

Assessing the Various Influences on Current Coffee Prices*

Coffee Pricing in Canadian Grocery Stores: Current Prices Strongly Shaped by Historical Trends

Yi Tang

December 3, 2024

This paper investigates the factors influencing coffee product prices in Canadian grocery stores, focusing on historical prices, vendor differences, and seasonal trends. Using a Bayesian multiple linear regression model, the study acknowledges that historical prices have the most significant positive impact on current prices, while vendor differences and seasonal variations also play measurable roles. These findings highlight how past prices serve as a benchmark for current pricing strategies, emphasizing the importance of historical data in retail decision-making. The results contribute to understanding the dynamics of coffee pricing and provide understanding for optimizing retail pricing strategies and future research directions.

Table of contents

1	Introduction	3
2	Data	4
2.1	Data Source	4
2.2	Inspiration on data organization	4
2.3	Measurement	6
2.4	Data Visualization	6
2.5	Outcome variables	9
2.6	Predictor variables	9
3	Model	10
3.1	Model set-up	10

*Code and data are available at: https://github.com/YiTang2/Canadian_Grocery_Analysis.git

3.2	Model justification	11
3.3	Model validation	11
4	Results	12
5	Discussion	14
5.1	Overall findings and Implication	14
5.2	Additional understanding from the Data Analysis	15
5.3	Shortcomes and Future outlook	16
5.3.1	Limitation	16
5.3.2	Weaknesses and next steps	16
	Appendix	18
A	Idealized Survey and Methodology	18
B	Model details	19
	References	24

1 Introduction

Consuming goods is an integral part of everyone’s daily life, with pricing being a major factor that influences consumer behavior.(Huang et al. 2014) People are particularly interested in the price of goods(Arnot, Boxall, and Cash 2006), and whether there is a price premium or discount. What determines these prices? Is it the purchasing power of consumers, or perhaps the cost associated with each stage of production? This paper seeks to answer some of these questions by combining product data from Canadian grocery stores with additional raw data to identify key factors that influence product pricing. Specifically, we focus on coffee products and use the current price as the outcome variable, trying to understand the influences behind it.

To explore this, I chose one of the most intuitive factors: the old price, since historical pricing often influences how vendors set current prices. The relationship between previous and current pricing can provide observation into whether vendors maintain consistency or adjust prices in response to market conditions. Additionally, I included data from two vendors—Metro and SaveOnFoods—located in different parts of Canada to investigate any differences in pricing strategies between the eastern and western regions. This distinction between vendors allows us to better understand regional variations and how local competition might impact pricing. The inclusion of the month variable allows us to explore potential seasonal trends in coffee pricing, assessing whether specific times of the year see significant price variations, possibly due to fluctuations in demand or promotional campaigns. By evaluating these factors, we can determine the extent to which historical pricing, regional differences, and seasonality drive current pricing decisions in the coffee market.

In this study of coffee pricing strategies in Canadian grocery stores, the focus is on estimating how historical prices (old price), time of year (month), and vendor characteristics (vendor) affect current coffee prices. Using a Bayesian multiple linear regression model, I estimated these effects to better understand the pricing dynamics of coffee products and to determine how much each factor contributes to pricing decisions in the market. This modeling approach provides a probabilistic framework that accounts for uncertainty in the estimates, allowing for more nuanced conclusions compared to traditional methods.

The findings reveal that the old price has the most significant positive impact on current coffee prices among all the predictors considered. This suggests that vendors heavily rely on historical prices when determining current prices, possibly to maintain a level of consistency that consumers expect. Consistency in pricing is often valued by consumers, as it fosters a sense of trust and predictability. Due to a lack of data for some months, it was challenging to establish a definitive analysis of seasonal trends, though this aspect remains an important factor for understanding pricing behavior comprehensively. Seasonal patterns can have substantial effects on consumer behavior, and thus their absence in the dataset may limit the scope of understanding regarding periodic fluctuations in coffee prices.

The paper is structured as follows: Section 2 describes the data used for analysis, providing details on data collection methods and any preprocessing required. Section 3 discusses how to set up, justify, and validate the Bayesian multiple linear regression model, explaining the rationale for selecting this approach and detailing the assumptions made. **?@sec-result** presents the findings derived from the data and model, highlighting the main effects and relationships observed. Finally, Section 5 delves into the implications of the results, addresses potential limitations, and outlines expectations for future research, considering how expanding the dataset or incorporating additional variables might enhance our understanding of coffee pricing dynamics.

2 Data

2.1 Data Source

Price of each month's coffee of different vendors data is provided by (Filipp 2024). This dataset records detailed sales about fast-moving consumer goods (FMCG) sold by various vendors, including volia, T&T, Loblaws, SaveOnFoods, Galleria, Metro, NoFrills and Walmart. It also includes product-level details, such as the product name, current price, historical price (old price), and the corresponding units and price per unit. The data also captures time-specific observations (2024-2-28 to 2024-6-22), with timestamps (nowtime) that can be used to analyze trends over days or months.

2.2 Inspiration on data organization

In order to simulate data, test simulated data, clean data, test cleaned data, exploratory data analysis and model data, we used R programming language (R Core Team 2023) to analyze the data and plot the graphs. The folder structure for this paper follows the startup folder created by (Wickham et al. 2019a). It is very helpful in keeping everything organized. I also referenced many of the examples and techniques provided by Telling Stories with Data (Alexander 2023), which helped me understand how to visualize the data and communicate the findings effectively. Specific libraries that assisted the analysis include `tidyverse` (Wickham et al. 2019b), `caret` (Kuhn and Max 2008), `janitor` (Firke 2023), `lubridate` (Grolemund and Wickham 2011), `knitr` (Xie 2014), `arrow` (Richardson et al. 2024), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2023), `here` (Müller 2020), `kableExtra` (Zhu 2024), `gridExtra` (Auguie 2017), `styler` (Walthert et al. 2024), `plumber` (Schloerke and Allen 2024) `modelsummary` (Arel-Bundock 2022), `rstanarm` (Goodrich et al. 2022).

The inspiration for my data processing came from my desire to study what factors would affect the current price of coffee products from two vendors in different regions of Canada, such as the current price of coffee products from two vendors, Metro and SaveOnFoods. Table 1 are

the data I selected after cleaning the data, and new variable month extracted and transformed from raw data also.

Since the variable nowtime only records 4 months, it is considered a lack of Long-Term Trends, which means it's difficult to identify long-term pricing or demand patterns by using short data periods. So I only extracted a new variable—month from date of nowtime, which can simplify temporal analysis and identify trends, such as seasonal monthly pattern with price changes or demand patterns. It allows grouping data for monthly aggregation and supporting seasonality-focused understanding or forecasting models.

Table 1: Variables for the Coffee Product Pricing.

Variable	Definition
vendor	The retailer selling the product in Canada
old price	The historical price of the product, showing previous pricing or discounts
current price	The price of the product at the time of observation
product name	The specific product being sold, providing product-level understanding
month	The month of data collection, extracted from 'nowtime'

To provide an preview of the coffee pricing with all potential factors that might affect it. Here, Table 2 simply explains the variation between current price and old price in June for Metro's coffee products.

