

Title*

Subtitle

Chenming Zhao

November 26, 2024

abstract

1 Introduction

2 Data

2.1 Overview

The product pricing data for the “No Name” brand sold by Loblaws, used in this analysis, is included in the dataset provided by (Filipp 2024). This dataset compiles detailed product information, including vendors, product names, brands, categories, and unit prices. Additionally, it includes time-series pricing records for each product, capturing timestamps of price changes, current prices, unit prices, and quantities sold. These records are periodically updated on the Hammer platform to reflect the latest pricing trends. This dataset forms the foundation for analyzing price trends, identifying unit variations, and categorizing products into specific types for further analysis. Table 1 previews the information for the target products included in this analysis, starting from June 2024.

To simulate, test, analyze, and clean the polling data, the statistical programming language R was employed (R Core Team 2023). Key libraries that supported the data analysis include `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2023), `arrow` (Richardson et al. 2024), `brms` (Bürkner 2017) `knitr` (Xie 2023), `ggplot2` (Wickham 2016), `kableExtra` (Zhu 2024) and `patchwork` (Pedersen 2024). Additionally, the `rstanarm` (Goodrich et al. 2022) library was utilized to implement multilevel modeling for prediction purposes. Also

*Code and data are available at: <https://github.com/ChenmingZ2000/Loblaws-NoNameBrand-Grocery-Price-Trend-Analysis.git>. We gratefully acknowledge Jacob Filipp’s groceries dataset used in this analysis. The data is available at: <https://jacobfilipp.com/hammer/>

this analysis inherits the folder structure from Professor Rohan Alexander from University of Toronto (Alexander 2024).

2.2 Measurement

In the real world, pricing data for consumer goods is a dynamic reflection of market trends, regional preferences, and competitive strategies. The prices of products fluctuate based on supply chain disruptions, seasonal demand, and retail strategies, making it essential to track and analyze this information systematically. The Hammer dataset serves as a resource in understanding these trends, offering a detailed collection of product pricing data sourced from leading retailers like Loblaws, Walmart, and Metro. This dataset is embedded with complexities arising from the data collection process, competitive pricing strategies, and evolving consumer behavior.

Although the data collection process for (Filipp 2024) data is not explicitly documented, based on the emphasis on “related” content on the Project Hammer website and standard industry methods, it is highly likely that common techniques such as web scraping, API integration, and data standardization were employed. These are typical approaches used to construct structured datasets from online retail platforms. That is, retailers provide public access to their product inventories and prices through their websites, making them prime candidates for automated data extraction. For example, web scraping tools are deployed to extract product details such as names, brands, vendors, categories, unit prices, and quantities, while timestamps capture the exact moment of data retrieval to ensure time-sensitive accuracy. However, even with automated methods, the process is not without its challenges. Retailers often introduce barriers like CAPTCHA systems or dynamic content loading, which require robust solutions such as headless browsers or API integrations to ensure smooth data acquisition.

Sampling and representativeness are considerations in Hammer’s data collection process. Since the dataset aims to reflect national pricing trends, it includes a diverse range of vendors and product types to capture regional and competitive variations. However, inherent biases may still exist. For instance, not all vendors update their online product catalogs in real time, which may introduce lag in capturing price changes. Additionally, regional pricing disparities might not always align with the dataset’s collection schedule, potentially skewing the data toward more actively monitored regions or vendors. Once the raw data is collected, it undergoes cleaning and standardization. Product names, for instance, often include variations due to marketing strategies (e.g., “Club Size” or “Family Pack”), which are standardized to ensure consistency across entries. Similarly, data with missing values—such as prices or unit information—are flagged for further investigation.

Finally, the cleaned and standardized data is stored in a structured database, organized by product ID, timestamps, and categorical variables such as vendor, brand, and product category. This structured approach allows researchers to track pricing trends over time, analyze variations across regions or vendors, and categorize products into specific types. Hammer’s

database not only facilitates trend analysis but also enables broader comparisons and aggregations, allowing researchers to examine shifts in consumer behavior or retailer strategies.

2.3 Analysis Data

This analysis focuses on Loblaw's "No Name" brand products, with the goal of understanding pricing trends. We only selected observations where the vendor is Loblaw and the brand is "No Name" because Loblaw announced a price freeze on more than 1,500 "No Name" products until January 31, 2023, to address the pressures of rising food prices. However, with the price freeze period now over in 2024, we are particularly interested in analyzing the price changes of "No Name" products (see more details in Loblaw Companies Limited 2022a). At the same time, we prioritized high-quality data by filtering out records with missing or incomplete price information, as we have plenty of price records for a product in one day and unreliable data could introduce significant bias and reduce the accuracy of our trend analysis. In researching price trends and product variations, we selected specific variables that are impactful for analyzing pricing behavior affected by product categorization. Below is an overview of the data used for this analysis Table 1, along with the selected outcome and predictor variables.

2.3.1 Outcome Variables

In this analysis of Loblaw's "No Name" brand products 2024, the primary outcome variable is the `current_price`, which represents the retail price of a product at the time of the record. Understanding the pricing trends of these products allows us to analyze consumer affordability, retailer pricing strategies, and competitive positioning in the market. By focusing on `current_price`, we aim to identify influences that influence price fluctuations over time and assess how product size, unit and type variations contribute to these trends. This metric serves as the foundation for evaluating the dynamics of product pricing and finding differences across product categories.

2.3.2 Predictor Variables

In this analysis, the predictor variables are the variables that we believe can influence the current price of products in the Loblaw grocery dataset. These variables are chosen based on their potential to explain variations in product pricing. Below are the predictor variables included in our model, along with an explanation of their significance.

