Beef Pricing Dynamics in Canadian Supermarkets*

Bo Tang

December 3, 2024

This study examines beef pricing trends in two Canadian supermarkets, T&T and Loblaws, using data from June to November 2024. We found that prices were stable overall, with seasonal increases in winter and significant differences between the supermarkets' pricing strategies. These findings reveal how cultural and operational factors, alongside seasonality, influence consumer pricing. This contributes to a better understanding of market dynamics and economic behavior in Canada.

Table of contents

*Code and data are available at: [https://github.com/Botangeric/Beef-Pricing-Dynamics-in-Canadian-Supermarkets).

1 Introduction

Beef is one of the most widely consumed staples in Canada, making it a critical product for understanding consumer behavior, market dynamics, and economic trends. Pricing trends in beef products provide insights into broader economic forces, including the effects of supply chains, consumer preferences, and seasonal variations. Despite its significance, a detailed analysis of beef pricing strategies in major Canadian supermarkets is limited, particularly when considering cultural differences and post-pandemic economic recovery.

This paper focuses on analyzing beef pricing trends in two leading Canadian supermarkets, T&T and Loblaws. These supermarkets cater to distinct consumer groups, with T&T focusing on Asian demographics and Loblaws serving a broader North American audience. By comparing these two supermarkets, we aim to uncover patterns in pricing strategies, seasonal variations, and factors influencing beef prices. The study utilizes a dataset spanning June to November 2024, capturing detailed pricing information across multiple products and months.

A key gap addressed in this paper is the limited understanding of how supermarket-specific strategies and broader economic factors interplay in shaping beef prices. While previous studies have focused on aggregate market trends, few have explored the cultural and operational distinctions between retailers and their implications for pricing. Our analysis fills this gap by incorporating a Bayesian linear regression model to evaluate the impact of vendor type, seasonal trends, and historical pricing on current beef prices.

The results of this study reveal that beef prices were relatively stable during the study period, with limited fluctuations and even discounted prices in some cases. However, significant differences in pricing strategies between T&T and Loblaws reflect cultural and operational distinctions. Furthermore, seasonal variations indicate that winter prices tend to be higher than summer prices, driven by increased demand during holidays and colder weather. These findings underscore the importance of considering both macroeconomic trends and retailer-specific strategies when analyzing pricing dynamics.

The remainder of this paper is structured as follows: Section 2 details the dataset and methodology used in this analysis. Section 3 describes the Bayesian linear regression model and its setup. Section 4 presents the results, highlighting key trends and insights. Finally, Section 5 discusses the implications of these findings, the limitations of the study, and potential directions for future research.

1.1 Estimand

We want to estimate our current price using vendor, month, and old price.After a quick review of the raw data, I believe that the estimated current price should ultimately be lower than the old price.

2 Data

2.1 Overview

The data for this study came from Project Hammer(Jacob 2024).The datasets were cleaned and analysed using the statistical programming software `R` (R Core Team 2023) along with the help of `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2014), `ggplot2` (Wickham 2016), `here` (Müller 2020), `dplyr` (Wickham et al. 2023), `rstanarm` (Goodrich et al. 2024), `arrow` (Richardson et al. 2024), `broom.mixed` (Bolker and Robinson 2022) and `kableExtra` (Zhu 2024).The raw data covers eight major large supermarkets in Canada and the changes in their products between February 28, 2024, and November 26, 2024.In the following section, I will explain the components of my dataset, the time period it covers, and other detailed information.

2.2 Measurement

The data was collected by Jacob from the product prices on each supermarket's website, and these prices are all for "in-store pickup".The data we obtained is all based on the prices of products visible to consumers, and we cannot observe the prices at which retailers acquire the goods.

- **Nowtime:** Timestamp indicating when the data was gathered.
- **Vendor:** One of the 7 grocery vendors.
- **Product Id:** An internal ID for a product - it represents 1 unique product at 1 vendor. It will allow you to track prices on the same product over time.
- **Product Name:** Product name. May include Brand, may include Units in it.
- **Brand:** Brand name. May be blank for some vendors.
- **Units:** Units (grams, kg, number of items in package). May be blank for some vendors/products.
- **Current Price:** Price at time of extract.
- **Old Price:** An "old" struck-out price. This indicates that there was a Sale on. This is how you can differentiate a price-drop advertised as a Sale vs. a "quiet" decrease without a Sale.
- **Price Per Unit:** Price per unit, as shown on the vendor website. May not be consistent with real calculation of "current price" divided by "units".
- **Other:** Other details that appeared on the listing. Values may be "Out of stock", "SALE", "Best seller", "$5.00 MIN 2".