Table 2: Sample of Analysis Data Showing Products Sold by Both Vendors

vendor	product_name	current_price	old_price	month
Metro	Non-Dairy Vanilla Flavoured Latte Coffee Cream	7.49	8.99	6
Metro	Vanilla And Caramel Flavoured K-Cup® Coffee Capsules	9.99	12.99	6
Metro	Classic Black K-Cup® Coffee Capsules	9.99	12.99	6
Metro	Cold Brew Unsweetened Iced Coffee	7.49	7.99	6
Metro	Limited Edition Coffee Whitener, Coffee Mate	4.99	6.99	6
Metro	Italian Blend Dark Roast K-Cup Coffee Pods	6.49	6.99	6
Metro	Classic Roast Ground Coffee	8.99	12.49	6
Metro	Medium Roast Decafreinated K-Cup Coffee Pods, Pike P...	22.99	26.99	6
Metro	Medium Roast House Blend K-Cup Coffee Pods, Organic	6.49	6.99	6
Metro	Classic Decaf Ground Coffee	8.99	12.99	6

2.3 Measurement

The dataset from Hammer provides a wide representation of real-world retail activities, capturing key elements such as product details, vendor listings, and price updates. When vendors make updates to product information, including pricing changes or availability adjustments, Hammer gathers and organizes this data into a structured dataset. This ensures that the dataset remains accurate and reflective of current retail conditions.

The dataset includes essential fields such as vendor, product name, current price, old price, and a timestamp referred to as nowtime. The data is collected through scraping and is systematically structured to support an in-depth analysis of retail trends, pricing strategies, and overall market dynamics across different periods. Each record captures a snapshot of a product's market presence at a given time, enabling targeted analyses like monitoring price fluctuations for specific products, such as coffee. By offering such detailed and time-specific information, the dataset provides a important foundation for understanding the evolving dynamics of retail pricing and vendor strategies.

2.4 Data Visualization

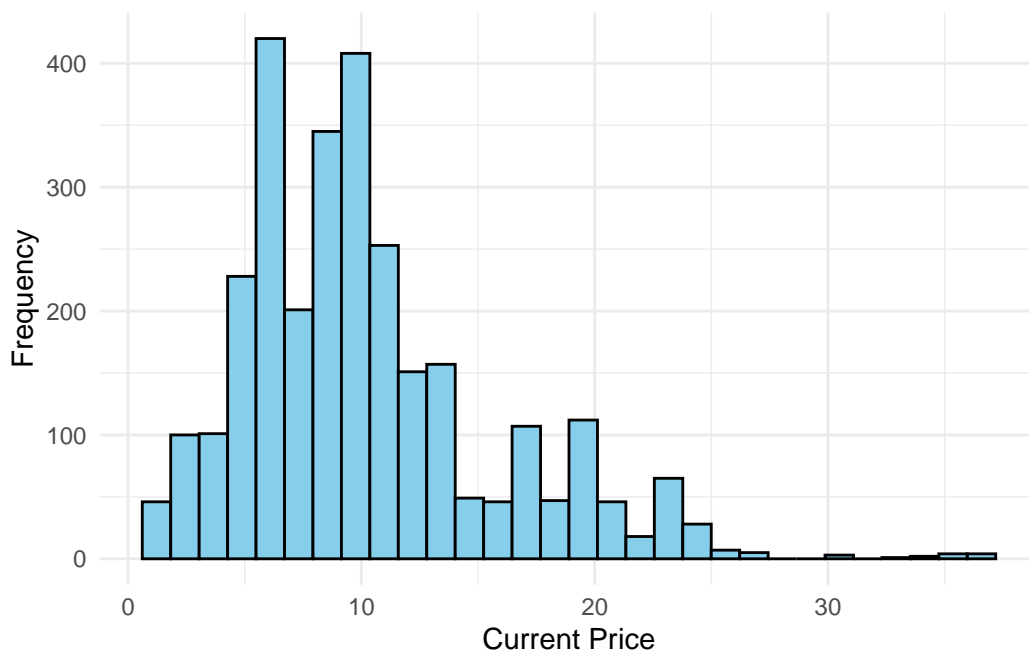


Figure 1: Right-skewed distribution of Current Price

Figure 1 shows a notable distribution with significant concentrations around 10 and 20 units. This suggests a pricing tier system where certain types of coffee products might be grouped

by price points due to quality, brand, or other market factors. The dual peaks could indicate two main categories of coffee products, possibly differentiated by premium versus regular products.

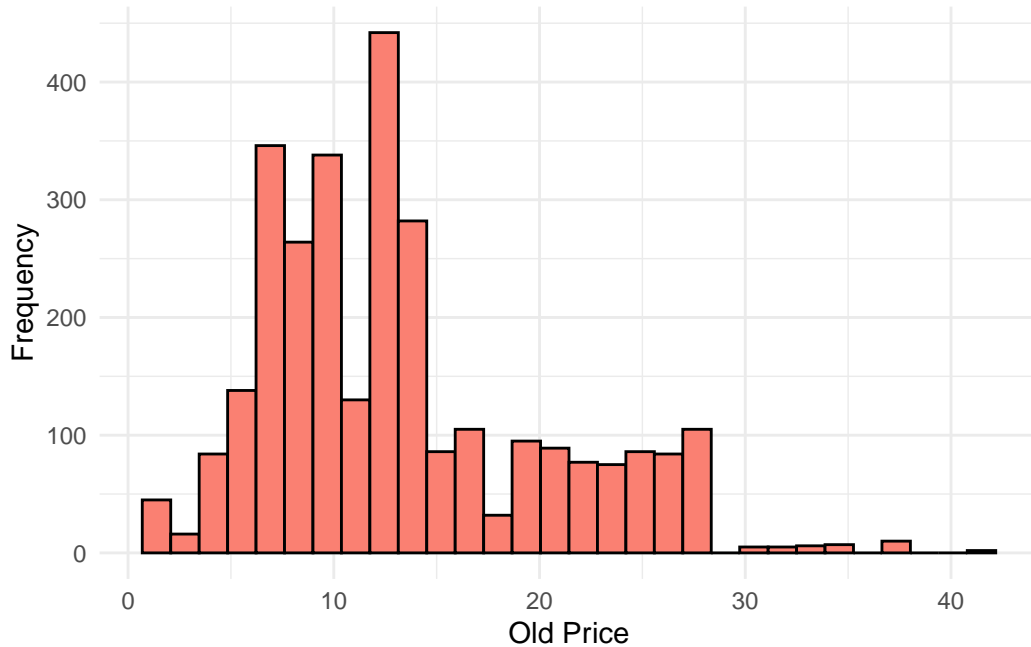


Figure 2: Right-skewed of Old Price

According to Figure 2, the distribution is mainly focused between 5 and 20 units, pointing to a past market strategy where products were clustered around these prices. The consistency in this price range might reflect a stable market before any recent pricing adjustments influenced by external factors like supplier changes or inflation.