Size: Product size is a determinant of pricing, as larger sizes generally cost more due to the increased quantity. However, pricing strategies such as bulk discounts or price-per-unit variations can create non-linear relationships between size and price, making size an important factor to analyze.

Type: Product type (e.g., snacks, seasoning, dairy) reflects category-level differences in pricing strategies. Different types of products may have varying production costs, supply chain logistics, and market demand, all of which contribute to price fluctuations.

Date: Time plays an important role in grocery pricing due to reasons like seasonal demand, inflation, supply chain disruptions, and promotions. Analyzing the impact of time allows us to capture temporal trends and shifts in pricing behavior, such as rising costs during inflationary periods or discounts during specific seasons.

2.3.3 Exclusions and Considerations

Other variables (e.g., vendor or brand-specific promotions) could also influence pricing. However, in this analysis, we focus specifically on the effects of **size**, **type**, and **date**, as they represent the most consistent and accessible predictors in the dataset. Additional variables, while potentially relevant, may not be consistently available or may introduce unnecessary complexity into the model. For example, promotions or regional pricing differences were not included in this analysis due to large amounts of data unavailability. Additionally, while **old_price** might influence pricing strategies, this dataset does not provide enough details. By focusing on the variables above, we aim to construct a reliable and interpretable model for understanding price dynamics within Loblaw’s “No Name” brand products.

Table 1: The First Few Lines of Cleaned Grocery Price Analysis Data

product_id	date	current_price	product_name	type	size
1295619	2024-06-11	5.49	Blueberry Waffles	Snacks	Large
1295620	2024-06-11	2.50	Blueberry Waffles	Snacks	Small
1757033	2024-06-11	3.99	Microwave Butter Flavour Popcorn	Snacks	Large
1870035	2024-06-11	5.99	100% Pure Canola Oil	Cooking Oil	Large
1870037	2024-06-11	3.99	100% Pure Canola Oil	Cooking Oil	Small
1870038	2024-06-11	8.49	100% Pure Canola Oil	Cooking Oil	Extra Large
1870043	2024-06-11	17.99	100% Pure Corn Oil	Cooking Oil	Extra Large

2.3.4 Relations Between Variables

The following figures, Figure 1 and Figure 2, present various aspects of data related to Loblaw’s “No Name” grocery prices, helping us understand how item prices change with different variables. In the left panel of Figure 1 and both panels in Figure 2, the x-axis is marked with dates categorized by month. The chart analyzes the relationships between product prices, size, type, and date, showing how these variables may interact and influence pricing trends. Product size and type have an impact on pricing trends, and we anticipate that incorporating these variables into the predictive model will improve the accuracy of price predictions under different product and market conditions.

While we observe certain significant fluctuations in the right panel of Figure 2, which may indicate abnormal prices being recorded or the presence of products with extreme prices in specific categories during certain periods, we cannot directly remove this data as it still contains information we need. In other words, the prices of these products can also explain some of the reasons behind price fluctuations. Completely removing these observations would weaken the model’s ability to reflect price trends accurately. Therefore, we choose to retain these observations and mitigate the biases they introduce by combining other variables to build a more robust predictive model.

The boxplot on the right part of Figure 1, shows the pricing differences across product sizes. Compared to “Small” and “Large” products, “Extra Large” products have a higher median price and a wider interquartile range. This reflects inherent pricing strategies where larger quantities are priced higher, but not always linearly (e.g., discounts for bulk purchases). However, the presence of outliers, especially in the “Extra Large” and “Large” categories, indicates that even within the same size group, different product types or brands can significantly impact final pricing. The narrow boxplot for “Small” products suggests a smaller range of price variation, which aligns with the lack of price fluctuations observed in the left panel of Figure 1 for small-sized products.

Figure 2 shows how average prices vary over time, segmented by product **type** (left) and **size** (right). In the left panel, we observe distinct pricing patterns across product types. For instance, “Seasoning” demonstrates higher volatility compared to “Flour/Rice,” likely due to differences in demand elasticity. The “Milk” category shows occasional price dips, possibly influenced by the introduction of new products or promotional events. The right panel shows price trends segmented by size. “Extra Large” products generally have higher average prices but exhibit more unstable trends over time. In contrast, “Small” products show minimal price variation, possibly due to their inherently lower pricing or because their unit price is relatively higher compared to larger sizes.

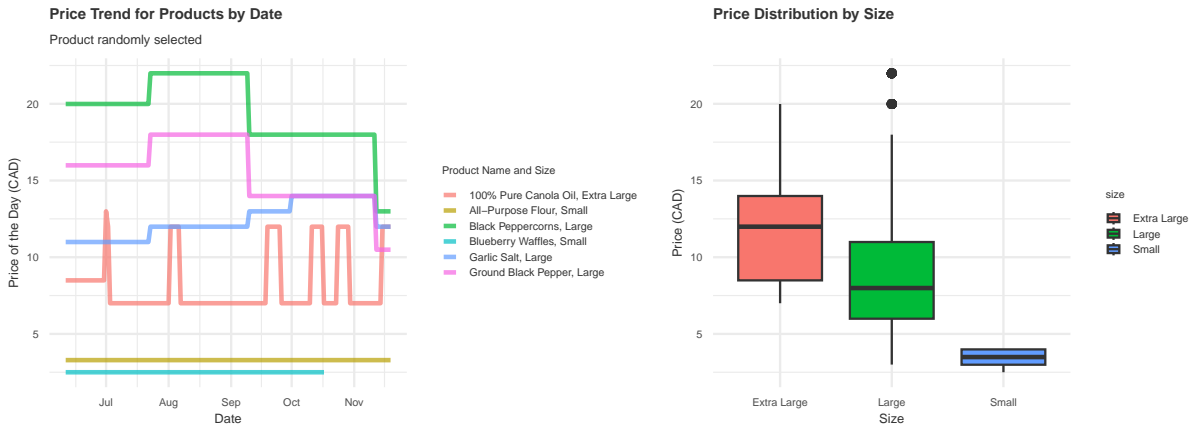


Figure 1: Price Trend for Six of the Loblaws “No Name” Groceries Over Time (left) and Price Distribution by Size (right).

