time period
Disposable income	Refers to total income that is available to individuals for consumption after deducing taxes
Healthcare	The category reports on total expenditure on healthcare services, including medical treatments, medicine cost, physician services among other services
Services	This category encomapasses a variety of services, including education, transportation, utilities, hospitality and many others

increase in income leads to increase in expenditure in food. Multiple linear regression evaluates the same topic, however, controlling for various other predictors.

3.1.1. Simple linear regression

Define y_i as the expenditure in food in year i . Then $income_i$, is level of disposable income in year i , both in billions of US dollars.

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

$$\mu_i = \beta_0 + \beta_1 \times income_i + \epsilon \quad (2)$$

$$\beta_0 \sim \text{Normal}(0, 279) \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 0.17) \quad (4)$$

$$\sigma \sim \text{Exponential}(0.009) \quad (5)$$

$$(6)$$

3.1.2. Multiple linear regression

Define y_i as the expenditure in food in year i . Then $income_i$, is level of disposable income in year i . Model further controls for durable goods with $durable_i$, non-durable goods with $nondurable_i$, levels

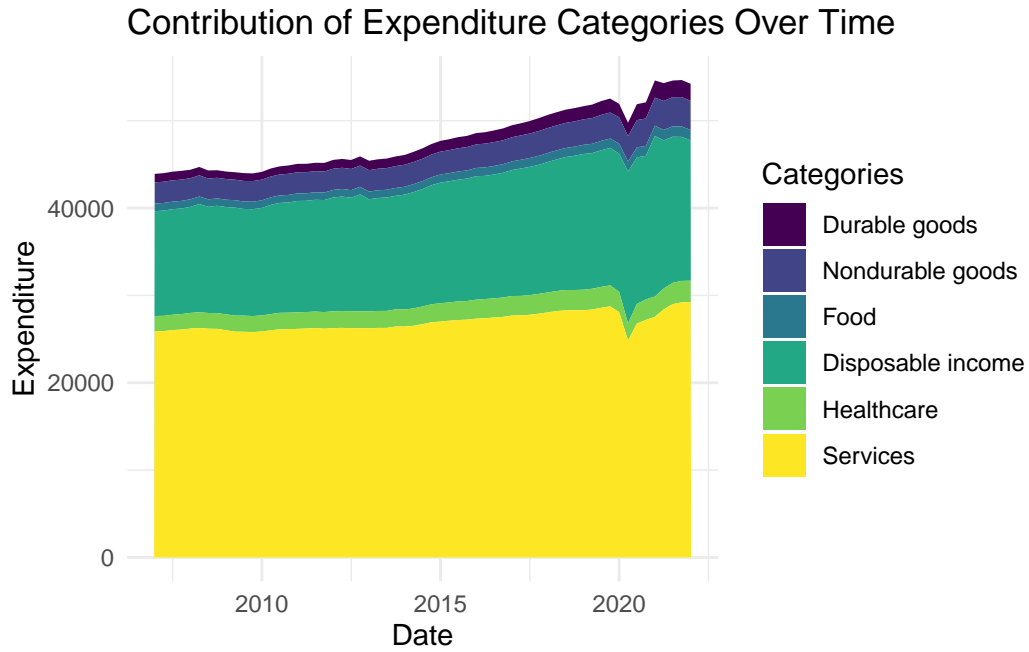


Figure 2: Expenditure for different categories, in billions of 2017 dollars

Table 2: Cleaned data showing real expenditure by different categories (in billions of 2017 US dollars)

