

Analyzing and forecasting for different vendors' relationships and pricing strategies*

By historical price, vendor type and nature of goods

Ruikang Wang(1008238872)

December 3, 2024

This study explores how old prices and different suppliers influence the current price of goods. By using Bayesian linear regression, the findings suggest that higher old prices often lead to slightly lower current prices. Supplier-specific effects also show varying strategies in adjusting price levels, but overall trends indicate eventual reductions.

1 Introduction

There is an old Chinese saying, “Hunger breeds discontentment”, which means that food is almost the most important thing in people’s lives. In modern society, people’s means of obtaining food has changed from hunting to consumption, and therefore the price of food has become a major criterion for reflecting the social situation. The analysis of food prices plays a vital role in understanding the Canadian consumer as well as the Canadian market. This paper focuses on the relationship between the prices of the different items and their own properties, supplier strategies, and historical prices. Four food items that are inseparable from breakfast, lunch, dinner, and snacks are chosen for this study: bread, milk, bacon, and eggs. Their pricing is affected by various factors including, but not limited to, supplier strategies, and historical prices. In terms of suppliers, all eight suppliers in the dataset are selected for this paper, as they cover the Chinese supermarket “TandT”, the Korean supermarket “Galleria”, and the more popular North American” and European supermarkets. Walmart” in China, ‘Galleria’ in Korea, and the more popular ‘Walmart’ in North America and ‘Voila’ in Europe. This makes the study more comprehensive and generalizable. The estimand of this paper is the current price of the commodity, which is determined based on the historical price, the type of supplier, and the nature of the commodity itself. By examining the effect of these predictor variables on the current price, an attempt is made to derive the relationship between suppliers and the

*Code and data are available at: <https://github.com/Rickenn/unfair-competition-or-not>.

pattern of commodity pricing. By using Bayesian linear regression model, the findings suggest that higher old prices often lead to slightly lower current prices, as suppliers adjust prices to stay competitive (Payne et al. 2021). Supplier-specific effects also show varying strategies in adjusting price levels, but overall trends indicate eventual reductions. While some suppliers (e.g., Metro, NoFrills, Voila) exhibit similar reductions, which might suggest possible collusion, the price trends over time do not show consistent alignment. Thus, the data does not provide sufficient evidence to support coordinated pricing behavior among these suppliers.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2 describes the clean data, and explains the variables for this research, Section 3 briefly indicates the basic of the model, Section 4 shows the prediction of current price, and justifies the relationship between different vendor, Section 5 explains the result, discusses the weakness and possible improve in the future.

2 Data

2.1 Data Source

The dataset used in this research is from Hammer (Filipp 2024), a publicly available repository of information designed to provide pricing data from many different retail chains. The dataset contains detailed information about product attributes such as product name, vendor, current price, date of changing price, brand and so on. Since it covers eight different suppliers and a variety of different commodities, we can use it to compare pricing between different suppliers and the timing of price changes between commodities, so that we can think about whether there is a pattern between the pricing of the various suppliers, and if so what that pattern is.

2.2 Measurement

The data for this study was systematically downloaded, cleaned, analyzed, modeled and visualized using R (R Core Team 2023), a extensive statistical programming language. The following packages were used for this study.

- **tidyverse** (Wickham et al. 2019): Data visualization.
- **kableExtra** (Zhu 2024): Construct Complex Table.
- **ggplot2** (Wickham 2016): Used for creating various types of visualizations.
- **dplyr** (Wickham et al. 2023): Data Manipulation.
- **here** (Müller 2020): A Simpler Way to Find Your Files.

- **knitr** (Xie, Yihui 2023a): A General-Purpose Package for Dynamic Report Generation in R.
- **rstanarm** (Goodrich, Ben and Gabry, Jonah and Ali, Iram and Brilleman, Sam 2023): Providing a straightforward way to fit Bayesian regression models using Stan.
- **modelsummary** (Arel-Bundock 2022): Data and Model Summaries in R.
- **readr** (Wickham, Hester, and Bryan 2024): Read Rectangular Text Data.
- **lme4** (Bates et al. 2015): Fitting Linear Mixed-Effects Models.
- **tinytex** (Xie, Yihui 2023b): Helper Functions to Install and Maintain ‘TeX Live’.
- **reshape2** (Wickham 2007): Reshaping Data.
- **arrow** (Neal, Weston and Urbanek, Simon and others 2023): Used for efficiently reading and writing large datasets.
- ***Telling Story With Data*** (Alexander 2024): This book provide a lot of example for code and explain the statistic concepts to help us analysis our study.

Table 1: Voila - Lactose-Free 2% Milk

nowtime	product_id	current_price	price_per_unit	other	vendor	product_name
2024-08-06 10:39:00	712154	6.49	\$0.32/100ml	NA	Voila	Lactose-Free 2% Milk
2024-08-07 08:47:00	712154	6.49	\$0.32/100ml	NA	Voila	Lactose-Free 2% Milk
2024-08-08 13:24:00	712154	5.99	\$0.30/100ml	SALE	Voila	Lactose-Free 2% Milk
2024-08-09 13:00:00	712154	5.99	\$0.30/100ml	SALE	Voila	Lactose-Free 2% Milk
2024-08-10 11:09:00	712154	5.99	\$0.30/100ml	SALE	Voila	Lactose-Free 2% Milk

Table 1 is a segment that I specifically cut out to best represent the dataset used in this research. Table 1 indicates that between the time points “2024-08-05 11:18:00” and “2024-08-10 11:09:00”, the distributor “Voila” decreased the price of the product “Lactose-Free 2% Milk” from 6.49 to 5.99, and the price per 100ml decreased from 0.32 to 0.30, both unit is dollars.

2.3 Outcome variables

2.3.1 Current Price

This variable represents the price of the commodity at the time of data collection. A model is built based on this variable to investigate what variables affect vendor pricing and how these variables affect the price of goods. The study of this variable allows consumers or other suppliers to make predictions and references before consuming. In this study, in order to model *current_price*, statistics on product categories, suppliers, and historical prices were conducted.

