The Effect of Amazon's Acquisition of Whole Foods

Backwards Design Assignment: Causal Inference

Team 3
Caleb, Junbo, Nikhil, Rhayoung, Xiaohan

1 Topic

What is your project about? What problem are you seeking to solve, or in which domain do you think you can contribute meaningfully?

In 1978, John Mackey started the first Whole Foods in Austin Texas and grew the business by merging with other local grocery stores. In recent times, however, the business went through multiple hurdles. Competitors started providing cheaper organic food and more players were entering the market. Walmart became the largest organic food provider in the states and investors were concerned that Whole Foods might fall behind the competition. In June 2017, Amazon acquired the business for $13.7 billion, showing its ambition towards the offline grocery business. Amazon immediately reduced prices, started selling one of its best-selling products, Amazon Echo, integrated Amazon Prime Membership, and rolled out a delivery service. After being acquired by Amazon, the revenue growth of Whole Foods increased tremendously.

2 Project Question

What specific question are you seeking to answer with this project? For this project, this must be a causal question.

For our backward design project, we aim to seek the effect of Amazon's acquisition of Whole Food in 2017. Our team targets to quantify the effect of Amazon's acquisition of Whole Foods by comparing the financials with other grocery stores that did not go through the treatment effect. This study can shed light on the business effect of acquisitions in the grocery industry.

3 Ideal Experiment

If you were a god, what experiment would you run to answer your question? Define both your treatment variable, and your outcome of interest.

Ideally we would conduct the experiment to maximize both internal as well as external validity. In our experiment, the treatment variable is the effect of Amazon's acquisition of Whole Foods Market, and our outcome of interest is the revenue of Whole Foods Market.

If we were gods, we would like to create a parallel universe just like the one we live in currently. However, in the parallel universe, all the grocery stores have the same market share and revenue as they do in the real world. The two universes would also develop at the same rate. The only difference would be that in the parallel universe, Amazon did not acquire Whole Food Market or any other grocery stores. I would also eliminate all the potential factors that would influence the revenue of Whole Food Market, including but not limited to the COVID-19 pandemic and underlying consumer trends. Since the pandemic would influence people's financial status and behavior, therefore influence the revenue of Whole Food Market. Then we will record the quarterly revenues of Whole Food Market from 2017 (when Amazon acquired Whole Food Market in the real world) to 2027 in both universes and compre the difference. The difference in revenues would be the causal inference of Amazon's acquisition of Whole Foods Market.

4 Pick a Study Context

Where can you get data that (a) measures your outcome variable, and (b) includes variation in your treatment variable?

For this project, the outcome variable is the revenue of Whole Foods before and after it's acquisition by Amazon in 2017. This information is publicly available through the company's quarterly 10k. Other grocer's that are similar to Whole Foods and are public also post their quarterly revenues in a similar fashion. Variations in the treatment variable such as the number of stores the company operates, percent of sales through digital mediums, areas of operation, wholesale operations, and subscription based are variables that can be collected manually through research online or through the 10ks discussed earlier.

5 Project Design

Given the context you want to study (and data you can find), what design do you think would be feasible?

In order to study the effect that Amazon's acquisition had on Whole Foods revenue, we believe that a difference in difference analysis makes the most sense. This is time series data that we will be looking at from 2009 - 2019/20 on a quarterly basis. Whole foods is a grocer, and given that there are a number of publicly traded grocery stores in the United States, we anticipate that we will have a number of relevant and reasonable comparisons, which is a key component in an effective difference in difference.

There is a clear delineation in time when we can say the treatment effect took place. Amazon's acquisition of Whole Foods was announced in June 2017, which is in the midst of fiscal quarter 2. We will leave Q2 2017 as a gap between when the effect took place, and look at all grocery financials up until Q1 2017, and from Q3 2017 onwards. Our hope is that trends pre-acquisition will look fairly similar between Whole Foods and our comparisons, and that after the acquisition we see a clear shift in revenue trendline for Whole Foods.