Date	Durable goods	Non-durable goods	Food	Disposable income	Healthcare	Services
2007-01-01	969.90	2435.00	871.80	11995.90	1736.72	25900.00
2007-04-01	980.10	2429.90	867.30	12055.30	1745.91	25913.00
2007-07-01	992.10	2437.50	868.10	12075.60	1762.55	26024.00
2007-10-01	999.50	2435.50	871.60	12090.30	1770.73	26088.00
2008-01-01	967.20	2416.60	866.90	12141.60	1788.97	26202.00
2008-04-01	960.30	2420.40	864.70	12391.20	1793.99	26273.00
2008-07-01	927.90	2385.10	850.40	12152.80	1800.01	26186.00
2008-10-01	859.60	2362.30	838.30	12291.70	1804.94	26167.00
2009-01-01	861.10	2361.40	841.40	12282.00	1817.91	26012.00
2009-04-01	854.90	2347.30	846.70	12364.40	1838.01	25848.00
2009-07-01	896.40	2354.70	851.90	12214.70	1849.31	25830.00
2009-10-01	875.30	2362.30	857.00	12232.60	1840.44	25795.00

^a Note: The table displays data from 2007-2009 to illustrate the of the aggregated clean data. Including the entire dataset would be redundant and unnecessarily clutter the display.

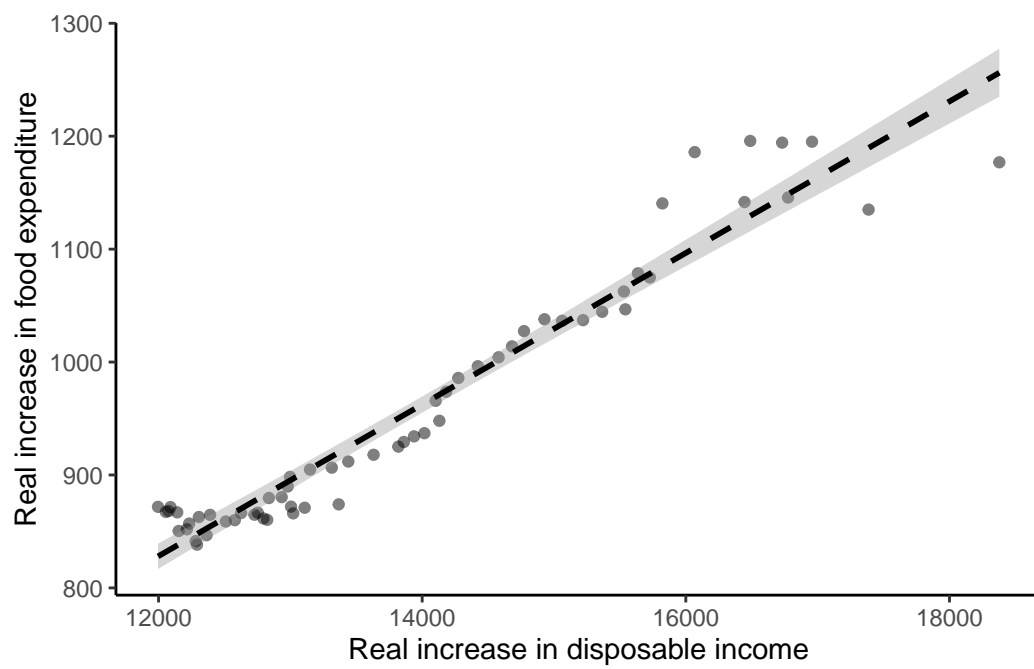


Figure 3: Relationship between increases in disposable income and increases in food expenditure, between 2007 and 2022.

Table 3: Summary results for both models

	Model 1	Model 2
(Intercept)	24.07 (30.88)	−97.50 (114.73)
income__expenditure	0.07 (0.00)	0.02 (0.00)
durable__expenditure		−0.15 (0.04)
nondurable__expenditure		0.47 (0.04)
healthcare__expenditure		0.00 (0.04)
services__expenditure		−0.01 (0.01)
Num.Obs.	61	61
R2	0.936	0.992
R2 Adj.	0.933	0.990
Log.Lik.	−289.419	−224.943
ELPD	−292.9	−233.0
ELPD s.e.	8.0	8.0
LOOIC	585.7	466.0
LOOIC s.e.	16.0	16.0
WAIC	585.6	464.3
RMSE	27.49	10.36