Table 2: Statistics for Current Price by Product

product_name	min	max	mean	median	sd
100% Whole Grain Ancient with Quinoa Bread	2.97	5.49	4.305259	4.49	0.5991119
Lactose-Free 2% Milk	4.99	7.99	6.671615	6.49	0.6100897
Omega 3 Eggs	4.37	7.43	6.823885	6.99	0.5908875
Smoked Classic Cut Bacon	3.99	9.99	8.304713	8.97	1.2722422

Table 2 summarizes the current price statistics for four products that this study chosen. “100% Whole Grain Ancient with Quinoa Bread” has the lowest price variability, while “Smoked Classic Cut Bacon” shows the highest variability with the widest price range. The average prices for the products are fairly close to their median prices, indicating a generally even distribution of prices.

2.4 Predictor variables

The predictor variables in this study are *product_name*, *vendor* and *old_price*. These variables can reflect the differences and similarities in the pricing strategies of different vendors for different kinds of goods.

2.4.1 Product Name

The predictor variable *product_name* is a categorization of four selected representative products, which are “Smoked Classic Cut Bacon”, “Lactose-Free 2% Milk”, “100% Whole Grain Ancient with Quinoa Bread”, and “Omega 3 Eggs”. “100% Whole Grain Ancient with Quinoa Bread”, and “Omega 3 Eggs”. These four items were chosen because they are common enough to be found in any large supermarket, not just the vendors in this study, which makes this report universal and allows the results of this study to be applied to most people’s lives.

And these four items can be “complementary” - an increase in the sales of one item can lead to an increase in the sales of the other items (Bansak, Simpson, and Zavodny 2015), which is an economic concept. As an example, breakfast is usually milk and bread with fried eggs and bacon. By studying them, it may also be possible to find out whether the such nature of goods (complementary) has an effect on vendor pricing.

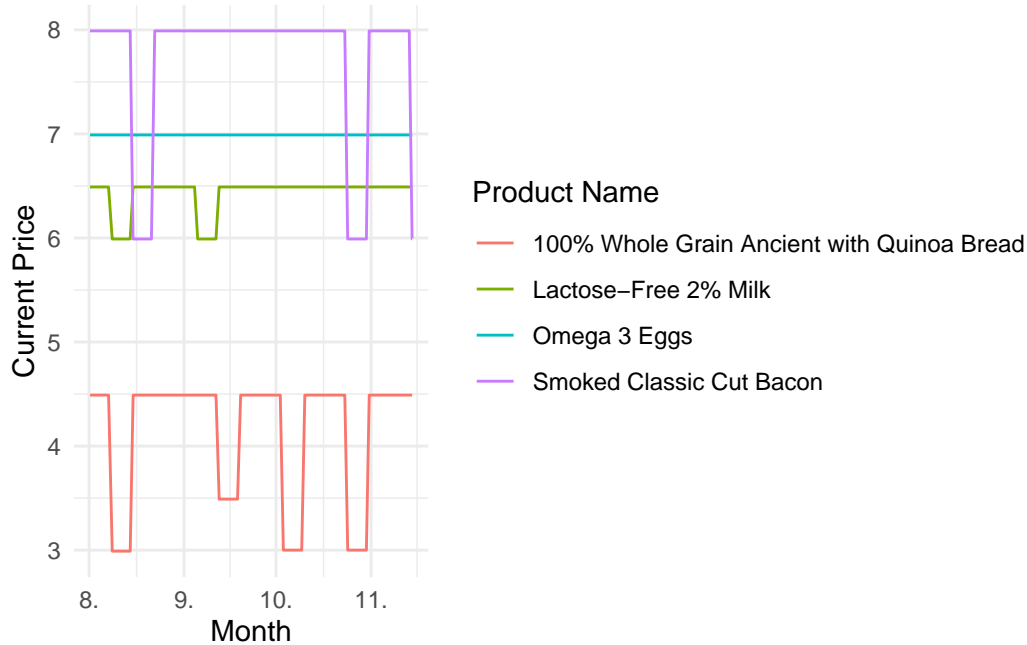


Figure 1: Current Price Over Time for Vendor Voila

Figure 1 represents price distribution of different products of *vendor* “Voila”. The “Smoked Classic Cut Bacon” showing the biggest price change, between 6 dollars and 8 dollars. The price of “100% Whole Grain Ancient with Quinoa Bread” changed most frequently, four times between August and November. On the other hand, the price of “Omega 3 Eggs” is incredibly stable, and according to the chart it has never changed.

2.4.2 Vendor

The predictor variable *vendor* is the name of eight the larger supermarkets in Toronto, which are “Voila”, “Galleria”, “Loblaws”, “Metro”, “NoFrills”, “SaveOnFoods”, and “Loblaws”. “Voila”, ‘Galleria’, ‘Loblaws’, ‘Metro’, ‘NoFrills’, ‘SaveOnFoods’. “TandT”, ‘Walmart’. They represent the consumer preferences and pricing strategies of Chinese, Korean, European, and North American supermarkets for different food products.

Figure 2 shows the price distribution with time of Omega 3 Eggs, different color means different *vendor*. “NA” is the supplier “SaveOnFoods” because they has the same “product_id”. Base

