Causal Inference and Econometrics - Measuring ROI on Sponsored Search Ads - Bazaar Company

Business Overview

Bazaar.com, a prominent online retailer in the United States, has established a significant presence in digital advertising. This includes a strong focus on display advertising as well as search engine advertising, with a particular emphasis on running paid search ads across platforms like Google and Bing. The paid advertisements by Bazaar are categorized primarily into two distinct groups based on the keywords used: branded and non-branded. Branded keywords incorporate the 'Bazaar' brand name, examples of which include 'Bazaar', 'Bazaar shoes', 'Bazaar clothes', and similar phrases. In contrast, non-branded keywords are those like 'shoes', 'dress', and other terms that do not feature the 'Bazaar' brand name.

Bazaar.com conducted an experiment involving a paid campaign with branded keywords, specifically those containing the word 'Bazaar'. The aim was to estimate the number of customers visiting their website through both sponsored and organic advertisement links. However, during week 10 of the Google campaign, a technical glitch occurred, preventing the capture of customer traffic data through sponsored advertisements. As a result, the team used data from weeks 1 to 9 to calculate the ROI for the sponsored ads.

Regarding the traffic data from Google and Bing, Bob, a member of Bazaar's marketing analytics team, calculated a 320% Return on Investment (ROI) on the company's expenditure for sponsored ads. However, there is skepticism surrounding his result. The main concern arises from the fact that individuals searching with the term 'Bazaar' are likely already inclined to visit Bazaar.com. This casts doubt on the actual effectiveness of the branded keyword ads.

However, Myra, one of the executives, raised concerns about potential overestimation in the ROI calculations. She posited that customers already searching for 'Bazaar' were naturally inclined to visit their website, regardless of whether they were exposed to sponsored ads or not. In her view, these customers would likely find their way to the site via organic links even in the absence of sponsored ads. This led to a decision to analyze the campaign data to assess the real impact of sponsored ads on web traffic. They planned to compare the total web traffic from both sponsored and organic advertisements before and after the week 10 glitch — essentially, up to week 9 and the subsequent period.

The primary objective now is to decipher the true causal effect of these search ads on business outcomes. To achieve a comprehensive understanding, we will conduct an in-depth analysis by addressing the following key questions:

1. What's wrong with Bob's ROI analysis?
2. What is the treatment and control of the experiment?
3. Is the First Difference reliable to estimate the causal effect of the treatment?
4. How should we compare with Difference-in-Difference estimation or Pre-Post estimation?

5. Based on our new treatment effect, what should be the corrected ROI? How is it compared with Bob's ROI?

To accurately gauge the effect of the absence of sponsored ads (due to the glitch) on the total number of advertisement clicks across various platforms, the team employed a 'difference in difference' approach. This method was chosen to more precisely estimate the actual ROI of the sponsored advertisements

Executive Summary

Diving deep into Bob's ROI calculation, we discovered two flaws in the calculation that make the figure inaccurate and not presentative - 1) Overvaluation of revenue and 2) Unaccounted opportunity cost

Overvaluation of revenue

The current ROI calculation assumes that everyone who clicked on the sponsored ads is casually driven by the sponsored ads, the error suggests the portion of people who have searched by the branded keywords already has the intent to visit Barzaar.com, this portion of people logically is not driven by the sponsored ads. Instead, they would still visit the website via organic link no matter if the sponsored ads existed or not, so they should not be included in the ROI calculation.

Therefore, if Bob included 'branded keyword search' clicks from these users visiting through the sponsored ads to be causal, his ROI calculation would have overestimated the revenue brought by the sponsored ads. We found that actually only 81% of total traffic via sponsored ads was a true result of displaying a sponsored ad.

Opportunity cost is considered as excluded

From our analysis, we discerned that only 81% of the traffic to Bazaar.com is genuinely driven by sponsored ads. The company is incurring $0.6 per click costs for the remaining 19% of traffic from branded keyword searches, which would still arrive via organic links without sponsored ads. This represents an opportunity cost, as these funds could be allocated to other marketing efforts.

To accurately assess the impact of sponsored ads, we employed a Difference-in-Difference approach with these steps:

We calculated the initial difference in weekly average traffic (combining Ads and Organic) before and after the glitch. This provided the immediate effect of Google's sponsored ads, rather than the overall treatment effect.

We compared this initial difference for Google with the same for other search engines. This comparison allowed us to isolate the true incremental effect, accounting for external factors like seasonal changes and market shifts.

Our findings show an average loss of 9.9K clicks per week without sponsored ads, accounting for about 81% of weekly traffic from these ads. We concluded that the remaining 19% of traffic comes through organic searches, and this portion should be excluded from the original ROI calculations.