of health care expenditure with $healthcare_i$ and levels of expenditure in services with $services_i$. All the variables are in billions of US dollars.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (7)$$

$$\mu_i = \beta_0 + \beta_1 \text{income}_i + \beta_2 \text{durable}_i + \beta_3 \text{nondurable}_i + \beta_4 \text{healthcare}_i + \beta_5 \text{services}_i + \epsilon \quad (8)$$

$$\beta_0 \sim \text{Normal}(0, 279) \quad (9)$$

$$\beta_1 \sim \text{Normal}(0, 0.17) \quad (10)$$

$$\beta_2 \sim \text{Normal}(0, 0.84) \quad (11)$$

$$\beta_3 \sim \text{Normal}(0, 0.95) \quad (12)$$

$$\beta_4 \sim \text{Normal}(0, 1.25) \quad (13)$$

$$\beta_5 \sim \text{Normal}(0, 0.27) \quad (14)$$

$$\sigma \sim \text{Exponential}(0.009) \quad (15)$$

3.2. Model justification

We anticipate a positive correlation between income and food expenditure. As depicted in Figure 3, an increase in disposable income corresponds to a similar rise in food expenditure. Additionally, Figure 1 compares these variables over time, revealing parallel growth patterns. Both exhibit a consistent, exponential increase between 2010 and 2020, followed by a divergence after 2020, likely due to the COVID-19 downturn.

Sekhampu (2012) analyzes data from South Africa and employs a multiple linear regression model. The study concludes that household income, size, age, and employment status exert a significant positive influence on food expenditure. Likewise, Liu, Renwick, and Fu (2019) utilizes Chinese data and a two-stage least squares estimator. Their findings reveal that for every 1,000 yuan increase in per-capita income, food expenditure rises by 61 yuan. Drawing from our exploratory data analysis and existing theories, we hypothesize a positive correlation between food expenditure and income levels. Therefore, we believe linear models would be appropriate for modelling our hypothesis.

We utilize two linear models in our analysis: simple linear regression and multiple regression. Simple linear regression enables us to easily quantify the linear association between food expenditure and disposable income, as illustrated by Figure 3. However, this model has limitations. Most notably, it does not account for confounding variables, and it requires us to make unrealistic assumptions about the independence of the error term (ϵ) and the predictor variable, income. This assumption implies that food expenditure and income are entirely independent of each other, which is problematic. Factors like economic downturns, income levels, and other variables can correlate with our model variables.

To address these issues, we employ a multiple linear regression model. This model helps us control for potential confounding variables, such as expenditure on non-durable goods, which may influence consumer expenditure on food. Using multiple linear regression allows us to better isolate the effect of income on food expenditure.

Between the two models, the multiple linear regression model may be more preferable than the simple linear regression model. Simple linear regression requires unrealistic assumptions that the multiple linear regression model relaxes. Furthermore, posterior predictive checks conducted for both models suggest that the multiple regression model may offer a better fit than the simple linear regression model. For more details, please refer to Section B.

4. Results

Our findings are outlined in Table 3. In simple linear regression, we observed an estimated intercept of 24.07 billion USD. Additionally, for every 1 billion USD increase in disposable income, there's a corresponding rise in food expenditure of \$70 million USD. The small standard deviation suggests a low level of variability and uncertainty.

In multiple linear regression, we found an estimated intercept of -97.40 billion USD. With each \$1 billion USD increase in income, food expenditure increased by 20 million USD. Similar to the simple regression model, the standard deviation is negligibly small.

Both models exhibit similar adjusted R^2 values, yet the second model slightly outperforms the first with a higher value (0.993 vs 0.990), indicating a superior fit. Notably, the root mean square error (RSME) for the second model is substantially lower than that of the first (10.36 vs 27.49), indicating enhanced predictive accuracy.

