Understanding Walmart's Pricing Dynamics: Stability, Proportionality, and Key Anomalies in Product Prices*

Consistent Pricing Strategies with Outliers Highlighting Dynamic Market Adjustment

Zhengjie Zhu

November 21, 2024

We analyzed Walmart product prices to uncover trends and anomalies in historical and current pricing data. Using statistical models, we found a strong positive relationship between old and current prices, indicating consistent pricing strategies over time, with a few notable outliers. These findings reveal how large retailers adapt prices in response to market conditions, providing insights into pricing stability and variability. This research enhances our understanding of retail pricing dynamics, benefiting consumers, competitors, and policymakers alike.

1 Introduction

The pricing strategies of major retailers like Walmart in Canada play a critical role in shaping market dynamics, influencing consumer behavior, and reflecting economic trends. Understanding how product prices evolve over time provides insight into the consistency, proportionality, and anomalies in pricing practices. While studies have explored broad market pricing trends, few focus on a single retailer's longitudinal pricing data, leaving a gap in our understanding of how large-scale retailers adapt to changing economic conditions. This paper addresses this gap by analyzing Walmart's historical and current product pricing, using a dataset that captures the prices of individual products across different time periods. By systematically exploring these trends, we aim to uncover patterns, identify outliers, and improve Walmart's pricing strategies.

Our analysis focuses on estimating the relationship between historical and current prices of Walmart products, with a particular emphasis on understanding proportional changes over

*Code and data are available at: https://github.com/nathan-1022/Walmart-price-analysis.git.

time. The estimand of interest is the expected change in current price for a unit change in the historical price, measured on a logarithmic scale to capture proportional effects. Additionally, we examine deviations from this relationship to identify pricing anomalies and assess their potential causes. This approach enables us to quantify pricing consistency while highlighting areas where dynamic adjustments occur.

Our findings reveal a strong positive correlation between historical and current prices, with the majority of products adhering to a proportional pricing trend. The log-transformed analysis shows that a unit increase in historical price is associated with a consistent increase in current price, confirming the stability of Walmart's pricing strategy over time. However, notable outliers were identified, representing products with disproportionate price changes. These outliers often reflect external market influences or strategic pricing decisions. Summary statistics and visualizations further demonstrate that most products exhibit minimal deviation from the trendline, highlighting the overall consistency in Walmart's pricing practices.

This analysis has important implications for consumers, competitors, and policymakers. For consumers, understanding pricing trends provides predictability, aiding in budgeting and purchasing decisions. Competitors can use insights about Walmart's pricing strategies and outliers to identify opportunities for competitive advantage. For policymakers, the findings offer a case study on market pricing behavior, contributing to efforts to monitor fair practices in the retail sector. By uncovering consistent pricing trends and highlighting anomalies, this research not only enhances understanding retail pricing dynamics but also provides actionable insights for stakeholders across the economic spectrum.

The remainder of this paper is structured as follows. In Section 2 we describe the dataset, its sources, and preprocessing steps. In Section 3, we present the results of our analysis, including key visualizations and statistical summaries. Section 4 discusses the implications of these findings for various stakeholders, and concludes the paper with a summary of insights and recommendations.

2 Data

2.1 Overview

We use the statistical programming language `R` (R Core Team 2023) to analyze and model our data, which was obtained from (Filipp 2024). The processed dataset focuses on Walmart product prices] and includes variables such as old and current product prices. These variables capture the historical and current pricing structures for Walmart products in Canada, providing a foundation for exploring pricing trends, stability, and outliers. Following Alexander (2023), we aim to communicate a clear and comprehensive sense of the data, ensuring transparency and replicability in our analysis.

The analysis makes extensive use of `R` packages, including `tidyverse` (Wickham et al. 2019), which provides a cohesive set of tools for data cleaning, wrangling, and visualization. Using `dplyr`, we filtered, grouped, and summarized key variables to examine pricing trends, while `ggplot2` was employed to create scatterplots and other visualizations that highlight proportional pricing trends and deviations. Additionally, we used `testthat` (Wickham 2011) to implement unit tests and ensure the reliability and reproducibility of all data transformations and calculations. We used `here` to facilitate data import and file path management to ensure reproducibility. These packages streamlined our workflow, enabling efficient exploration and validation of the dataset.