In Figure 3, the frequency of products available from Metro significantly surpasses that from SaveOnFoods, suggesting that Metro has a larger share of the market or a wider variety of coffee products. This dominance in product offerings could provide Metro with a competitive edge in attracting a broader customer base.

The positive relationship shown in Figure 4 here indicates that current prices are influenced by their historical prices, maintaining a proportional increase or decrease. This trend suggests a pricing policy that adjusts prices based on previous benchmarks while taking into account factors like cost adjustments or market demand.

Figure 5 highlights how each vendor prices their products within the market. It shows that both vendors offer a wide range of prices, yet the spread and density of the data points may indicate Metro's pricing strategy targets both lower and upper market segments, whereas SaveOnFoods might be focusing on a specific niche.

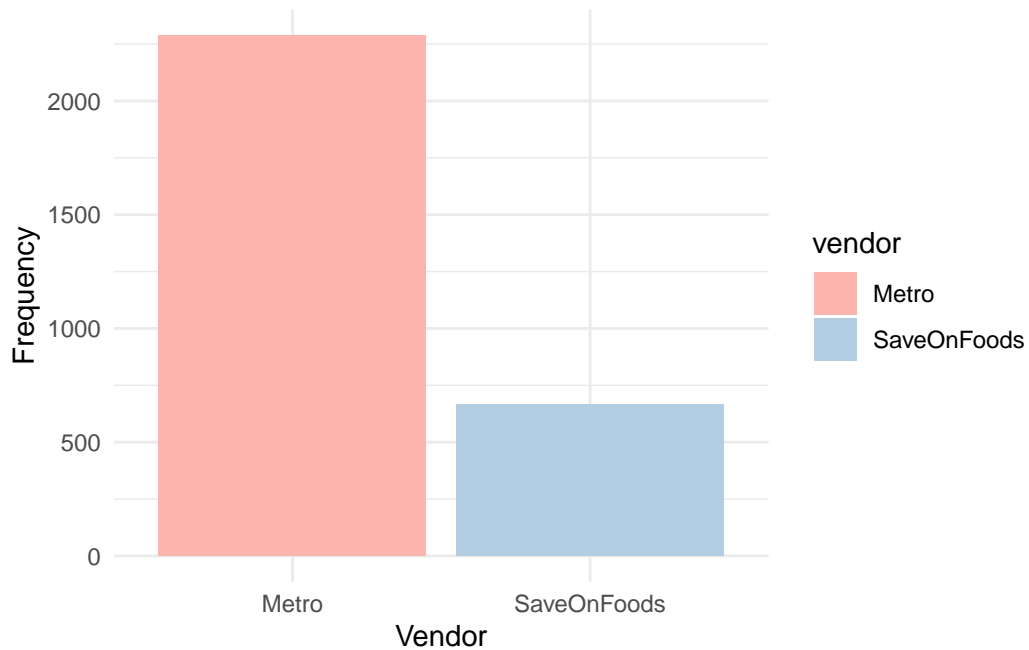


Figure 3: Higher frequency on Metro compared to SaveOnFoods

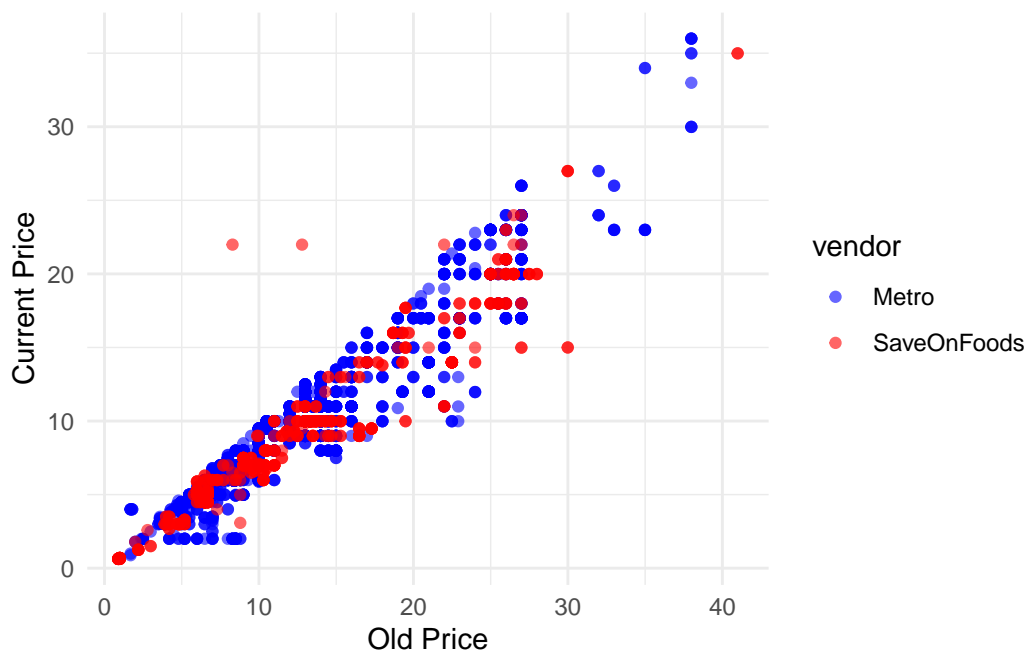


Figure 4: Strong positive correlation between old and current prices for both vendors

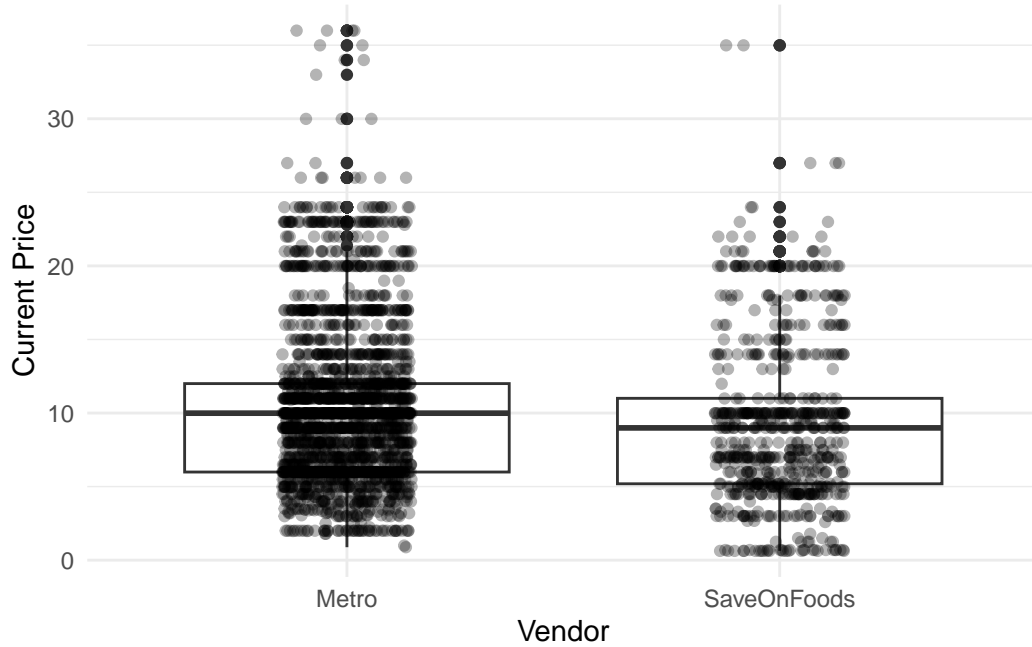


Figure 5: Similar median prices for both vendor and a wider range plus more outliers in Metro

