

# SEO @ Bazaar

*Brady Engelke, Minya Na, Yidan Gao, William Wu, Chia-Hsuan Chou*

*2/22/2020*

## Contents

0.0.1	Executive Summary . . . . .	1
0.0.2	Exploratory Analysis & Experimental Design Checks . . . . .	1
0.0.3	Difference-in-Differences . . . . .	5
0.0.4	Regression Discontinuity . . . . .	7
0.0.5	Final Recommendations . . . . .	10

### 0.0.1 Executive Summary

Bazaar's marketing team runs paid search ads on two search engines: Google and Bing. They leverage branded and non-branded keywords to target their ad copies. Bazaar's marketing strategy is identical in terms of the keywords chosen to advertise on, and ad copies used across both Google and Bing since the mix of customers on each platform is roughly identical. Management hypothesizes that many customers who click on sponsored ads on branded keyword searches would have come to Bazaar's site anyway by clicking on an organic ad copy. Thus, management wants to leverage the technical glitch in the 12-week ad campaign that halted sponsored ads on Google from weeks 10 - 12 to understand the true impact of sponsored advertising on click-through-rate from Google.

After some exploration of the data, it was found that clicks on organic ad copies did rise after week 9. However, total traffic and average profits from Google dropped substantially. The econometric methodology of Difference-in-Differences was first applied to isolate the impact sponsored ad copies targeted at branded keyword searches on Google had on Bazaar's expected profits. Substantiative results were found but the primary assumption of a parallel trend between Google and Bing could not be validated, leading to the methodology of Regression Discontinuity on only the data collected on conversions from Google for weeks 7 - 12 to be utilized. It was assumed that the treatment effect would not slowly manifest itself and could be observed immediately after week 9. This analysis provided evidence that by not utilizing sponsored ad copies for branded keyword searches the expected profits per week from Google dropped by \$18,048 for Bazaar. These sponsored ad copies are persuading consumers to come to Bazaar's site and are providing a substantial boost to the ROI of Bazaar's SEO budget.

### 0.0.2 Exploratory Analysis & Experimental Design Checks

With data at the weekly traffic volume level coming from consumers clicking on sponsored ad links as well as organic links that appear on Google, Bazaar's marketing team has calculated that 20% of traffic came from sponsored ads on branded keyword searches shown on Google during the weeks 1 - 9. First, these facts were verified, the data was checked for cleanliness and outliers, and trends over time were explored. Management was concerned that weeks 1 - 9 were systematically different from weeks 10 - 12 which needed to be carefully considered as well.

```
library(tidyverse)
library(pwr)
library(aod)
library(naniar)
library(plm)
library(rddtools)
library(rdrobust)
library(rdd)
library(ggpubr)
```

```
setwd("~/MSBA/spring/Econometrics/hw/2")
data <- read.csv('did_sponsored_ads.csv')
```

```
miss_var_summary(data)
```

```
glimpse(data)
```

```
google <- data %>% filter(search_engine == 'goog' & week < 10)
round(sum(google$avg_spons) / sum(google$tot_traffic) * 100, 3)
```

```
## [1] 20.004
```

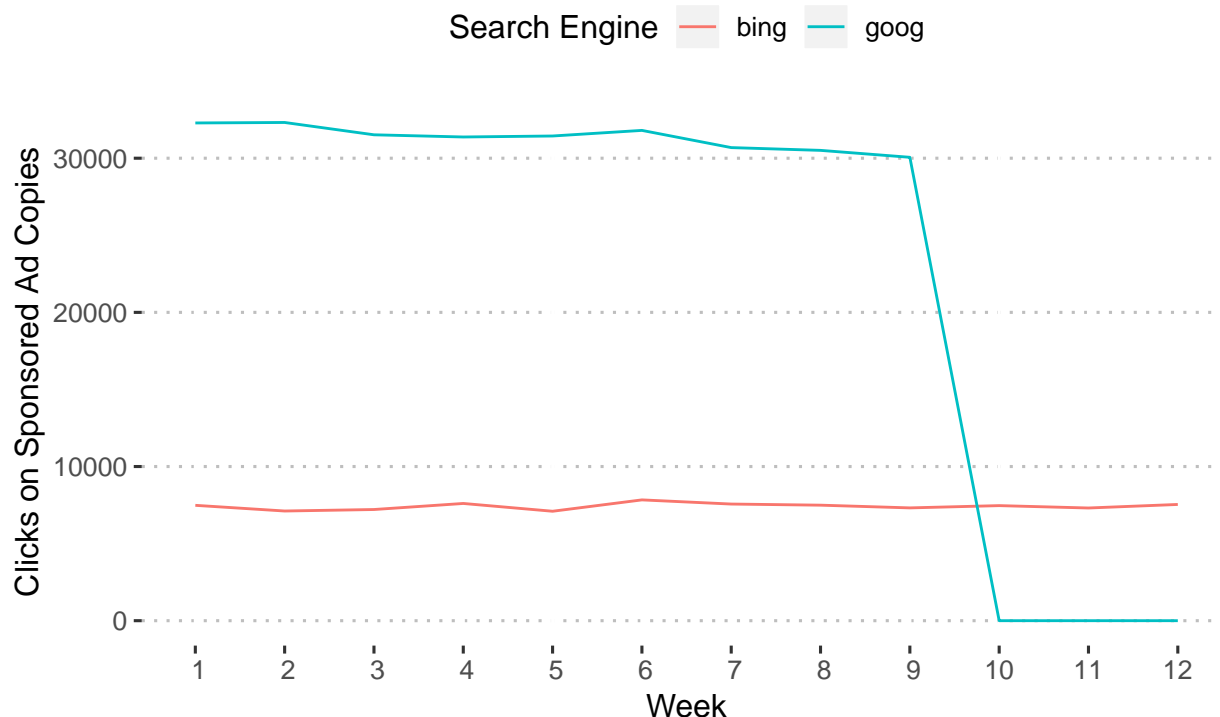
Indeed, sponsored ad copies accounted for 20% of click throughs from Google during weeks 1 - 9.

```
google_bing <- data %>% filter(search_engine == 'goog' | search_engine == 'bing')
google <- data %>% filter(search_engine == 'goog')
```

```
ggplot(google_bing) + geom_line(aes(x = week, y = avg_spons, color = search_engine)) +
  labs(title = 'Sponsored Ads Were Suspended on Google After Week 9',
        subtitle = 'Figure 1', color = 'Search Engine') +
  ylab('Clicks on Sponsored Ad Copies') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(1:12))
```