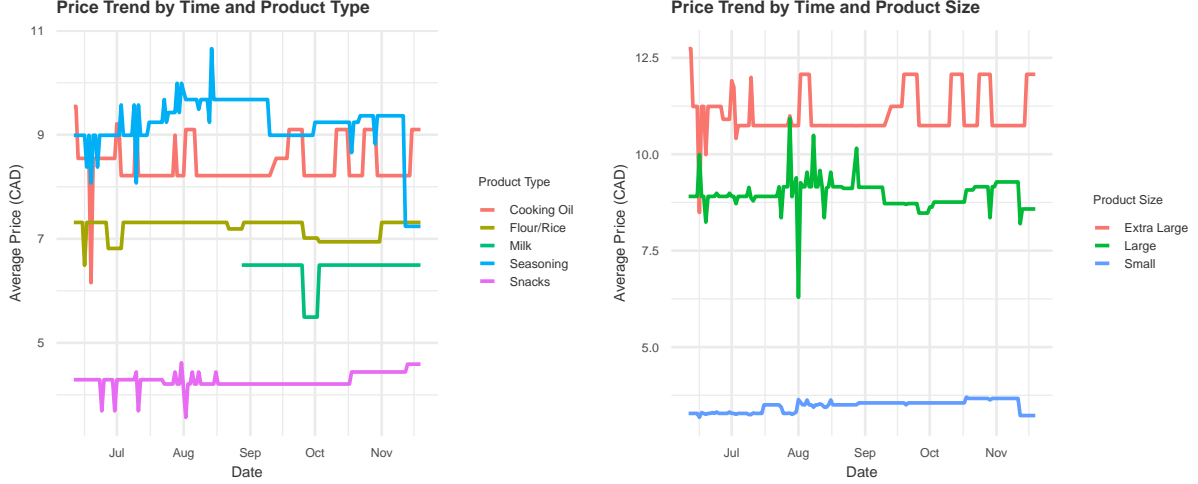


Figure 2: Average Support Rate for Donald Trump from 2024-07-21 to 2024-10-23 reported by different pollsters

3 Model

Our model references the analytical approach proposed by Chen, Aravkin, and Martin (2018), adopting a hierarchical Bayesian framework to model the price trends of Loblaw’s “No Name” grocery products. Specifically, the model is based on a Gamma regression framework with a log-link function, designed to handle the positively skewed distribution of product prices. By using the Gamma distribution, we can capture the inherent variability in prices while adhering to the non-negative nature of price data. The Gamma distribution is particularly well-suited for modeling continuous, non-negative, and positively right-skewed data, which are common in economic contexts such as insurance claims and survival data (Chen, Aravkin, and Martin 2018).

The goal of our modeling strategy is to build a pricing model aimed at predicting changes in `current_price` over time while considering key variables that influence price fluctuations. The structure of the model combines fixed effects and random effects and includes multiple variables, such as `product_id`, `product size`, and `product type`. This modeling approach is designed to balance the complexity of the data while ensuring robust predictions. The hierarchical structure of the model allows us to capture variations at the level of individual products while identifying overall trends across different product sizes and types. It also enables us to forecast future price trends of Loblaw’s “No Name” groceries, observing whether the retailer’s pricing behavior aligns with their commitments. Here, we provide a brief overview of the Bayesian analytical model used to investigate the price trends of Loblaw’s “No Name” products. Background details, model assumptions, and diagnostics are included in Appendix A.

3.1 Model set-up

This model is a generalized linear mixed model (GLMM) within a Bayesian framework, specifically designed to handle positively skewed data:

$$y_i \sim \text{Gamma}(\mu_i, \phi)$$

Define y_i as the current price of the product in the i -th observation, assumed to follow a Gamma distribution with mean μ_i and shape parameter ϕ . This model accounts for variations across different product sizes, types, and temporal trends, with random intercepts for each product ID.

The linear predictor for the mean (μ_i) of the Gamma distribution is defined as follows:

$$\log(\mu_i) = \alpha + \beta_1 \cdot \text{size}_i + \beta_2 \cdot \text{type}_i + \beta_3 \cdot \text{date}_i + u_{\text{productID}[i]}$$

where:

- y_i is the observed current price for the i -th product.
- μ_i is the predicted mean price for the i -th product, modeled on a log scale.
- $\alpha \sim \text{Normal}(2, 5)$ is the global intercept representing the baseline price.
- $\beta_1 \sim \text{Normal}(0, 5)$ captures the effect of product size on pricing (e.g., Small, Large, Extra Large).
- $\beta_2 \sim \text{Normal}(0, 5)$ represents the effect of product type (e.g., Snacks, Flour/Rice, Cooking Oil) on pricing.
- $\beta_3 \sim \text{Normal}(0, 5)$ accounts for the temporal trend in prices over time.

Random Effects: $u_{\text{productID}} \sim \text{Normal}(0, \sigma_{\text{productID}})$

where $u_{\text{productID}[i]}$ is the random intercept for each product ID, capturing product-specific variations in pricing.

3.2 Model justification

3.2.1 Model Prior Selection, Assumptions, and Validation

Our prior selection was informed by the research in gamma modeling Veazie et al. (2023), we aim to provide sufficient flexibility for the model to adapt to the data. For example, the prior for the intercept was set as $\text{Normal}(2, 5)$, reflecting the assumption that the baseline log-transformed price of Loblaws “No Name” products fluctuates around 2 while allowing for some uncertainty. For the effect coefficients of variables such as **size**, **type**, and **date**, we chose priors of $\text{Normal}(0, 5)$, representing the expectation of moderate impacts of these variables on pricing. These priors enable the model to adjust based on observed data.

