Assignment 2 :Measuring RoI on Sponsored Search Ads

Yufan Li, Zecong Zhao, Priyanka Bhosale, Manish Jain

Contents

Executive Summary 1
Introduction 2
Experimental Design 4
Analysis 5
Conclusion 12

Executive Summary

Bazaar is a leading online retailer in the United States. In their online marketing efforts, Bazaar uses both display advertising and search engine advertising, running paid search ads on major search engines, Google, Bing, Yahoo, and Ask.Bazaar's senior marketing analytics team wants to review the ROI do their marketing campaigns. Bob, provides an estimate of 320% ROI on the campaign. This ignores the endogeneity and does not consider the real impact of sponsored ads on website traffic.

Due to a technical glitch, there a natural experiment occurred where sponsored ads were not shown on one of the platforms. This presented an opportunity to use causal inference to estimate the real impact of running sponsored ads. Considering Google as the test group and other platforms as control, the treatment of not having sponsored ads was used. Using in

Difference-in-Difference the real impact of sponsored ads was nearly 67% of the total clicks.

Based on the estimate of average treatment effect, the real ROI on sponsored ads is 182.45%.It shows that the sponsored ads are in-fact effective. Thus it is recommended that Bazaar continue sponsored ads on branded keywords.

Introduction

Bazaar is a leading online retailer in the United States. In their online marketing efforts, Bazaar uses both display advertising and search engine advertising, running paid search ads on major search engines, Google, Bing, Yahoo, and Ask. Just as any search engine marketing implementation, Bazaar releases its ads in response to keywords used by online customers.These keywords are classified into two categories, Branded and Non-branded.

Bazaar's senior marketing analytics team want to reviewing recent marketing reports relating to paid search efforts. Specifically, they want to take a deeper dive into ROI calculations on the sponsored search ads for keywords on Google. The team looks at the sponsored ad campaigns that were supposed to be run by 12 weeks on all 4 platforms. However, due to a technical glitch, the campaign only ran for the first 9 weeks on Google.

Looking at the data from this campaign, they receive 20% of the traffic for the branded keywords through the sponsored ads they run on Google. As the majority of population clicks on organic ads as the customer who searches with a branded keyword is likely to click on the Bazaar.com website with or without the sponsored ad, the team is curious to see if running sponsored ads on branded keywords makes sense or not.

There initial analysis, presented by Bob, concluded an amazing ROI of 320%. He used the average cost per click for a sponsored ad at $0.60. Once a customer lands on the platform, their probability of making a purchase is 12% and the average margin per conversion is $21. Thus the average revenue per click is 0.12 x $21, or $2.52, which implies an ROI of ($2.52 - $0.60)/($0.60), or 320%.

(Question A) What is Wrong with Bob's RoI Calculation?

The biggest problem is that Bob ignores endogeneity and does not consider the real impact of sponsored ads on website traffic. In the absence of sponsored ads, the customers would click on organic links as they show a inherent affinity to make a purchase with Bazaar considering a branded keyword search. Therefore, the real effect of sponsored ads is the incremental clicks over the clicks obtained from organic link when running the sponsored ads.

Examining this from the perspective of the above ROI calculation,

ROI = Net Return on Investment / Cost of Investment $\times$ 100%

Net Return on Investment = (Total Sponsored ads Clicks from Google * % of incremental clicks from Sponsored ads * Conversion rate from clicks * average margin per conversion)

Cost of Investment = (Total Sponsored ads Clicks from Google * Cost per Click)

As the conversion rate and margin per conversion is on total traffic after it enters the platform, Bob did not account for the real incremental effect of sponsored ads over the organic clicks.Thus overestimating the number of actual clicks gained from sponsored ads, ultimately leading to the overestimated ROI.

Besides, according to the estimated ROI of Google Ads shared by Google, the average number should be 100%, which makes Bob's ROI of 320% hardly convincing [1].

We need to estimate the true impact of sponsored ads in-order to accurately calculate the real ROI. In the particular scenario above, for the last three weeks, the Google sponsored ads have not run due to a glitch in the system. However, the campaign was successful on the other 3 platforms. This has presented a natural experiment for us to compare the true impact of Google ads.