2.3 Outcome variables

The outcome variable, current price, represents the most recent market price of products in the dataset. It serves as a critical metric for understanding pricing strategies, consumer behavior, and market trends. By analyzing current price, we can evaluate the impact of factors such as discounts (old price), vendor differences, and temporal trends (e.g., month) on product pricing.The line graph Figure 1 illustrates the average monthly current price trend across the dataset. The most notable pattern is the significant drop in prices between months 6 and 8, reaching the lowest average in month 8. This is followed by a steady upward trend, with prices gradually increasing and peaking in month 11.

This trend likely reflects seasonal or market dynamics that influence product pricing. The decline in prices during the mid-year months could indicate promotional events, seasonal sales, or a decrease in consumer demand during that period. Conversely, the rise in prices from months 9 to 11 may correspond to increased demand, possibly due to holiday preparations or end of year market adjustments.

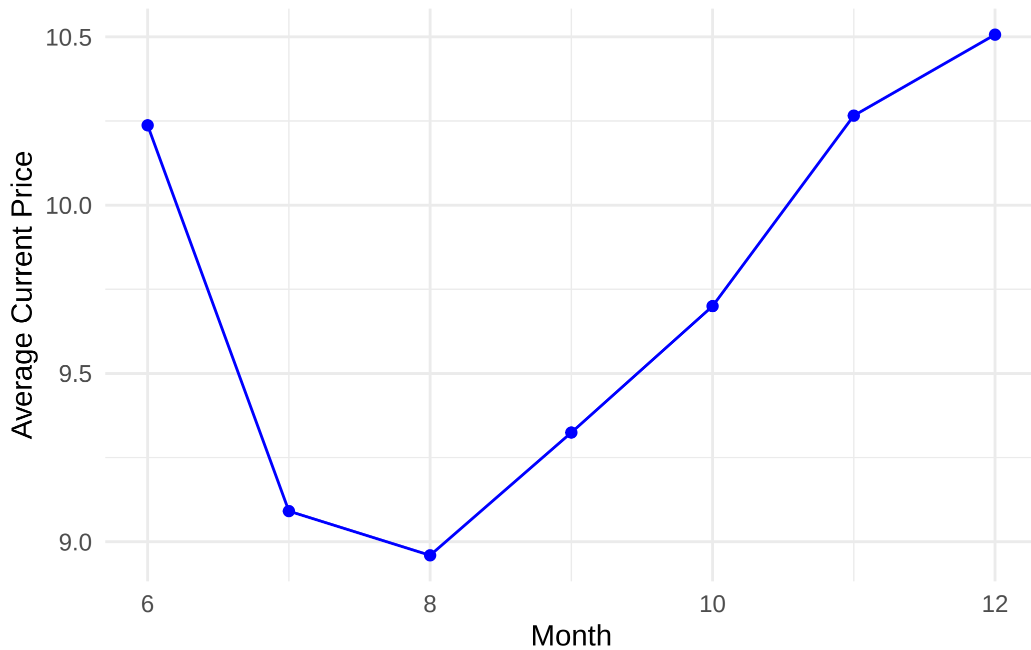

Figure 1: Current Price Distribution Across Months

The summary shows that the current price in the dataset ranges from 0.39 to 101.15, with most prices concentrated in the lower range, as indicated by a median of 6.87 and a third quartile of 10.99. The average price is 9.56, while the standard deviation of 11.39 suggests significant variability, likely influenced by a few high outliers.

	Statistic	Value
Min.	Minimum	0.390000
1st Qu.	1st Quartile	3.000000
Median	Median	6.870000
Mean	Mean	9.557173
3rd Qu.	3rd Quartile	10.990000
Max.	Maximum	101.150000
	Standard Deviation	11.391188

Summary Statistics for Current Price

2.4 Predictor variables

Our main predictor variables are vendor, month, and old price.

Vendor:This variable represents the seller or distributor of the product.

Month:This variable reflects the time of year (e.g., seasonality) and allows us to analyze temporal trends in pricing. It can help identify seasonal effects or promotional periods that impact product prices.

Old price:This variable represents the product's previous price. It serves as an indicator of discounts or price changes, which are key factors in determining the current price. It provides a direct measure of pricing adjustments over time.

2.4.1 Vendor

The boxplot Figure 2 illustrates the price distribution of beef products by vendor. Loblaws shows a wider range of prices, with several outliers exceeding $100, indicating higher variability in pricing. The median price for Loblaws is lower, with the interquartile range suggesting a focus on moderately priced products, but the presence of outliers points to occasional high-priced items. In contrast, T&T exhibits a more concentrated price distribution, with fewer outliers and a higher median price. The interquartile range for T&T is narrower, indicating a more consistent pricing approach with fewer extreme values. These patterns highlight the differences in the spread and central tendency of prices between the two vendors.

