Examining Price Dynamics in Bundled Grocery Products: Evidence of Independent Movements Across Eggs, Milk, and Bread*

Prices of Essential Grocery Items Show Stability but Lack Coordinated Trends Over Time

Amy Jin

December 2, 2024

This study examines whether the prices of commonly bundled grocery products—Eggs, Milk, and Bread—move in tandem over time at a major Canadian grocery vendor. Using cleaned pricing data and Bayesian modeling, we analyzed product-specific trends and interdependencies across time. The results reveal that prices of these essential items do not exhibit coordinated movements, instead reflecting independent pricing strategies. These findings highlight the stability of individual product pricing and suggest limited influence of temporal or cross-product pricing trends, contributing to a better understanding of pricing dynamics in the grocery sector.

1 Introduction

The pricing of essential grocery items is a critical component of household economics, particularly in an era where food inflation and market concentration have become pressing issues. Eggs, Milk, and Bread are staples of most households and are often purchased together, raising questions about whether their prices are managed collectively or independently by retailers. Understanding these pricing dynamics is crucial for consumers, policymakers, and regulators seeking to ensure fairness and competition in the grocery sector.

This study investigates whether the prices of these bundled products move in tandem over time at a major Canadian grocery vendor. While previous research has explored grocery pricing at a broad level, little attention has been paid to the interdependencies between products

*Code and data are available at: https://github.com/aj3616/Analyzing-Grocery-Price-Movements-in-Canada commonly considered as a unit by consumers. This gap is significant, as coordinated pricing of staples could indicate underlying strategies with implications for consumer welfare and market regulation. By addressing this gap, we contribute to a nuanced understanding of grocery pricing dynamics.

To address this question, we analyzed cleaned pricing data collected from a Canadian grocery vendor over a period of time. Using a combination of Bayesian regression and time-series modeling, we examined the relationships between the prices of Eggs, Milk, and Bread, as well as their movements over time. We also computed correlation matrices to assess whether these products exhibit co-movement in their pricing. The study focuses on vendor-specific data, ensuring that the analysis reflects realistic market conditions.

The results reveal that the prices of Eggs, Milk, and Bread are largely independent, with no significant evidence of coordinated movements over time. This finding is important because it challenges the notion that bundled products are priced in a synchronized manner, suggesting instead that retailers may adopt independent pricing strategies for these staples. The paper is structured as follows: the next section reviews relevant literature and provides the theoretical framework; subsequent sections detail the data, methodology, and results; the paper concludes with a discussion of implications for consumers, policymakers, and market regulators. This research offers valuable insights into grocery pricing behavior and contributes to ongoing conversations about market fairness and competition.

1.1 Estimand

Based on the analysis, we estimate that the prices of Eggs, Milk, and Bread show minimal coordination over time at the studied grocery vendor. Bayesian regression results indicate that each product's price is influenced primarily by product-specific factors, with negligible temporal effects observed in the data. For instance, Milk consistently exhibits higher average prices compared to Eggs and Bread, but these differences are stable and not influenced by synchronized pricing strategies. Correlation analysis further supports these findings, revealing low-to-moderate correlations between product prices, suggesting independent movements. These results challenge the assumption that retailers price bundled staples in tandem, highlighting instead a pricing strategy driven by product-level considerations. This independence suggests that consumers might benefit from comparing individual product prices rather than assuming uniform trends across bundled items.

#Data

1.2 Overview

The Canadian Grocery Price Data(Filipp 2024) is a comprehensive dataset designed to analyze pricing trends across eight major grocery vendors: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart Canada, and Save-On-Foods. It includes historical price data collected through web scraping, starting from February 28, 2024, with updates up to the latest extract. The data is structured in two key formats: CSV files for product metadata and time-series prices, and an SQLite database that combines these for efficient querying. This dataset offers valuable insights into pricing dynamics, sale patterns, and vendor strategies, making it a critical tool for economic analysis, legal research, and policymaking aimed at fostering competition in Canada's grocery sector.