Taking these factors into account, we recalculated the ROI for Bazaar.com's sponsored branded ads. After removing the overestimated revenue from sponsored ads and considering the adjusted opportunity costs, the previously calculated ROI of 320% is now a more accurate 241%. This revised figure still signifies a substantial return on investment for sponsored ad campaigns.

Experimentation

We used 12 weeks of data including different platforms(Google, Bing, Yahoo, Ask), average sponsored branded keywords, and average organic search results. The technical glitch occurred on week 10-12, leading to no sponsored ads on Google platform. On the other hand, the other platforms run the keyword ads uninterruptedly for 12 weeks. Given the situation, we want to understand the causal effect of the ads. Treatment is the interrupted sponsored ads in Google during week 10-12. Google acts as the treatment group and other search engines such as Bing, Yahoo, and Ask act as the control group.

Threats to Causal Inference Establishment

Selection Bias: For an experiment to be free from selection bias, the participants should be comparable to each other and the larger population they represent. In our scenario, since the advertising strategies, keyword bidding, and customer profiles are consistent across Google and other search engines, we can reasonably assume that our experimental design doesn't introduce selection bias.

Omitted Variable Bias: There's a possibility that unaccounted external factors could influence our experiment's outcomes. For instance, Google's search engine might experience unrelated technical issues, or Bazaar's rivals might launch an intensive short-term marketing campaign on a specific search engine, both of which are not included in our dataset.

Simultaneity Bias: Our experiment might not fully consider the two-way interaction between the variables. Specifically, we've looked at how ad clicks impact website visits, but the reverse - website visits influencing ad clicks - is also plausible. Customers often compare Bazaar products with competitors before making a purchase decision, which could lead to more ad clicks.

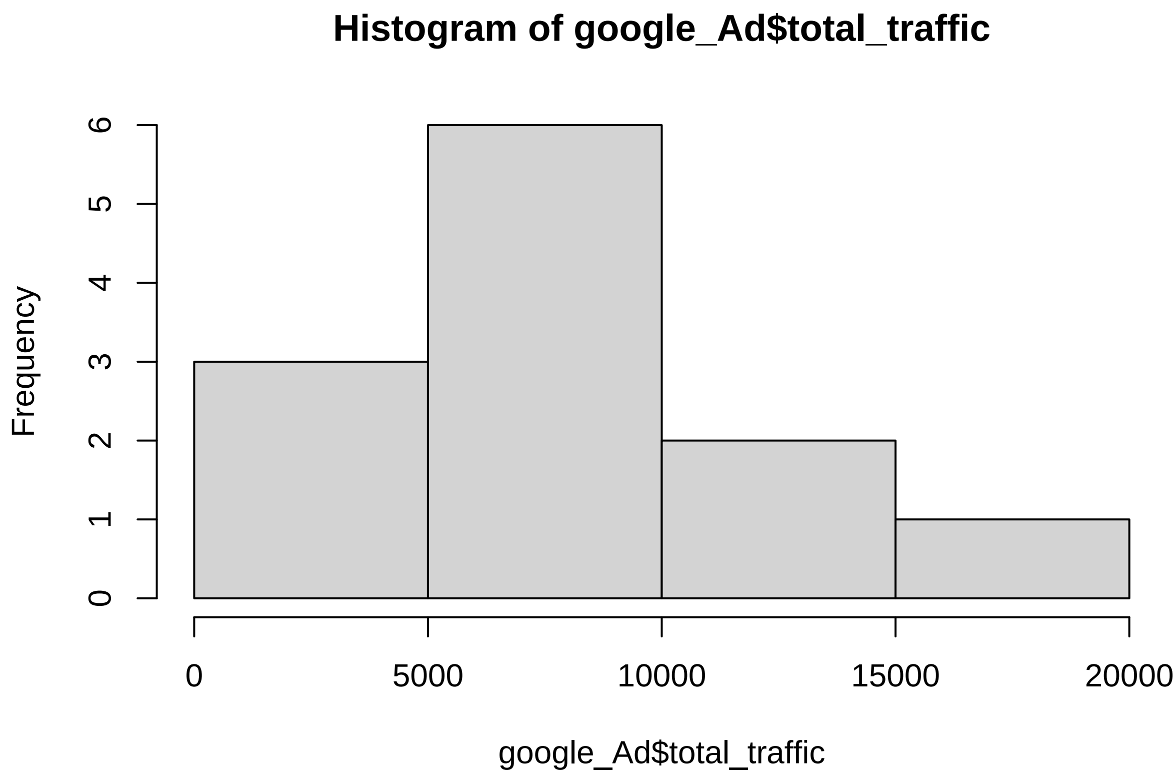
Measurement Error: The accuracy of our click-through data for sponsored ads might be questionable. There's a chance that consumers click on an ad accidentally and then quickly exit the landing page, which wouldn't be a true reflection of ad engagement.