2.4.2 Month

Figure 3 illustrates the distribution of current prices across months, showing variations in the median price, interquartile range, and outliers for each month. The median price remains relatively consistent throughout the months, with slight increases observed in later months (e.g., 10 and 11). The interquartile range is narrow across all months, indicating that most prices are clustered within a specific range. However, outliers are present in every month, with several extreme values exceeding 50 and even 100 in some cases, particularly in months 7 and 10. This suggests that while the majority of prices are stable, there are occasional high-priced products influencing the distribution. The overall pattern suggests minor fluctuations in price distribution across the observed months.

2.4.3 old price

The scatterplot compares current prices against old prices for products, showing a strong positive relationship between the two variables. Most points cluster near the red dashed line (representing equality between current and old prices), indicating that current prices often

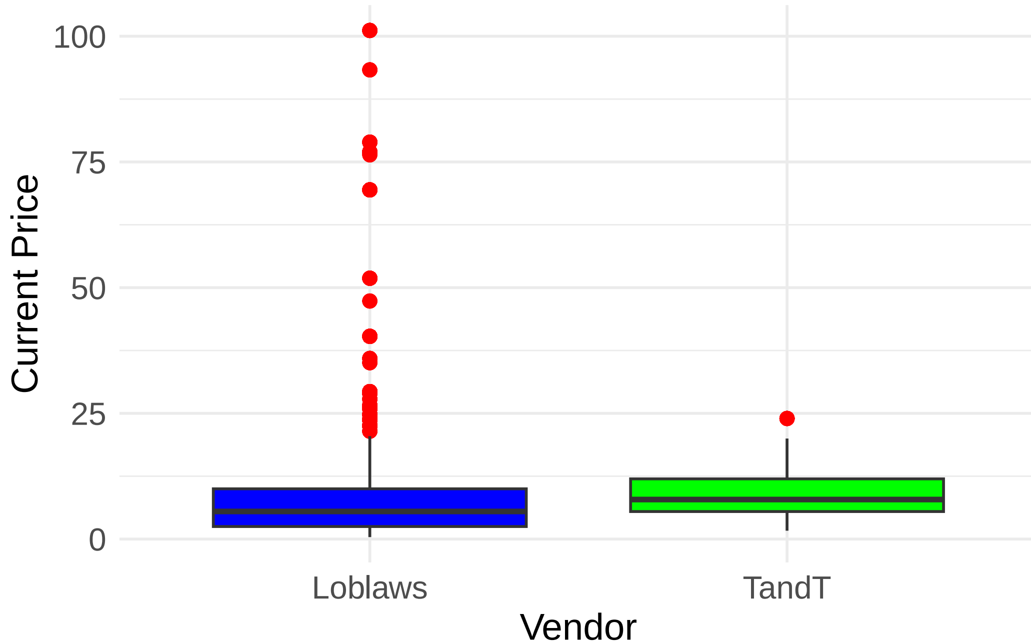

Figure 2: Price Distribution of Beef Products by Vendor

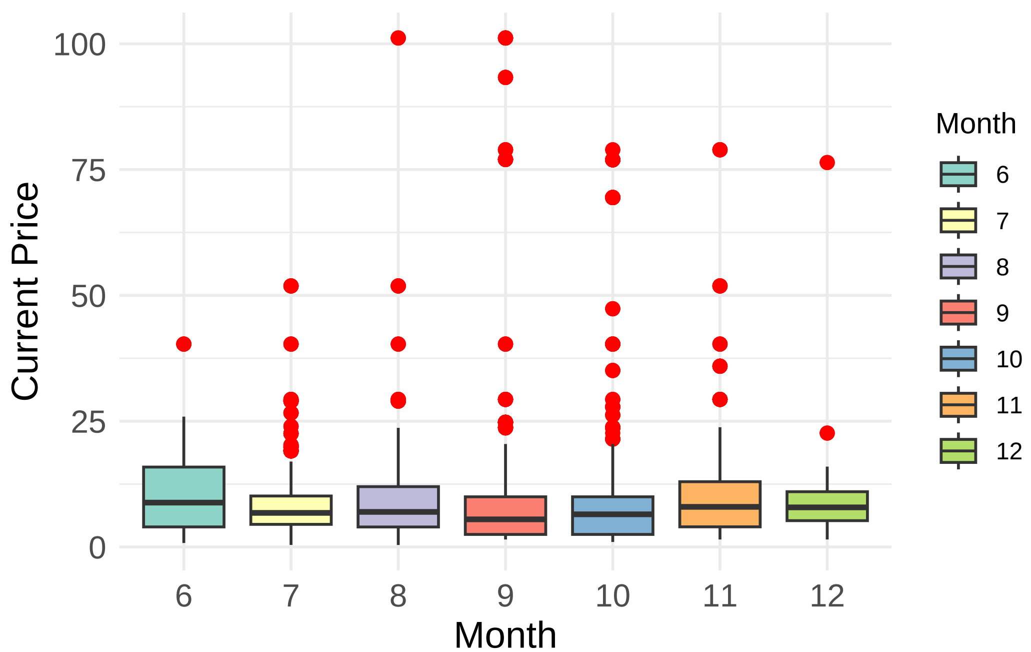

Figure 3: Current Price Distribution Across Months

closely follow old prices. However, there are noticeable deviations where current prices are lower than old prices, which suggests discounts or price reductions. A few outliers, particularly for old prices exceeding 100, show significant variance, with current prices remaining below the old prices. This pattern highlights that while prices generally follow historical trends, discounts and occasional large price differences exist.