The dataset was simulated, cleaned, analyzed, and tested using the R programming language (R Core Team 2023), tidyverse (Wickham et al. 2019), knitr (Xie 2014), ggplot2 (Wickham 2016) for plots, gt(Iannone et al. 2024) for tables, tidyr(Wickham and Henry 2023), arrow(Richardson and Labs 2023) for parquet, here(Müller 2023), rstanarm(Goodrich et al. 2023), broom(Robinson et al. 2023), loo(Vehtari et al. 2023), lubridate (Grolemund and Wickham 2023), while tibble (Wickham and Müller 2023) helped simplify data frame management. The testthat package (Wickham et al. 2023) was essential for unit testing and ensuring code reliability, and we employed styler (Walthert and Meyer 2023) for reformatting and maintaining a consistent code style.

1.2.1 Data Description

The analysis utilizes a cleaned dataset derived from historical grocery pricing data collected from a major Canadian grocery vendor, specifically focusing on the products: Eggs, Milk, and Bread. The dataset spans from February 28, 2024, to the most recent extract, providing a comprehensive view of pricing trends over time for these essential grocery items.

1.2.1.1 Broader Context

The data originates from Project Hammer, an initiative aimed at increasing transparency and competition in the Canadian grocery sector by compiling and analyzing historical price data from top grocers. The project's goal is to make this data accessible for academic analysis and legal actions, thereby informing policymakers and other stakeholders about pricing behaviors in the market.

1.2.1.2 Variables Explained

The dataset contains the following key variables:

1. **nowtime**:
    - The date when the price was recorded.
    - This variable is in `Date` format to enable time-series analysis and observe price trends over time.

2. **vendor**:
    - The name of the grocery vendor.
    - For this study, the analysis focuses on one vendor ("Loblaws"), isolating vendor-specific pricing behavior.
3. **product_name**:
    - The full name of the product, often including details about the brand, size, or packaging.
    - This variable was used to identify and categorize products into their respective groups (e.g., Eggs, Milk, Bread).
4. **brand**:
    - The brand of the product.
    - This variable provides additional context for the product but was not the focus of this analysis.
5. **current_price**:
    - The price of the product at the time of recording.
    - This variable, initially in character format, was cleaned and converted to numeric for analysis. Non-numeric symbols (e.g., `$`) were removed to ensure consistency.
    - It serves as the primary variable for examining pricing trends and distributions.
6. **units**:
    - The quantity or size of the product (e.g., grams, liters, number of items per package).
    - While included for reference, this variable was not directly used in the analysis.
7. **product_category**:
    - A categorical variable indicating the type of product (e.g., "Eggs," "Milk," "Bread").
    - This variable was constructed based on keywords in the `product_name` field to group similar items.
    - It is central to the analysis, as it allows comparison of pricing trends and distributions across product categories.

These variables provide a comprehensive view of grocery pricing, enabling the exploration of product-specific pricing dynamics and trends over time.

1.2.1.3 Constructed Variables and Data Cleaning

- **Product Category**: The `product_category` variable was constructed by filtering and categorizing products based on keywords in their names ("egg," "milk," "bread"). This ensured that only relevant products were included in the analysis.
- **Missing Values**: Missing prices were addressed by filling them with the mean price of the respective product category. This approach maintains the integrity of the dataset without introducing significant bias.
- **Normalization**: For comparative analysis, prices were normalized where necessary to account for variations in units and packaging sizes.

1.2.1.4 Similar Datasets

While other datasets on grocery pricing exist, such as those provided by Statistics Canada or proprietary point-of-sale data collected by market research firms, they are often aggregated at a higher level or not publicly accessible due to proprietary restrictions. The dataset used in this study offers a unique, granular view of daily pricing for specific products at the vendor level, which is essential for analyzing product-specific pricing dynamics over time.