The model assumes that price variations at the product level are non-negative, right-skewed, and that the price differences associated with size and type are random and normally distributed. While these assumptions simplify computation and enhance interpretability, they may not fully capture the impact of market conditions, regional competition, or seasonal trends on pricing dynamics. For instance, the model assumes that the effect of size is consistent across all product types, but in reality, certain product types may be more significantly influenced by supply chain issues or promotional activities. Additionally, the model incorporates random effects for individual products to capture specific variations for each product, assuming these effects remain consistent over time.

The model was implemented using the `rstanarm` package and employed MCMC sampling for Bayesian inference. To validate the model, we conducted out-of-sample testing and calculated the predictive accuracy using RMSE (Root Mean Square Error), which was 0.83 on the log-transformed price scale. In the context of grocery price data, this level of error is acceptable, as market forces and external factors inherently introduce noise. Convergence diagnostics (e.g., \hat{R}) showed values close to 1 for all parameters, indicating successful convergence of the chains. The effective sample size (`n_eff`) for key parameters exceeded 500, ensuring the robustness of the inferences. Sensitivity analysis on priors revealed that minor adjustments to prior specifications did not significantly affect the model’s predictions, demonstrating high model stability.

We also compared this model to alternative approaches. A simpler linear regression model without random effects failed to account for product-level variations, particularly for products with fewer observations, resulting in significantly higher prediction errors. Conversely, a more complex model incorporating additional factors such as promotional pricing or competitor prices achieved only marginal improvements in predictive accuracy but at a substantial increase in computational cost, making it overly complex. Therefore, the current model, which integrates a hierarchical structure and focuses on size, type, and temporal trends, strikes a good balance between predictive accuracy, interpretability, and computational efficiency.

3.2.2 Expectation of relationships between variables

For the variable `type`, we anticipate that different product types will influence the estimation of current prices. While we cannot precisely determine which product categories incur higher costs due to supply chain or consumer demand, it is reasonable to assume that categories such as “Snacks” or “Flour/Rice” may exhibit distinct pricing dynamics. For instance, “Snacks” might experience greater price volatility due to internet trends or elastic demand, whereas “Flour/Rice” could reflect a more stable trend. These category-specific variables introduce variability into the overall pricing estimates.

For the variable `date`, we hypothesize that the prices of Loblaw’s “No Name” products fluctuate over time, reflecting seasonal trends, market dynamics, and promotional activities. This hypothesis is based on historical observations, where prices often vary due to promotions or

stock-outs. Long-term trends, on the other hand, may reflect broader economic conditions or changes in production costs.

The variable size captures the impact of product quantity on pricing. We expect larger-sized products, such as “Extra Large,” to generally have higher prices due to increased quantities. However, economies of scale may result in lower unit prices for larger sizes, while smaller sizes, despite lower overall costs, may have higher unit prices. This pricing strategy is common in retail to ensure accessibility for different consumer groups while optimizing profit margins.

To account for variations among individual products, we introduced a random intercept ($1 \mid \text{product_id}$) for each product. This intercept captures product-specific price differences, such as brand effects, unique demand characteristics, or production costs.

By modeling the interaction between size and type ($\text{size} * \text{type}$), the model allows for a more detailed understanding of relationships, such as pricing differences for “Extra Large” products between “Cooking Oil” and “Milk.” This interaction term helps identify combined effects that cannot be fully explained by individual variables. Together, these variables enable the model to capture the dynamic pricing structure of Loblaws “No Name” products, reflecting time trends and structural differences across product categories and sizes.

4 Results

4.1 Model Output

Our model results are summarized in Table ???. This summary shows the effects of product attributes, including size and type, on price predictions. For instance, products categorized as “Large” are associated with a price reduction of 0.47 CAD compared to the baseline (Extra Large), while “Small” products show an even larger reduction of 0.98 CAD. Similarly, “Snacks” exhibit a negative effect of 0.39 CAD on price compared to the baseline type (Cooking Oil), while “Flour/Rice” shows a negligible effect. The model achieves an R^2 value of 0.967, indicating that it explains 96.7% of the variation in the data. This shows that the model effectively captures the main patterns in the data. The marginal R^2 , which excludes the random effects, is 0.742, suggesting that fixed effects alone explain a substantial portion of the variance. The Intraclass Correlation Coefficient (ICC) is 0.1, indicating that only 10% of the variability is attributable to clustering by `product_id`, suggesting relatively low within-group correlation. Additional metrics support the model’s robustness. The WAIC (6585.0) and LOOIC (6585.1) values confirm that the model balances predictive accuracy with generalizability, avoiding overfitting.

4.2 Prediction Output

Our analysis aims to predict the price trends of Loblaw’s “No Name” products from November 20, 2024, to February 2025. The results of our model are illustrated in Figure 3 and Figure 4. Figure 3 presents the predicted trends categorized by product type and size, providing a comparison across different products. The model predicts overall moderate price stability, with some variation between product types and sizes. Figure 4 shows the prediction distribution categorized by product size.

At the product category level, the left panel of Figure 3 shows that the price of Snacks remains relatively stable, while the other four categories exhibit comparatively larger fluctuations. These trends align with historical data. However, compared to historical data, the model underestimates the price volatility in certain categories. For instance, the historical data shows significant short-term price fluctuations in Seasoning, which are not fully reflected in the predicted results.