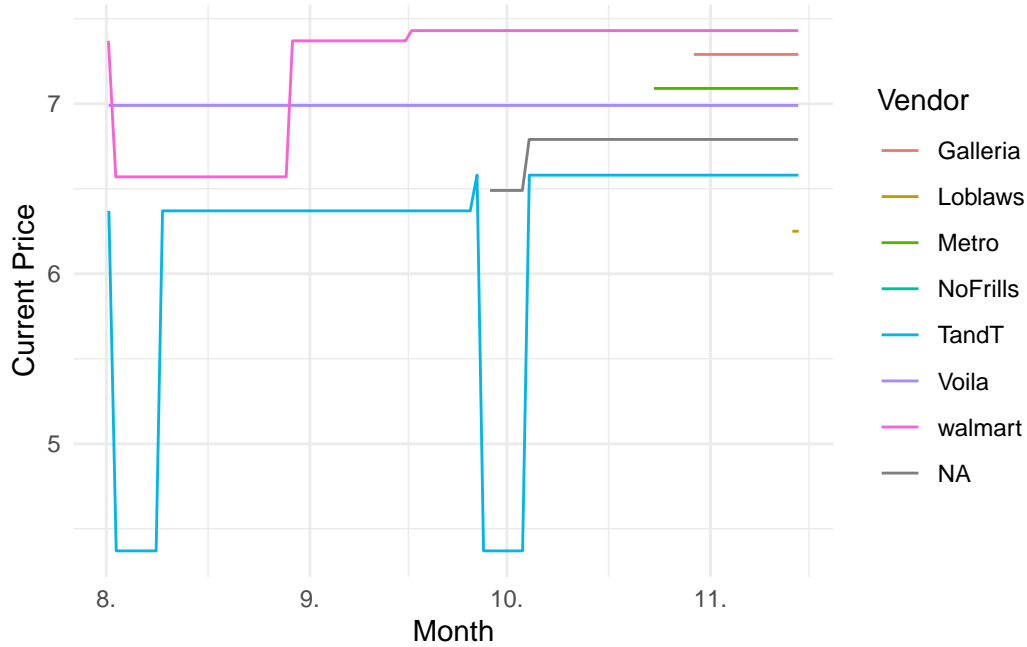


Figure 2: Current Price for Omega 3 Eggs by Vendor

on the data distribution, *vendor* “TandT” changed prices most frequently and by the largest magnitude, but in the end the prices stayed at their lowest. “Walmart” also changed the price but not frequent, and stayed at the highest price of whole market, around 7.99 dollars, after the end of August. “Voila” has the most stable price for Omega 3 Eggs, this can be corroborated with the previous “Voila” price distribution by commodity, which is more evidence of the reliability of the data.

2.4.3 Old Price

The predictor variable *old_price* is the price of each commodity on the day before the price change. Combining it with *current_price* provides insight into the price change habits of each vendor. Combining it with the predictor variable *vendor* allows us to study the response of each vendor to the strategies of other vendors. For example, do they change their prices when other vendors raise their prices in order to make higher immediate profits and stabilize the market, or do they choose Malicious price cutting to expand the market and achieve monopoly status.

Table 3: Statistics for Old Price by Product

product_name	min	max	mean	median	sd
100% Whole Grain Ancient with Quinoa Bread	2.97	5.49	3.887333	3.99	0.7705885
Lactose-Free 2% Milk	4.99	7.19	6.573265	6.49	0.4988670
Omega 3 Eggs	4.37	7.37	6.206667	6.49	1.1100901
Smoked Classic Cut Bacon	3.99	8.99	7.178696	7.99	1.6241874

Table 3 summarizes the old prices for the same four products. Compared to the summary statistics for current prices, “Smoked Classic Cut Bacon” consistently has the highest price, and both its old and current prices exhibit the largest variability (highest standard deviation). Meanwhile, “100% Whole Grain Ancient with Quinoa Bread” has the lowest variability in both current and old prices. The trends indicate that there have been moderate changes in average prices across products, but the relative relationships between products have remained similar.

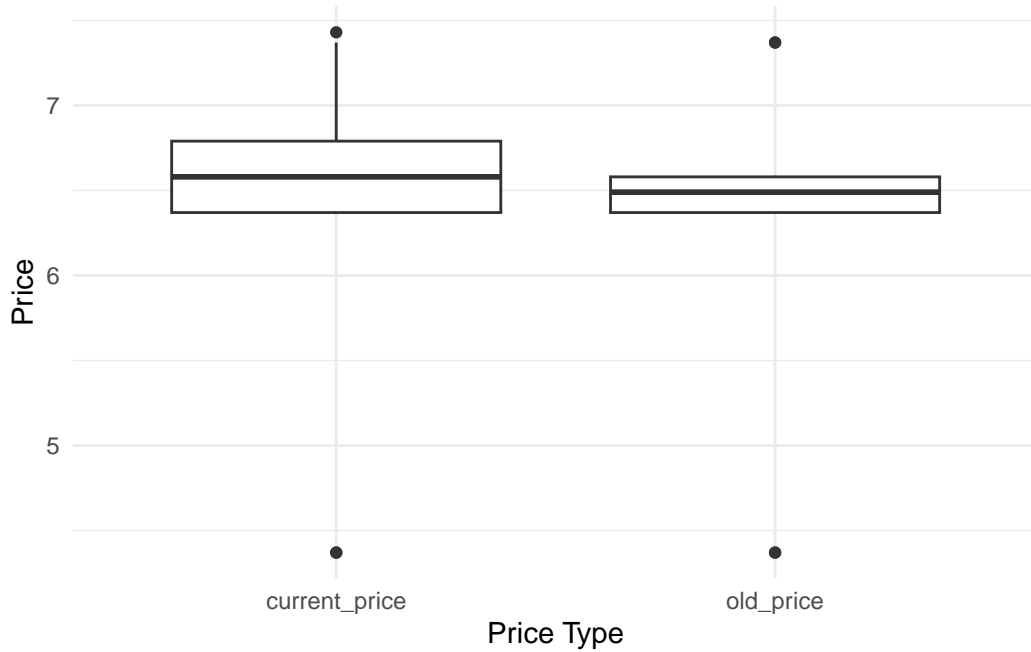


Figure 3: Comparison of Current Price and Old Price for Omega 3 Eggs

The boxplot Figure 3 compares the current price and old price for “Omega 3 Eggs”. The current price has more variability, as indicated by a wider box, while the old price is more consistent. The median of the current price is 6.6 which is slightly higher than that of the old price, showing a general increase in price. The overall trend suggests that in the whole market, the price of

“Omega 3 Eggs” has increased, with more fluctuation in the current price compared to the previous price. The interesting thing is that the lowest of the old and new prices remain the same, still 6.49 dollars, even both except their outlier.

3 Model

Here we briefly describe the Bayesian analysis model used to investigate how different initial prices between different commodities from different suppliers can affect subsequent price changes. Background details and diagnostics are included in Appendix [.2](#).