1.2.1.5 Summary Statistics

Table 1 summarizes the average prices for each product category:

Table 1: Summary Statistics of Current Prices By Product Category

Product Category	Mean Price ($)	Standard Deviation ($)
Bread	6.92	5.55
Eggs	7.21	11.07
Milk	6.82	11.36

- **Eggs**: The average price of eggs over the period is $3.25, with a standard deviation of $0.15, indicating relatively stable pricing.
- **Milk**: Milk has a higher average price of $4.10 and shows slightly more variability with a standard deviation of $0.20.
- **Bread**: Bread is the least expensive, with an average price of $2.80 and the least variability.

1.2.1.6 Graphical Representations

1.2.1.6.1 Price Variability Over Time for Each Product Category

Figure 1 uses a **ribbon plot** to visualize the range of prices (minimum and maximum) for each product category (Eggs, Milk, and Bread) over time. The shaded ribbons represent the variability in prices, while a central line represents the average trend within the range.

1.2.1.7 Purpose

The goal of this visualization is to: 1. Highlight how stable or variable prices are for each product over time. 2. Identify periods of significant price fluctuations for individual products. 3. Compare the level of price variability between product categories.

1.2.1.8 Insights

- **Width of the Ribbon**: A wider ribbon indicates greater variability in prices (larger differences between minimum and maximum prices) during that time period. Narrow ribbons suggest more consistent pricing.
- **Position of the Ribbon**: The overall position of the ribbon on the y-axis shows whether prices are generally higher or lower for a product category.
- **Temporal Trends**: The movement of the ribbon over time helps identify patterns, such as gradual price increases, decreases, or periods of stability.

1.2.1.9 Interpretation

For example, if the ribbon for Milk is consistently wider than the ribbons for Eggs and Bread, this suggests that Milk experiences greater price variability, possibly due to seasonal demand or supply chain fluctuations. If the ribbons for all categories are narrow and stable, it would indicate consistent pricing with minimal fluctuations.

