Measuring the Causal Effect of Data Capture on Retail Sales

MSBA 6440 - Section 070

Group 3
Michael Brucek, Kevin Grady, Anthony Meyers, Danny Moncada, Jack Quick

CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

I. Executive Summary

This analysis was designed in response to management's request for more information about how the new collection of promotional markdown data has impacted sales revenue.

The company began recording markdown information at 45 locations in early November of 2011 to leverage it during their price-setting strategies. Leadership’s goal for the data collection initiative was to give local managers a new tool for increasing weekly sales revenue and to broaden the corporate understanding of effective promotions at the regional level. Their assumption was that monitoring markdown efficacy would allow local managers to note trends in consumer response, then apply that knowledge to maximizing revenue.

Average sales revenue per week increased after markdown data collection began, but the increase cannot be definitively attributed to the change in data collection. Average weekly sales had been trending upwards long before the change. The company deployed the data collection strategy at all locations simultaneously, which prevents analysts from comparing sales between stores that made a change and those that did not.

Analysts retroactively devised an experimental model that allows management to explore the impact of data collection. Stores were grouped into four segments based on similar traits, such as weekly revenue, store size, regional economic indicators, and others. Each segment contained stores that were alike, and the segment profiles differed from each other in meaningful ways, such as size of store, CPI and sales volume. The segments allowed analysts to create a baseline group for comparison by combining three of the four segments into a single group, then weighting it to mirror pre-change behavior of the remaining original segment. The synthetic group served as a baseline expectation from which to compare against the original segment. A review of average weekly sales between the two groups from before data collection and after showed the impact of markdowns. Simply stated, researchers looked for any changes between the baseline group and the original segment. If sales were higher than the baseline group, it can be reasoned that data collection was the cause since all other aspects of the groups were identical.

The study found no meaningful evidence of increased weekly sales revenue as a result of markdown data collection. The baseline group and the original segment saw identical revenue trends after the change. This study only included 144 weeks of observation, only 50 of which included markdowns, which challenges the reliability of this conclusion. Lack of a pure control group is also problematic. It is recommended that the company work with analysts to design a more controlled study that provides clearer evidence of how markdowns are impacting sales.

Introduction / Question

Problem Statement

Did the company effectively leverage data assets from product markdown strategies to generate an accretive ROI in the post-treatment period?

Background

A large retailer provided sales data from forty-five anonymized store locations, covering a period of thirty-three months (February 5, 2010 - November 1, 2012) with a daily breakdown of sales across the departments for each store. In November 2011 the retailer started capturing markdown data for each store. Our goal was to determine whether a causal relationship could be proven between the collection of markdown data over this one-year treatment period, and an accretive ROI in the post-treatment period.

Does capturing data on markdowns enable managers to leverage meaningful insights into how sales (and potential sales) are impacted? By conducting this analysis, we want to determine if store managers for the large-scale retailer are using this data to enhance their decision making in the amount and frequency of sales.

What is a Markdown?

Markdown refers to the practice of lowering a product's price for the purpose of accumulating sales. Unlike promotions and discounts, the price of a product that has been marked down has been reduced *permanently*. The calculation involves subtracting the actual selling price from the original selling price. For example, a large-scale retailer might purchase jeans at $10 a pair and then offer them at their store for $15; however, if sales for that item turn out to be lackluster, the retailer can markdown the price to $11 to increase sales.

Markdowns can be used in a variety of ways to save companies from the problem of product selling, like slow-moving inventory or the inability to provide informed customer decisions. They are not always a last resort; marking down prices of specific products can aid in clearing stock to replace them with more popular items and cutting down the costs of items can make them more attractive to buyers looking for a good bargain. With more information available to consumers than ever before, an effective markdown strategy is extremely important for retailers looking to maintain a consistent cost of goods sold.

