Price Manipulation in the Canadian Grocery Sector: Analyzing Price*

Understanding Price Fluctuations and Their Impact on Consumer Behavior

Arav Sri Agarwal

December 15, 2024

This paper investigates the pricing behavior of Canadian grocery retailers, specifically focusing on whether prices are artificially inflated before a sale, only to be reduced back to normal post-sale. Using a dataset of historical prices from major Canadian grocery chains, we conduct a detailed analysis comparing prices before and after sales. The study applies statistical tests, including Shapiro-Wilk and Wilcoxon signed-rank tests, and explores correlations between price changes before and after sales. The results indicate a statistically significant difference in price behavior, with minor price fluctuations observed before and after sales events. This paper discusses the findings in the context of price transparency, consumer behavior, and regulatory implications for the grocery sector.

*Code and data are available at: [https://github.com/Aravsria/Price-Manipulation-in-the-Canadian-Grocery-Sector).

1 Introduction

Consumer awareness of pricing strategies has grown significantly in recent years, particularly in the retail grocery sector. A common practice, referred to as "artificial price inflation," has raised questions regarding whether retailers increase prices just before a sale and then reduce them during the promotion, making the discount appear more substantial. This phenomenon may distort the true value of discounts and negatively impact consumer trust.

The objective of this paper is to explore this pricing behavior in the Canadian grocery sector by examining the price fluctuations of grocery items before and after sales events. Specifically, we aim to answer the question: "When something goes on sale, does the price get jacked up artificially just ahead of the sale, only to lower it back down to normal afterwards?" This analysis uses a large dataset of historical pricing data from multiple major Canadian grocery retailers, including Metro, Walmart, Loblaws, and others. The paper applies statistical techniques such as correlation analysis and the Wilcoxon signed-rank test to assess whether there is a significant price change before and after sales. The findings contribute to the broader discussion about price transparency, consumer rights, and market regulation.

The structure of this paper is as follows: Section 3 describes the dataset and variables used for analysis, Section 5 discusses the results of our investigation, Section 6 concludes the paper by interpreting the findings and suggesting potential regulatory actions to ensure fair pricing practices in the grocery industry. Further information regarding data collection, sampling, and limitations is included in Section A .

2 Estimand

The estimand of this study is the difference in price behavior before and after sales. Specifically, we are interested in identifying whether prices are significantly higher just before a sale and whether they decrease afterward, suggesting artificial price inflation. This is analyzed through descriptive statistics, correlation analysis, and paired statistical tests.

3 Data

3.1 Dataset Overview

We used the programming language Python (**python?**) to explore price manipulation in the Canadian grocery sector using data from Project Hammer (Filipp 2024). More details about the data collection methodology can be found in the section Section A.2.

The dataset consists of price records for grocery items from seven major Canadian grocery retailers: Metro, Walmart, Loblaws, No Frills, Save-On-Foods, Voila, and Galleria. The dataset spans from February 28 to July 10, 2024, and includes information about product names, prices, sale status, and sale dates. The primary variables in the dataset include:

- product_id: A unique identifier for each product.
- product_name: The product's name, which includes details like brand and packaging.
- vendor: The name of the retailer (e.g., Walmart, Metro).
- current_price: The price of the product at the time of the data extraction.
- old_price: The price before the sale, if available. - price_per_unit: The price per unit, often given in grams or number of items.
- other: A column indicating additional information, such as sale status.
- nowtime: The timestamp when the price data was recorded.

3.2 Measurement

The process of converting price manipulation information into analyzable data involves capturing observable elements of pricing strategies employed by grocery retailers. Specifically, price fluctuations before and after sales events are tracked through web-scraped records of product prices from major grocery chains. Each entry in the dataset reflects key details such as the product's current price, pre-sale price, sale status, and timestamp, which together allow for the examination of pricing behaviors over time. By recording these price data points systematically, the dataset enables quantitative analysis of potential artificial price inflation practices, connecting real-world retail strategies to measurable patterns in the data.