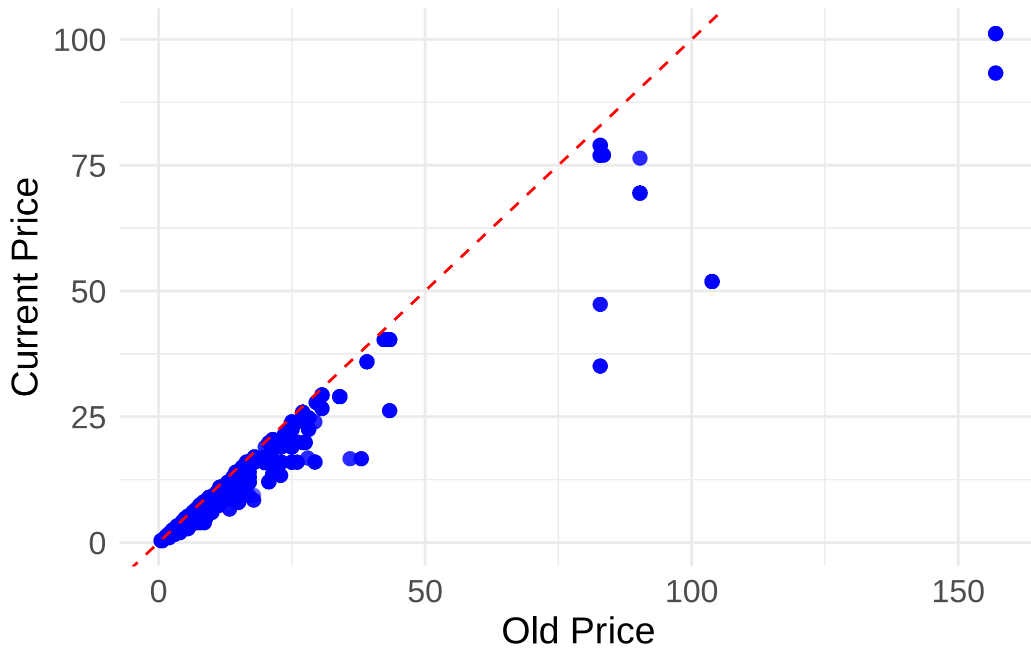

Figure 4: Comparison of Current Price vs Old Price

3 Model

3.1 Model overveiw

The model is a Bayesian linear regression that predicts the current price of beef products. It uses the month and old price as a numerical variable and includes vendor as categorical variables. This setup allows the model to account for the effects of historical prices, differences between supermarkets, and seasonal variations.

3.2 Model set-up

We set current price to be Y,Old price to be X,Vendor to be Z,month to be H,The linear regression model I will be using is:

$$Y = \beta_0 + \beta_1 \cdot X + \beta_2 \cdot Z + \beta_3 \cdot H$$

$$\begin{aligned} \beta_0 &\sim \text{Normal}(0, 2.5) \\ \beta_1 &\sim \text{Normal}(0, 2.5) \\ \beta_2 &\sim \text{Normal}(0, 2.5) \\ \beta_3 &\sim \text{Normal}(0, 2.5) \end{aligned}$$

Where,

- β_0 The intercept term, indicating the baseline price when all predictors are at their reference levels.
- β_1 The coefficient for old_price, representing the linear effect of the old price on the current price.
- β_2 The coefficient for vendor, representing the effect of different vendors on the current price (a categorical variable encoded as dummy variables).
- β_3 The coefficient for month, representing the effect of different months on the current price.

In my model, normal priors with a mean of 0 and a standard deviation of 2.5 are used for the coefficients. This choice implies no specific expectation regarding the direction or magnitude of the relationships between the predictors and the dependent variable. I selected these priors because I do not have strong prior knowledge about the likely effect sizes in this context.

The standard deviation of 2.5 represents a moderately wide prior, reflecting uncertainty in the prior beliefs. This ensures that the priors are weakly informative, allowing the data to play the dominant role in determining the posterior distributions.

The chosen priors also serve to regularize the model by limiting the potential for extreme coefficient estimates, which can arise due to noise or limited sample sizes. This regularization helps stabilize the model and improves its ability to generalize to new data, particularly when there are potential multicollinearity issues or when the dataset is small.