Importance of Causal Question

In today's world, with consumers accessing information about products every day and in real time, leveraging data analytics has never been more important for retailers. A Bain & Company research study of 400 large companies suggests that only **4**% of companies are "really good at analytics”. [Wegener and Sinha, 2013] The same study also found that companies that are leveraging data analytics effectively in their decision making are two times as likely to be in the **top** quartile of financial performance within their industry, three times more likely to execute business decisions as intended, and *five* times more likely to make decisions with increased speed and accuracy. [Wegener and Sinha, 2013]

When retailers miscalculate inventory and have products out-of-stock or sitting in warehouses unsold, it can be extremely costly to their bottom line. For example, in 2014 Walmart executives admitted that the company was leaving $3 billion a year because out-of-stock inventory [Fantini, 2019]. In 2018, Swedish clothing company H&M had managed to accumulate unsold clothing valued at $4 billion because of weaker than expected sales [Fantini, 2019]. Leveraging data and analytics teams effectively can drive extreme revenue generation. At the same time, measuring the efficacy of this data acquisition and analytics team is extremely important, as not leveraging the data effectively can create enormous costs to the business.

The temptation for many companies is to simply cut their prices to rid themselves of excess stock, but this solution brings additional complications to an effective sales strategy. Retailers and brands who markdown inventory too much or too frequently run the risk of conditioning customers to expect discounts and potentially devaluing their brand. A study by the National Retail Federation conducted in 2016 uncovered that one in three shoppers **only** purchased sale items during their holiday shopping. [National Retail Federation, 2016]

Every retail business should budget some dollars towards markdowns. There is no one formula to reference because the nature of the product determines the degree and frequency of markdowns. The more time or season sensitive the merchandise is, the faster and more frequent the markdowns. For example, clothing is a fashion and season sensitive product and typically has a four to six-week full margin selling period. After that the markdowns start because the retailer must move it out to make room for the next season’s merchandise.

How companies address the causal question

There are a few companies that are already leveraging emerging big data and data analytics technologies to help large scale retailers make more informed business decisions. For example, Evo Markdown [Fantini, 2019] is a company that offers software using big data and machine learning to help retailers find and maintain a balance on discounts for each product offering to help sales strategies be more effective. The goal of the software is collecting enough data to accurately predict consumer trends, adjust pricing strategy with improved accuracy, and remove a lot of the waste from unsold products sitting in inventory rooms.

Another example is Solum ESL [Solum blog, n.d.], a company that offers retailers electronic shelf labels (ESL), a modern form of traditional price tags that are equipped with features to automate most aspects of day-to-day operations for maintaining inventory. ESLs enable price tags for each product in a store to be connected to a central server so that retailers can electronically update price and information in real-time.

The dilemma of finding the most effective markdown strategy is becoming easier thanks to the rise of big data and AI. Retailers should always prepare a detailed plan for figuring out how and when they will mark items down to keep sales strategies consistent and ensure the highest likelihood of success. They should also use historical sales data and patterns to inform future decision making; it can provide retailers helpful insights and benchmarks to use when planning price markdowns. Using software like Evo Markdown and ESLs can help retailers harness big data and wield it to their advantage, keeping overstocks, out-of-stocks and excessive markdowns from unnecessarily ruining their profit margins.

II. Research Design & Data

The Ideal Experiment

Our experiment is crippled by a lack of data. One can assume that the company did in fact offer markdowns before November 11, 2011, but there is no data to suggest to what extent. Only after November 11, 2011, when the company began collecting markdown data, would we be able to examine the causal impact of offering markdowns on weekly sales. With a lack of pre-treatment period, we instead look at the impact of *collecting* markdown data, rather than the markdowns themselves.

In the ideal experiment, we would have access to a random sample of data collected over the 33-month observation period. Full markdown data at a department-level, or even item-level, would be provided detailing the original retail price, markdown price, and total markdown amount in dollars percent. The treatment, composed of products or clusters of products receiving a markdown, would be randomly assigned to half of the sample stores to act as the treatment group, while the remaining half will act as the control group. The control group will not collect any markdown data during the entire duration of the experiment. After a set amount of time, we will measure the change in weekly sales in the treatment group after the markdown data is captured and compare it to the change in weekly sales in the control group in the after-treatment period. If there is a causal effect in capturing markdown data, there will be a significant difference between the two differences.