2.5 Outcome variables

The outcome variable for this analysis is the current price of coffee products. This variable is significant for understanding the dynamics of pricing in the retail market, as it reflects the final cost faced by consumers. Analyzing the current price helps uncover patterns in how vendors adjust prices over time, respond to market conditions, and compete in a fluctuating retail environment. By focusing on the current price, we can measure the impact of historical trends, vendor-specific strategies, and seasonal changes on pricing decisions.

2.6 Predictor variables

The predictor variables selected from the raw dataset are old price, month, and vendor. The old price serves as a predictor because it provides beneficial understanding into how past pricing decisions influence current prices. Vendors often make incremental adjustments based on historical pricing data, consumer behavior, and market reactions, which makes it a key factor in understanding pricing strategies. Month is included to account for possible seasonal trends in pricing, as consumer demand and promotional activities often vary throughout the year, potentially impacting coffee prices. Lastly, vendor is used to examine pricing variations between Metro and SaveOnFoods. Each vendor may have distinct pricing approaches based on

factors like target market, cost structures, or regional positioning, and including this variable enables a comparative analysis of vendor-specific pricing dynamics.

3 Model

The goal of our Bayesian multiple linear regression is to investigate the factors that influence the current price of coffee in our dataset. Specifically, we try to understand how historical pricing, vendor differences, and seasonal monthly pattern affect current coffee prices.

3.1 Model set-up

Define y_i as the current price of coffee for the i -th observation in the dataset. The predictors include:

- $X_{\text{Old Price},i}$, the old price of the coffee,
- $X_{\text{Vendor},i}$, dummy variable for the vendor, where:
 $X_{\text{Vendor},i} = 1$: Vendor is “SaveOnFoods”; $X_{\text{Vendor},i} = 0$: Vendor is “Metro”,
- $X_{\text{Month},i}$, the numeric month variable.

The model is formulated as follows:

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma), \quad (1)$$

$$\mu_i = \alpha + \beta_1 \cdot X_{\text{Old Price},i} + \beta_2 \cdot X_{\text{Vendor},i} + \beta_3 \cdot X_{\text{Month},i}, \quad (2)$$

$$\alpha \sim \text{Normal}(0, 2.5), \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 2.5), \quad (4)$$

$$\beta_2 \sim \text{Normal}(0, 2.5), \quad (5)$$

$$\beta_3 \sim \text{Normal}(0, 2.5), \quad (6)$$

$$\sigma \sim \text{Exponential}(1). \quad (7)$$

This model describes the relationship between the current price of coffee y_i and three predictors: the old price of coffee $X_{\text{Old Price},i}$, a categorical vendor variable $X_{\text{Vendor},i}$ indicating whether the vendor is “Metro” or “SaveOnFoods,” and a numeric variable for the month $X_{\text{Month},i}$. The response variable y_i is modeled as normally distributed with mean μ_i and standard deviation σ . The mean μ_i is defined as a linear combination of these predictors, with coefficients β_1, β_2 , and β_3 , and an intercept α . Prior distributions for the parameters are specified, including normal priors for α and the coefficients, and an exponential prior for σ . Intercept α represents the baseline mean current price for Metro if $X_{\text{Vendor},i} = 1$; otherwise when $X_{\text{Vendor},i} = 0$, it represents the mean current price for SaveOnFoods. Also, when old

price and month is equal to 0, the intercept is not meaningful. Coefficient β_1 captures how changes in the old price affect the current price. Coefficient β_2 measures the difference in the mean coffee price between SaveOnFoods ($X_{\text{Vendor},i} = 1$) and Metro ($X_{\text{Vendor},i} = 0$). Coefficient β_3 reflects how the month influences current pricing, potentially capturing seasonal effects.

To implement this Bayesian model, we use the `rstanarm` package (Goodrich et al. 2022) in R (R Core Team 2023).

3.2 Model justification

We considered an alternative model, the frequentist model, but ultimately did not select it. We conducted a training and testing dataset split and compared their prediction power, particularly looking at the Mean Squared Error (MSE). The results showed that the Bayesian Multiple Linear Regression (MLR) had a lower MSE, making it a better fit for our data. Additionally, the Bayesian approach provides greater flexibility.

The Bayesian Multiple Linear Regression (MLR) model is a suitable choice for analyzing the relationship between the current price (the dependent variable) and the predictors in the dataset. The dependent variable is continuous, and the Bayesian framework assumes a normal distribution for the response, which aligns well with the nature of coffee prices. This model captures the linear relationships between the old price (continuous), vendor (categorical, represented as a dummy variable), and month (numeric). These predictors are assumed to have additive effects on the response, fitting the linear regression framework.

Logistic regression is used when the outcome variable is binary (e.g., 0 or 1). However, in our dataset, the dependent variable, current price, is continuous. Since logistic regression cannot model continuous outcomes, it is unsuitable for this analysis. Similarly, Poisson or negative binomial regression is typically applied when the response variable represents count data (e.g., the number of events occurring in a fixed period). Since the current price represents continuous pricing data rather than counts, these models do not align with the nature of the dependent variable.

The final results indicate that the Bayesian MLR model outperforms the frequentist model in terms of both flexibility and predictive performance, which led us to choose the Bayesian MLR model for our analysis.

3.3 Model validation

Figure 6 shows that our model accurately captures the central tendency of the data, though there are some deviations in the tail ends suggesting that the fit could be improved for extreme values. Additionally, the parameter estimation comparison chart highlights that most parameter estimates closely align with their priors, indicating a strong influence of prior settings on the

estimates, especially under limited data. This is particularly evident with the vendorSaveOnFoods parameter, where its posterior distribution significantly diverges from others, hinting at potential anomalies in data sources or unique behaviors that warrant further investigation.

Figure 8 illustrate that our Bayesian model parameters are converging and demonstrating stability, essential for robust statistical inference. The slight oscillation of the intercept around -0.5 indicates minor variability. Meanwhile, the old price coefficient shows remarkable consistency at approximately 0.77, underlining dependable estimates. The month parameter’s minor fluctuations suggest a subtle yet consistent temporal effect. The vendorSaveOnFoods coefficient consistently remains near -0.6, indicating a persistently negative price influence compared to Metro. Finally, sigma’s stability around 1.75 ensures the model’s error variability is well-accounted for. These observations collectively affirm that the model’s parameters are effectively calibrated, offering a reliable foundation for understanding the influences on coffee prices.