3.3 Model Limitatioin

A limitation of the model is that it is entirely trained on a single dataset without splitting it into training and testing subsets. While this approach ensures all data contributes to parameter estimation, it limits the ability to validate predictions on unseen data, potentially impacting generalizability. Additionally, the model only considers three predictors—old price, vendor, and month—while other factors, such as production costs, supply chain disruptions, or regional economic conditions, may also influence beef prices but are not included. Furthermore, the dataset covers only a six-month period in 2024, restricting the model's applicability to other timeframes or broader market contexts.

3.4 Model justification

The Bayesian linear regression model was selected for this analysis due to its flexibility and ability to incorporate prior knowledge. This model allows us to estimate the effect of historical prices, vendor strategies, and seasonal trends on beef pricing while accounting for uncertainty in the estimates. The choice of predictors, including scaled old prices, vendor, and month, aligns with established theories in market dynamics and consumer behavior, ensuring the model is grounded in sound reasoning.

A Bayesian approach was particularly suitable for this dataset, as it balances interpretability with predictive accuracy. By using priors, the model prevents overfitting and provides robust estimates even with limited data. This framework also enables us to quantify uncertainty in the predictions through posterior distributions, offering richer insights compared to traditional regression methods. Additionally, the inclusion of categorical variables (vendor) and a continuous predictor (old prices and month) makes the model versatile for capturing both fixed and dynamic aspects of pricing. These attributes make Bayesian linear regression the optimal choice for this analysis.

4 Results

Figure 5 illustrates the coefficients and their corresponding 95% credible intervals for the predictor variables in the Bayesian model. A 95% credible interval means that there is a 95% probability that the true parameter lies within the interval, given the observed data and the model assumptions.

Each point represents the estimated coefficient for a predictor variable, with the horizontal lines indicating the credible interval around the estimate. Variables with coefficients to the right of zero indicate a positive association with the outcome variable, suggesting that an increase in the predictor variable corresponds to an increase in the outcome variable. Conversely, coefficients to the left of zero indicate a negative association, implying that an increase in the predictor variable is associated with a decrease in the outcome variable. Variables whose credible intervals include zero indicate that their effect on the outcome is uncertain or not statistically significant.

The results show that old price scaled is a significant positive factor influencing the current price, with its credible interval entirely to the right of zero, indicating that higher old prices are associated with higher current prices. In contrast, the coefficients for vendor "TandT" and most month variables have credible intervals that include zero, suggesting their effects on current price are weak or uncertain. Additionally, the intercept has a high estimate with a credible interval far from zero, highlighting the significant role of the baseline price in the model.

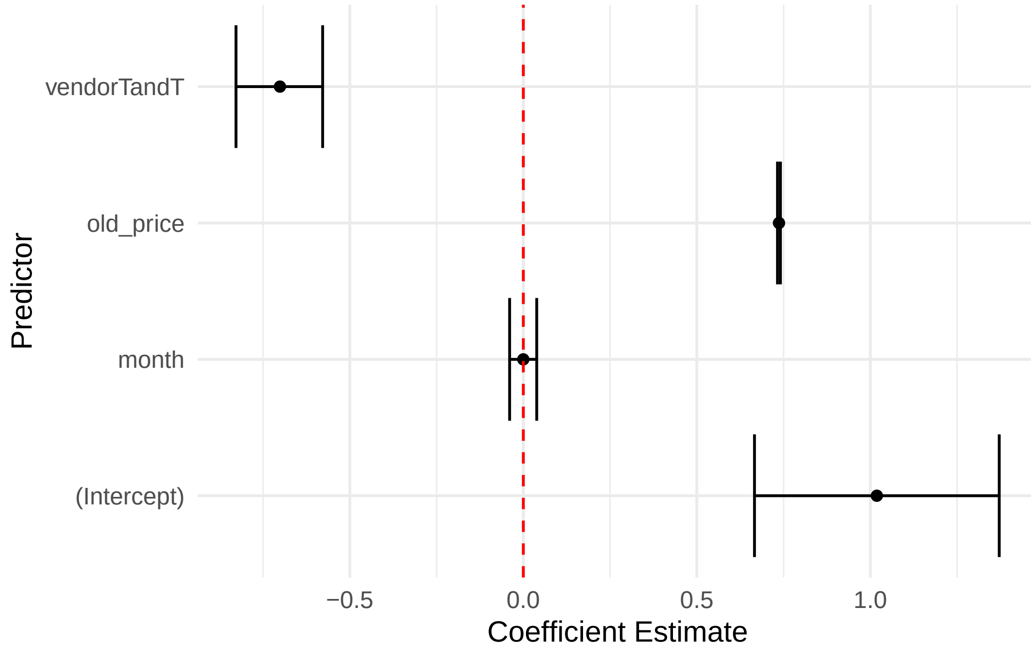

Figure 5: Model Coefficients with 95% Credible Intervals

Figure 1 In the data section, we can observe that beef products exhibit clear seasonal price patterns. During summer, the prices of beef products are significantly lower compared to winter, which may be due to seasonal demand changes and the impact of holidays. For example, in winter, the demand for beef products is typically higher, possibly driven by a preference for high-calorie foods during colder weather. Additionally, most major holidays, such as Christmas and Thanksgiving, occur in winter, further boosting the demand for beef products and leading to price increases.