Confounders to the Ideal Experiment

We are provided data on 45 stores out of roughly 8,000 US based locations [O'Connell, 2020], without insight as to how or why these stores were chosen, so any conclusions drawn from this experiment may not be generalizable to the overall population. We would ideally draw a large enough random sample of stores from the general population that would allow us to observe a significant difference in weekly sales after the treatment is applied.

Every store in the dataset began collecting markdown data at the exact same time and continued to collect markdown data for the entirety of the treatment period. There were no holdout stores in the dataset, that is, stores that did not collect markdown data during the observation period, effectively eliminating the possibility of a control group from the experiment. A lack of distinct control group makes it difficult to measure the causal impact of collecting markdown data on weekly sales, as there will be no baseline for sales had the larger retailer *not* begun collecting markdown data.

As previously mentioned, we would ideally have markdown data subcategorized by department or product, allowing us to discern at a store department or item level whether access to markdown data during the treatment period impacts sales.

Data

We are provided three data files detailing historical sales, store-level information and regional demographics for 45 large-scale retail stores. The data was split into three files:

- Stores.csv
  - Anonymized information about the 45 stores, indicating type and size of store.
- Sales.csv
  - Historical sales data by store and department, February 2010 - November 2012.
- Features.csv
  - Regional data for each store detailing temperature, fuel price, consumer price index, and unemployment for each week during the experiment.

The treatment in this experiment is defined as collecting markdown data. Each store in the experiment began collecting markdown data at the exact same time on November 11, 2011. The outcome variable for the experiment is defined as weekly sales. This is provided at a weekly level for each store and department, but we will eventually attempt to estimate weekly sales for distinct clusters of stores outlined in the Results & Conclusions section.

Experiment Components

- The unit of measurement are the 45 stores included in the dataset.
- Our treatment variable in this analysis is data capture of markdown sales for the post treatment period (weeks 93+)
- Our outcome variable is the weekly total sales by store type and cluster.

Threats to Causal Inference, and how they could be addressed

<u>Stability of the data generating process</u>: Given the potential for changes in leadership at the company over time, there is risk that the data generating process that collects markdown data

could change. Care will need to be taken by the company's Data Analytic team to ensure data is collected the same way over time so that the data collection process remains stable.

Omitted Variable Bias: While some demographic factors like CPI and fuel prices are included within this dataset, there are clearly omitted variables in this dataset. Significant markdown detail information is absent, such as lack of markdown information at a departmental level or by product. By not describing what products are being marked down or even describing the relative cost reduction, this leaves room for an enormous amount of omitted variable bias. By enhancing the data collection to include very detailed information around the products being marked down along with information around the relative decreases, more sophisticated models can be designed that leverage this level of detail.

Selection Bias: While inference from small sample sizes is possible, in this case 45 stores out of, roughly, 8,000 is not an adequate sample size to capture all the endogenous factors around economic forecasting. Economic forecasting can be driven by regional, even micro regional, factors that the current data collection process does not capture. With a more sophisticated data capture process, it would be advisable to extend the data collection to significantly more stores.

Simultaneity Bias: There is the potential for simultaneity bias, as well-performing stores are less likely to need to utilize markdown strategies to drive sales. Simultaneity can be addressed with improved data collection as well, as it will enable real time adjustments to markdown prices as the sales data streams in. Matching techniques becomes more viable as the data collection is expanded in breadth and improved in quality as well, so measurement of causal activity will become more reliable.

III. Results & Conclusions

Fixed Effects Model

With all stores beginning markdown data collection at the same exact point in time, November 11, 2011, there is no control group with which we can compare the effect of collecting markdown data. A series of fixed effect models were fit in order to account for confounders that are constant within a single store. The coefficients for the final effect model using a store fixed effect and the socioeconomic and geographic data to model total sales is provided in the table below. With all coefficients at the 0.05 significance level, the coefficient for the "Treatment" variable suggests that expected sales increase by $56,749.90 on average across all stores once they begin collecting markdown data.

Generally, the coefficients of the estimates pass the logical sense check:

- Sales increase as prices decrease.
- As temperature increases, relatively less sales occur because people want to be at the beach over in a store

- As other expenses become more expensive, fuel / CPI, sales decrease.