3.1 Model set-up

Define outcome variable y_i as the *current_price* of products at time i , expressed in dollars. Then define predictor variables are x_1 that is the *old_price* and also expressed in dollars, x_2 that is the *vendor*, and x_3 that is the *product_name*, which x_2 and x_3 are categorical variables.

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

$$\mu_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \alpha \quad (2)$$

Explanation of each coefficient: - β_1 can indicate how old prices affect current prices, considering the impact of historical prices is helpful in determining whether past prices have an ongoing impact on current pricing decisions. - β_2 represents the impact of vendors, which is a categorical variable. It helps us to recognize the different pricing strategies of different vendors and make unique predictions based on different vendors. - β_3 is also a categorical variable. It exhibits the effect of different types of goods on prices. Based on it, we can find out what kind of goods are price sensitive and thus understand the pricing strategy of the merchants for different goods. - α is the intercept term that represents the standard price when all the predictors are at the reference level.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich, Ben and Gabry, Jonah and Ali, Iram and Brilleman, Sam (2023). We use the default priors from `rstanarm`.

3.2 Model Priors

For the model’s intercepts, a normal prior distribution is typically used with a mean of 0. This keeps the intercept results of the model from being too close to the center. Similarly, the coefficients in the model are assigned a normal prior distribution with a mean of zero. The standard deviation was set at 2.5 to increase the flexibility of the experimental results without making the results meaningless due to excessive differences.

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

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

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

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

This priors are designed to be weakly informative, allowing the model to draw conclusions based almost entirely on the data.

3.3 Model justification

By using *vendor* and *product_name* as categorical variables, we can not only analyze how different vendors and different products affect prices, but also analyze the nature of the products. The same product with different suppliers can be considered as a “substitute” - this type of product can be used as a alternative choice for each other (Bansak, Simpson, and Zavodny 2015), so if the sales of one product go down, the sales of the other one will go up accordingly, this is an economic concept. The different commodities selected for this study are “complementary” to each other, as explained in the “data” section. The analysis of the variable *vendor* also allows us to explore whether there is bad bidding and collusion among suppliers in order to monopolize the market. The reason for using *old_price* as a numerical variable is that, according to price inertia, past prices may affect new pricing.

3.3.1 Assumption

The model is assumed to be linear, both between the response variable and the predictor variable. This makes the model simple and easy to understand. And the model can visualize the relationship between the variables. It is more likely to improve the consistency of the model under specific conditions. However, assuming linearity may lead to model mis-specification. This means that important interactions or curvatures in the data are ignored, leading to biased or incorrect estimates and predictions. And linear regression requires assumptions about residual errors. If these assumptions are violated, the predictions of future price may become unreliable.

3.3.2 Different model comparison

I have tried to do a linear model before because it was assumed at the outset that there was a linear relationship between old price and current price, but the linear model came up with predictions that were far from the true value. This may be due to the fact that the linear model is monotonic in the direction of change, its prediction is a straight line with only one direction, either up or down, but this is not the case with price movements. Although the study concludes that the price is generally going down (which will be analyzed in more detail in a later section), there are often rises in the process of decline, just with different magnitudes and frequencies, so the linear model is often subject to a great deal of error. Bayesian linear regression has an advantage over simple linear models in dealing with uncertainty by providing flexibility for hierarchical modeling and the ability to deal with complex model structures, so that the current price predicted by the Bayesian model, although in error from the actual price, is almost has identical turning point with actual prices, and these small errors cannot hide the powerful analytical ability of the Bayesian model. This is also the reason why I finally chose the Bayesian model.

4 Results

4.1 Graph analysis

The three figures depict the price changes over time for lactose-free milk, whole grain bread, and smoked bacon sold by different vendors.

In Figure 4, most sellers (Loblaws, NoFrills, and Walmart) ended with prices between 6.0 and 6.5, while TandT remained above 7.0, indicating it was an outlier with a consistently higher price. However, based on the curves of Loblaws and NoFrills, we find that when the distribution of their price changes is reversed, one party chooses to reduce prices when the other increases prices, and vice versa, which is consistent with our definition of “substitution”. Figure 5 shows that vendors like Loblaws, Metro, NoFrills, and Voila ended with prices in a narrower range (4.25 to 4.75), suggesting a convergence of pricing strategies for whole grain bread. In Figure 6, Walmart and SaveOnFoods maintained stable prices of \$9.00 throughout, while NoFrills also eventually increased its price to match this level.

Overall, the prices, regardless of their initial differences, tended to converge and stabilize across most sellers over time, indicating possible market influences that drive similar pricing across vendors. Although the prices end up converging in a close range, the paths of price changes are all different, so there is no reason to judge the suppliers as having cooperative behavior. And it is true that a change in the price of a commodity will have the opposite effect on its substitute, whether that change is up or down.