The dataset was chosen over similar publicly available data because it offers a unique focus on Walmart's pricing trends over time, which are essential for understanding broader market dynamics.Variables were examined and constructed as necessary to align with our research objectives. For example, log-transformed variables for pricing were created to capture proportional changes effectively. High-level cleaning tasks included removing missing values, handling outliers, and ensuring consistency in data formatting, which ensured the dataset was analysis-ready without altering its inherent structure.

To communicate the characteristics of the dataset, we include visualizations that explore the relationships between variables. Scatterplots, such as those comparing historical and current prices, reveal proportional trends and highlight outliers, Relationships between variables, such as the correlation between historical and current prices, are discussed to uncover underlying trends and patterns. This comprehensive approach ensures that the dataset's structure, trends, and unique features are well understood before delving into more detailed analyses.

2.2 Measurement

The journey from observing a phenomenon in the world to its representation as an entry in the dataset involves several critical steps of conceptualization, data collection, and preprocessing. In this case, the phenomenon of interest is the pricing trends of Walmart products over time, a real-world dynamic influenced by market forces, consumer demand, competition, and retailer strategies. Capturing this phenomenon requires translating abstract concepts like "pricing behavior" or "price consistency" into measurable variables that can be systematically recorded and analyzed.

The first step in this process involves identifying the specific aspects of the phenomenon to be captured. Here, we focus on two key variables: the historical ("Old Price") and current ("Current Price") values of Walmart products. These variables were chosen because they allow for a direct comparison of pricing trends, enabling us to study patterns, outliers, and potential anomalies. Additional metadata, such as product categories, dates of price changes, and product identifiers, could also be collected to provide context and facilitate more nuanced analyses.

Data collection then transforms this conceptual framework into tangible entries. Prices for Walmart products are recorded from publicly available data sources, such as historical price tracking databases or retailer pricing archives. Ensuring accuracy and completeness during this stage is critical, as any gaps or errors could compromise the integrity of the dataset. The collected data is then structured into a format suitable for analysis, with each entry representing a unique product and its associated pricing details.

Once collected, the data undergoes preprocessing to refine its representation and ensure usability. For example, raw price values might include inconsistencies, such as formatting errors or missing values, that must be addressed. Additionally, the prices are transformed into logarithmic values to better reflect proportional changes and facilitate meaningful comparisons. Through these steps, the abstract concept of "price change" in the real world is systematically captured, cleaned, and structured into a dataset that can be analyzed to uncover insights about Walmart's pricing behavior.

Ultimately, this process allows us to bridge the gap between the complexity of real-world phenomena and the structured simplicity of a dataset. By carefully conceptualizing the phenomenon, collecting relevant data, and preprocessing it for analysis, we create a dataset that serves as an accurate and reliable representation of the underlying world dynamics we aim to study.

3 Results

Figure 1 focused on Walmart product prices, reveals a clear and strong positive correlation between historical ("Old Price") and current ("Current Price"), with both axes on a logarithmic scale. The clustering of green data points along the orange trendline suggests that Walmart employs a consistent pricing strategy over time, with proportional changes in product prices. However, the scatter also highlights deviations from this trendline, where certain products have experienced disproportionate price changes. These outliers could indicate dynamic pricing adjustments driven by market factors such as demand fluctuations, supply chain disruptions, seasonal trends, or strategic pricing shifts for competitive positioning. Many products, however, exhibit minimal deviation from the trendline, reflecting stable and predictable pricing practices for the majority of Walmart's offerings.

The implications of this analysis are significant for various stakeholders. For consumers, the observed pricing consistency provides predictability, aiding in budgeting for frequently purchased items, though outliers may warrant attention for unexpected price fluctuations. For Walmart, these findings highlight the effectiveness of its pricing strategies while identifying outliers as potential opportunities for optimization or re-evaluation. Competitors can leverage this insight to target products with significant price deviations to gain a competitive edge. Finally, for policymakers, the graph underscores the value of such analyses in monitoring pricing trends to ensure fair market practices, particularly in essential goods. By providing

transparency in pricing strategies, this analysis supports the broader goals of fair competition and consumer trust in the retail sector.