Experimental Design

This was a natural experiment with the treatment considered as the technical glitch of sponsored ads not running on the Google platform for the last 3 weeks of the sponsored ads campaign.

(Question B) Define the Treatment and Control.

As the campaign ran successfully on all the 3 other platforms, they are part of the control group. Unit of measurement in our analysis is the total clicks from a platform in any given week. We take total clicks as due to treatment on Google in the last 3 weeks, traffic may have redirected to organic links and hence should be considered in our analysis.

In our analysis, First, we look at the percentage change in web traffic arriving from Google and get the pre-post difference in the treated cohort. However, the first difference estimate does not consider counter factuals; if Bazaar did not stop putting sponsor ads due to technical glitches, how much will web traffic from Google change. Thus, we decide to conduct difference in difference.

There are two assumptions before conducting DID.

- **Parallel trends assumption**- Without the treatment, treated subjects would have continued parallel with the control because the control group needs to be a good counter factual for the treated group. We can get a rough sense from the graph that this assumption is valid. To get a more rigorous proof, we will use dynamic DID to test this assumption.
- **No interference**- The change of web traffic from one search engine depends only on whether it received treatment and not on other search engines' treatment status. This assumption is valid because stopping sponsor ads on Google will not change web traffic arriving from other websites since people are unlikely to change their searching behavior, so user groups from different websites are relatively constant. Also, even

if Bazaar stops sponsor ads on Google, users still can land on Bazaar's website from organic links.

Analysis

(Question C) Consider a First Difference Estimate.

By taking first difference estimate we are just trying to see the effect of the treatment on pre-post clicks counts on the treated cohort. In this case, we do a simple linear regression on the log of total clicks from Google Platform

```
##
## Call:
## lm(formula = log(total) ~ after, data = ad_Google)
##
## 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
```

```
##  (Intercept)        after
## 6.524719e+05 1.306972e-01
```

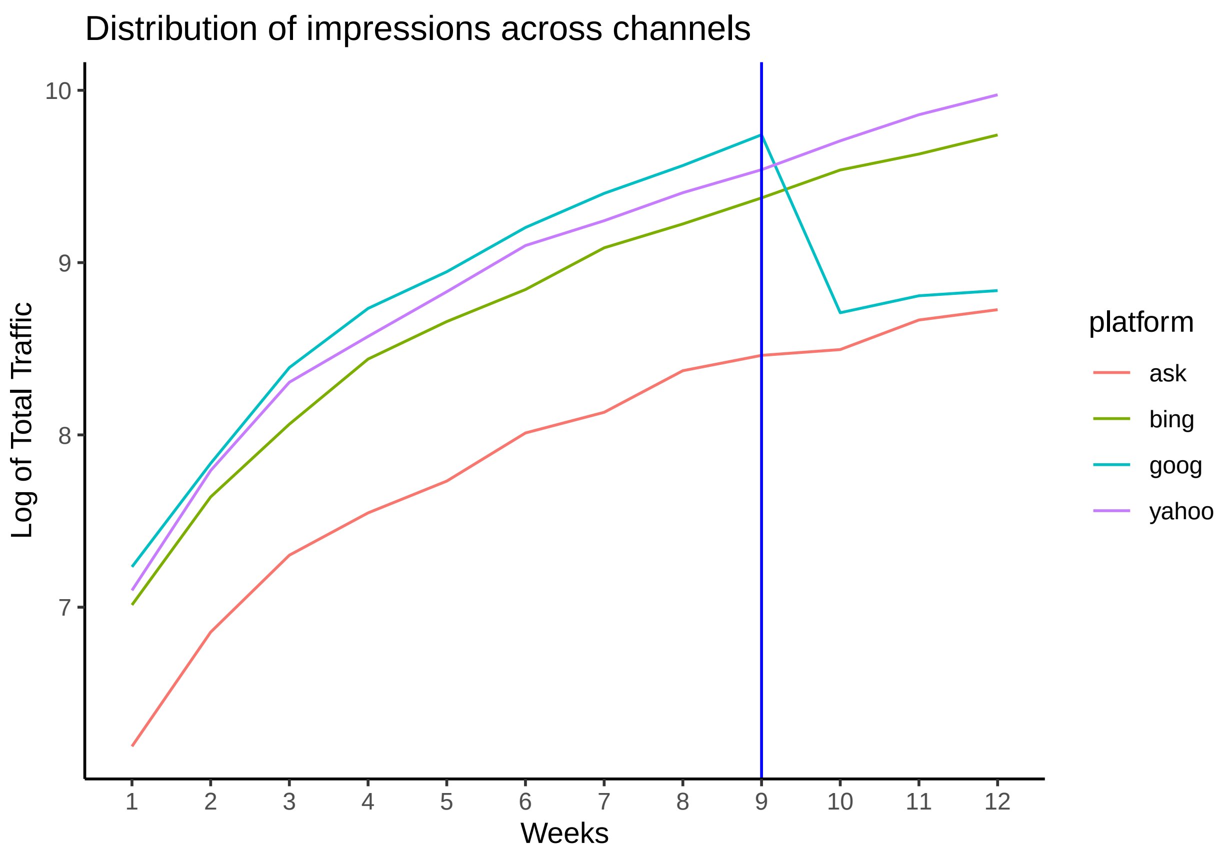
Interpretation He we are trying to find the first difference estimate using a simple linear regression. Using the log of the dependent variable, the coefficient of the model gives the percentage change after the treatment. In our result, the model indicates that there is a 0.13% increase in total clicks from Google website after week 9 when the sponsored ads stopped, which implies that sponsored ads might be cannibalizing the organic clicks.