Feature	Coefficient
Treatment	$56,749.90
Temperature	-$825.312
Fuel Price	-$28,924.70
CPI	-$3,632.31
Unemployment	-$9,505.96

All significant at the 0.05 level

However, regression diagnostics of the fixed effects model suggest that the coefficients and interpretations should not be used due to several reasons. We first test for endogeneity in the fixed effect model by conducting a Durbin-Wu-Hausman test. The results of the Durbin-Wu-Hausman test are inconclusive at the 0.05 significance level, meaning we do not have enough evidence to conclude that there is no endogeneity present in our model.

A Breusch-Godfrey test for serial correlation suggests significant autocorrelation in the error term of the fixed effect model. This test suggests there is significant correlation between a store's sales for one week and the same store's sales after some amount of lag in time. The results of this test would suggest that there is some temporal presence that is not being accounted for in the model.

The error terms of the fixed effect regression model were tested for heteroskedasticity by conducting a Breusch-Pagan test. The Breusch-Pagan test suggests that there is significant heteroskedasticity present in the error term, meaning that as the predicted sales amount for a store increases, the magnitude of the error term will also increase.

Test	p-value
Endogeneity (*Hausman*)	0.9353
Serial Correlation (*Breusch-Godfrey*)	2.2e-16
Heteroskedasticity (*Breusch-Pagan*)	2.2e-16

Coefficient results should not be used

Synthetic Control Method

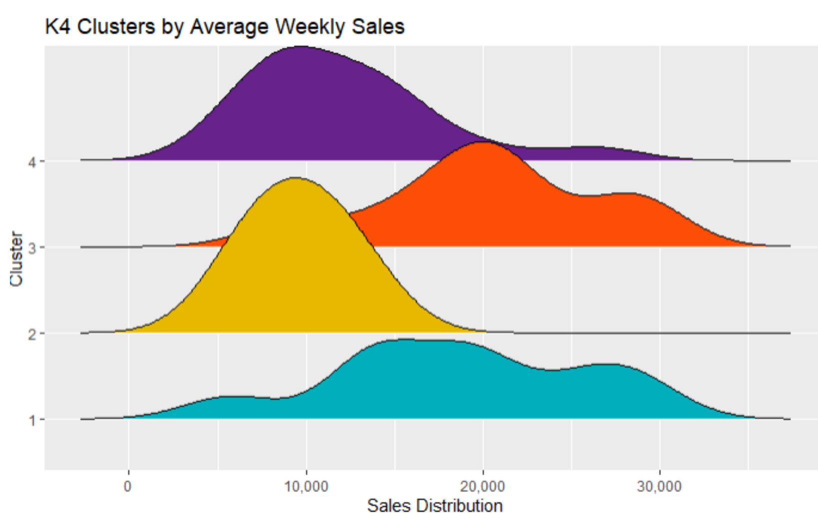
Due to the lack of a true control group in the experiment, it is difficult to examine the causal impact of capturing promotional data using the data provided. Using Synthetic Control Methods, we can synthesize how the weekly sales in these stores might have looked in the post-treatment period if there was no causal impact of capturing markdown data. Using the results of the clustering analysis outlined above as well as markdown data, predictions were generated on a cluster level for the post-treatment period.

In order to generate more accurate predictors and get away from looking at store-level information, stores were clustered using K-Means algorithm to control for unobserved confounders not present in the data such as:

- Regional shopping habits
- Regional cost levels
- Regional distribution of stores
- Microeconomic considerations