In Figure 9 analysis of coffee product pricing in Canada, the use of the \hat{R} values to assess model convergence explains that all parameters have \hat{R} values below 1.05, indicating excellent convergence of the model. This result validates the reliability of our model in estimating coffee prices and ensures the robustness of the analysis outcomes. It allows us to trust the model outputs, providing a solid foundation for further strategic decision-making and market analysis.

4 Results

The analysis of coffee pricing in Canadian grocery stores several key observations based on the Bayesian multiple linear regression model. The model effectively captures the dynamics of pricing strategies employed by vendors and highlights the influence of historical price, vendor differences, and seasonal trends on current prices. Table 3 indicates a strong fit with a high R^2 of 0.900, indicating that approximately 90% of the variance in coffee prices is explained by the included predictors. This strong fit suggests the chosen factors—historical prices, vendor characteristics, and month—are significant in determining current coffee pricing. The coefficient for the historical price variable was found to be 0.77, indicating a strong positive effect on the current price. This means that higher past prices tend to lead to higher current prices, suggesting that vendors are consistent in maintaining pricing trends over time. The strong relationship between historical and current prices suggests that coffee vendors may to ensure price stability, which could be beneficial for maintaining customer loyalty and trust, as consumers often prefer predictable and steady pricing.

The model shows a significant difference between the two vendors, Metro and SaveOnFoods. Specifically, SaveOnFoods has a negative association with current price levels, with a coefficient of -0.61. This indicates that SaveOnFoods tends to offer coffee products at lower prices compared to Metro, which may point to different pricing strategies or market positioning.

SaveOnFoods may focus on affordability and competitive pricing to attract cost-sensitive consumers, whereas Metro might position its products as more premium, thereby maintaining higher price points. The coefficient for the month variable, 0.06, suggests that there is a slight but consistent increase in prices over time, reflecting potential seasonal variations in demand or pricing adjustments throughout the year. This indicates that factors like holidays, consumer buying habits, and promotional events could have an impact on pricing patterns, although, due to the short data collection period, definitive conclusions regarding seasonal trends are limited. Expanding the data collection period to include more months or years would provide a clearer understanding of these seasonal trends and their impact on pricing.

Figure 7 illustrates the posterior distributions for the parameters were consistent with the priors, providing stability in the estimates. This consistency indicates that the model’s assumptions and data align well, reinforcing the reliability of the estimated effects. Notably, the vendor parameter for SaveOnFoods diverged more significantly, which might indicate unique pricing behaviors that warrant further investigation. SaveOnFoods may be employing specific pricing tactics that differ considerably from Metro, such as frequent promotional discounts or region-specific pricing strategies. Understanding these unique behaviors could help to explain why SaveOnFoods consistently offers lower prices. The model’s reliability is further validated by a low RMSE of 1.75 and satisfactory WAIC and LOOIC scores, which collectively support the model’s accuracy in predicting current prices. The greatness of these metrics suggests that the model can effectively be used to understand pricing strategies in the broader retail context.

The distribution of current prices shows significant concentration around specific price points, potentially due to product quality or brand categorization. For instance, certain brands or types of coffee might be priced consistently across different vendors, creating clusters of similar price points. Similarly, historical prices display consistency, suggesting stable pricing strategies among vendors, which could be influenced by factors like cost of production, supply chain considerations, and consumer preferences. The frequency analysis highlights Metro’s larger product offering compared to SaveOnFoods, which may provide Metro with a competitive edge in variety and customer attraction. A greater variety of products allows Metro to cater to a wider range of consumer preferences, potentially attracting both premium and budget-conscious customers. This variety might enable Metro to occupy different segments of the market, thereby enhancing its overall market presence.

The results suggest that vendors rely heavily on historical prices to set current prices, maintaining a level of consistency that is likely focused on fostering consumer trust. Consumers often associate consistent pricing with reliability, which can help build brand loyalty and repeat purchasing behavior. The pricing differences between Metro and SaveOnFoods may also reflect distinct regional market strategies or consumer demographics. For example, SaveOnFoods might operate in regions with a higher sensitivity to pricing, leading it to adopt more competitive price levels. Metro, on the other hand, may cater to areas where consumers are willing to pay a premium for perceived quality or brand reputation. Seasonal factors, although not extensively captured due to data limitations, do show some influence on pricing

adjustments over time. Seasonal trends such as increased demand during the holiday season or promotional events might cause temporary price hikes, which could be better understood with a more extended dataset. Incorporating data from a full year or multiple years could provide more substantial evidence of these seasonal effects, enhancing the understanding of how temporal factors shape pricing strategies.

Overall, this analysis underscores the importance of historical pricing, vendor differences, and seasonal influences in determining coffee prices in Canadian grocery stores. The reliance on historical data helps vendors ensure price stability, which can be a key factor in maintaining customer satisfaction. Vendor-specific strategies, such as those seen with SaveOnFoods' lower pricing, indicate that different retailers may adopt distinct approaches to appeal to their target markets. Understanding these dynamics can help vendors refine their pricing strategies, optimize inventory management, and ultimately enhance their market competitiveness. Future research could benefit from a longer data collection period and the inclusion of additional variables such as promotional activities, consumer demographics, and regional economic conditions to provide a more view of coffee pricing dynamics.

5 Discussion

5.1 Overall findings and Implication

This research performed an extensive examination of coffee product pricing across major Canadian grocery store chains, focusing on Metro and SaveOnFoods. By employing a regression model, the study analyzed how historical price data (by month) serves as a predictor of current coffee prices. This method helped in identifying pricing patterns and provided a structured approach to understand the economic factors influencing those patterns. The analysis involved scrutinizing various pricing aspects, ensuring that each factor contributing to the fluctuations was carefully assessed to draw meaningful conclusions.

The findings highlight a dynamic relationship between past and present pricing strategies, emphasizing the role of historical prices in shaping current market behaviors. The analysis indicates that historical prices strongly influence current prices, suggesting that vendors often rely on previous data when determining future pricing decisions. This approach appears to be focused at ensuring stability in the market and optimizing profit margins, which is important in a highly competitive retail environment. By maintaining consistency in pricing and leveraging historical trends, vendors might be focusing on strategies to attract and retain customers, avoid sudden price changes, and manage costs effectively. The reliance on historical pricing data allows vendors to make informed decisions that minimize risks associated with unexpected market changes. Additionally, it suggests that these pricing strategies are designed to foster consumer trust by providing predictable price points, which is particularly important in sustaining customer loyalty and enhancing brand perception in a highly volatile market.

Moreover, the study also provides observation into how price stability can help in inventory management and supply chain optimization. By using historical data as a benchmark, vendors can better predict demand patterns, allowing them to streamline their logistics and reduce overhead costs. This, in turn, helps maintain a balance between supply and demand, ultimately leading to enhanced operational efficiency. In a market where even small fluctuations can have a significant impact on consumer purchasing decisions, understanding the relationship between historical and current prices is quite necessary for developing effective pricing strategies that ensure both profitability and competitiveness.