6 Model Results

The strategy used for causal inference analysis is difference-in-difference. We will compare the pre and post-treatment difference in revenue of Whole Foods to that of other grocery stores that have not been treated (ie. other grocery stores that have not been acquired). The difference-in-differences estimator $\hat{\delta}$ is defined as:

$$\hat{\delta} = (\bar{y}_{T=1,\,post} - \bar{y}_{T=1,\,pre}) - (\bar{y}_{T=0,\,post} - \bar{y}_{T=0,\,pre}),$$

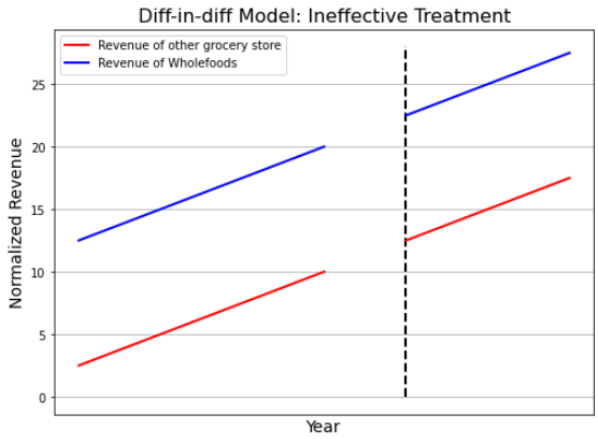
where $y$is the average revenue. If the acquisition by Amazon had a positive causal effect on Whole Foods' revenue, $\hat{\delta}$ would be positive; if the acquisition had a negative causal effect, $\hat{\delta}$ would be around 0 or negative.

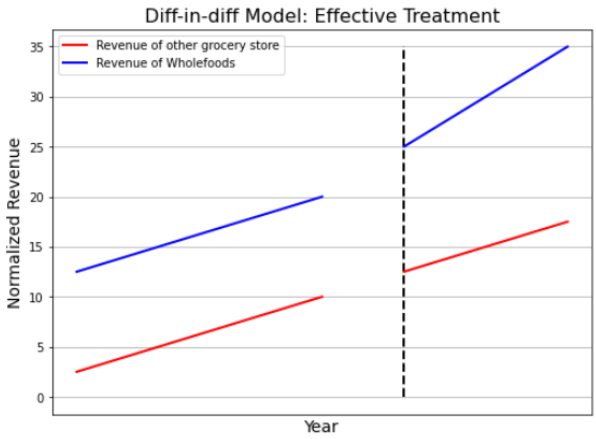
Difference-in-difference regression would also be adopted to aid the analysis. The regression formula is defined as:

$$Y = \alpha + \psi_c + \beta_1 * post\,treatment + \beta_2 * post\,treatment * store,$$

where Y is the revenue of a store, ψ_c are fixed effects, $post\ treatment$ is an indicator for whether we are in a period after the acquisition, and $store$ is an indicator for whether an observation is in the treatment group (ie. is Whole Foods). Our difference-in-difference estimate would be the coefficient β_2. Still, β_2 would be positive if the acquisition has a good influence on revenue.

Additionally, regression plots would be adopted. As the figure below shows, when the acquisition exerts no causal effect, Whole Foods' post-treatment revenue trend would be similar to that of other grocery stores. On the contrary, when the acquisition exerts a causal effect, Wholefoods' post-treatment revenue trend would show a steeper slope.





7 Final Variables Required

Now that you've specified what an answer to your question looks like, what data do you need to generate that answer?

Since we're interested in finding out whether or not amazon's acquisition of wholefoods had a positive impact on Whole Foods' revenue in a difference in differences analysis, we will need to have the quarterly revenue of Whole Foods and comparable grocery chains before and after 2017 (time of acquisition). More specifically, the following grocery chains' have been identified to be possible comparisons to Whole Foods, but also with publicly available revenue data: Costco, Walmart (Grocery Only), Sam's Club (Grocery Only), Target (Grocery Only), SuperValu, and Sprouts.

Besides the response, additional control variables such as the number of grocery stores per chain, digitalization metrics, countries and/or American regions of operation, if it's a wholesale retailer, and whether or not a grocery chain is subscription will need to be collected.

8 Data Sources

Finally, given the variables you need for your analysis, what actual data sources do you think will you have the data you need?

Since the grocery stores mentioned earlier are publicly traded, each company's respective 10k has quarterly revenues with respect to their grocery sales. Additionally, variables such as the number of stores and percentage of digital sales are available in each 10k. Other control variables such as whether or not the company is a wholesale retailer or needs a subscription can be manually created based on research. The dataset as whole will likely need to be generated by manually collecting information from these various sources.