Figure 1 illustrates quarterly changes in disposable income and food expenditure from 2007 to 2022. Both variables show consistent growth trends, with 33.9% increase in disposable income and 32.5% increase in expenditure in food in this period. We see a notable surge in disposable income in 2020 due to COVID-era relief efforts. However, this increase was followed by a decline as restrictions eased, allowing for increased spending towards the end of 2021.

Figure 2 displays the distribution of expenditure across categories, highlighting durable and non-durable goods as the largest expenditures. This aligns with expectations given the nature of these commodities. Following closely are food expenditure and disposable income, reflecting a strong correlation evident in Figure 3, where a clear linear relationship is observed with small standard deviations. Finally, healthcare and services expenditure are clustered together.

5. Discussion

Table 3 has discovered that for every billion-dollar rise in disposable income, the average expenditure on food increases by an amount ranging from 20 million USD to 70 million USD. This validates our hypothesis, suggesting a direct correlation between income and food consumption. According to the models, this translates to an economic significance of $\$70,000,000/\$1,000,000,000$, equivalent to a 7% increase. In the following section, we talk about the reasons behind relatively low economic significance, exploring the impact of savings, consumer behavior and elasticities, followed by potential weaknesses in our paper.

5.1. Accounting for savings

A factor that might clarify the low economic significance is that consumers tend to opt for saving their additional income rather than spending it. This aligns with existing economic theories, as suggested by (Krugman, Obstfeld, and Melitz 2022), which indicate that consumption increases only slightly in response to income growth, as individuals prefer to set aside funds for “rainy day”. This tendency becomes apparent during periods of heightened economic uncertainty, when people tend to prioritize saving as a precautionary measure for the future. Mody, Ohnsorge, and Sandri (2012) explores economic downturns and savings rates further, finding that in advanced economies, income uncertainty correlates with higher household savings. Additionally, Lusk and McFadden (2021) explores spending behavior, particularly focusing on food expenditures during the COVID-19 pandemic, revealing a typical decrease in food spending during recessions. We see this concept playing out in Figure 2, where expenditure decreases on all consumption sectors, right around the start of COVID-19 pandemic in 2020. These findings suggest that an elevated level of saving could indeed be a reason why individuals opt to reduce their expenditure on food and prioritize saving for future consumption.

5.2. Interpreting consumer behavior

Current research suggests that higher income tends to result in increased expenditure on durable goods, surpassing that of other consumable products. Consumers often allocate the majority of their additional income towards items that provide utility over an extended period rather than those consumed in a single instance. Using a linear probability function, Lee (1964) showed that consumers typically increase their spending on durable goods in response to income growth and resist reducing it even when income decreases. This pattern was evident during the COVID-19 pandemic, where heightened disposable income led to above-average spending on durable goods such as furniture, appliances, motor vehicles, and recreational items (Tauber and Van Zandweghe 2021). Therefore, it is plausible that food expenditure remains relatively stable compared to expenditures on durable goods, which may fluctuate more significantly in response to changes in income.

5.3. Factoring in elasticities

The income and price elasticities of food are crucial factors in understanding why there's not much change in food spending despite changes in income. Income elasticity measures how much food demand shifts when income changes, while price elasticity assesses the impact of price changes on food demand.

A study by Almås, Haushofer, and Kjelsrud (2019) in Kenya, which randomly allocated cash transfers, found the income elasticity of food to be 0.87. Building on this, Purcell and Raunikaar (1967) found that income elasticities for items like poultry, fish, seafood, and fruits were below 0.5. This indicates that food is a normal commodity: as income increases, the demand for food also increases, but only slightly.

Moreover, Andreyeva, Long, and Brownell (2010) conducted a meta-analysis on the price elasticities of various food categories and discovered that essential items such as fish, beef, poultry, fruits, and vegetables had elasticities below 0.75. This suggests that when food prices rise, the decrease in demand is only slight.