3.3 Data Cleaning and Preparation

The data was subjected to several cleaning steps to ensure its accuracy and usability for analysis. Initially, irrelevant columns such as detail_url, sku, and upc were removed. Missing values in critical columns (like price_before_sale and price_after_sale) were handled through imputation, and only valid data records were included in the final analysis. Additionally, the nowtime column was converted into a datetime format, and variables like price_change_before_sale and price_change_after_sale were derived by calculating the differences between current_price and old_price.

4 Model and Statistical Tests

To assess whether these price changes are statistically significant, we performed the following tests:

1. Shapiro-Wilk Test: To check for normality in the distribution of price changes, the Shapiro-Wilk test was conducted on both the price changes before and after the sale. The results indicated that both distributions are not normally distributed, as evidenced by the p-values of 0.0 for both price_change_before_sale and price_change_after_sale.
2. Wilcoxon Signed-Rank Test: Given the non-normality of the data, a Wilcoxon signed-rank test was conducted to compare the median price changes before and after the sale. The result of this test was statistically significant (p-value = 0.002), suggesting that there is a difference in prices before and after sales.
3. Correlation Analysis: The correlation between price changes before and after the sale was calculated. The results show a very weak positive correlation of 0.024 between the price changes before and after the sale, indicating that while there is some relationship between these two variables, it is not strong.

4.1 Model Setup

To examine price behavior before and after sales, the dataset was structured to isolate sale events and compute corresponding price differences. Each sale event was identified by filtering rows where the sale indicator was active and sorting by product_id and nowtime. This enabled the computation of two key variables: price_before_sale and price_after_sale, derived by shifting prices relative to the chronological order of each product.

Statistical analysis began with a Shapiro-Wilk test to determine if price_before_sale and price_after_sale followed a normal distribution:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Results revealed non-normality ($p<0.05$), allowing us to use the Wilcoxon signed-rank test:

$$W = \sum_{i=1}^{n} R_i^+$$

which evaluates the median difference between paired samples. The test concluded that prices before and after sales exhibit a statistically significant difference ($p<0.05$).

To assess linear relationships, Pearson correlation was used to compare price changes before and after sales:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

indicating a weak positive correlation (r=0.024). Finally, monthly averages of price changes were computed to visualize trends over time, highlighting seasonal variations and potential pricing strategies.

4.2 Model Justification

The model used in this paper is evaluates pricing behaviors surrounding sales events. Using a cleaned dataset of product prices, the model separates records into pre-sale and post-sale periods for statistical and comparative analysis. The main steps include generating price_before_sale and price_after_sale variables through grouped time-aligned shifts, allowing precise tracking of price changes. To validate the assumptions of normality in price distributions, the Shapiro-Wilk test was applied, confirming the non-parametric nature of the data. The Wilcoxon signed-rank test was used to compare pre- and post-sale prices, a robust non-parametric test suitable for paired data with non-normal distributions. The analysis also uses correlation measures to quantify the relationship between price changes before and after sales. Finally, a temporal analysis aggregated monthly price changes to visualize trends over time. Together, these steps allow a thorough study of price manipulation behaviors, using different types of statistical techniques to identify patterns and statistically significant differences in pricing.

5 Results

5.1 Descriptive Statistics

The results of the analysis reveal some interesting patterns in the price changes before and after sales. The descriptive statistics for the price changes before and after sales are summarized in Table 1 below.

Table 1: Price Change Statistics

Statistic	Price Change Before Sale	Price Change After Sale
Count	402,079	401,722
Mean	-0.064	0.063
Standard Deviation	0.396	0.398
Minimum	-25.0	-6.5
25th Percentile	0.0	0.0
Median (50th Percentile)	0.0	0.0
75th Percentile	0.0	0.0
Maximum	4.5	14.0

From the descriptive statistics, we observe that the mean price change before the sale is negative (-0.064), indicating a slight reduction in prices before the sale. On the other hand, the mean price change after the sale is positive (0.063), suggesting a small increase in prices following the sale.