For product size, the right panel of Figure 3 and Figure 4 both indicate clear price stratification among Small, Large, and Extra Large products, with larger-sized products having higher average prices. The model successfully replicates a reasonable stratification structure, but the predicted price range is narrower, and the price volatility is lower compared to the historical observations.



Figure 3: Prediction of average price trend for Loblaw’s No Name Grocery by type and by size

5 Discussion

5.1 Price Forecast

In this paper, we focus on forecasting the price trends of Loblaw’s “No Name” products from November 20, 2024, to February 15, 2025. The analysis emphasizes the dynamics of prices

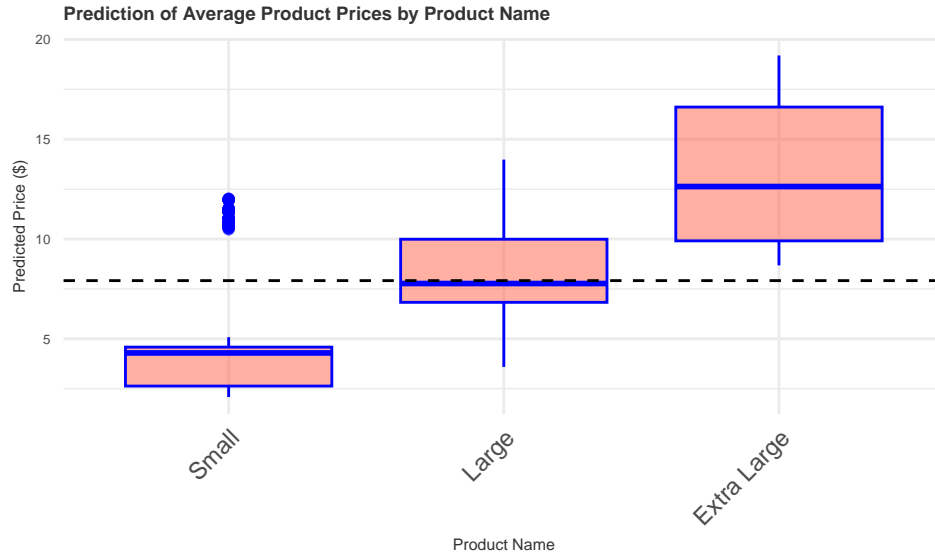


Figure 4: Prediction of average price distribution for Loblaws’ No Name Grocery by size

categorized by product type and size, identifying behavioral patterns and differences between categories. We incorporated historical data like in Figure 1 and Figure 2 to ensure that the predictions align with observed past trends. Additionally, we designed the prediction dataset with constraints to simulate reasonable price changes—prices adjust at least every 14 days, with fluctuations ranging between 10% and 30% of the previous price. The results in **Figure 3**, indicate that while the prices of Loblaws “No Name” products are expected to vary over time, the overall increase or decrease remains relatively limited. Both the predicted and historical data demonstrate similar trends across different categories. This suggests that retail pricing adjustments and predictions for similar product categories may follow patterns that are relatively consistent over time.

From the projected price fluctuations and distributions, the forecast suggests that there will be no significant price increases for Loblaws “No Name” products analyzed in this study from late November 2024 to mid-February 2025. This aligns with Loblaws’ stated commitment to “maintain price stability” for No Name products whenever possible.

5.2 Limitations of Existing Data

This section discusses the limitations of the data, data processing, and the proposed model, aiming to assess its robustness and the challenges in capturing the complexities of price changes. Each step, from data collection to model construction, involves certain assumptions and potential biases. Below is a detailed discussion of these aspects:

Firstly, the raw data used for observation and modeling is extensive and complex, with numerous duplicates and classification errors, as detailed in [Appendix -sec-data-details]. One of the major challenges for the model is providing accurate predictions under limited accuracy and highly complex categorization. The complexity of the data makes it challenging for the model to provide clear interpretations. For example, the model might struggle to distinguish whether price fluctuations are caused by random events (such as promotions) or long-term trends (such as seasonal demand changes). This reduces the model’s practical applicability, especially when decision-makers need straightforward conclusions. The high variability in the data means that the model may struggle to accurately capture key trends and patterns. With a high level of noise, parameter estimates may be significantly affected, leading to wider prediction intervals and reduced reliability. For example, frequent fluctuations or outliers in the price data for certain products could cause the model’s mean predictions to deviate from actual median trends. Misclassified product prices make it difficult for the model to correctly assess the impact of product characteristics on price. This may lower the model’s ability to differentiate between categories, resulting in predictions that deviate from reality.

To reduce bias, we opted to clean duplicates and classification errors before modeling, attempting to standardize the data format. However, this greatly reduced the usable data volume. Unfortunately, since the true state of the data is unknown, the cleaned data may still contain inaccuracies. In the future, when faced with highly variable and complex categories, the model may need to incorporate additional random effects or interaction terms to account for data heterogeneity. However, such complexity may lead to overfitting, where the model fits the training data too closely but performs poorly on new data. This might also increase computational demands to impractical levels.

Loblaw Companies Limited (2022b) also emphasized the importance of regional and independent retailers in maintaining market competitiveness, particularly in non-urbanized areas. However, due to the lack of geographic data, our model assumes uniform price behavior across regions, which may oversimplify the actual dynamics of the Canadian grocery market. If geographic information were incorporated, introducing regional effects or hierarchical modeling could enhance prediction accuracy, especially in capturing price trend differences between urban centers and less competitive markets.