Data Overview

The data set is weekly average traffic data through four platforms for 12 weeks. After importing the data, we created a "treatment" variable for test vs. control groups, an "after" variable that indicates 1 for treatment weeks and 0 for pre-treatment weeks and a "total_traffic" variable to aggregate the sponsored and organic traffic.

```
setwd ("/Users/Trung/Downloads")
data = read.csv('did_sponsored_ads.csv')
treatment_week = c(10,11,12)
data <- data %>% mutate(treatment = ifelse(platform == 'goog',1,0),
                       after = ifelse(week %in% treatment_week,1,0),
                       total_traffic = avg_spons + avg_org)


google_Ad = data %>% filter (platform=="goog")
#Plotting the histogram for total clicks on Google
hist(google_Ad$total_traffic)
```



What is Wrong with Current RoI Calculation

Sponsored advertisements play a crucial role in the survival and growth of businesses, often accounting for a significant portion of their expenses. However, accurately assessing the effectiveness of ad spending is a complex challenge. We identified that the primary issue in Bob's ROI calculation is the inflation of revenue estimates. Specifically, Bob failed to account for a key behavioral aspect: not all customers who click on sponsored ads are influenced by these ads to visit Bazaar's website and make a purchase. Some would have entered the site through organic links regardless of the presence of sponsored ads, meaning the revenue they generate isn't actually attributable to the ads. Consequently, treating all clicks from sponsored ads as direct contributors to sales leads to an inflated estimation of revenue and ROI.

Illustrative Example: Consider a scenario where 100 users reach Bazaar through sponsored ads. Initially, Bob estimated the ROI based on all 100 users, assuming an average purchase probability of 12%, an average margin per customer of $21, and an average cost per click of $0.6.

ROI = (($21 * 0.12 * 100) - (100 * $0.6)) / (100 * $0.6) = 320%

```
ROI = (((21 * 0.12 * 100) - (100*0.6) )/ (100 * 0.6))*100
ROI
```

```
## [1] 320
```

However, in reality, let's assume only 40 of these visits were genuinely influenced by the ads. The adjusted ROI would then be calculated as follows:

```
Adj_ROI = (((21 * 0.12 * 40) - (100* 0.6)) / (100 * 0.6))*100
Adj_ROI
```

```
## [1] 68
```

Define the Treatment and Control

As per this experiment, stoppage of sponsored ads serves as a way to test the causality, hence in our analysis (say in difference in differences we would be testing difference of change in Google click through ads vs other platforms) Google would serve as treatment group (as there was impact of outage here) and Bing , yahoo and ask would serve as control groups (as there was no impact of outage)

Consider a First Difference Estimate

In our analytical approach, we focus on determining whether there's a significant change in total web traffic following the ad outage compared to the period before the outage. This comparison of pre- and post-outage data is crucial for gaining insights into the causal impact of the outage on website traffic. Given that the data is heavily skewed — with 75% of the data showing fewer than 10,000 clicks — we employ a log transformation. This transformation is a standard technique to normalize data, making it more suitable for model building. By using a log transform, we can mitigate the effects of the skewness in the data, leading to more reliable and interpretable analytical results

```
# Try a simple pre-post estimator
# Simple pre-post estimator
google_data<- data %>% filter(platform == "goog")
model <- lm(log(total_traffic) ~ after, data = google_data)
summary(model)
```

```
##
## Call:
## lm(formula = log(total_traffic) ~ after, data = google_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.54933 -0.15495  0.03784  0.46975  0.95834
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.783506   0.248968  35.280 7.94e-12 ***
## after       0.001306   0.497936   0.003    0.998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7469 on 10 degrees of freedom
## Multiple R-squared:  6.88e-07,   Adjusted R-squared:   -0.1
## F-statistic: 6.88e-06 on 1 and 10 DF,  p-value: 0.998
```

Interpretation:

The p-value of 0.998 suggests no significant insights from our test. While there appears to be a 13.06% increase in web traffic from Google, the high p-value indicates this is not statistically significant. Our current analysis, based on pre-post comparison within the treatment group, assumes constant market conditions, which might not hold true in scenarios like holiday seasons. To overcome these limitations, we propose using the Difference in Difference (DiD) method. DiD considers both treatment and control groups, offering a more reliable approach by accounting for external variations.