In essence, these findings show that there's only a modest increase in food demand as income rises and a moderate decrease as food prices go up. Consequently, changes in income don't lead to significant shifts in food demand.

5.4. Weaknesses and next steps

Our model's weaknesses stem from the data we utilized—quarterly data spanning from 2007 to 2022. This period was chosen to explore changes in the relationship between food expenditure and income, however, it also captures economic impact of COVID-19, which was not our intention. With the pandemic's onset in late 2020, disposable income surged due to social benefit programs (see Figure 1b), while food expenditure remained relatively stable due to its essential nature (see Figure 1a). This led to erratic swings in income and predictable changes in food spending, distorting our data.

This distortion poses challenges for accurately estimating our target variable or estimand. Unforeseen economic fluctuations during the pandemic may skew the true relationship we seek to understand, reducing the model's external validity. The unique nature of the pandemic's shocks further limits the model's generalizability, especially considering variations in social benefit policies and lockdown measures among different countries. Consequently, applying our model to another country within the same timeframe may yield unreliable results due to differing policy responses to COVID-19.

In future iterations, a narrower timeframe excluding such large-scale global shocks would be prudent. This approach aims to enhance the model's external validity and clarify the estimand we are pursuing.

Appendix

A. Additional data details

A.1. Datasheet

Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

It was created in order to analyze how food expenditure changes when income fluctuates. While all the datasets can be found individually on FRED website, they however, are not complied together and my datasets fills that void. The dataset complies different types of expenditure into one data frame, which I believe can help to answer the proposed question.

Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?

Ahnaf Alam, an undergraduate student at University of Toronto.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

No direct funding was received for this project.

Any other comments?

No.

Composition

What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

Each row of the dataset composed on valuation in billions of US dollars, on a specific date. The data provides information of cumulative spending habits by Americans throughout the year.

How many instances are there in total (of each type, if appropriate)?

There are about 366 instances in total.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).

The dataset is a sample, however it isn't random. The larger dataset consists of all the observations for expenditure on a daily basis from the time datasets were made available to FRED database. In that sense, sample is representative of the larger dataset as we see the patterns of consumptions and expenditure over 25 year period match the patterns we see our sample.

What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.

Each instance consists of value of expenditure in billions of USD, across different categories.

Is there a label or target associated with each instance? If so, please provide a description

Yes, the unique consists of specific date

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.

There are no missing values in the dataset.

Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

Yes, using the 'year' column.

Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

No.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

No.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

The data is self-contained. The data will exist and won't change over time.

Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.

No. These are publicly available data, released by public organizations

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

No.

Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

No.

Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.

No. The data reports on national level.

Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.

No.

Any other comments?

No.

Collection process

How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

The data was collected from FRED website, using **fredr** package. The data was not directly observable but based on audits, taxes and surveys.

What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?

We downloaded the data using **fredr** package on R.

If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?

The dataset was collected based on specific years on a quarterly basis.

Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?

Ahnaf Alam and no one else.

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

The dataframe was created over period of one week.

Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

No.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?

No. We relied on third-party website for in all cases.

Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

No.

Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

No.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

Consent was not needed as we are dealing with data on country level.

Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

No.

Any other comments?

No.

Preprocessing/cleaning/labelling

Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.

Yes. The data was cleaned. There were no missing values in the dataset. From the raw data, we only selected columns that were pertinent to our question.

Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data.

No. However, if one were to run the codes available on 01-download_data.R file, they can access the raw data.

Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.

R was used.

Any other comments?

No.

Uses

Has the dataset been used for any tasks already? If so, please provide a description.

Not that I am aware of.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

No.

What (other) tasks could the dataset be used for?

we could use different types of model to see how they compare to the ones we have performed in our paper.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?

No.

Any other comments?

No.

Distribution

Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

The dataset will be available on Github for later uses.

How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

It will be distributed through Github.