5.3 Real-World Impact

Another limitation of our analysis lies in its reliance on historical pricing data as the primary input. While this approach ensures consistency, it risks overlooking real-time market signals, such as sudden shifts in consumer preferences due to trends, unexpected supply chain disruptions for certain goods, or the introduction of new products with prices significantly different from similar items in the same category. This can result in predictions lagging behind actual market conditions, leading to inaccuracies in forecasting price volatility and trends and diminishing the model’s real-time applicability. Additionally, the model does not account for

external factors like inflation, competitor pricing, or macroeconomic indicators. These variables also influence retail prices, but due to a lack of data, they were excluded. Furthermore, the relatively narrow predicted price range suggests that the model may rely too heavily on prior assumptions, reducing its sensitivity to changes in price dynamics.

In future research, if sufficient data becomes available, we plan to enhance the model by incorporating dynamic factors or real-time updating mechanisms to respond quickly to market changes. This could include forecasting pricing for new products and simulating their potential impact on the market. However, like mentioned in Loblaw Companies Limited (2022b) price categorization in retail is often determined by consumer demand, packaging specifications, or branding strategies, which may not directly align with physical dimensions or averages of the products. This challenge is difficult to resolve, as we cannot predict changes in consumer preferences, meaning the results may not fully reflect actual market characteristics. The predictions will always lag behind real-world outcomes, with the goal being to move as close to reality as possible.

5.4 Loblaws' Pricing Strategy

In Cadloff (2022), the article highlights fairer supplier agreements, with one of the core goals of the code of conduct being to reduce conflicts and non-transparent practices between suppliers and retailers. This could provide greater market access for small and medium-sized suppliers, increasing the flexibility of private label supply chains, and would enable Loblaws to maintain price stability at lower supplier costs (see more details in Cadloff 2022). Combined with the low price volatility observed in our prediction results, this dynamic could further support Loblaws in offering competitive prices through its “No Name” brand, as these products primarily emphasize affordability. The code of conduct could help Loblaws maintain or improve No Name product quality without significantly increasing prices by reduced costs.

Combining contents from Hawley (2023) and Lan, Sexton, and Xia (2006), we infer that Loblaws positions its private label, “No Name,” at a lower price point to attract price-sensitive consumers while getting greater bargaining power within the supply chain. Through this strategy, Loblaws may stabilize or even reduce supplier costs while maintaining customer price satisfaction. Additionally, promotional activities such as discounts and loyalty rewards significantly influence consumer purchasing behavior and brand preferences. This aligns with Loblaws' approach of engaging customers through the PC Optimum program. By offering discounts, reward points, and maintaining stable prices for its private label products, Loblaws can partially offset consumers' negative perception of price increases on other products. This strategy allows Loblaws to preserve its market position and sustain profitability without losing its overall customer base, even in the face of rising prices for other product categories.

Appendix

A Model details

R package `bayesplot` (Gabry J 2024) is used for posteriors.

Due to space limitations, we have chosen to display only the posterior and prior comparisons for a subset of variables in the posterior analysis. If we were to display all variables included in the model, the chart would become overly complex and difficult to interpret. Although the effects of multiple variables differ, focusing on this subset may give the impression that the model’s sensitivity to variations across different factors is limited. Thus, we have selected representative variables, including the intercept, size categories (“sizeLarge” and “sizeSmall”), and product types (“typeSnacks,” “typeFlour/Rice,” and “typeSeasoning”). The output results for other variables not shown are generally consistent with the trends observed here. Additionally, in regression models (such as `stan_glm`), categorical variables (factors) are typically automatically converted into dummy variables. One category is set as the reference category (baseline), and its effect is included in the intercept, while the coefficients for other categories represent their offsets relative to the baseline. For example, in our model, if **Extra Large** is set as the baseline (default), it will not appear as a separate coefficient. The coefficients for `sizeLarge` and `sizeSmall` would then represent their deviations relative to **Extra Large**.

A.1 Posterior predictive check

Figure 5 displays a comparison between the distribution of observed data y and the distribution of simulated data y_{rep} generated by our model. The dark solid line represents the observed data’s actual distribution, while the multiple light lines correspond to distributions from various simulated datasets. If the distribution of simulated data y_{rep} closely resembles that of the observed data y , it suggests that the model effectively captures the underlying characteristics of the observed data.

In Figure 5, we observe that the simulated data distributions (light lines) align closely with the real data distribution (dark line) across the entire range, particularly at the peaks and troughs of the observed data. This indicates that the model has captured the distributional features of the observed data, including its skewness and multimodal structure. The ability to replicate these features demonstrates the model’s strong predictive capacity and its reliability in estimating posterior distributions. This result builds confidence in the model’s suitability for further inference and prediction.

Figure 6 displays the MCMC sampling trace plots for six key parameters in our model, including the (Intercept), size variables `sizeLarge` and `sizeSmall`, and type variables `typeSnacks`, `typeFlour/Rice` and `typeSeasoning`. Each plot includes four chains from Chain