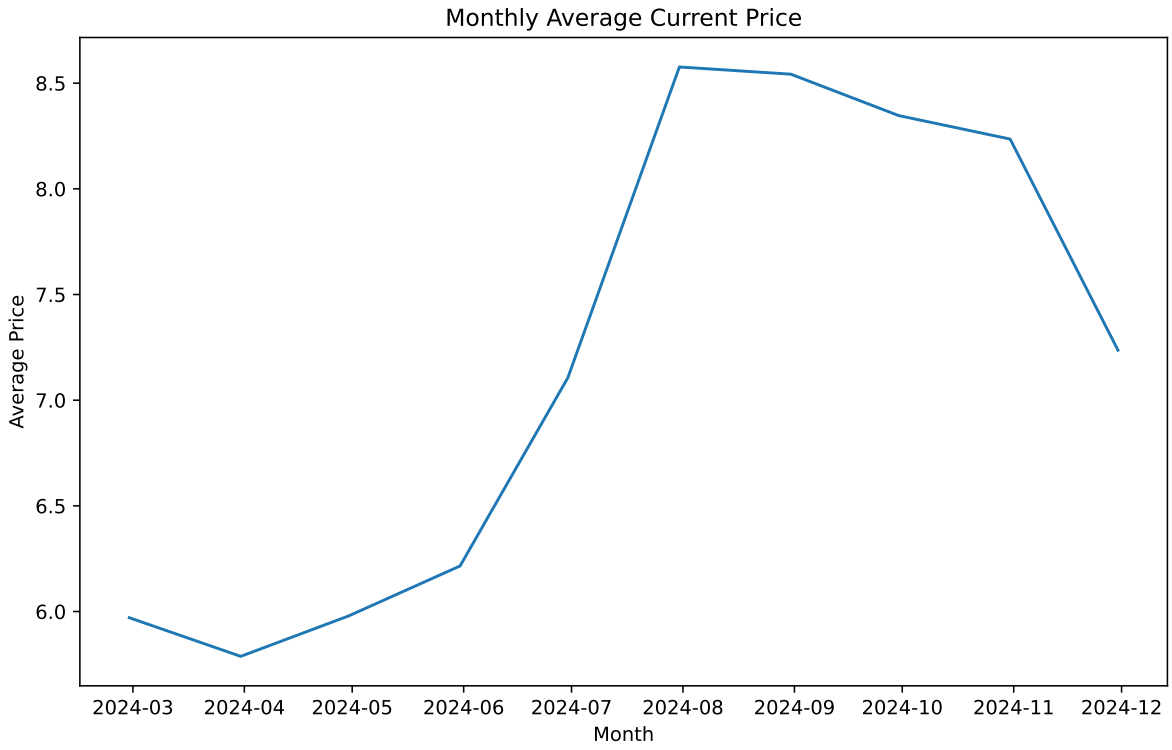

Figure 1: Monthly Average Current Price

Figure 1 above shows the trend of monthly average prices for grocery items over a 10-month period in 2024. The lowest average price was in April 2024 (~6.0 units), suggesting a potential period of higher price competition or promotions aimed at attracting consumers. The peak average price was in August 2024 (~8.5 units), suggesting potential seasonal factors such as increased demand during summer months or strategic price increases by retailers. The consistent increase in prices from March to August highlights a possible seasonal pattern or planned retailer strategy, where prices gradually rise toward peak demand periods. Following the peak in August, prices begin to decline, with a notable drop in November and December 2024. This could be attributed to holiday promotions, sales events, or efforts to clear inventory.

These trends suggest that retailers might be using pricing strategies that account for consumer purchasing habits or seasonal demand. For example, higher prices in summer months could reflect increased demand for fresh produce and outdoor-related food products. The sharp price drop toward the end of the year raises questions about whether these reductions are genuine discounts or price resets after artificial inflation during previous months. The significant variations in prices could warrant further investigation by regulatory bodies to ensure transparency in pricing practices. Monitoring such fluctuations can help protect consumers from misleading promotional activities.

The graph highlights the importance of continuous monitoring of monthly price trends to

understand retailer behaviors better and anticipate periods of higher or lower prices.