When will the dataset be distributed?

The dataset is available now.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

No.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

None.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

No.

Any other comments?

No.

Maintenance

Who will be supporting/hosting/maintaining the dataset?

Ahnaf Alam

How can the owner/curator/manager of the dataset be contacted (for example, email address)?

ahnaf.alam@mail.utoronto.ca

Is there an erratum? If so, please provide a link or other access point.

No.

Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?

No.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.

Not applicable.

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.

No.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.

Pull request on Github.

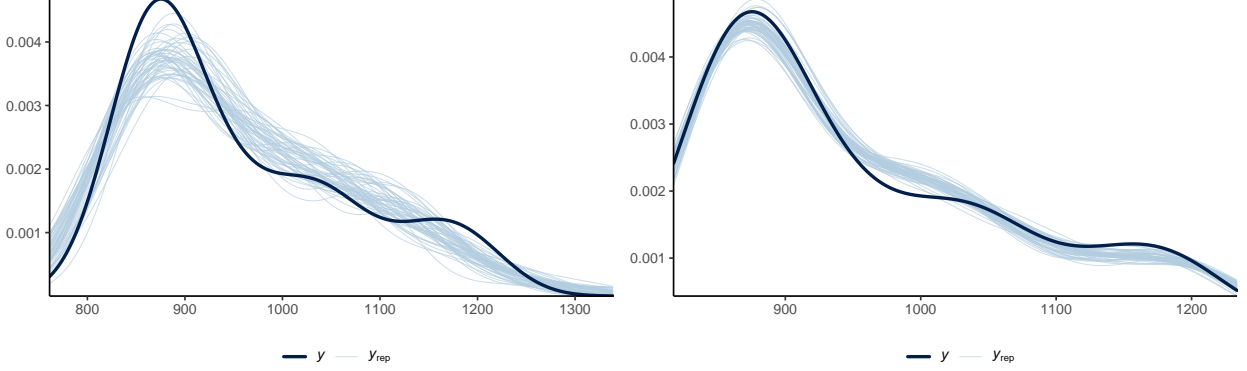
Any other comments?

No.

B. Model details

B.1. Posterior predictive check

We conducted posterior predictive checks for both models, with simple linear regression on Figure 4a and multiple linear regression on Figure 4b. We wanted to understand the fit of the model and whether, can explain the observed data. For both graphs, y is the observed data and y_{rep} is the predicted data. Figure 4a shows that the Bayesian model fits the observed data well, however, there few spots where observed data does not track predicted data very well. We do not see this issue in multiple linear regression in Figure 4b, where predicted and observed track each other well. Therefore, multiple linear regression may be more preferable than simple linear regression.

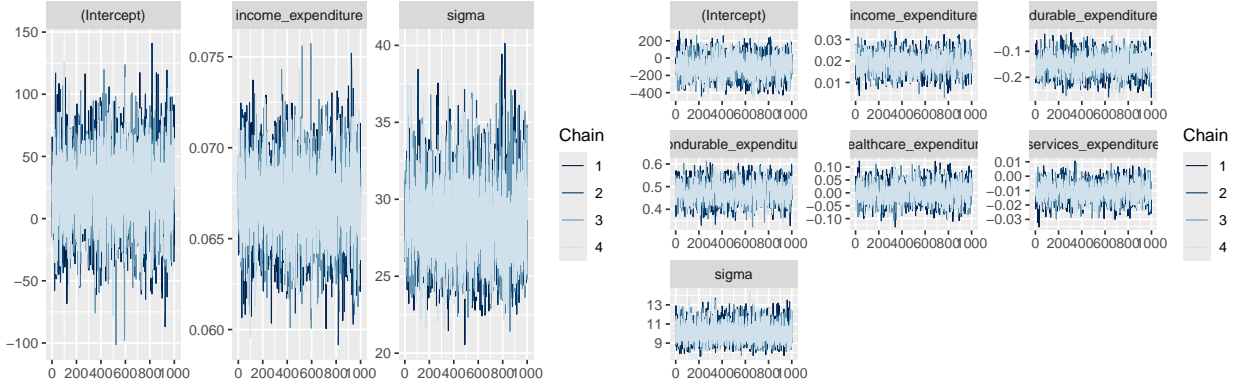


(a) Posterior prediction check of simple linear regression (b) Posterior prediction check of multiple linear regression