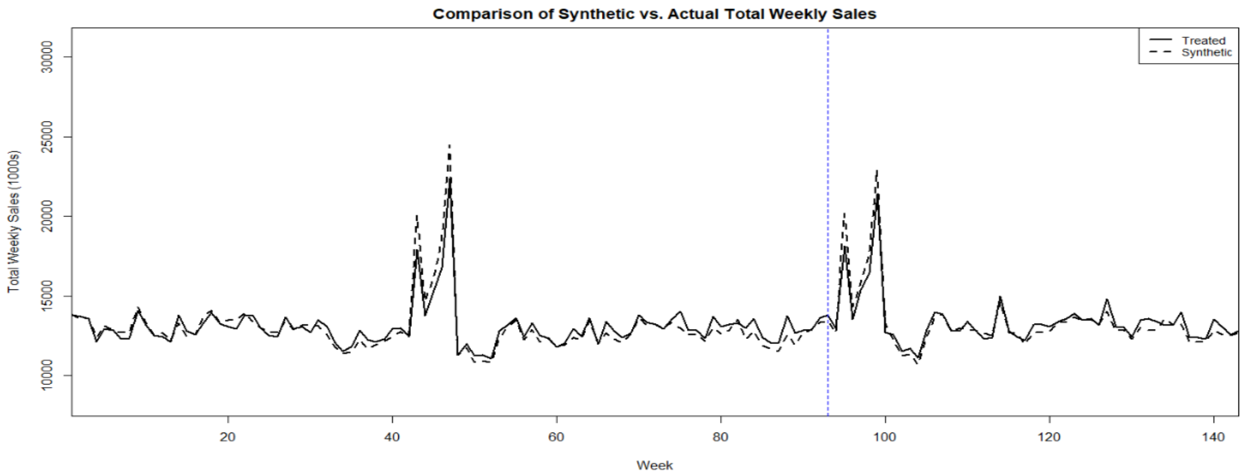
Store-specific features were used to create four distinct store profiles. These features were aggregated at an average or total level for a store over the entire time period and normalized before being included in the clustering solution. The features are as follows:

- Size
- Temperature
- Fuel price
- Consumer Price Index (CPI)
- Unemployment
- Weekly sales
- Department number



Four distinct clusters identified

The predictions generated using the Synthetic Control Method generated good fit to actual data, with a mean prediction error around 4%. As a result, Synthetic Control Methods show no significant increase in weekly sales after promotional data is captured.



MPE: 542,822

Conclusion

Despite implementing several methods to gauge collecting markdown data's effect on weekly sales, no causal relationship was found between simply capturing data and an increase in sales. As exhibited by the statistical testing that was performed, there is simply too much endogeneity left within the data to perform any reliable causal inference. Additionally, it may simply be the fact that the management teams have not yet acted on the data science teams recommendations.

Many of the solutions to addressing the threats to causal inference should be considered in order to aid future analysis of the markdown data to generate effective sales techniques. This retailer should also focus upon incorporating markdown data from significantly more stores than the 45 locations presented in this dataset.

Furthermore, inclusion of markdown data at a store department level could lead to additional insights to help shape markdown strategy effectiveness. Additional experiments should be geared towards altering specific markdown strategies in ways that allow for causal impacts to be more easily measured. Costs of acquiring, retaining and analyzing more robust store-level promotional data should be considered for future experiments given the cost potential of acquiring this data as well as funding analytics teams.

IV. Appendix

A. Clustering Analysis (see attached PDF)
B. Causal Effect of Data Capture on Retail Sales (see attached PDF)

V. References

[1] Fantini, Fabrizio. (2019, January 31). A better approach to retail markdown strategy. *LinkedIn*. Retrieved from https://www.linkedin.com/pulse/better-approach-retail-markdown-strategy-fabrizio-fantini

[2] How Retailers Can Create a Markdown Strategy. (n.d.). Retrieved from https://solumesl.com/how-retailers-can-create-a-markdown-strategy/

[3] O'Connell, L. (2020, April 2). Total number of Walmart stores worldwide 2008 - 2020. *Statista.* Retrieved from https://www.statista.com/statistics/256172/total-number-of-walmart-stores-worldwide/

[4] Retailers made Black Friday irresistible for consumers with great deals, online and in-store. (2016, November 27). *National Retail Federation*. Retrieved from https://nrf.com/media-center/press-releases/retailers-made-black-friday-irresistible-consumers-great-deals-online

[5] Rosenblum, Paula. (2014, April 15). Walmart's Out of Stock Problem: Only Half the Story? *Forbes*. Retrieved from https://www.forbes.com/sites/paularosenblum/2014/04/15/walmarts-out-of-stock-problem-only-half-the-story/#36cfbc657369

[6] Wegener, R. and Sinha, V. (2013, September 17). The value of Big Data: How analytics differentiates winners. *Bain & Company*. Retrieved from https://www.bain.com/insights/the-value-of-big-data