Understanding Regional and Temporal Variations in Food Prices: Insights from US Consumer Price Index and Economic Trends*

— TODO: CHANGE — A Bayesian Approach Reveals the Impact of Economic Indicators and Geographic Disparities on Price Dynamics

Tanmay Shinde

November 25, 2024

This paper examines how regional and temporal variations influence food-at-home prices across the United States, and the impact of economic factors such as purchase volume, food categories, and the Consumer Price Index (CPI) on these trends. — The analysis reveals that while food prices have generally increased iver time, certain categories, such as dairy, fresh produce, and meats, tend to exhibit greater volatility. Factors such as purchase volume, food categories, and the CPI have a considerable influence on these price trends. Notably, higher CPI values correlate with increased food prices, and food categories with higher demand or limited supply exhibit more substantial price fluctuations — . These findings provide actionable insights for policymakers and stakeholders in food economics, emphasizing the impact of region-specific market conditions and consumer purchasing behaviors on food pricing trends.

Table of contents

*TODO - CHANGE THE LINK - Code and data are available at: https://github.com/Tanmay-Shinde/Week10Reflection.

1 Introduction

Food prices are a critical component of economic and social well-being, directly affecting food security, diet quality, and household expenditures. The Food-at-Home Monthly Area Prices (F-MAP) dataset, developed by the USDA Economic Research Service, offers detailed insights into food pricing trends across the United States. Covering the years 2012 to 2018, the dataset provides monthly price data for 90 food categories across 15 geographic areas, making it a valuable resource for understanding regional and temporal variations in food costs. This paper leverages the F-MAP data to analyze how regional disparities and time-based trends influence food-at-home prices and explores the role of economic factors such as purchase volume, food categories, and the Consumer Price Index (CPI) in shaping these patterns.

The primary focus of this analysis is to estimate how regional differences, food categories, and time affect food prices in the U.S. The estimand is the expected food price for a given food category, time period, and region, conditional on factors such as purchase volume and CPI. By modeling these variations, we aim to uncover the drivers of price changes and predict trends in food costs.

Using a Bayesian hierarchical model, this study analyzes monthly price data from the F-MAP dataset to uncover the drivers of food price variations. The model incorporates random effects to capture regional disparities and fixed effects to analyze the impact of economic factors, such as CPI, purchase volume, and food categories, on price trends. The analysis also highlights categories with greater price volatility and quantifies the influence of these factors on regional and national pricing dynamics. The findings reveal that food prices have generally increased over time, with categories like dairy, fresh produce, and meats experiencing higher price volatility. Economic factors such as purchase volume, food categories, and CPI significantly influence these trends, with higher CPI values strongly correlating with increased prices. Furthermore, food categories with higher demand or limited supply exhibit more substantial price fluctuations, underscoring the importance of understanding market-specific dynamics.

Understanding food pricing trends is essential for addressing issues related to food affordability and access. These insights are particularly valuable for policymakers and stakeholders aiming to reduce regional disparities, promote equitable access to food, and mitigate the effects of inflation on low-income households. By analyzing the drivers of price variations, this study provides a framework for informed decision-making in food policy and economic planning.

The remainder of this paper is structured as follows: Section 2 discusses the data sources, the F-MAP dataset and its variables, and pre-processing methods. Section 3 explains the Bayesian hierarchical model and methodology used for analysis. Section 4 presents the results, followed by a discussion of the key findings and conclusion of the study, as well as the limitations of the data in Section 5. Finally, **?@sec-appendix** — TODO: COMPLETE WHAT THE APPENDIX INCLUDES —.

2 Data

2.1 Source and Coverage

The Food-at-Home Monthly Area Prices (F-MAP) data product (U. S. Department of Agriculture 2024) is a comprehensive and detailed data product developed by the USDA Economic Research Service (ERS) that provides monthly U.S. food price data for 90 food-at-home (FAH) categories across 15 geographic areas of the United States. The dataset includes two primary price measures for each food group, geographic area, and month: (1) a mean unit value price (dollars per 100 grams) and (2) price indexes derived using advanced index formulas. These measures enable researchers to track food price trends at a granular level and compare them across geographic and temporal dimensions, while accounting for economic factors such as consumer purchasing volume, store characteristics, and inflation metrics like the Consumer Price Index (CPI). By utilizing Circana OmniMarket Core Outlets retail scanner data, the F-MAP captures detailed consumer purchasing data from over 50,000 retail stores annually, including grocery stores, supercenters, and convenience stores.

The F-MAP provides data across the following dimensions:

- Monthly, 2012–18
- 15 geographic areas
  - Nationally
  - 4 Census regions: Midwest, Northeast, South, West
  - 10 metropolitan areas: Atlanta, Boston, Chicago, Dallas, Detroit, Houston, Los Angeles, Miami, New York, and Philadelphia
- 90 ERS Food Purchase Groups (EFPGs) (Note: A full list of the 90 EFPGs and their descriptions can be found in Appendix A)
  - 8 groups for grains
  - 23 groups for vegetables
  - 8 groups for fruit
  - 8 groups for dairy and plant-based milk products
  - 14 groups for meat and protein foods
  - 4 groups for prepared meals, sides, and salads
  - 25 groups for other foods

For each of these month, area, and food group combinations, the F-MAP includes the following value variables:

- Purchase_dollars_wtd: Total weighted sales in U.S. dollars (nominal, i.e., not adjusted for inflation)
- Purchase_dollars_unwtd: Total unweighted sales in U.S. dollars (nominal, i.e., not adjusted for inflation)
- Purchase_grams_wtd: Total weighted quantities in grams
- Purchase_grams_unwtd: Total unweighted quantities in grams
- Number_stores: Number of stores in geographic area
- Unit_value_mean_wtd: Weighted mean unit value per 100 grams
- Unit_value_se_wtd: Standard error of weighted mean unit value
- Unit_value_mean_unwtd: Unweighted mean unit value per 100 grams
- Price_index_GEKS: Weighted price index value, constructed using Gini-Eltetö-Köves-Szulc (GEKS) formula

The F-MAP dataset is designed to align closely with the USDA Dietary Guidelines for Americans, facilitating research into food affordability, diet quality, and food security. Unlike other datasets, F-MAP offers monthly frequency data, making it particularly valuable for tracking short-term and seasonal trends. The dataset's hierarchical structure—spanning individual food categories, metropolitan regions, and national aggregates—supports diverse research applications. For example, it can be used to model the effects of policy interventions such as soda taxes or subsidies on dietary behavior and public health outcomes.

While other datasets such as the Bureau of Labor Statistics (BLS) Consumer Price Index (CPI) and the USDA Purchase to Plate National Average Prices (PP-NAP) provide useful insights, they fall short in capturing the comprehensive geographic and categorical detail offered by F-MAP. For instance, the CPI lacks subnational comparability across regions and provides limited food category detail, while the PP-NAP is focused on prepared foods and lacks the temporal granularity needed for trend analysis. The F-MAP dataset bridges these gaps by offering a more detailed, frequent, and regionally comparable resource.

The data is available to download in .xlsx format on the USDA Economic Research Service website. Specifically, we use the "Food-at-Home Monthly Area Prices, 2012 to 2018" dataset for our analysis.

2.2 Data Processing and Cleaning

We use the statistical programming language R (R Core Team 2023).... Our data (U. S. Department of Agriculture 2024).... Following Alexander (2023), we consider...

2.3 Measurement

The Food-at-Home Monthly Area Prices (F-MAP) dataset is built using high-frequency retail scanner data sourced from approximately 50,000–60,000 retail establishments annually. These include grocery stores, supercenters, club stores, drug stores, and convenience stores. The scanner data capture weekly sales in nominal dollars (not adjusted for inflation) and the quantities of food items sold. Weekly sales data are aggregated into monthly intervals to align with the dataset's temporal structure. If a sales week spans two months, the sales values and units are proportionately allocated based on the number of days in each month, ensuring temporal consistency.

To standardize the data, outliers in unit values are removed using the interquartile range (IQR) method, which identifies extreme values beyond 1.5 times the IQR from the 25th and 75th percentiles of the price distribution. This step eliminates inaccuracies that might arise from reporting errors or anomalous transactions. Package weights are converted into grams to ensure uniformity, using standard conversion factors (e.g., grams per ounce, fluid ounce, or pound). Prices are then expressed as unit values per 100 grams, providing a consistent measure of price across products of varying sizes.

The categorization of products into 90 detailed food categories is based on the USDA Economic Research Service (ERS) Food Purchase Groups (EFPG) system. This classification system organizes foods by their characteristics, such as ingredients, nutritional content, and convenience level. It aligns closely with the Dietary Guidelines for Americans and enables researchers to aggregate, disaggregate, or customize categories for specific research needs. These EFPG classifications are foundational to understanding price trends within and across food categories.

Retail sales data from certain retailers are reported at a broader Retailer Marketing Area (RMA) level rather than individual store locations. To ensure granularity, these RMA-level sales are disaggregated to individual stores using proportionate weighting methods, based on store-level weights developed specifically for the scanner data. These weights adjust the sales data to reflect the population of stores nationally and regionally, ensuring that the dataset is representative of real-world purchasing behaviors. Both weighted and unweighted unit value estimates are included, enabling diverse analytical approaches.

The dataset also includes price indexes, which measure the cost of a basket of goods over time and across locations. The GEKS multilateral price index, the primary index in the F-MAP, is constructed to compare prices dynamically while accounting for product substitution and turnover. This index is based on the geometric mean of bilateral indexes (Laspeyres, Paasche, Fisher Ideal) and employs a 1-year rolling window to maintain transitivity and minimize chain drift. By capturing the cost of goods relative to a base period (2016–2018 national averages), these price indexes provide a robust measure for temporal and spatial price comparisons, making the dataset suitable for inflation and affordability analyses.

Through this rigorous process, real-world phenomena such as regional price variations, inflationary trends, and food category-specific dynamics are translated into structured data entries. The combination of high-frequency retail data, standardized unit values, and multilateral price indexes ensures that the F-MAP dataset is both comprehensive and precise, supporting its use in economic research and policy-making. A more detailed overview of the measurement process, including the data acquisition, adjustments, and methods for calculating the price index can be found in Appendix A.

2.4 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Talk more about it.

Talk way more about it.

2.5 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

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 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 **??**.

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

Please don't use these as sub-heading labels - change them to be what your point actually is.

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

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected
by, the data

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algo-
rithm

Please include an Appendix where you focus on an aspect of surveys, sampling or observational data, related to your paper. This should be an in-depth exploration, akin to the "idealized methodology/survey/pollster methodology" sections of Paper 2. Some aspect of this is likely covered in the Measurement sub-section of your Data section, but this Appendix would be much more detailed, and might include aspects like simulation, links to the literature, explorations and comparisons, among other aspects.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

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

U. S. Department of Agriculture, Economic Research Service. 2024. "Food-at-Home Monthly Area Prices [Data Product]." U.S. Department of Agriculture.