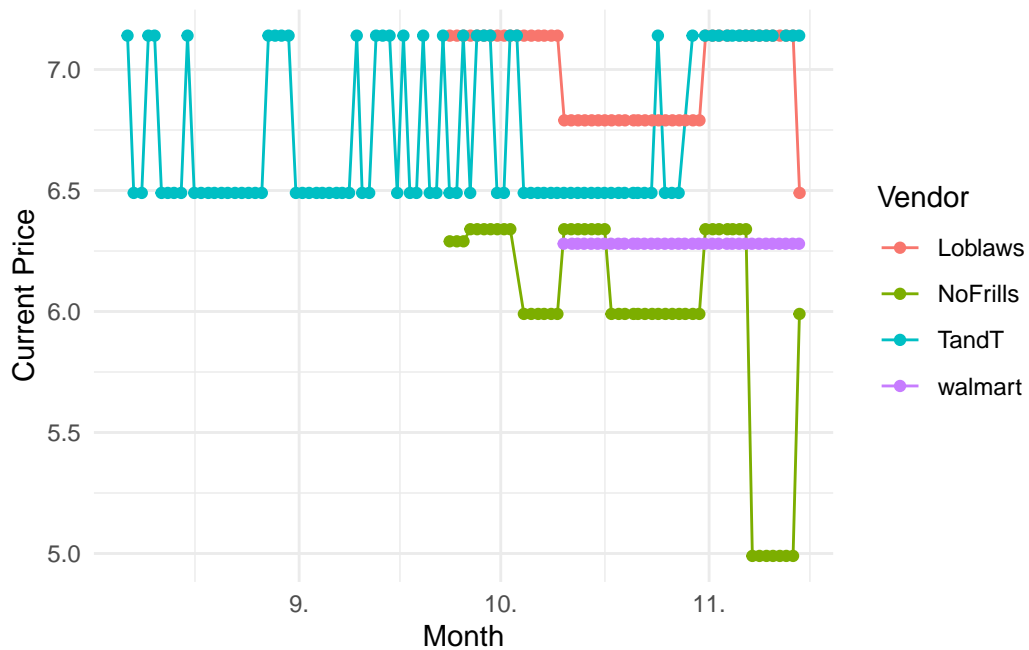


Figure 4: Price Trends Over Time for Lactose-Free 2% milk

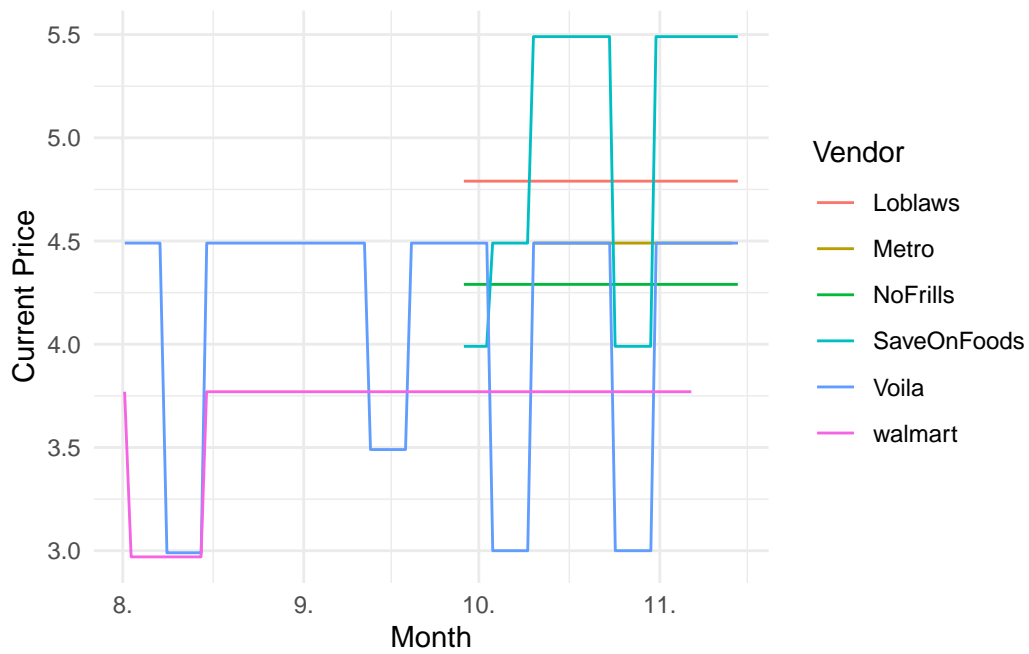


Figure 5: Price Trends Over Time for 100% whole Grains Ancient with Quinoa Bread

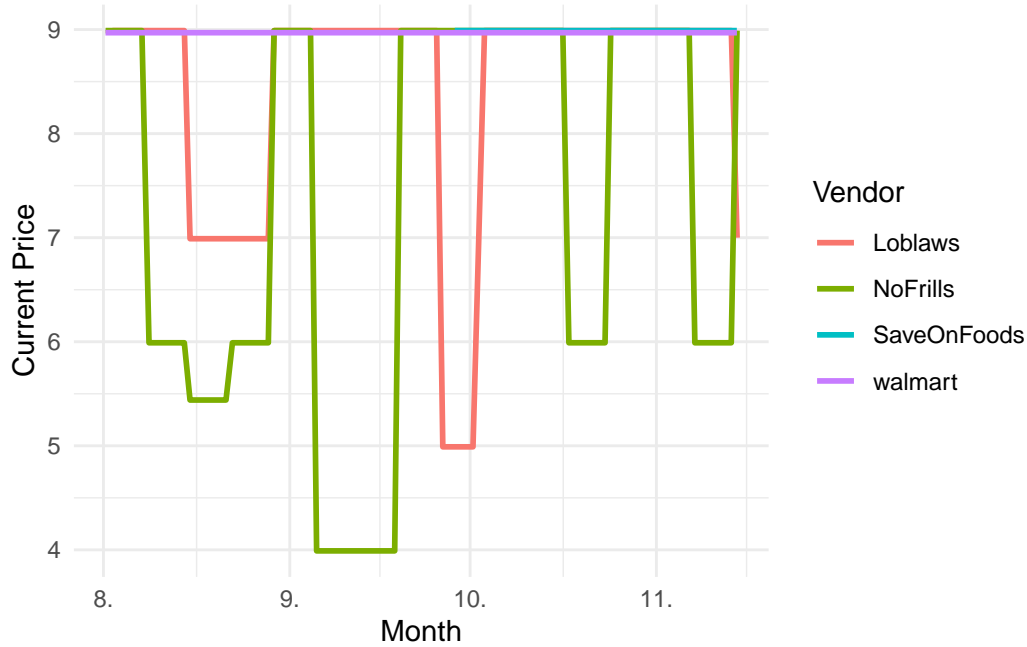


Figure 6: Price Trends Over Time for Smoked Classic Cut Bacon

4.2 Modeling

Table 4 demonstrates the findings of the Bayesian model used to analyze current price and helps to analyze the parameter estimates and diagnostic measures. This model explains how the predictor variables of old price, vendor, and product type affect the current price.

The Intercept 2.10 is the starting point, it means the current price starts at 2.10 when old price is 0 and other situation remain the same.

The coefficient of old Price is -0.10 which means that the old price and current price are negatively correlated. If the old price goes up by 1 unit, the new price is expected to decrease slightly by 0.10 units.