5.2 Additional understanding from the Data Analysis

The coffee product prices are influenced by more than just historical trends; they also react to various external economic factors. This study shows how changes in coffee prices align with broader economic indicators, such as inflation rates and shifts in consumer demand. These patterns are essential for understanding the complexities of the coffee market, implying that vendors adjust their pricing strategies in response to both past price trends and the changing economic landscape. The analysis illustrates how vendors take into account inflation, fluctuations in demand, and other economic conditions when determining prices. These considerations help us understand the broader market environment in which these businesses operate, highlighting the strategic choices vendors make to navigate economic pressures, maintain competitiveness, and adapt to consumer expectations.

Furthermore, the influence of external factors, such as changes in import tariffs, fluctuations in exchange rates, and variations in production costs, also plays a significant role in determining coffee prices. Vendors must continuously monitor these factors to adjust their pricing strategies accordingly. For instance, an increase in production costs due to higher raw material prices may necessitate a corresponding increase in retail prices to maintain profit margins. Similarly, changes in consumer behavior, such as a shift towards premium coffee products or an increased preference for organic options, can also impact pricing strategies. Vendors need to adapt to these trends to remain relevant in the market and meet evolving consumer expectations.

The study also points out how macroeconomic indicators, like interest rates and employment levels, indirectly affect coffee pricing. Higher interest rates may lead to reduced consumer spending power, prompting vendors to adjust their pricing strategies to maintain sales volume. On the other hand, lower unemployment rates could lead to increased disposable income, allowing vendors to explore opportunities for premium pricing. These dynamic interactions between external economic factors and pricing decisions underscore the need for a flexible and responsive pricing strategy that can adapt to changes in the broader economic environment. By understanding these relationships, vendors can develop more resilient pricing models that not only respond to immediate market conditions but also anticipate future economic shifts.

5.3 Shortcomes and Future outlook

5.3.1 Limitation

This study involved integrating two separate raw datasets to build the analytical dataset, which presented challenges due to differing variable names and formats. The integration process required aligning these variables accurately to ensure consistency in the final dataset used for modeling. This emphasizes the importance of proper data preparation to ensure reliable results from statistical analyses. The process of integrating data from different sources is inherently complex and requires careful attention to detail to avoid errors that could compromise the validity of the findings. Ensuring that the data is clean, consistent, and accurately aligned is fundamental to producing meaningful and trustworthy results.

A major limitation of this study is the absence of geographic information for the vendors. While the analysis provides useful observations on the pricing strategies of Metro and SaveOnFoods, it does not include information that could show how pricing varies by geographical region across Canada. Metro and SaveOnFoods operate in different parts of the country, potentially representing Eastern and Western Canada. It is possible that differences in regional economic conditions, levels of competition, and consumer demographics affect pricing strategies. The lack of geographic data prevents a deeper analysis of regional pricing variations and limits the ability to fully understand how local factors influence pricing decisions. Moreover, this study only covers coffee product pricing data from June to November, meaning that we miss the impact of full-year seasonality on pricing.

In addition to the absence of geographic data, another limitation is the lack of information regarding promotional activities and discounts. Promotions play a significant role in influencing consumer purchasing decisions, and without this data, it is challenging to fully understand the factors driving price changes. Promotional campaigns, such as discounts or bundle provides, can have a considerable impact on sales volume and pricing strategies, and their absence from the dataset limits the scope of the analysis. Including this data in future research would provide a more broad view of the factors influencing pricing decisions and help to better understand the interplay between promotional activities and price adjustments.

5.3.2 Weaknesses and next steps

The model provides useful observations; however, there are certain limitations due to the dataset's short time frame, spanning only from June to November. This short time frame restricts our ability to analyze long-term trends or seasonal impacts beyond this period. Additionally, the model assumes linear relationships between the variables, which may not fully represent the complexity of price fluctuations driven by diverse factors, including consumer behavior and market conditions. The linearity assumption may oversimplify the relationships between variables, potentially overlooking non-linear dynamics that could provide a deeper understanding of the factors influencing pricing.

Future research should extend the data collection period to cover multiple years, which would allow for a more thorough understanding of long-term trends and seasonal price variations. Including additional variables, such as promotional activities, competitor pricing, and broader economic indicators, could further enrich the analysis. Exploring more complex models, such as non-linear methods or machine learning techniques, may also reveal deeper connections within the data and offer a more detailed understanding of the pricing strategies used by different vendors. Future studies should also prioritize obtaining geographic data to allow a detailed study of location-specific influences on pricing. This would enable a more subtle analysis of how regional factors, such as differences in consumer preferences and local market competition, impact pricing strategies. By incorporating geographic data, future research could provide observation into more localized or region-specific strategies that vendors use, helping to better understand how they adapt to regional differences and tailor their strategies accordingly.

Expanding the scope of the research to include qualitative data, such as consumer feedback and vendor interviews, could also provide a richer context for understanding pricing strategies. Qualitative observation could help uncover the motivations behind pricing decisions, offering a perspective that quantitative data alone cannot provide. Additionally, incorporating machine learning algorithms could enhance the predictive accuracy of the model by capturing non-linear relationships and complex interactions between variables. Machine learning approaches, such as decision trees or neural networks, could offer a more sophisticated analysis of pricing dynamics and help identify hidden patterns that are not immediately apparent through traditional regression models.

Future research could also explore the impact of external shocks, such as supply chain disruptions or sudden changes in consumer preferences, on pricing strategies. By analyzing how vendors respond to unexpected events, researchers could gain a better understanding of the resilience and adaptability of pricing models in the face of uncertainty. This would provide helpful understanding into the strategies that vendors use to mitigate risks and maintain stability in times of economic turbulence. Furthermore, examining the role of consumer loyalty programs and their effect on pricing strategies could add another layer of depth to the analysis, as these programs often influence purchasing behavior and price sensitivity.

By addressing these limitations and expanding the scope of the research, future studies could provide a more overall understanding of the factors that influence coffee pricing in the Canadian retail market. This would not only enhance our understanding of vendor strategies but also provide beneficial information for policymakers, industry stakeholders, and consumers looking to navigate the complexities of the coffee market.

Appendix

A Idealized Survey and Methodology

To address the collection of data regarding coffee prices through an idealized survey and methodology, we have accurately designed a survey that aligns with best practices in research and data collection (Filipp 2024). The survey begins with an introductory section, welcoming participants to the “Coffee Price perspective Survey.” This introduction serves to inform respondents of the purpose of the survey, which is to gather data on coffee prices across different regions to analyze market trends and pricing strategies. The introduction reassures participants of the value and importance of their input and provides an estimated completion time of 5 minutes.

Contact information for the survey overseer is clearly listed, with Dr. Jane Brewer as the point of contact, providing her email and a contact number to ensure respondents can reach out with questions or require further assistance. This transparency builds trust and aids in increasing response rates.