This result is completely at odds with the initial finding of Bob. However, it is too soon to conclude the quality of this result as, looking at first difference only ignore the effect of counterfactuals. There might be other variables affecting the clicks at a give point of time and not just the presence of the sponsored ads. We should compare our result with the control groups to remove any outside effect on the clicks and estimate the actual average treatment effect. We leverage Difference-in-Difference to achieve the same.

(Question D)(d) Calculate the Difference-in-Differences

In order to perform Difference in difference, we need to first evaluate the assumptions of required for the method. We evaluate the parallel trends assumption. It states that the trends of both treatment and control should be similar before the treatment is applied.

```
ggplot(ad, aes(x=factor(week), y=log(total), group = platform)) +
geom_line(aes(color = platform)) +
theme_classic() +
ylab("Log of Total Traffic") + xlab("Weeks") +
ggtitle("Distribution of impressions across channels") +
geom_vline(xintercept=9, color = "blue")
```



From the graph above, we can clearly see that the trends of the 4 platforms is similar in the pre-treatment period. We can go ahead and perform our difference-in-difference test.

Basic Difference-in-Difference

In order to perform a basic DiD analysis, we regress the total clicks on treatment and after variables for all the 4 platforms. Google is the test group which was treated in the after period which is the last 3 weeks of the campaign. We also add an interaction term in the regression, which is the estimate of the average treatment effect

```
BasicDiD = lm(log(total) ~ treat + after + treat*after, data=ad)
summary(BasicDiD)
```

```
##
## Call:
```

```
## lm(formula = log(total) ~ treat + after + treat * after, data = ad)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0608 -0.5442  0.1414  0.5811  1.2861
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.2532     0.1543  53.505  < 2e-16 ***
## treat         0.5303     0.3085   1.719 0.092629 .
## after         1.1176     0.3085   3.623 0.000751 ***
## treat:after  -1.1163     0.6170  -1.809 0.077241 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8015 on 44 degrees of freedom
## Multiple R-squared:  0.2415, Adjusted R-squared:  0.1898
## F-statistic:  4.67 on 3 and 44 DF,  p-value: 0.006435
```

```
exp(coef(BasicDiD))-1
```

```
##  (Intercept)        treat        after  treat:after
## 3838.7756797    0.6995053    2.0576347   -0.6725224
```

Interpretation

Looking at the above result, the interaction term is significant if confidence interval is 90%. As we take the log of the total clicks are our DV, the coefficient of interaction term gives us a decrease of 67.25% in total clicks after the treatment is applied. This is starkly different from the initial finding in the first difference case as now we compare our results with the control and remove any confounding effects just to analyze the total impact of sponsored ads, which contribute to 2/3rd of the total clicks brought on the platform.