Vendor Effects: Different stores affect the price differently compared to the reference store. For example, Metro has a value of -0.53, meaning products in Metro tend to have a price that is 0.53 units lower compared to the reference store. Other stores like Loblaw's (-0.32), NoFrills (-0.50), etc., also show negative values, meaning their prices are lower compared to the baseline.

Product Effects: Lactose-Free 2% Milk (0.83), Omega 3 Eggs (0.65), and Smoked Classic Cut Bacon (0.98): These numbers mean these products tend to have higher prices compared to a baseline product. For example, Smoked Classic Cut Bacon has a value of 0.98, meaning it's more expensive compared to the reference product.

Table 4: Explanatory models of Current price based on Old price and vendor and Product type

	Current Price model
(Intercept)	2.10 (0.23)
old_price	−0.10 (0.02)
vendorLoblaws	−0.32 (0.20)
vendorMetro	−0.53 (0.21)
vendorNoFrills	−0.50 (0.20)
vendorSaveOnFoods	−0.14 (0.21)
vendorTandT	−0.36 (0.20)
vendorVoila	−0.43 (0.20)
vendorwalmart	−0.24 (0.22)
product_nameLactose-Free 2% Milk	0.83 (0.10)
product_nameOmega 3 Eggs	0.65 (0.12)
product_nameSmoked Classic Cut Bacon	0.98 (0.12)
Num.Obs.	95
Log.Lik.	−124.707
ELPD	−133.5
ELPD s.e.	3.9
LOOIC	266.9
LOOIC s.e.	7.8
WAIC	265.8
RMSE	0.70

Model Fit Metrics: • Log Likelihood (Log.Lik.) = -124.707: This is a number that shows how well the model fits the data. A less negative value is better and means the model fits the data more closely. • Expected Log Predictive Density (ELPD = -133.5), Leave-One-Out Information Criterion (LOOIC = 266.9), and Widely Applicable Information Criterion (WAIC = 265.8): They can help us understand if the model is good at predicting new data. The low values means this model predicts with fancy performance. • Root Mean Squared Error (RMSE = 0.70): This number tells us how far off our predictions are on average. An RMSE of 0.70 means there is some difference between what the model predicted and what actually happened, but it's not too bad.

5 Discussion

5.1 Explanation of Results

Research findings show that if the old price of a good increases, the current price tends to decrease slightly. This is because price reduction is the best way to expand when there is a change in the market. The impact of different suppliers is also taken into account. Each store has its own impact on prices compared to the baseline, but overall, regardless of the initial price increase, individual suppliers will eventually lower the price of an item given enough time. And different suppliers may have their own unique pricing strategies that will adjust the nodes, frequency, and magnitude of price reductions differently. Overall, this study helps explain how old prices and store selection affect the final price of a product, highlighting pricing trends across different suppliers. This is a useful way to understand how current pricing changes based on past data and where products are sold.

5.2 Market Collusion or not

All the vendors listed (e.g., Voila: -0.43, Metro: -0.53, NoFrills: -0.50, SaveOnFoods: -0.14, etc.) have negative coefficients compared to the reference category, indicating that the current price is generally reduced for each of these vendors relative to the baseline. A similar trend across vendors could indicate coordinated efforts to keep prices low for competitive advantage. However, this observation alone is not enough to confirm collusion, as many market factors could explain these pricing strategies. If the vendors have similar levels of reduction in prices (such as NoFrills, Metro, and Voila with coefficients close to each other), this may raise suspicion of coordinated pricing behavior, especially if it is observed consistently over time. Collusion often manifests in uniform price movements among competitors to avoid direct competition. Therefore NoFrills, Metro, and Voila are the three vendors most likely to collude. However, if we look back at the images of the price changes over time for each commodity we see that the trends in the price movements of these three commodities are not similar. So based on the available data and modeling, we cannot accept the hypothesis that they are colluding with each other, and their similar coefficients may just be coincidental.

5.3 Prediction for future

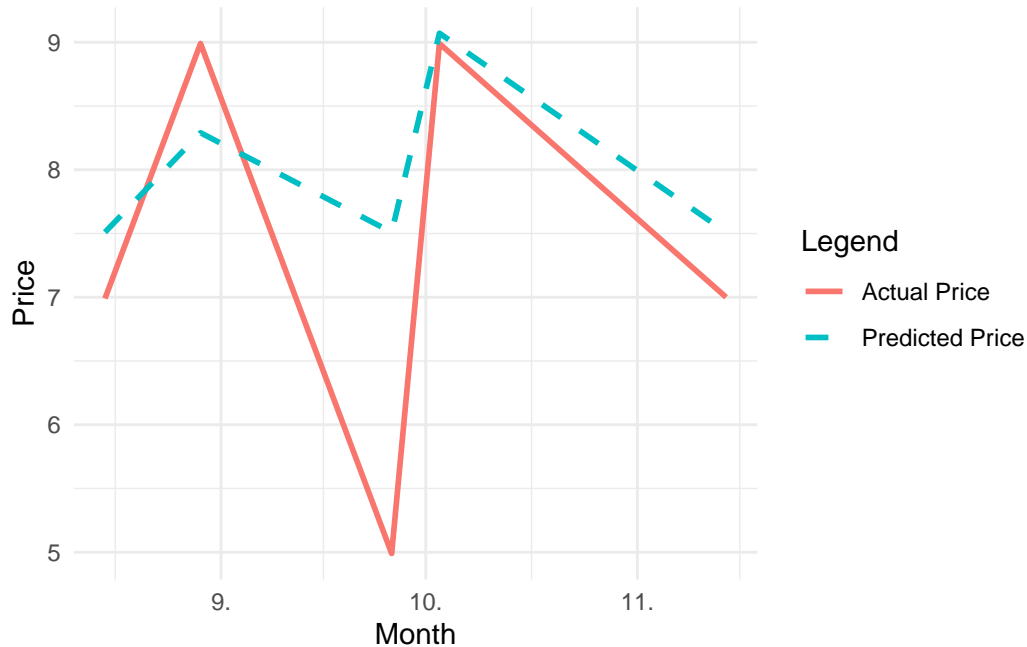


Figure 7: Comparison of Actual and Predicted Prices for Classic Cut Bacon at Loblaw's