The survey itself comprises a series of well-constructed questions designed to gather all data while being easy to understand and respond to. It starts with basic multiple-choice questions regarding where respondents buy their coffee and how frequently. This is followed by checkbox options for the type of coffee they purchase, allowing for multiple selections to capture all preferences. The survey progresses to more detailed questions, such as ranking their preferred coffee brands and disclosing their monthly coffee expenditure through an open-ended question. Additionally, a simple yes/no question about noticing seasonal price variations is included to gather data on temporal pricing trends.

Each question is strategically ordered from the most straightforward to the more detailed to prevent respondent fatigue and ensure logical flow, maintaining engagement throughout the survey. The varied types of questions (multiple-choice, checkboxes, ranking, open-ended, and binary) cater to different data needs and enhance the richness of the data collected, making the analysis more robust.

The survey concludes with a thank-you section, expressing gratitude to the respondents for their time and understanding, which reinforces the value of their contribution and ends the survey on a positive note.

For the methodology, a stratified sampling strategy is employed to ensure a diverse and representative sample by dividing the population into subgroups based on geographical regions. This approach not only improves the generalizability of the results but also allows for detailed analysis of regional variations in coffee prices. Data collection is conducted via an online survey platform, chosen for its wide reach and efficiency, as well as its built-in data validation features to ensure accuracy in the responses gathered.

Data analysis involves quantitative techniques to assess pricing trends and consumer behavior, utilizing statistical software for precision. Qualitative responses from open-ended questions undergo thematic analysis to uncover deeper understanding. This analysis approach ensures that both numerical data and rich textual data are effectively utilized.

The overall survey design and methodology are grounded in the literature on survey research, drawing from established frameworks and studies that underscore the effectiveness of such approaches. Additionally, preliminary simulations using historical data are conducted to test and refine the survey and sampling strategy, ensuring that the actual data collection is as effective and efficient as possible.

This idealized survey and methodology framework is not only methodically sound and strategically designed but also extensively planned to achieve detailed understanding into coffee pricing, thereby fulfilling the objectives set out for this research initiative.

B Model details

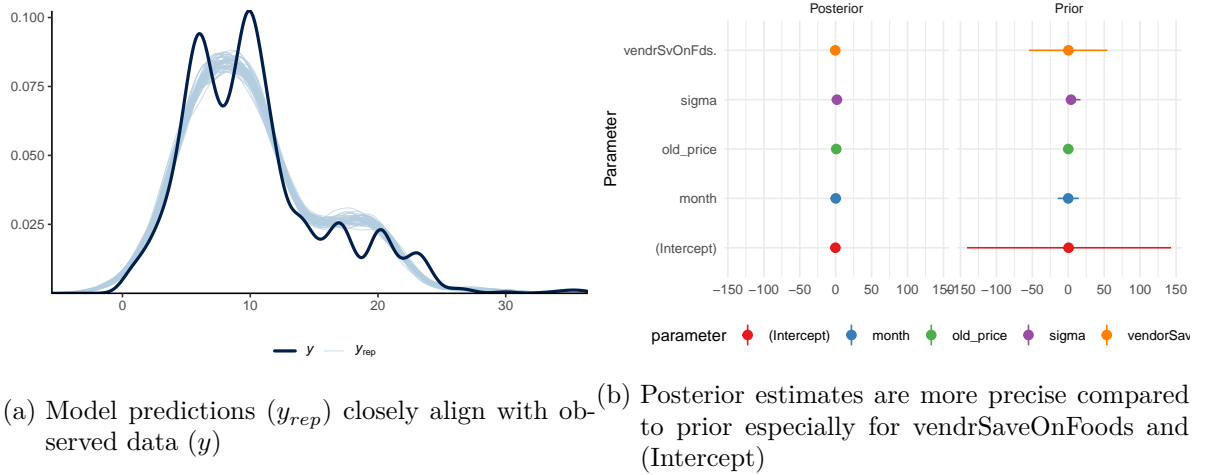


Figure 6: PPcheck & Posterior vs Prior

Table 3: Good prediction accuracy with R^2 of 0.900 and RMSE of 1.75

Coffee product pricing	
(Intercept)	−0.41
old_price	0.77
month	0.06
vendorSaveOnFoods	−0.61
Num.Obs.	2954
R2	0.900
R2 Adj.	0.900
Log.Lik.	−5837.981
ELPD	−5843.3
ELPD s.e.	71.3
LOOIC	11 686.6
LOOIC s.e.	142.6
WAIC	11 686.6
RMSE	1.75

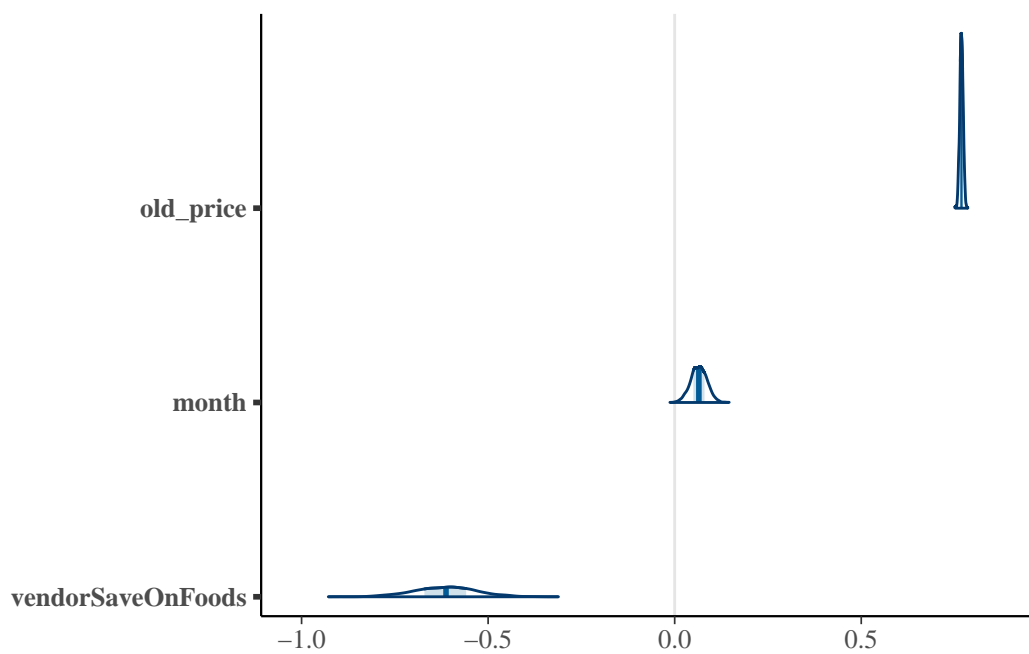


Figure 7: Credible intervals indicate high certainty for old price and lower certainty for other parameters