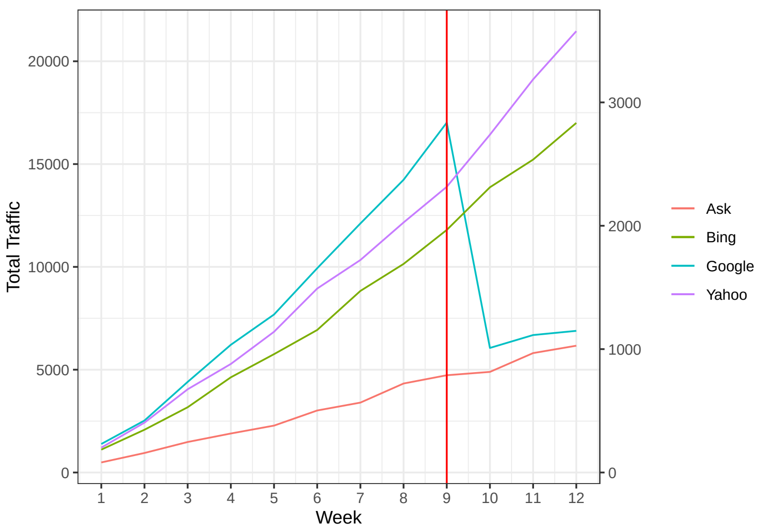
Calculate the Difference-in-Differences

Prior to conducting a Difference-in-Differences analysis, it's crucial to verify the existence of parallel trends beforehand. This step ensures that the differences we're comparing are, in fact, meaningful. If there's already a decreasing trend between the differences, it would hinder our ability to derive accurate insights.

Visualization of parallel trend

```
temp1 = data %>%  filter(platform %in%  c('bing')) %>% select(week, total_traffic)
temp2 = data %>%  filter(platform %in%  c('yahoo')) %>% select(week, total_traffic)
temp3 = data %>%  filter(platform %in%  c('ask')) %>% select(week, total_traffic)

ggplot(data %>% filter(platform == 'goog'), aes(x=week, y= total_traffic, color = 'Google')) +
  geom_line() +
  geom_line(aes(x=week, y= total_traffic, color = 'Bing'), data = temp1) +
  geom_line(aes(x=week, y= total_traffic, color = 'Yahoo'), data = temp2) +
  geom_line(aes(x=week, y= total_traffic, color = 'Ask'), data = temp3) +
  geom_vline(xintercept = 9,color='red') +
  scale_y_continuous(sec.axis = sec_axis(~./6)) +
  scale_x_continuous(breaks = seq(1, 12, by = 1)) +
  labs(y = "Total Traffic", x = "Week") +
  theme_bw() +
  theme(legend.title = element_blank())
```



The graph shows no parallel trends, but rather an increasing divergence until week 9, followed by a decrease in Google's click-through numbers compared to other platforms, indicating convergence. Therefore, while DiD analysis remains valuable, it requires cautious application and careful interpretation of the results.

```
model_did <- lm(total_traffic ~ treatment + factor(week) + treatment * factor(week),data=data)
summary(model_did)
```

```
##
## Call:
## lm(formula = total_traffic ~ treatment + factor(week) + treatment *
##     factor(week), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -8710.7  -111.8    87.3  1422.3  6586.3
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  936.3     2465.2   0.380 0.707414
## treatment                    449.7     4930.3   0.091 0.928087
## factor(week)2                881.3     3486.3   0.253 0.802574
## factor(week)3               1964.7     3486.3   0.564 0.578291
## factor(week)4               2998.3     3486.3   0.860 0.398274
## factor(week)5               4023.3     3486.3   1.154 0.259840
## factor(week)6               5361.0     3486.3   1.538 0.137190
## factor(week)7               6584.7     3486.3   1.889 0.071069 .
## factor(week)8               7940.0     3486.3   2.278 0.031955 *
## factor(week)9               9204.3     3486.3   2.640 0.014337 *
## factor(week)10             10794.3     3486.3   3.096 0.004932 **
## factor(week)11             12445.3     3486.3   3.570 0.001550 **
## factor(week)12             13940.3     3486.3   3.999 0.000529 ***
## treatment:factor(week)2      259.7     6972.5   0.037 0.970600
## treatment:factor(week)3     1055.3     6972.5   0.151 0.880960
## treatment:factor(week)4     1826.7     6972.5   0.262 0.795571
## treatment:factor(week)5     2274.7     6972.5   0.326 0.747075
## treatment:factor(week)6     3187.0     6972.5   0.457 0.651723
## treatment:factor(week)7     4140.3     6972.5   0.594 0.558196
## treatment:factor(week)8     4909.0     6972.5   0.704 0.488177
## treatment:factor(week)9     6424.7     6972.5   0.921 0.365997
## treatment:factor(week)10   -6122.3     6972.5  -0.878 0.388613
## treatment:factor(week)11   -7146.3     6972.5  -1.025 0.315616
## treatment:factor(week)12   -8437.3     6972.5  -1.210 0.238030
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4270 on 24 degrees of freedom
## Multiple R-squared:  0.6819, Adjusted R-squared:  0.3771
## F-statistic: 2.237 on 23 and 24 DF,  p-value: 0.0278
```