The Figure 7 shows the comparison between the actual and predicted prices for “Classic Cut Bacon” at Loblaw's over time. The actual prices demonstrate significant fluctuations, including sharp increases and rapid drops, suggesting high volatility in the pricing of this product. Specifically, the price appears to vary between 5 and 9 dollars, with extreme changes observed within a short time period. In contrast, the predicted price trend, shown as a dashed line, is much smoother and indicates a gradual upward movement, followed by a peak and a slight downward adjustment. The model predicts a more stable price trajectory compared to the actual volatile pattern. The discrepancy between the actual and predicted trends suggests that while the historical data shows significant instability, the model expects this instability to smooth out over time. Notably, towards the end of the observed period, the predicted price exceeds the actual price, which implies that the model anticipates an increase in prices after the recent decline. This predicted upward trend indicates a possible recovery or price correction that may occur in the near future. In summary, while the actual price data reflects considerable volatility for “Classic Cut Bacon” at Loblaw's, the model suggests that the future price trend will be more stable, with moderate upward movement. This prediction may be optimistic, assuming that the factors driving the extreme price changes previously will stabilize. It is essential to acknowledge that price predictions are based on historical patterns, and external factors such as market promotions, supply issues, or unexpected changes in consumer demand could result in future prices deviating from these predictions. Overall, the graph suggests that

the price of “Classic Cut Bacon” is expected to rise and stabilize, but real-world dynamics could still impact actual outcomes.

5.4 Weaknesses and next steps

5.4.1 Weaknesses

The model uses a linear approach, which might not capture complex relationships between variables. Real-world pricing may involve non-linear effects or interactions that this model does not capture. The model also does not include interaction terms, which means it does not account for situations where the effect of one variable depends on another. For example, the effect of a vendor on price might depend on the specific product, which this model cannot capture without interaction terms.

5.4.2 Next steps

To address the limitations of linear modeling and the absence of interaction terms, consider the following improvements. First, to capture non-linear relationships, you can include polynomial terms or use Generalized Additive Models (GAMs), which provide more flexibility. Tree-based models like Random Forest or Gradient Boosting Machines can also effectively capture non-linear and complex relationships without assuming a specific form for the data. Second, to account for interaction effects between variables, add explicit interaction terms, such as between the vendor and product name, to capture scenarios where one variable’s impact depends on another. Generalized Additive Mixed Models (GAMMs) can also be used to combine smooth functions with random effects to account for complex interactions. Next steps include experimenting with different models, adding interaction terms, and conducting model diagnostics to determine which approach yields the best performance. Evaluating different methods using metrics such as RMSE, cross-validation, and residual analysis will help identify the most suitable model to capture real-world pricing behaviors accurately. By incorporating non-linearities and interaction effects, the model will become more robust and better represent the intricacies of the dataset.

Appendix

.1 Additional data details

The dataset “price_change_data” used to construct the model starts with two numeric variables *price_change* and *old_price*, but because the code for building the Bayesian model cannot have negative numbers, and as we all know, prices cannot only go up but not down, so we have to do something about “price_change_data”. By adding a new variable *current_price* to make the data usable, because the price of any commodity will not be negative. However, this new data is used to generate the model and all the analysis in this paper is still based on the unchanged “price_change_data” and “analysis_data”. The formula for *current_price* is: $current_price = old_price + price_change$.

.2 Model details

.2.1 Posterior predictive check

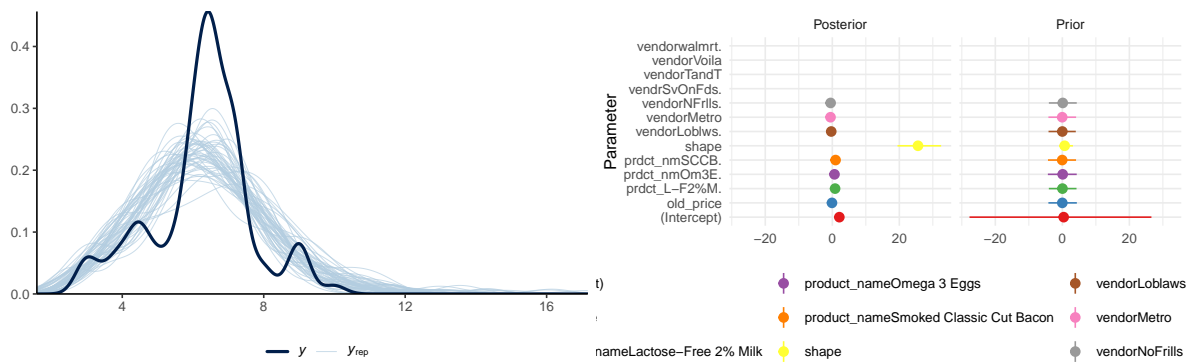
In Figure 8a we implement a posterior predictive check. The thick dark line represents the actual data distribution (y). The lighter lines represent the distributions from posterior predictive samples (y_{rep}). This graph represents that the predicted data roughly follow the same pattern as the actual data. This suggests that the model is doing a reasonable job in predicting the shape of the data, but there are some discrepancies at the tails, indicating potential areas for improvement.

In Figure 8b we compare the posterior with the prior. This shows that the posterior estimates (colored dots) have diverged significantly from the priors for several parameters, the *product_name* and *vendor* are good example, which suggests that the data was informative for these estimates. It indicates that the model has learned from the data, as seen by the divergence of the posteriors from the priors. The influence of data is strong on several parameters.

.2.2 Diagnostics

Figure 9a is a trace plot. It also called MCMC (Markov Chain Monte Carlo) sampling process. It shows the sampling chains for each parameter in the model. Each chain looks fuzzy like a caterpillar, and both of them show good mixing and a consistent pattern. This suggests that the MCMC chains have converged and are sampling effectively from the posterior distribution, so the chains are exploring the parameter space effectively and are converging.