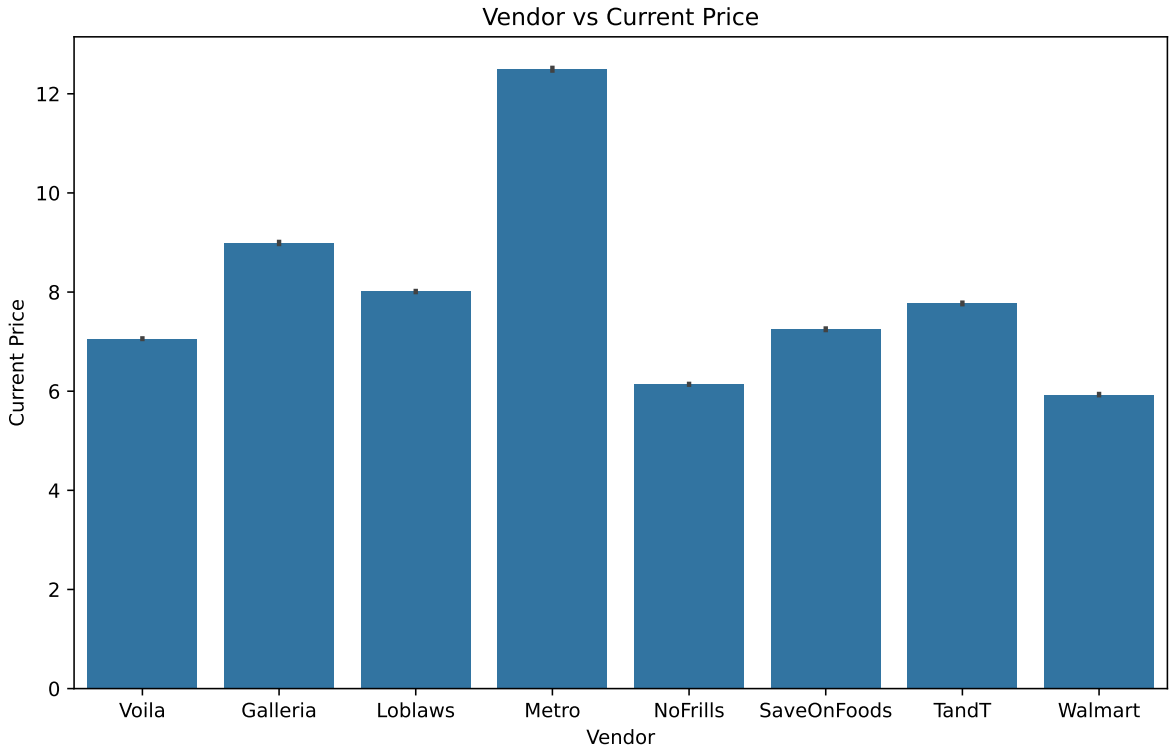

Figure 2: Vendor vs Current Price

Figure 2 above shows the distribution of average current prices across different vendors. Metro has the highest current prices, averaging over 12 units. Walmart exhibits the lowest current prices among the vendors, averaging below 8 units. Other vendors, such as Loblaws, SaveOn-Foods, and TandT, have moderately high prices, falling between 8 and 10 units. The significant price differences highlight the varying pricing strategies among these vendors, likely driven by their target customer base, product quality, and competitive positioning in the market. Premium retailers, such as Metro, may justify higher prices through better service, higher quality products, or exclusive offerings. On the other hand, value-focused retailers like Walmart strive to offer lower prices.

This graph can also serve as a tool for regulatory bodies to study potential price manipulation or disparities in the grocery sector. Policymakers might investigate whether higher-priced vendors are providing fair prices or simply charging a premium without justification.