Although our initial assumption was not confirmed, we proceeded with a Difference in Difference (DiD) regression between the treatment and control groups. This approach aims to estimate the actual causal impact of the sponsored ads. The key independent variables in the DiD regression are Treatment, After, and the Interaction between Treatment and After.

```r
did <- lm(total_traffic ~ treatment + after + treatment * after, data=data)
summary(did)
```

```
##
## Call:
## lm(formula = total_traffic ~ treatment + after + treatment *
##     after, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8437.7 -3231.0  -510.5  3591.6  8630.0
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5265.0      882.5   5.966 3.79e-07 ***
```

```
## treatment         3124.9     1765.0   1.770  0.08357 .
## after             8064.7     1765.0   4.569 3.94e-05 ***
## treatment:after  -9910.6     3530.0  -2.808  0.00741 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4586 on 44 degrees of freedom
## Multiple R-squared:  0.3274, Adjusted R-squared:  0.2816
## F-statistic: 7.141 on 3 and 44 DF,  p-value: 0.0005211
```

By stopping sponsored ads on Google, Bazaar experiences an average loss of 9,910 clicks per week. This finding, derived from comparing the new treatment effect with the control group, highlights the causal influence of sponsored ads. This method is superior to the pre-post estimate because it allows us to analyze the behaviors of both control and treatment groups in a single model, adjusting for time-related variations and minimizing the impact of seasonality

Fix RoI Calculation, given New Treatment Effect Estimate

Since sponsored ads were turned off on google (treatment) from week 10 till 12 ("after" period from our data set) portion of users that would actually click through sponsored ads would have gone to organic search results (as they already have intent to use bazaar), for these users sponsored ads aren't effective, they just add cost, hence we need not consider them while calculating revenue. We can get these average numbers from below model

To compute the correct ROI, we needed a more accurate estimate of the proportion who were causally driven by the sponsored ads. In the previous part, we already calculated the total traffic causally driven by the ads, this step we needed to determine the traffic that would have used organic search results, granted that the sponsored ads were not running anymore, to arrive at the proportion of traffic that was causally driven by ads.

```r
did_organic <- lm(avg_org ~ treatment + after + treatment * after, data=data)
summary(did_organic)
```

```
##
## Call:
## lm(formula = avg_org ~ treatment + after + treatment * after,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1928.78  -847.92   -52.67   825.00  2067.33
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1489.7      215.4   6.917 1.51e-08 ***
## treatment          777.0      430.7   1.804   0.0781 .
## after             1984.1      430.7   4.607 3.49e-05 ***
## treatment:after   2293.2      861.4   2.662   0.0108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1119 on 44 degrees of freedom
## Multiple R-squared:  0.6043, Adjusted R-squared:  0.5773
## F-statistic:  22.4 on 3 and 44 DF,  p-value: 5.881e-09
```

In the absence of running sponsored ads on Google, 2293 would have used organic search results to visit

Bazaar.com.
Total clicks from sponsored ads (C) = clicks truly motivated by sponsored ads (A) + clicks by customers who would still visit Bazaar.com in the absence of sponsored ads (B)

A = 9,910 (new treatment effect in did) B = 2,293 (new treatment effect in did_org)

The model reveals that the interaction coefficient of 'After' and 'Treatment' is 2293, suggesting an average increase of 2293 clicks per week on Google after sponsored ads were discontinued. Previously, we noted a decrease of 9910 clicks when ads were stopped, implying that out of the total 12203 clicks (9910 + 2293), 2293 clicks were naturally occurring via organic search and didn't benefit from sponsored ads. This accounts for 81.2% of the total. Calculating ROI with these figures gives us:

```
New_ROI = ((21 * 0.12 * 0.8120954 - 0.6)/0.6)*100
New_ROI
```

```
## [1] 241.0801
```

Thus, the adjusted ROI is 241%, significantly lower than the initial 320% but still substantial. Our recommendation is to continue with sponsored ads based on these ROI calculations. This approach not only maintains profitability but also prevents competitors from capturing our intended customer base, especially if they initiate similar campaigns."