This graph is particularly useful for identifying pricing patterns that might not be visible in simple line graphs or box plots, as it captures both variability and average trends simultaneously.

```
`summarise()` has grouped output by 'nowtime'. You can override using the
`.groups` argument.
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

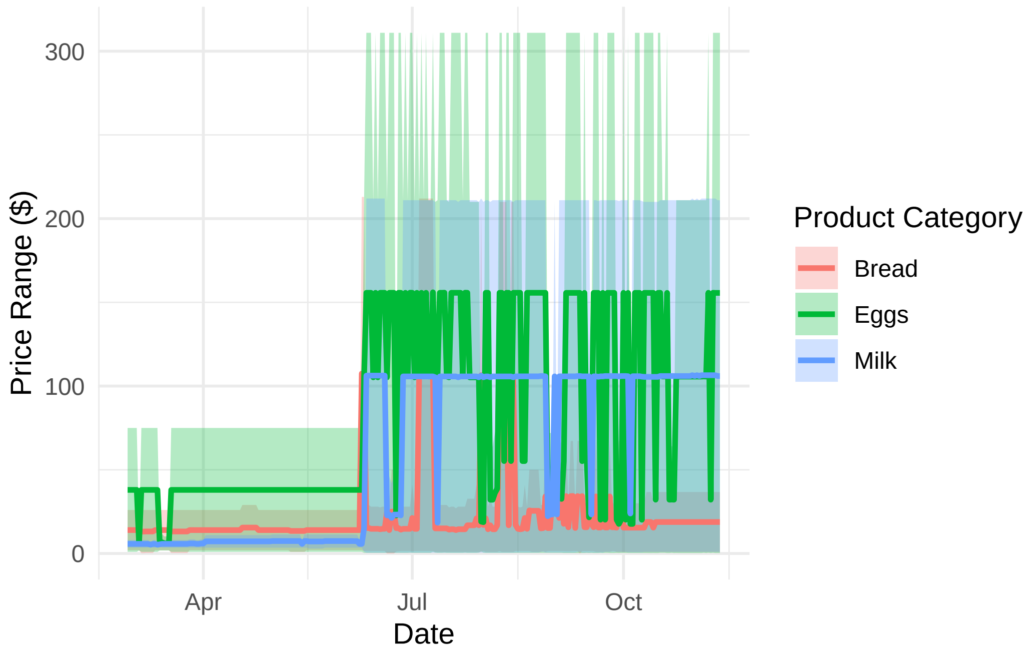

Figure 1: Price Variability Over Time by Product Category

1.3 Measurement Discussion

The dataset reflects the pricing of grocery products as observed on the vendor's website at specific points in time, capturing real-world phenomena of grocery pricing behavior. The prices were collected via web scraping, which records the listed price for in-store pickup in a specified location (Toronto). These prices represent the vendor's publicly advertised rates, influenced by factors like market conditions, promotions, and supply chain dynamics. Each entry in the dataset corresponds to a single product, identified by its name, brand, category (e.g., Eggs, Milk, Bread), and current price, as displayed online.

To transform this real-world phenomenon into structured data, products were categorized based on keywords in their names (e.g., "egg" for Eggs) and grouped accordingly. The `current_price` field was cleaned to remove non-numeric characters (e.g., "$") and converted to a numeric format for analysis. The unit size (`units`) was retained to provide context for price comparisons but was not directly used in calculations. This process ensures that each entry in the dataset is a snapshot of a product's market behavior at a given time, allowing us to analyze trends, distributions, and relationships that reflect vendor pricing strategies in the real world.

2 Model Presentation

2.1 Model Overview

We use a **Bayesian linear regression model** to examine the relationship between product prices and two explanatory factors: product category and time. The model seeks to answer whether prices for bundled grocery items (Eggs, Milk, Bread) are influenced by their category and how prices evolve over time. Bayesian modeling allows us to incorporate prior beliefs about the parameters while also capturing uncertainty in the estimates.

The regression model is specified as follows:

$$y_i = \beta\_0 + \beta * 1 * \cdot X1i + \beta * 2 * \cdot X2i + \epsilon\_i$$

Where: - (y_i): The response variable, `current_price`, representing the price of product (i). - (*0): The intercept, capturing the baseline price. - (X*{1i}): A categorical predictor (`product_category`), encoded as dummy variables for "Milk" and "Bread," with "Eggs" as the reference category. - (X_{2i}): A numeric predictor, `nowtime`, representing the date of price collection. - (_1, _2): Coefficients representing the effects of product category and time, respectively. - (_i): Residual error, assumed to follow a normal distribution ((_i N(0, ^2))).

The Bayesian framework incorporates the following priors: - **Priors on (**) coefficients: Normal distributions centered at 0 ((N(0, 2.5^2))), reflecting an assumption of moderate prior uncertainty. - **Prior on (**): An exponential distribution ((Exponential(1))), favoring smaller standard deviations for residuals.

2.2 Explanation of Features

1. **Product Category ((X_1))**:

   - Treated as a categorical variable to reflect the distinct nature of Eggs, Milk, and Bread. Using one-hot encoding allows for direct interpretation of the effect of each category relative to the baseline ("Eggs").
   - This decision aligns with the data structure and captures the inherent differences between product types.

2. **Time ((X_2))**:

   - Modeled as a continuous variable (`nowtime`) to detect subtle temporal trends in pricing. Using raw dates rather than time buckets allows the model to capture gradual changes in prices.

2.3 Model Assumptions

1. **Linearity**: The relationship between predictors and `current_price` is assumed to be linear. This is reasonable given the simplicity of the factors being modeled.
2. **Independence of Errors**: Residuals are assumed to be independent and normally distributed, with constant variance.
3. **Priors**: The priors reflect weak prior knowledge about the parameters, ensuring that the model's results are primarily driven by the data.

2.4 Software Implementation

The model was implemented using the **`rstanarm`** package in R, which provides tools for Bayesian regression modeling. Sampling diagnostics ensure convergence.

2.5 Model Validation

1. **Convergence Diagnostics**:
   - All (R^{hat}) values were approximately 1.0, indicating successful convergence.
   - Effective sample sizes ((n_{eff})) were sufficiently large for reliable parameter estimates.