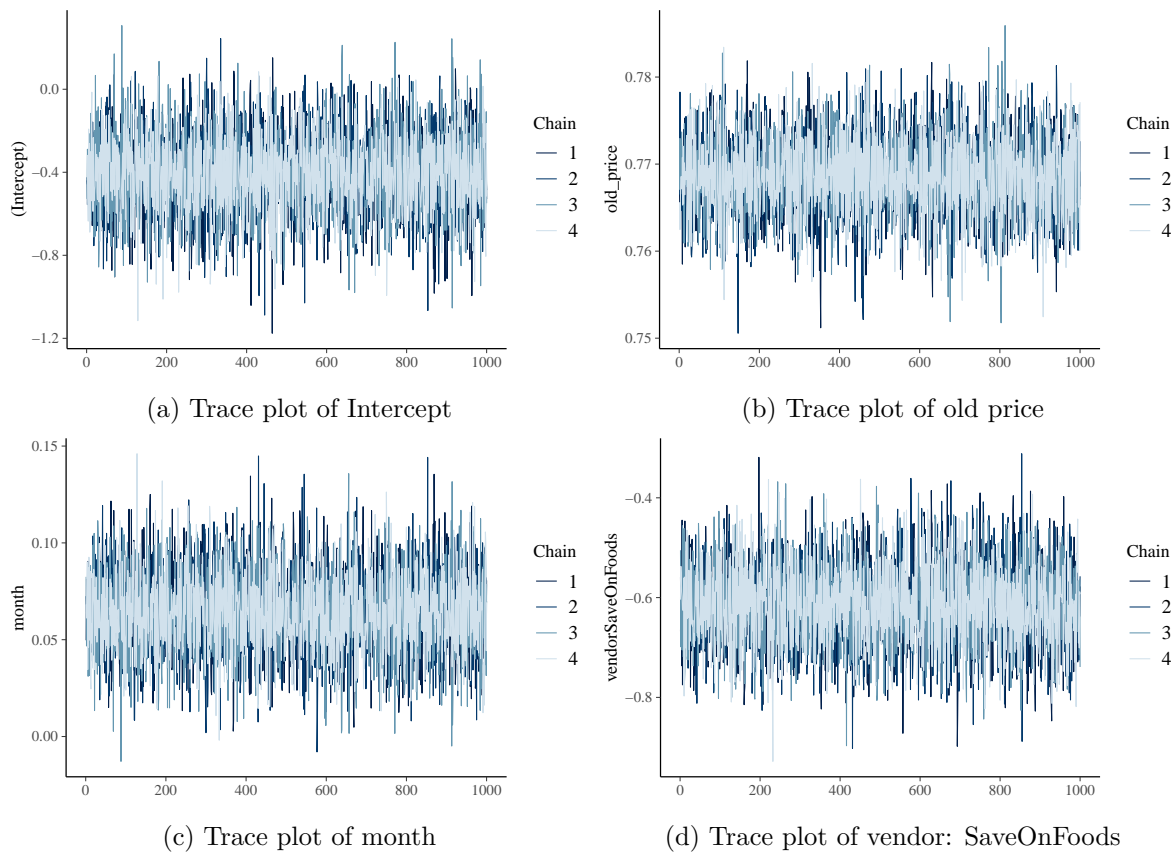


Figure 8: Trace plot of variables

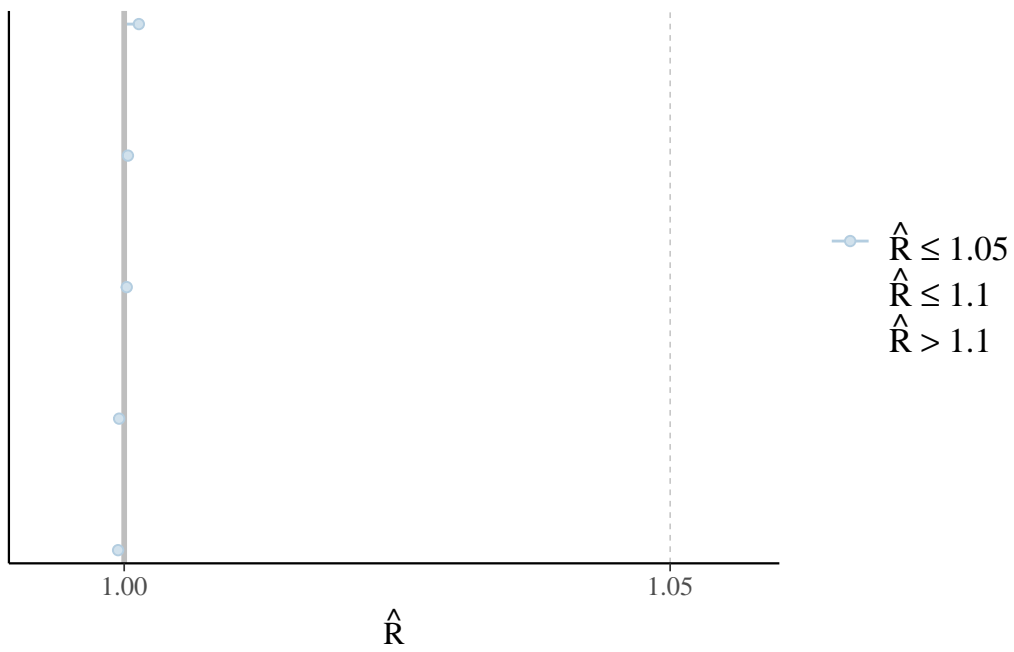


Figure 9: Good convergence with Rhat values close to 1

References

- Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. Chapman; Hall/CRC.
- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Arnot, Chris, Peter C Boxall, and Sean B Cash. 2006. “Do Ethical Consumers Care about Price? A Revealed Preference Analysis of Fair Trade Coffee Purchases.” *Canadian Journal of Agricultural Economics/Revue Canadienne d’agroeconomie* 54 (4): 555–65.
- Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for “Grid” Graphics*. <https://CRAN.R-project.org/package=gridExtra>.
- Filipp, Jacob. 2024. “Hammer Dataset.” <https://jacobfilipp.com/hammer/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Huang, Hui-Chun, Ya-Ting Chang, Che-Yi Yeh, and Chung-Wei Liao. 2014. “Promote the Price Promotion: The Effects of Price Promotions on Customer Evaluations in Coffee Chain Stores.” *International Journal of Contemporary Hospitality Management* 26 (7): 1065–82.
- Kuhn, and Max. 2008. “Building Predictive Models in r Using the Caret Package.” *Journal of Statistical Software* 28 (5): 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, 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://CRAN.R-project.org/package=arrow>.
- Schloerke, Barret, and Jeff Allen. 2024. *Plumber: An API Generator for r*. <https://CRAN.R-project.org/package=plumber>.
- Walthert, Lorenz, Kirill Müller, Nicholas Tierney, Henrik Bengtsson, Gábor Csárdi, Aaron Jacobs, and Jennifer Bryan. 2024. *Styler: Non-Invasive Pretty Printing of r Code*. <https://CRAN.R-project.org/package=styler>.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019b. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

- , et al. 2019a. “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.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with ‘kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.