Figure 4: Examining how the model fits, and is affected by, the data

B.2. Diagnostics

We use ran trace plots for both models and we are interested in understanding whether the algorithm ran into any issues (Alexander 2023). Looking at the plots for both models, we do not encounter any out of the ordinary events.



(a) Trace plot of Model 1

(b) Trace plot of model 2

Figure 5: Checking the convergence of the MCMC algorithm

References

- Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. CRC Press, Taylor; Francis Group.
- Almås, Ingvild, Johannes Haushofer, and Anders Kjelsrud. 2019. “The Income Elasticity for Nutrition: Evidence from Unconditional Cash Transfers in Kenya.” National Bureau of Economic Research.
- Andreyeva, Tatiana, Michael W Long, and Kelly D Brownell. 2010. “The Impact of Food Prices on Consumption: A Systematic Review of Research on the Price Elasticity of Demand for Food.” *American Journal of Public Health* 100 (2): 216–22.
- Boysel, Sam, and Davis Vaughan. 2021. *Fredr: An r Client for the 'FRED' API*. <https://CRAN.R-project.org/package=fredr>.
- Friend, Irwin. 1946. “Relationship Between Consumers’ Expenditures, Savings, and Disposable Income.” *The Review of Economics and Statistics* 28 (4): 208–15.
- Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, et al. 2024. *viridis(Lite) - Colorblind-Friendly Color Maps for r*. <https://doi.org/10.5281/zenodo.4679423>.
- “Getting to Know FRED.” 2024. *Getting To Know FRED*. <https://fredhelp.stlouisfed.org/fred/data/understanding-the-data/how-are-missing-values-treated-in-average-sum-and-end-of-period-aggregation-methods-2/>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Krugman, Paul, Maurice Obstfeld, and Marc Melitz. 2022. *International Economics: Theory and Policy, Global Edition*. Pearson Education Limited.
- Lee, Maw Lin. 1964. “Income, Income Change, and Durable Goods Demand.” *Journal of the American Statistical Association* 59 (308): 1194–1202.
- Lewbel, Arthur. 2008. “Engel Curves.” *The New Palgrave Dictionary of Economics* 2: 1–4.
- Liu, Yuying, Alan Renwick, and Xinhong Fu. 2019. “Off-Farm Income and Food Expenditure of Rural Households in China.” *British Food Journal* 121 (5): 1088–1100.
- Lusk, Jayson L., and Brandon R. McFadden. 2021. “Consumer Food Buying During a Recession.” *Choices* 36 (3): 1–9. <https://www.jstor.org/stable/27098605>.
- Mody, Ashoka, Franziska Ohnsorge, and Damiano Sandri. 2012. “Precautionary Savings in the Great Recession.” *IMF Economic Review* 60 (1): 114–38.
- Purcell, JC, and Robert Raunika. 1967. “Quantity-Income Elasticities for Foods by Level of Income.” *Journal of Farm Economics* 49 (5): 1410–14.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sekhampu, Tshediso Joseph. 2012. “Socio-Economic Determinants of Household Food Expenditure in a Low Income Township in South Africa.”
- Tauber, Kristen, and Willem Van Zandweghe. 2021. “Why Has Durable Goods Spending Been so Strong During the COVID-19 Pandemic?” *Federal Reserve Bank of Cleveland, Economic Commentary*, no. 2021-16. <https://doi.org/10.26509/frbc-ec-202116>.

- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.