Fixed-effects DiD

Furthermore we try to replace the treatment with group and time-fixed effects to observe any changes in the effect size of the treatment.

```
g_did_fe = plm(log(total) ~ treat + after + treat*after, data = ad,
               index=c("id"), effect="individual", model="within")
summary(g_did_fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(total) ~ treat + after + treat * after, data = ad,
##     effect = "individual", model = "within", index = c("id"))
##
## Balanced Panel: n = 4, T = 12, N = 48
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.575180 -0.236438  0.037839  0.453689  0.958344
##
## Coefficients:
##             Estimate Std. Error t-value Pr(>|t|)
## after        1.11764    0.26744  4.1791 0.000145 ***
## treat:after -1.11634    0.53487 -2.0871 0.042985 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    28.708
## Residual Sum of Squares: 20.276
## R-Squared:      0.2937
## Adj. R-Squared: 0.20962
## F-statistic: 8.73248 on 2 and 42 DF, p-value: 0.00067414
```

```
exp(coef(g_did_fe))-1
```

```
##       after treat:after
##   2.0576347  -0.6725224
```

```
g_did_sfe_tfe = plm(log(total) ~ platform + after + platform*after, data = ad,
                    index=c("id", "week"), effect="twoway", model="within")
summary(g_did_sfe_tfe)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = log(total) ~ platform + after + platform * after,
##     data = ad, effect = "twoway", model = "within", index = c("id",
##         "week"))
##
## Balanced Panel: n = 4, T = 12, N = 48
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -0.093838 -0.022221  0.003899  0.021242  0.087408
##
```

```
## Coefficients:
##                       Estimate Std. Error  t-value  Pr(>|t|)    
## platformbing:after   0.146846   0.043598   3.3682  0.002091 ** 
## platformgoog:after  -1.005531   0.043598 -23.0636 < 2.2e-16 ***
## platformyahoo:after  0.185568   0.043598   4.2563  0.000188 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2.2102
## Residual Sum of Squares: 0.064152
## R-Squared:      0.97098
## Adj. R-Squared: 0.95453
## F-statistic: 334.533 on 3 and 30 DF, p-value: < 2.22e-16
```

```r
exp(coef(g_did_sfe_tfe))-1
```

```
##  platformbing:after  platformgoog:after platformyahoo:after 
##           0.1581760          -0.6341496           0.2039016
```

Interpretation

In both the cases the effect is -67.25% and -63.41% for the fixed group effects and fixed time effects. Therefore, our result was not affected much by the time specific and group specific factors

Therefore, we can conclude that the average treatment effect or the impact of sponsored ads on the total clicks is 67.25%.

(Question E) (e) Given Your Treatment Effect Estimate, Fix Bob's RoI Calculation.

In order to estimate the real ROI of sponsored ads, we calculate the Revenue generated and the cost of running the campaign. Using the formulas discussed above

```
# Average weekly sponsored clicks from google in the pre-treatment period
weekly_sponsored_ads <- mean(ad[ad$platform =="goog" & ad$after==0,"avg_spons"])

# Average weekly cost of running sponsored ads
Cost <- weekly_sponsored_ads * 0.6


#Based on the total treatment effect in clicks i.e., 67.25%.
#Revenue from sponsored ads is:


Revenue <- weekly_sponsored_ads * 0.6725 * 0.12 * 21


(Revenue - Cost)*100/Cost
```

```
## [1] 182.45
```

Conclusion

The real return on investment for sponsored ads is profit from sponsored click per the cost of sponsored click, which is 182.45%. Although the real ROI for sponsored ads is less than Bob's estimated 320%, the current estimated ROI of 182.45% is also a good number to show that the sponsored ads are in fact effective.

Thus it is recommended that Bazaar continue sponsored ads on branded keywords. Furthermore, the team should utilize an A/B testing framework to estimate the true effect of their ad campaigns rather than using heuristics in their calculations.