Figure 2 illustrates the price distribution of beef products by vendor using boxplots for Loblaws and TandT. The y-axis represents the current price, while the x-axis groups prices by vendor. Loblaws exhibits a much wider price range, with numerous outliers far beyond the interquartile range (IQR), and some prices exceeding 100 units. In contrast, TandT has a narrower distribution, with most prices concentrated within a smaller range and fewer extreme outliers.

As a local supermarket, Loblaws offers a diverse selection of beef products with significant price variation, catering to a broad spectrum of consumers, from those seeking affordable options to those purchasing high-end products. This results in a broader price distribution and a greater number of high-price outliers. TandT, an Asian supermarket, focuses on a more specific product line tailored to the preferences of Asian consumers, leading to a more consistent and concentrated price range with fewer outliers.

Figure 4 compares the current prices of products to their old prices. Most points are concentrated near the 1:1 red dashed line, indicating that current prices are generally close to old prices. However, there is a noticeable trend where current prices tend to be slightly lower, suggesting a general price reduction across products. A few outliers show significant deviations, where some products experienced substantial price cuts or increases. This indicates a mix of market factors, such as promotions or pricing adjustments. Overall, the data reflects a slight downward shift in prices, with a few exceptions warranting further investigation.

5 Discussion

5.1 The pricing situation of beef products

In this article, I aim to analyze the pricing of beef products in two major supermarkets, T&T and Loblaws. I chose beef products because they are one of the main staples consumed by people living in Canada. Changes in their prices can reflect not only market conditions but also broader economic trends.

After studying and analyzing the data, I found that during the time frame provided by the raw data (June to November 2024), the prices of beef products were relatively stable. Price fluctuations were minimal, and there was no significant increase. In fact, some products were sold at discounted prices, which were significantly lower than their original prices. Additionally, my research revealed substantial differences in pricing strategies between the two supermarkets. These differences can be attributed to the fact that the two supermarkets target distinct consumer groups—Loblaws caters primarily to a North American audience, while T&T serves an Asian demographic. These differences in cultural preferences and business strategies result in variations in the pricing of beef products between the two supermarkets.Moreover, beef is a highly seasonal product. As observed in the analysis above, the price of beef in winter is approximately $1 higher than in summer, and as temperatures warm up, the prices decrease again.

5.2 Post-Pandemic Economic Recovery and Rising Beef Prices

When I further researched online, I found that even in 2024, a year when beef prices remained relatively stable, they were still higher compared to previous years. Why does this happen? I believe it is because, after the pandemic, the global economy has been undergoing a recovery period, and the impacts of COVID-19 have not completely disappeared even today. Under such circumstances, factors such as labor costs, feed costs, and transportation costs—which all contribute to the price consumers pay for beef products—have increased. This, in turn, has indirectly forced consumers to spend more to purchase the same beef products.

5.3 Weaknesses and Limitations

In this study, we chose to focus on only two supermarkets, T&T and Loblaws, to control the sample size. While this approach helps streamline our analysis, it also limits the generalizability of our findings, as the data may not represent the broader Canadian beef market. Additionally, in handling missing values, we opted for direct removal, which, while simplifying the data processing, might have introduced bias and reduced the robustness of the dataset.

Our analysis of factors influencing beef prices focused on three variables: month, vendors, and old price. However, beef prices are shaped by a wide array of factors beyond these three. For instance, the costs associated with bringing beef products to market—such as labor, transportation, and feeding costs—as well as external influences like government market policies and supermarket business strategies, all play a role. By simplifying the analysis to these three variables, our study may lack the depth required to produce authoritative predictions about beef pricing.

Furthermore, our dataset is restricted to the year 2024, which limits the applicability of the model to other years. Beef prices vary significantly from year to year due to both predictable factors (such as seasonal trends) and unforeseen circumstances (such as economic disruptions or policy changes). Consequently, our model is best suited for predicting beef prices in 2024 within regions of Canada where T&T and Loblaws operate. Extrapolation to other years or locations should be approached with caution.

5.4 Future Directions

Future research could expand the scope to include more vendors, product categories, and regional variations to provide a broader understanding of market dynamics. Incorporating external factors, such as economic conditions, consumer spending habits, and supply chain disruptions, could enhance the model's explanatory power. Moreover, applying more granular time-series analysis or hierarchical modeling approaches could uncover detailed temporal patterns and interactions between predictors. Weaknesses and next steps should also be included.