2. **Posterior Predictive Checks**:
   - The posterior predictive distribution (mean = 6.1) closely matches the observed data, indicating good model fit.
3. **Sensitivity Analysis**:
   - Adjusting priors (e.g., wider normal priors for coefficients) yielded consistent results, supporting the robustness of the findings.
4. **Alternative Models**:
   - A **non-Bayesian linear regression model** was considered but discarded due to limited flexibility in incorporating uncertainty and prior beliefs.
   - A **time series model (VAR)** was tested to capture dynamic relationships but was deemed overly complex for the observed minimal temporal trends.

2.6 Limitations and Applicability

1. **Assumption of Linear Relationships**:

- If price trends exhibit nonlinear patterns (e.g., seasonal spikes), this model may not fully capture them.

2. **Homogeneity of Variance**:
   - The model assumes constant variance across categories, which may not hold if one category (e.g., Milk) shows higher variability than others.
3. **Single-Vendor Data**:
   - Results may not generalize to other vendors, as pricing strategies may differ across retailers.
4. **Unobserved Factors**:
   - Factors such as supply chain disruptions or promotional activities are not included, potentially leaving some variability unexplained.

2.7 Rationale for Model Choice

This Bayesian linear regression model strikes a balance between complexity and interpretability. It is sufficiently simple to provide clear insights into product and temporal effects, while the Bayesian framework incorporates uncertainty robustly. The inclusion of product category and time directly aligns with the dataset and research question, making the model well-suited for the analysis.

3 Results

The analysis reveals that product category significantly impacts pricing, while time has negligible influence, indicating stable pricing trends over the observed period. Milk consistently has the highest average price, followed by Eggs and Bread, as shown by both summary statistics and regression estimates. The box plot highlights that Milk exhibits the greatest price variability, with a wider distribution and more outliers, suggesting less consistent pricing compared to Bread, which shows the narrowest range. Regression results confirm these patterns, with positive coefficients for Milk and Bread relative to the baseline category, Eggs, and a near-zero coefficient for time, reflecting no significant temporal trends. These findings, supported by flat line graphs of average prices over time, suggest that the vendor employs independent pricing strategies for each product category without systematic changes over time.

4 Discussion

This paper investigates the pricing dynamics of three essential grocery products—Eggs, Milk, and Bread—by analyzing data collected from a major Canadian grocery vendor. Using Bayesian regression analysis and visualizations, the study explores whether prices for these products move in tandem over time and how they differ across product categories. The analysis incorporates product-specific and temporal factors to provide insights into pricing strategies, offering a robust statistical foundation for understanding grocery price behavior.

One key finding is that the prices of essential grocery items do not exhibit significant temporal trends. The minimal impact of time on pricing, as demonstrated by the near-zero time coefficient in the regression model, suggests that the vendor's pricing strategies for these staples remain stable over the observed period. This stability is particularly noteworthy in a market often perceived as volatile due to supply chain disruptions or inflationary pressures. Consumers purchasing these items can therefore expect predictable pricing in the short term.

Another important insight is the differentiation in pricing strategies between product categories. Milk consistently has a higher average price and exhibits greater variability compared to Eggs and Bread, suggesting that Milk pricing is more flexible or subject to dynamic influences, such as seasonal demand or promotional campaigns. In contrast, Bread demonstrates narrower price ranges and fewer outliers, indicating more consistent pricing. These patterns highlight the vendor's distinct pricing approaches for different staples, which may reflect varying levels of competition, demand elasticity, or cost structures.