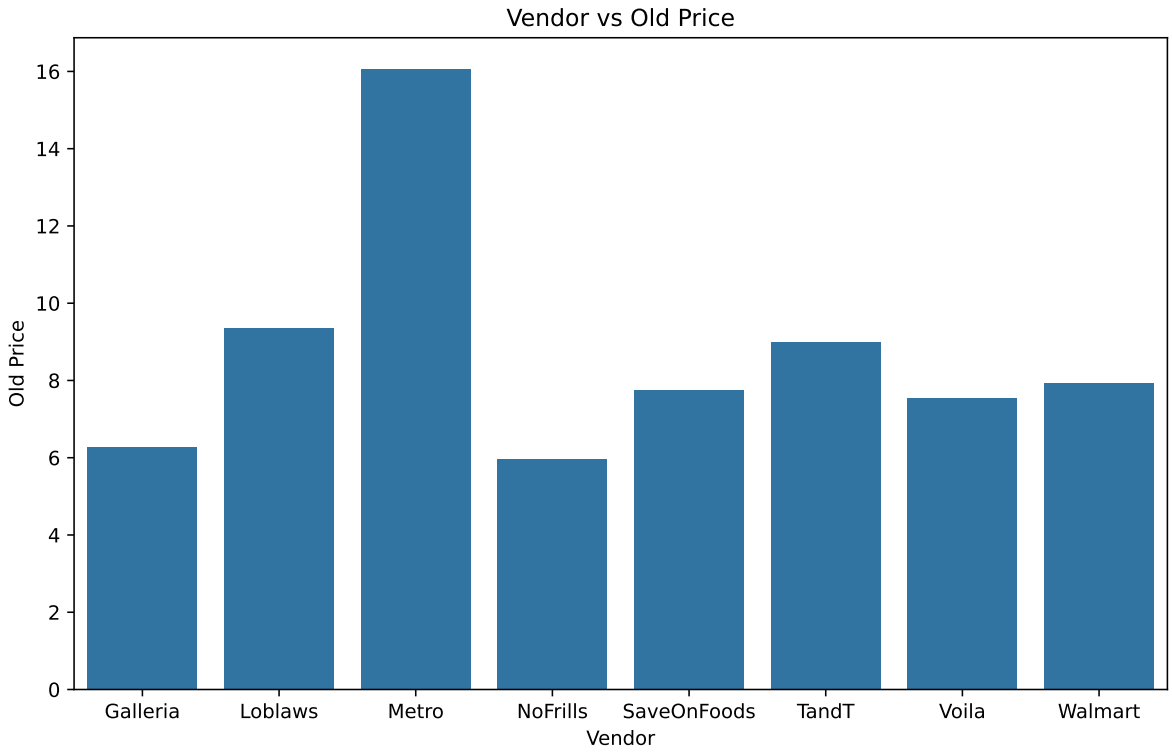

Figure 3: Vendor vs Old Price

Figure 3 above shows the distribution of the old prices across different vendors. Metro continues to have the highest old prices, exceeding 16 units. This finding reinforces Metro's position as a premium-priced vendor. NoFrills, on the other hand, has the lowest old prices, averaging around 6 units. This suggests that NoFrills consistently positions itself as an affordable option even before sales or discounts. Vendors like SaveOnFoods and TandT show moderately high old prices, similar to their current pricing trends.

The close alignment between old prices and current prices for most vendors suggests minimal artificial price inflation in the dataset. However, examining the relationship between old prices and sale prices could reveal whether some vendors artificially inflate prices before marking items on sale.

shows the correlations among the numeric variables in the dataset. This visualization allows us to visually see relationships between pricing variables.