Appendix

A Additional data details

A.1 Analysis dataset

Table 2: Analysis dataset

vendor	current_price	old_price	product_name	month
Loblaws	8.99	12.49	Kofta Beef Meatballs	7
Loblaws	8.99	12.49	Kofta Beef Meatballs	7
Loblaws	8.99	12.49	Kofta Beef Meatballs	7
Loblaws	8.99	12.49	Kofta Beef Meatballs	7
Loblaws	8.99	12.49	Kofta Beef Meatballs	7
Loblaws	8.99	12.49	Kofta Beef Meatballs	7

Table 2 provides a glimpse of the dataset used for analysis.

A.2 Data cleaning

After obtaining the raw data, I realized that the sample size was too large, with over 10 million rows. Therefore, I decided to focus primarily on the beef product prices of the two supermarkets, "T&T" and "Loblaws". These two supermarkets are more popular and offer a relatively rich variety of beef products. I chose beef products because, in daily life, beef is one of the main foods consumed by people, and changes in its price can reflect certain social conditions.

I also processed the dates, focusing primarily on monthly changes. I simplified the nowtime variable to retain only the month. Additionally, some data in the dataset contained missing values. However, when reviewing the raw data, I observed that for the same product within the same month, the current price and old price were relatively consistent. For example, if there was complete data on the 1st of the month, missing data on the 4th, and complete data again on the 10th, the values on the 1st and the 10th were identical. Therefore, to control the sample size, I chose to directly remove entries with missing values.After handling this, it will not have a significant impact on my monthly figures.

B Model details

B.1 Model summary

Table 3: Summary of the model

term	estimate	std.error	conf.low	conf.high
(Intercept)	1.02	0.22	0.67	1.37
old_price	0.74	0.00	0.73	0.74
vendorTandT	-0.70	0.08	-0.83	-0.58
month	0.00	0.02	-0.04	0.04

Table 3 summarizes the results of the Bayesian model, providing key insights into the model's parameters. The Estimate column represents the posterior mean of each parameter, indicating its effect size on the outcome variable. The Standard Error (SE) measures the precision of these estimates, reflecting the variability of the sample statistics. The 95% Credible Interval (CI) shows the range within which the parameter lies with 95% probability, given the observed data and prior distributions. Additionally, diagnostic metrics like Rhat assess the convergence of the model, while Effective Sample Size (ESS) ensures the reliability of the posterior samples. This table offers a comprehensive view of the model's performance and parameter estimates.

B.2 Posterior predictive check

Figure 6 The observed data (dark curve) and the predictive data (light curve) are largely aligned, suggesting that the model captures the central features of the data reasonably well. Deviations between the two curves may indicate areas where the model struggles to replicate certain patterns in the observed data.The peak in both distributions around lower values indicates that the majority of observed and predicted values are concentrated in this range.The predictive distribution's spread captures most of the observed data, indicating a good fit overall.

B.3 Diagnostics

Figure 7 and Figure 8 indicate that the Bayesian model has converged properly for all parameters. The consistent mixing and stable trajectories across chains suggest that the posterior distributions are well-estimated and that the results can be interpreted with confidence.Since rhat are close to 1,it confirms that Bayesian model has achieved proper convergence, and the parameter estimates are robust and can be confidently interpreted.