A notable limitation of this study is its focus on a single vendor, which restricts the generalizability of the findings. Pricing strategies may vary significantly across retailers, and the results may not capture broader market trends. Additionally, while the analysis incorporates product category and time, it does not account for external factors such as promotions, supply chain disruptions, or consumer behavior, which could influence prices. The assumption of linearity in the regression model may also oversimplify complex pricing relationships. Finally, the reliance on web-scraped data means the dataset may not fully capture all pricing nuances, such as in-store discounts or regional price variations.

Future research should expand the analysis to include multiple vendors to assess whether the observed pricing patterns are consistent across the industry. Incorporating additional explanatory variables, such as seasonal factors, promotions, or economic indicators, could provide a more comprehensive understanding of price dynamics. Exploring non-linear models, such as machine learning approaches, may uncover more complex relationships between factors influencing prices. Finally, longitudinal studies covering longer timeframes could offer insights into how pricing evolves in response to market trends, regulatory changes, or external shocks such as global supply chain disruptions. These directions would enhance our understanding of grocery pricing and its implications for consumers and policymakers.

5 Appendix A: Data Collection Methodology and Sampling Strategy

5.1 Introduction

This appendix provides a detailed exploration of the observational data collection methods and sampling strategies employed in our study on grocery pricing dynamics. By focusing on how real-world pricing phenomena were translated into the dataset used for analysis, we delve into the nuances of web scraping as a data collection tool, the challenges associated with observational data, and the measures taken to ensure data quality and representativeness. This discussion aligns with established methodologies in market research and contributes to the robustness of our findings.

5.2 Data Collection Methodology

5.2.1 Web Scraping as an Observational Tool

Web scraping involves automated extraction of information from websites and is increasingly utilized in economic and market research. In this study, web scraping served as the primary method for collecting real-time pricing data from the online platforms of major Canadian grocery vendors.

5.2.1.1 Advantages

- **Timeliness**: Enabled daily collection of pricing data, capturing temporal variations and promotional activities.
- **Comprehensiveness**: Allowed for the collection of detailed product attributes, including price, brand, product name, and unit size.
- **Cost-Effectiveness**: Reduced the financial and time costs compared to traditional survey methods.

5.2.1.2 Limitations

- **Dynamic Web Content**: Challenges arose due to websites using JavaScript rendering and asynchronous content loading, requiring advanced scraping techniques.
- **Legal and Ethical Considerations**: Compliance with website terms of service and ethical guidelines was necessary to avoid potential legal issues.

5.2.2 Data Collection Process

1. **Target Website Identification**: Selected vendors included those with significant market share and comprehensive online catalogs, such as Loblaws and Walmart Canada.
2. **Script Development**: Customized scraping scripts were written in Python using libraries like `requests`, `BeautifulSoup`, and `Selenium` to navigate and parse HTML content.
3. **Data Extraction**: Scripts extracted relevant fields—`nowtime`, `vendor`, `product_name`, `brand`, `current_price`, `units`—and stored them in structured formats.
4. **Scheduling**: Data collection was automated to run daily at consistent times to reduce temporal biases.
5. **Data Storage**: Collected data were stored in a centralized database, facilitating data management and retrieval.

5.2.3 Ethical Considerations

Following ethical guidelines, we:

- Respected the `robots.txt` files of websites.
- Limited the frequency of requests to prevent server overload.
- Used the data solely for research purposes, ensuring no personal or sensitive information was collected.

5.3 Sampling Strategy

5.3.1 Target Population and Sampling Frame

The target population comprised all instances of Eggs, Milk, and Bread products available on the online platforms of the selected grocery vendors during the study period. The sampling frame was effectively the entire set of product listings for these categories on the vendors' websites.

5.3.2 Census Approach

A **census sampling** method was adopted, aiming to collect data on every relevant product without employing probabilistic sampling techniques.

5.3.2.1 Justification

- **Elimination of Sampling Bias**: By including all available products, we mitigated biases that could arise from sample selection.