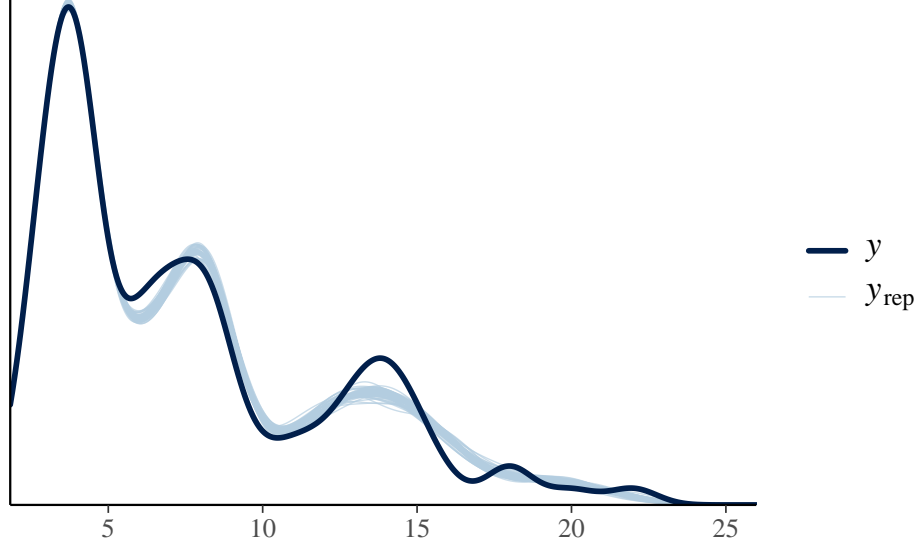


Figure 5: PP Check for Model

1 to Chain 4, representing independent MCMC sampling processes. In a well-converged model, our expectation is that each chain to stabilize and fluctuate around a central value without showing a clear trend (i.e., the mean of the samples converges over time). Additionally, the chains for a parameter should exhibit consistent behavior, indicating agreement among the chains.

From Figure 6, we observe that the sampling chains for all six parameters show a steady state, with fluctuations around a central value and no observable trends over the iterations. Moreover, the chains for each parameter are closely aligned, demonstrating consistency between the chains. This suggests that the model has reached convergence and the posterior distributions for these parameters are stable. These trace plots support the reliability of our MCMC sampling process and provide confidence in the validity of the posterior estimates.

Figure 7 shows the \hat{R} diagnostic statistic for our model, which is used to assess the convergence of MCMC sampling. Ideally, is that the \hat{R} value would be close to value 1, which indicates consistent variance across different MCMC chains, signifying that the model has reached a stable state. The horizontal axis represents the \hat{R} values, where values closer to 1 indicate better convergence. Our expectation is that, all parameters would have \hat{R} values below 1.05, which means minimal variance differences between chains and consistency across them. This plot shows that all \hat{R} values are concentrated around 1, and all \hat{R} values are lower than the threshold of 1.05 (indicated by the vertical dashed line on the right), suggesting that the sampling chains have converged well, providing thw reliability for our model results.

Figure 8 shows the posterior distributions of certain parameters in our model, including (Intercept), size variables `sizeLarge` and `sizeSmall`, and type variables `typeSnacks`, `typeFlour/Rice`, and `typeSeasoning`. The shape and location of these distributions help us

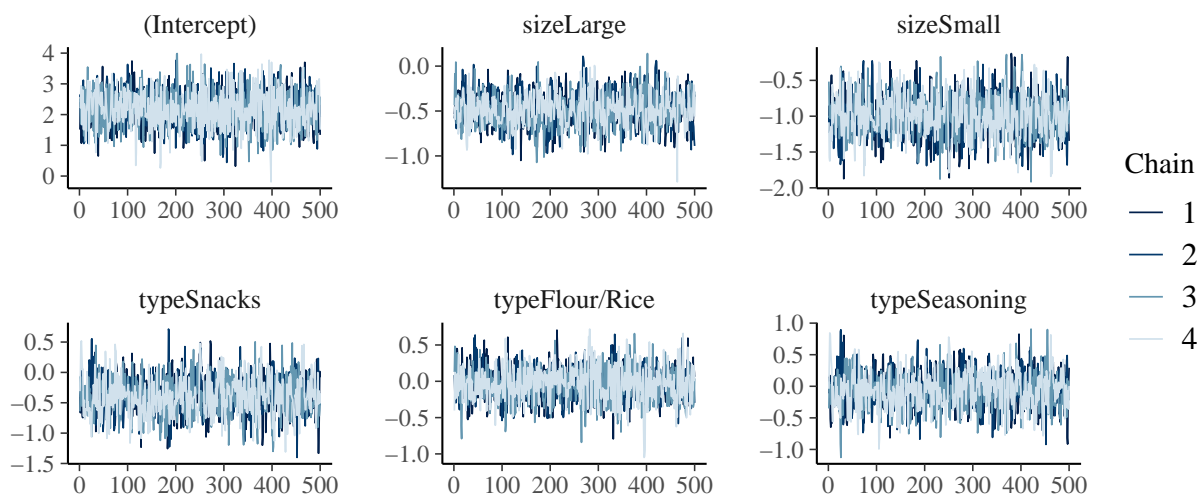


Figure 6: MCMC sampling trace plots for model

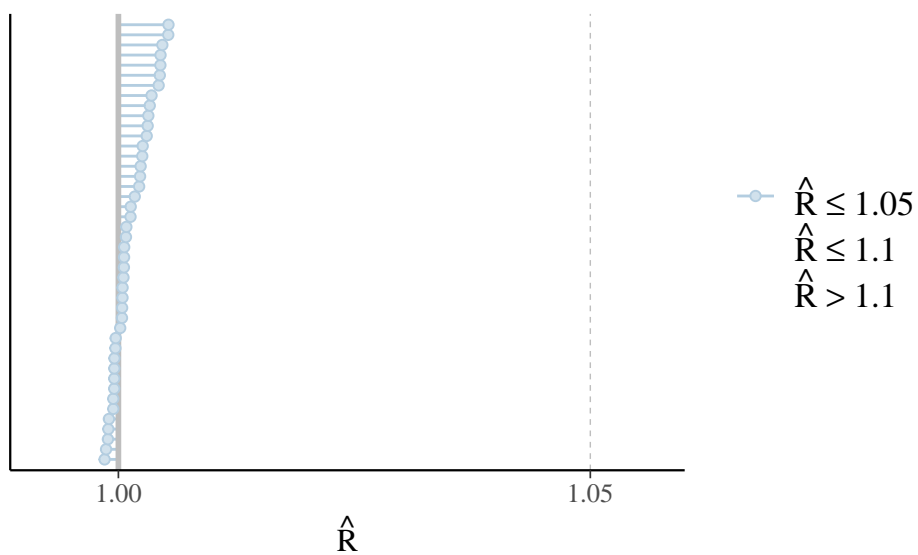


Figure 7: R hat Diagnostic statistic for model

understand the extent to which the observed data supports these parameter estimates and the degree of uncertainty associated with them.