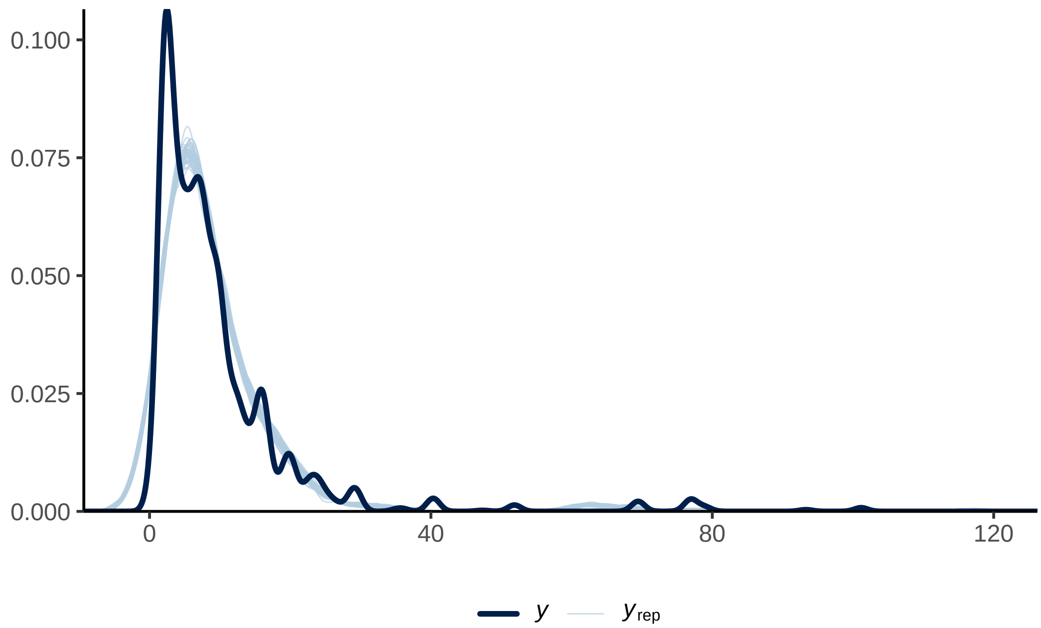

Figure 6: Posterior distribution for regression model

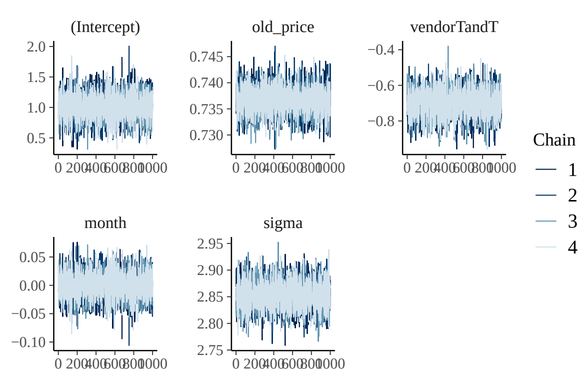

Figure 7: Markov chain Monte Carlo (MCMC) Convergence Check

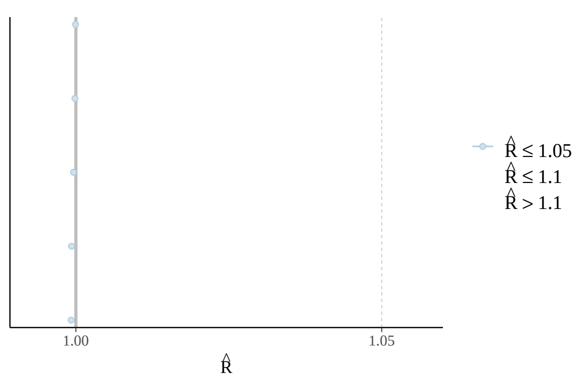

Figure 8: Markov chain Monte Carlo (MCMC) Convergence Check

C Data Collection for Beef Products

C.1 Introduction to Survey Design

We aim to design a survey to help Canadians better understand the price fluctuations of beef products across the country. This survey will support Canada in gaining a clearer grasp of beef pricing trends to inform policymaking and other decisions.

C.2 idealize Collection methods

Collecting daily product prices from the websites of major supermarkets, such as Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods, provides a valuable way to gather comprehensive pricing information. Focusing on large chain supermarkets ensures better data coverage and consistency. Using an automated approach, we can record daily prices of various products from different supermarkets and consolidate all the data into a single dataset.

The primary expense in this setup will be software development to automate the data collection and storage process. This system will streamline data gathering and ensure consistency over time. To achieve this, all activities must be conducted with the explicit consent of the supermarkets and in full compliance with ethical and legal standards, ensuring no infringement on the rights of individuals or groups. Through this system, we can obtain three key pieces of data: the product name, its location of sale, and its price. Additionally, we can collaborate with supermarkets to directly access their beef product pricing information. By complementing this with daily sales data, we can better analyze the demand for beef products.Building on this, we will divide the analysis into two major sections: rural and urban areas. The datasets for these two categories will be analyzed separately, as the demand and market size for beef products differ significantly between rural and urban areas. This approach will allow us to conduct a more targeted analysis of each group.

C.3 Survey Questionnaire

With the market survey and data collection complete, we will have gathered enough information to proceed. Next, our focus will shift to understanding the people themselves. By analyzing demand for beef products across different demographics—such as age, region, and gender—we aim to study how demand varies among different groups. This will allow us to gain deeper insights into consumer preferences and behaviors.

Section 1: Personal Information

1. Gender

- Male
- Female
- Other/Prefer not to say

2. Age Group

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 and above

3. Location

- Province/Territory: (Drop-down menu or text input for respondents to specify their province/territory in Canada)

4. Area:

- Urban
- Rural

Section 2: Beef Product Demand

1.How would you describe your demand for beef products?

- High demand (consume beef frequently, buy large quantities)
- Moderate demand (consume beef occasionally, buy moderate quantities)
- Low demand (rarely consume beef, buy small quantities)
- No demand (do not consume beef)

2.How often do you purchase beef products?

- Daily
- 2-3 times a week
- Weekly
- Biweekly
- Monthly
- Rarely/Never

Section Optional 3 Additional Feedback

1.Are there any factors that influence your beef product purchasing decisions?

- (e.g., price, availability, dietary preferences, cultural practices, etc.)

2.Would you like to be contacted for future surveys or updates on this research?

- Yes
- No

References

Bolker, Ben, and David Robinson. 2022. *Broom.mixed: Tidying Methods for Mixed Models*. https://CRAN.R-project.org/package=broom.mixed.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

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, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

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, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.