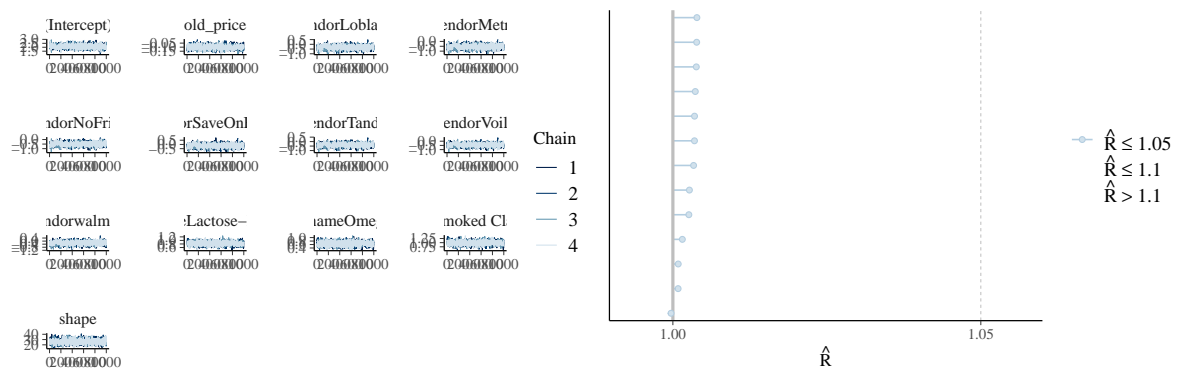
Figure 9b is a Rhat plot. It shows shows the potential scale reduction factor, commonly called R-hat, for each parameter. In this plot, all the R-hat values are equal to or below 1.05, which is a good sign that convergence has been achieved.



(a) Posterior prediction check

(b) Comparing the posterior with the prior

Figure 8: Examining how the model fits, and is affected by, the data



(a) Trace plot

(b) Rhat plot

Figure 9: Checking the convergence

A Idealized Methodology for a Survey

A.1 Overview

Idealized survey methods capture the perceptions of consumers, as well as employees and rule makers who actually work in this market, about the market's food pricing. The survey is designed to determine simulations of purchasing decisions, judgments about the fairness of the market, and preferred sales tactics for vendors. This approach allows for a more comprehensive understanding of the factors that influence the dynamics of food pricing in Canada thereby corroborating or modifying the analysis in this paper and laying the groundwork for providing a truly rational, useful analytical model.

A.1.1 Question Types

The survey focuses on self-identity, eating habits, living habits, and shopping styles, with multiple-choice and single-response questions, and also asks participants to think about the other person's point of view, so as to explore what the other person might be thinking about in the mind of each participant.

A.1.2 Question List

1. What self-identify you prefer?
 - Customer
 - vendor staff
 - vendor manager
2. How often you shopping?
 - Daily
 - Weekly
 - Monthly
 - Other
3. Which is your primary way of consumption?
 - Offline
 - Online
4. Which vendor do you often go? (multiple)
 - Voila
 - Metro

- TandT
 - Save On Foods
 - Loblaws
 - No Frills
 - Galleria
 - Walmart
 - other
5. Which factors do you most care for your shopping?
- Price
 - Quality
 - Environment
 - Convenience
6. Do you have breakfast in recently one month?
- Often
 - Occasionally
 - Barely
 - Never
7. If you have breakfast, which you will eat most frequent?
- Rice
 - Noodle
 - Bread
 - Cereal
8. If you have breakfast, which may be your source of protein? (multiple)
- Egg
 - Bacon
 - Steak
 - Protein bar
9. If you have breakfast, which is your drink?
- Water
 - Milk
 - Juice
 - Alcohol
 - Soft drink
10. If you are the manager for vendor, please make a combo by choose four of below food.

- Rice
- Noodle
- Bread
- Cereal
- Egg
- Bacon
- Steak
- Protein bar
- Water
- Milk
- Juice
- Alcohol
- Soft drink

11. If you are the customer, please make a combo by choose four of below food that you want be provided in vendor.

- Rice
- Noodle
- Bread
- Cereal
- Egg
- Bacon
- Steak
- Protein bar
- Water
- Milk
- Juice
- Alcohol
- Soft drink

12. If you are customer, will you still by this four foods together you choose before even no combo?

- Yes
- No
- Base on situation

13. If the price of foods other than the combo you choose drops, will you still buy these four foods?

- Yes
- No
- Base on situation

A.2 Recruitment Strategy

A.2.1 Online questionnaire

Extensive participant databases by age, income and geographic location, and distribute online questionnaires. This method of collection are cheap, convenient, and comprehensive, allowing for diverse responses. The data is automatically recorded.

A.2.2 offline interview/questionnaire

Distribute questionnaires in vendors, or staffs and managers do short interviews with customers as they check out. This way provide a personal touch, allowing for better rapport and engagement between the interviewer and the respondent, which can lead to more detailed and honest responses.

References

- Alexander, Rohan. 2024. *Telling Stories with Data*. "University of Toronto". <https://www.tellingstorieswithdata.com>.
- 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>.
- Bansak, C., N. Simpson, and M. Zavodny. 2015. *The Economics of Immigration*. Taylor & Francis. <https://books.google.ca/books?id=SHKhCAAAQBAJ>.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Filipp, Jacob. 2024. "Hammer." <https://jacobfilipp.com/hammer/>. <https://jacobfilipp.com/hammer/>.
- Goodrich, Ben and Gabry, Jonah and Ali, Iram and Brilleman, Sam. 2023. *rstanarm: Bayesian Applied Regression Modeling via Stan*. <https://mc-stan.org/rstanarm/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Neal, Weston and Urbanek, Simon and others. 2023. *arrow: Integration to Apache Arrow*. <https://CRAN.R-project.org/package=arrow>.
- Payne, J., W. R. Dorn, D. Pastore, and J. Ulrich. 2021. *Managing Indirect Spend: Enhancing Profitability Through Strategic Sourcing*. Wiley Corporate f&a. Wiley. <https://books.google.ca/books?id=wxouEAAAQBAJ>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- . 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://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Xie, Yihui. 2023a. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://CRAN.R-project.org/package=knitr>.
- . 2023b. *tinytex: Helper Functions to Install and Maintain 'TeX Live', and Compile 'LaTeX' Documents*. <https://CRAN.R-project.org/package=tinytex>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.