The **(Intercept)** is centered nearly around 2.5, reflecting the baseline level for product prices in our model when all other factors are held constant. The narrow and peaked distribution suggests strong support for this estimate with relatively low uncertainty. For **sizeLarge** and **sizeSmall**, the distributions are shifted slightly to the left, indicating that these sizes are generally associated with lower prices compared to the baseline category **ExtraLarge**. Type variables **typeSnacks**, **typeFlour/Rice**, and **typeSeasoning** exhibit posterior distributions centered close to zero, implying that their influence on product pricing is minimal or uncertain, as the data does not strongly support a directional impact.

Overall, the results suggest that the model successfully captures the baseline pricing level and the relative influence of product sizes, but it struggles to discern clear effects for product types, possibly due to overlapping pricing patterns or insufficient variation within these categories.

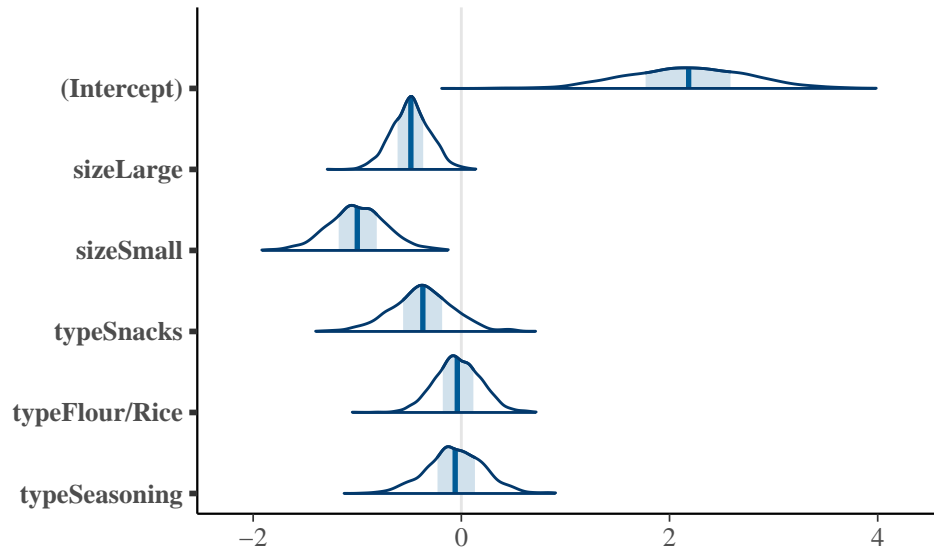


Figure 8: posterior distributions of four parameters in model

References

- Alexander, Rohan. 2024. “Starter Folder.” https://github.com/RohanAlexander/starter_folder.git.
- Bürkner, Paul-Christian. 2017. “brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software* 80 (1): 1–28. <https://doi.org/10.18637/jss.v080.i01>.
- Cadloff, Emily Baron. 2022. “Rising Food Prices: Could a Grocer Code of Conduct Help?” 2022. <https://thewalrus.ca/grocer-code-of-conduct/>.
- Chen, Xin, Aleksandr Y. Aravkin, and R. Douglas Martin. 2018. “Generalized Linear Model for Gamma Distributed Variables via Elastic Net Regularization.” <https://arxiv.org/pdf/1804.07780>.
- Filipp, Jacob. 2024. “Hammer Canadian Grocery Price Data.” <https://jacobfilipp.com/hammer/>.
- Gabry J, Mahr T. 2024. *Bayesplot: Plotting for Bayesian Models*. <https://mc-stan.org/bayesplot/>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Hawley, Ransom. 2023. “Brand Loyalty in the Age of Inflation.” *Canadian Grocer*. <https://canadiangrocer.com/brand-loyalty-age-inflation>.
- Lan, Li, Richard J. Sexton, and Tian Xia. 2006. “Food Retailers’ Pricing and Marketing Strategies, with Implications for Producers.” *Agricultural and Resource Economics Review* 35 (2): 221–38. <https://doi.org/10.1017/S1068280500006687>.
- Loblaw Companies Limited. 2022a. “Loblaw Hits the Brakes on Food Inflation by Freezing Prices on 1,500 No Name Products.” <https://www.loblaw.ca/en/loblaw-hits-the-brakes-on-food-inflation-by-freezing-prices-on-1500-no-name-products/>.
- . 2022b. “LOBLAW’s SUBMISSION TO THE COMPETITION BUREAU OF CANADA MARKET STUDY NOTICE ON OCTOBER 21, 2022: COMPETITION IN CANADA’s GROCERY SECTOR.” <https://competition-bureau.canada.ca/how-we-foster-competition/consultations/response-loblaws-consultation-market-study-retail-grocery>.
- Pedersen, Thomas Lin. 2024. *patchwork: The Composer of Plots*. <https://patchwork.data-imaginist.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/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *arrow: Integration to ‘Apache’ ‘Arrow’*. <https://github.com/apache/arrow/>.
- Veazie, Peter, Orna Intrator, Bruce Kinosian, and Ciaran S. Phibbs. 2023. “Better Performance for Right-Skewed Data Using an Alternative Gamma Model.” *BMC Medical Research Methodology* 23 (298). <https://doi.org/10.1186/s12874-023-02113-1>.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan,

- Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Xie, Yihui. 2023. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://cran.r-project.org/web/packages/knitr/index.html>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. <http://haozhu233.github.io/kableExtra/>.