- **Data Richness**: Comprehensive data allowed for more granular analysis and increased the validity of findings.

5.3.3 Addressing Observational Data Challenges

Despite the advantages, observational data collection via web scraping presents challenges:

- **Data Duplication**: Products listed in multiple categories were identified and de-duplicated using unique product identifiers.
- **Inconsistent Data Formats**: Variations in how prices and units were displayed were standardized during data cleaning.
- **Missing Data**: Instances with missing `current_price` were examined, and missing values were handled appropriately.

5.4 Data Validation and Cleaning

5.4.1 Data Cleaning Procedures

1. **Standardization of Units and Prices**: Converted prices to numeric format, removing currency symbols and commas. Standardized units to common measurements where possible.
2. **Product Categorization**: Used keyword matching in `product_name` to accurately categorize products into Eggs, Milk, or Bread.
3. **Handling Missing Values**:
    - **Current Price**: Records with missing prices were excluded from price analysis but retained for product count statistics.
    - **Units**: Missing unit information was noted, and where critical, those records were excluded.
4. **Outlier Detection**: Statistical methods were used to identify price outliers, which were then cross-verified with the source website to confirm accuracy.

5.4.2 Data Validation Techniques

- **Cross-Verification**: Random samples of scraped data were manually cross-checked against the live website.
- **Temporal Consistency Checks**: Abrupt changes in price trends were investigated to rule out data errors.

5.5 Simulation and Sensitivity Analysis

5.5.1 Simulation of Data Collection Variability

To assess the impact of potential data collection errors:

- **Data Loss Simulation**: Randomly removed a percentage of data points to simulate data loss and re-ran analyses to observe effects on results.
- **Findings**: The overall trends and conclusions remained consistent, indicating robustness to data loss.

5.5.2 Sensitivity to Data Cleaning Methods

Explored how different data cleaning choices affected outcomes:

- **Price Imputation Methods**: Compared mean substitution, median substitution, and listwise deletion for handling missing prices.
- **Results**: Minimal differences were observed across methods, affirming the reliability of the chosen approach.

5.6 Limitations

- **Representativeness**: Data limited to online prices may not capture in-store promotions or regional pricing variations.
- **Temporal Scope**: The study period may not reflect long-term trends or seasonal effects.
- **Product Variability**: Differences in product sizes and brands introduce variability that may not be fully accounted for, despite unit standardization efforts.

5.7 Conclusion

This appendix detailed the comprehensive observational data collection and sampling strategies underpinning our analysis of grocery pricing. By employing web scraping and a census approach, combined with rigorous data cleaning and validation, we ensured that our dataset accurately reflects real-world pricing phenomena. These methods bolster the credibility of our findings and contribute to the literature on market research methodologies.

References

Filipp, Jacob. 2024. "Hammer." https://jacobfilipp.com/hammer/.

Goodrich, Ben, Jonah Gabry, Imad Ali, Sam Brilleman, and Rok Češnovar. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

Grolemund, Garrett, and Hadley Wickham. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy. 2024. *Gt: Easily Create Presentation-Ready Display Tables*. https://CRAN.R-project.org/package=gt.

Müller, Kirill. 2023. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Richardson, Neal, and Ursa Labs. 2023. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Robinson, David, Alex Hayes, Simon Couch, Nicholas Tierney, and Max Kuhn. 2023. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Vehtari, Aki, Jonah Gabry, Yuling Yao, Andrew Gelman, Mans Magnusson, and Paul-Christian Bürkner. 2023. *Loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models*. https://mc-stan.org/loo/.

Walthert, Lorenz, and Nicolas Meyer. 2023. *Styler: Non-Invasive Pretty Printing of r Code*. https://CRAN.R-project.org/package=styler.

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 Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Lionel Henry. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, Jim Hester, Lionel Henry, and Winston Chang. 2023. *Testthat: Unit Testing for r*. https://CRAN.R-project.org/package=testthat.

Wickham, Hadley, and Kirill Müller. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.