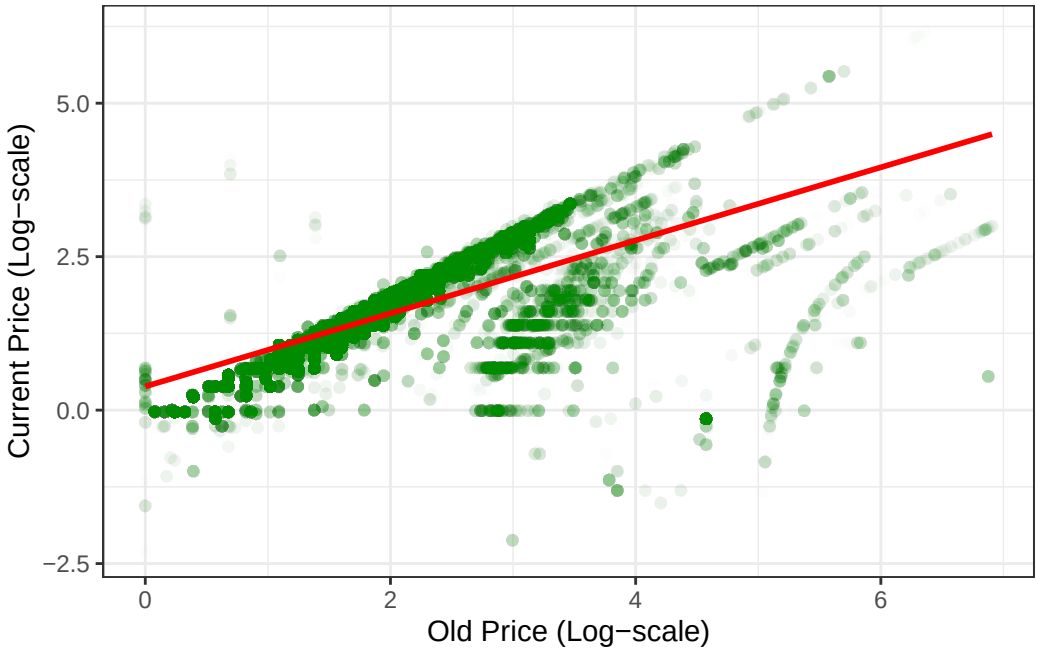

Figure 1: The relationship between current and old product prices of Walmart in Canada.

4 Discussion

4.1 Correlation vs. Causation

In analyzing the relationship between historical ("Old Price") and current ("Current Price") data, it is important to emphasize that correlation does not imply causation. While the scatterplot and trendline reveal a strong positive correlation, suggesting that current prices are influenced by historical ones, this relationship does not inherently prove causation. Other factors, such as market demand, inflation, supply chain constraints, or strategic pricing decisions, may also play a role in shaping current prices. Without a controlled experimental design or additional data on causal mechanisms, we cannot definitively attribute changes in current prices to historical pricing alone.

4.2 Missing Data

Missing data presents a potential challenge in ensuring the completeness and accuracy of our analysis. Gaps in pricing records, whether due to data entry errors, unrecorded products, or incomplete historical archives, could introduce bias or reduce the generalizability of findings. In our dataset, missing data was addressed during preprocessing by either imputing reasonable estimates or excluding incomplete entries, depending on the context and availability of information. While these measures mitigate some of the issues, it is essential to recognize the inherent limitations imposed by missing data and to interpret results with caution, particularly for products or time periods with limited coverage.

4.3 Sources of Bias

Bias in this analysis could stem from various sources, including the selection of products, data collection methods, and preprocessing decisions. For instance, the dataset may disproportionately represent products with consistent pricing trends while excluding items with irregular pricing histories, leading to skewed results. Additionally, if pricing data is collected only for specific regions or store formats, the findings may not generalize across all Walmart locations. Preprocessing decisions, such as removing extreme outliers or imputing missing data, could also introduce bias by altering the dataset's structure. Awareness of these biases is crucial for contextualizing the results and understanding the limitations of the analysis.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

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

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. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.