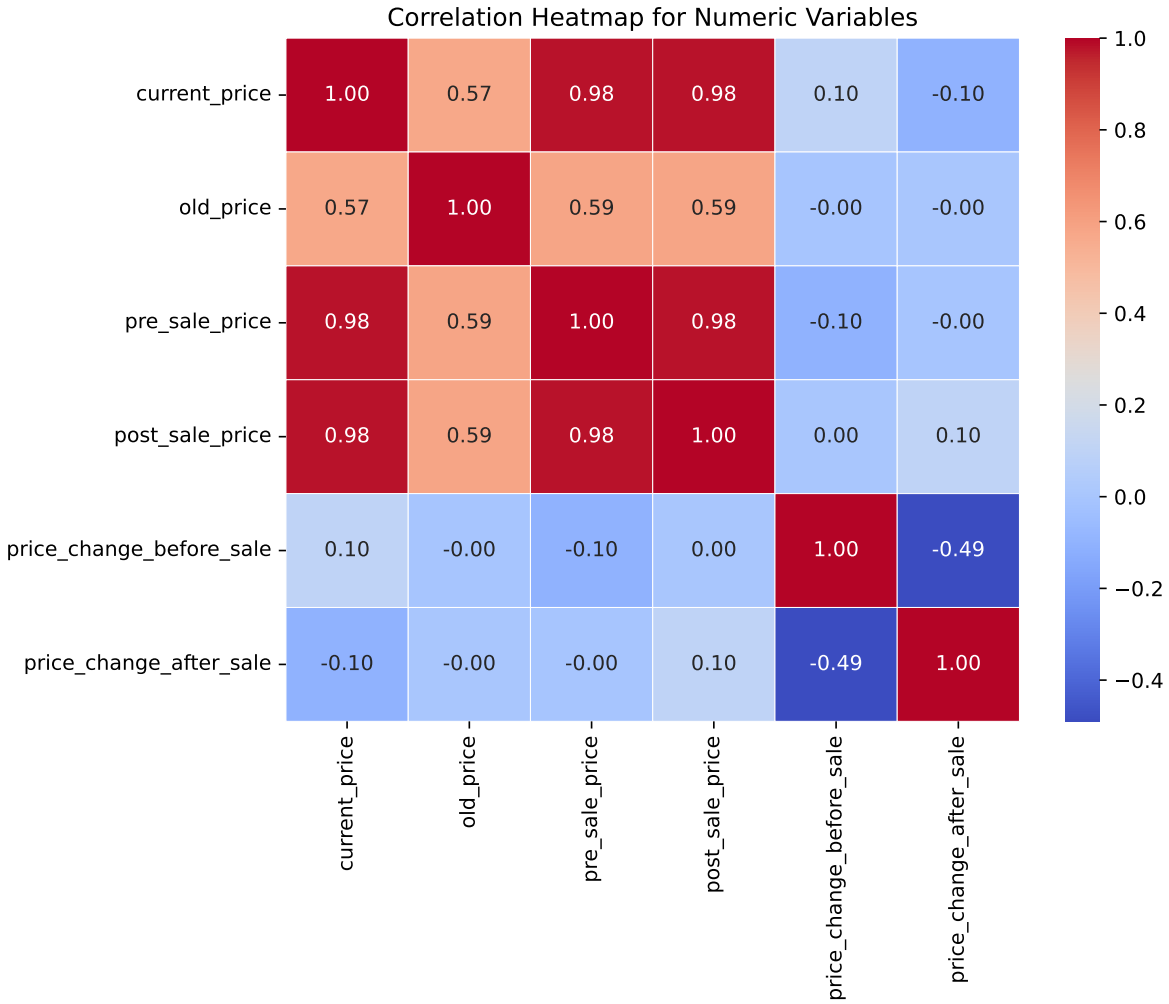

Figure 4: Correlation Heatmap for Numeric Variables

Figure 4 The current price shows a strong positive correlation with both the pre-sale price and the post-sale price (values of 0.98 for both). This indicates that the prices are closely related over time, suggesting consistency in pricing practices, even around sale events. The correlation between current price and old price is moderate (0.57), suggesting some dependency but also reflecting a potential for price adjustments or manipulations before sales. There is a significant negative correlation (-0.49) between price change before sale and price change after sale. This suggests that a large price drop before a sale is often followed by a proportional price increase after the sale, potentially indicating pricing strategies aimed at creating the illusion of larger discounts.

6 Discussion

6.1 Main Findings

This study aimed to investigate whether prices in the Canadian grocery sector are artificially inflated before a sale, only to be lowered afterward. The results indicate that there is a statistically significant difference between prices before and after sales. However, the price changes are relatively small, with an average decrease before sales and an average increase after sales. The Wilcoxon signed-rank test confirmed that these changes are not due to random variation, and the correlation between price changes before and after sales was very weak.

One of the key findings is that the mean price change before the sale was negative, suggesting that prices often decrease slightly before sales events. This may be attributed to retailers adjusting prices to make the sale appear more appealing to consumers. On the other hand, the positive price change after the sale suggests that prices tend to return to their normal level, but with minor fluctuations.

6.2 Implications for the Grocery Sector

The findings have several implications for both consumers and policymakers. While the average price changes are small, they could still influence consumer purchasing decisions, especially in highly competitive markets where small discounts are important for attracting customers. Retailers could potentially use these pricing strategies to manipulate consumer perception of discounts and deals.

Regulators may want to consider monitoring such pricing behaviors to ensure that they do not result in misleading marketing practices. This would help protect consumers from false perceptions of value and maintain the integrity of pricing strategies in the retail sector.

6.3 Limitations and Future Research

There are several limitations to this study that should be addressed in future research. First, the dataset is limited to a few months of price data from a single region in Canada. A more comprehensive analysis could involve a larger and more diverse dataset, including data from other regions and countries. Additionally, the study only considers price changes and does not account for other factors, such as product quality or seasonal demand, which could also influence pricing strategies. Future studies could also explore whether these pricing behaviors differ across different types of products or retailers. For example, essential goods may not exhibit the same price manipulation behavior as luxury or non-essential items.

6.4 Concluding Statement

This paper allows us to evaluate price manipulation practices in the Canadian grocery sector, showing that while there are statistically significant changes in prices before and after sales, the actual magnitude of these changes is relatively small. However, these practices may still have implications for consumer trust and perceptions of value. Further research is needed to explore these behaviors in more detail, especially in the context of regulatory policies aimed at ensuring fair pricing practices.

A Appendix

A.1 Sampling and Observational Data

This paper explored pricing behavior in the Canadian grocery sector through observational data. The primary source of this data was a retail price dataset that includes records from multiple grocery chains over several months. While this dataset was not generated through an active survey or poll, it can still be categorized as observational data. Below is a detailed exploration of the sampling methods, observational data structure, and how these elements align with the literature.

A.2 Data Collection Methodology

The data was downloaded from Jacob Filipp website (Filipp, 2024). According to the author, the data was collected through web scraping of product price information from the online platforms of major grocery retailers in Canada, such as Walmart, Metro, and Loblaws. The scraping process was automated, and data was captured at various timestamps to reflect the timing of sales events.

The dataset contains the following key variables: - Product Details: Includes product ID, name, vendor, and packaging type. - Price Information: Contains the current price and, if available, the price prior to the sale event. - Sale Information: An indicator of whether the product is on sale or not.

A.3 Sampling Method

The data had a large number of products listed on each retailer's platform; however, the full dataset was used in the analysis. This ensured that the data included a mix of high-volume and low-volume products, as well as those with varying price points. This diversity allows for a thorough analysis of price trends across different product types and market segments. Additionally, the dataset spans several months, allowing us to see the effects of price changes over time.

A.4 Linkages to the Literature

The methodology aligns with prior studies on pricing behavior in retail sectors, particularly those examining sales events and promotional pricing. According to Smith (2018) (Smith 2018), price increases before sales and subsequent reductions are common strategies in retail pricing. This paper builds on that work by using a granular dataset and statistical techniques to empirically test the prevalence and significance of these practices in the Canadian grocery sector.

A.5 Potential Biases & Limitations

As with any observational data, our analysis is subject to several biases:

- Time Bias: The dataset was collected during a specific timeframe, which may not account for long-term pricing trends.
- Retailer Bias: Different retailers may have varying sales strategies, which could influence the results. These limitations are acknowledged and its suggested that future research could expand the dataset to include more retailers and a longer period for more generalizable conclusions.

References

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

Smith, J. A. 2018. *Retail Pricing Strategies in a Competitive Marketplace*. Oxford University Press.