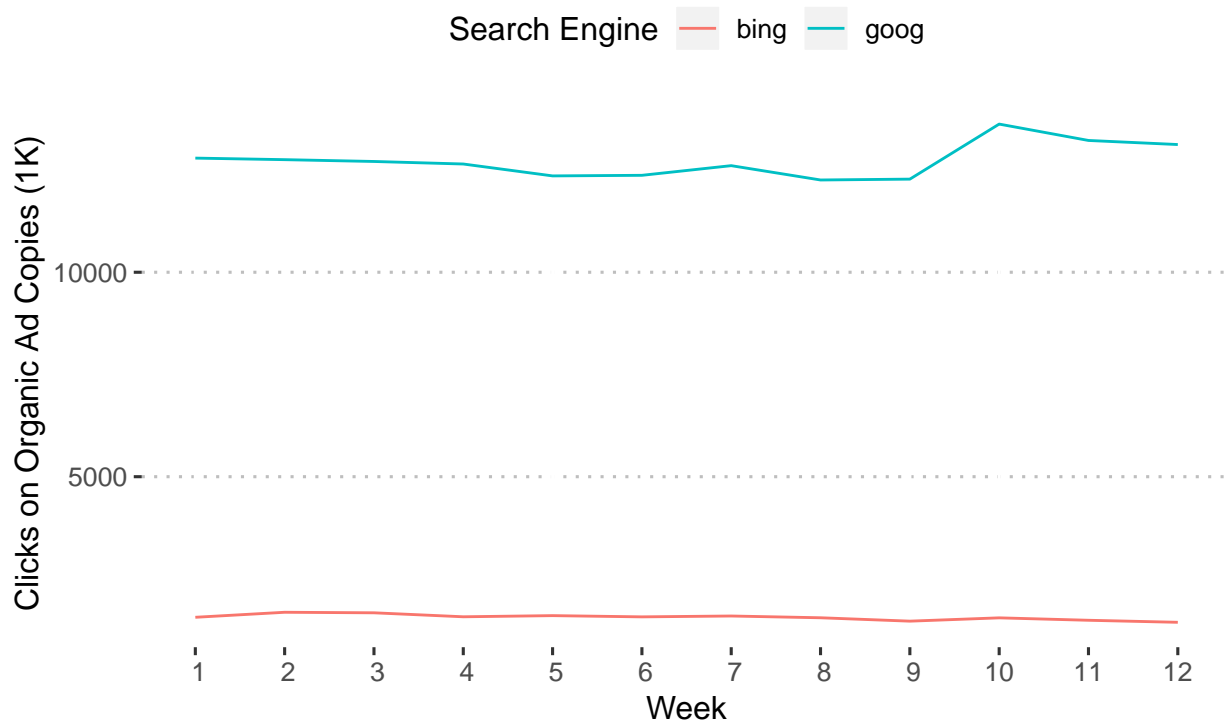
## Sponsored Ads Were Suspended on Google After Week 9

Figure 1



```
ggplot(google_bing) + geom_line(aes(x = week, y = avg_org / 10, color = search_engine)) +
  labs(title = 'Clicks on Organic Ad Copies On Google Jumped After Week 9',
        subtitle = 'Figure 2', color = 'Search Engine') +
  ylab('Clicks on Organic Ad Copies (1K)') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(1:12))
```

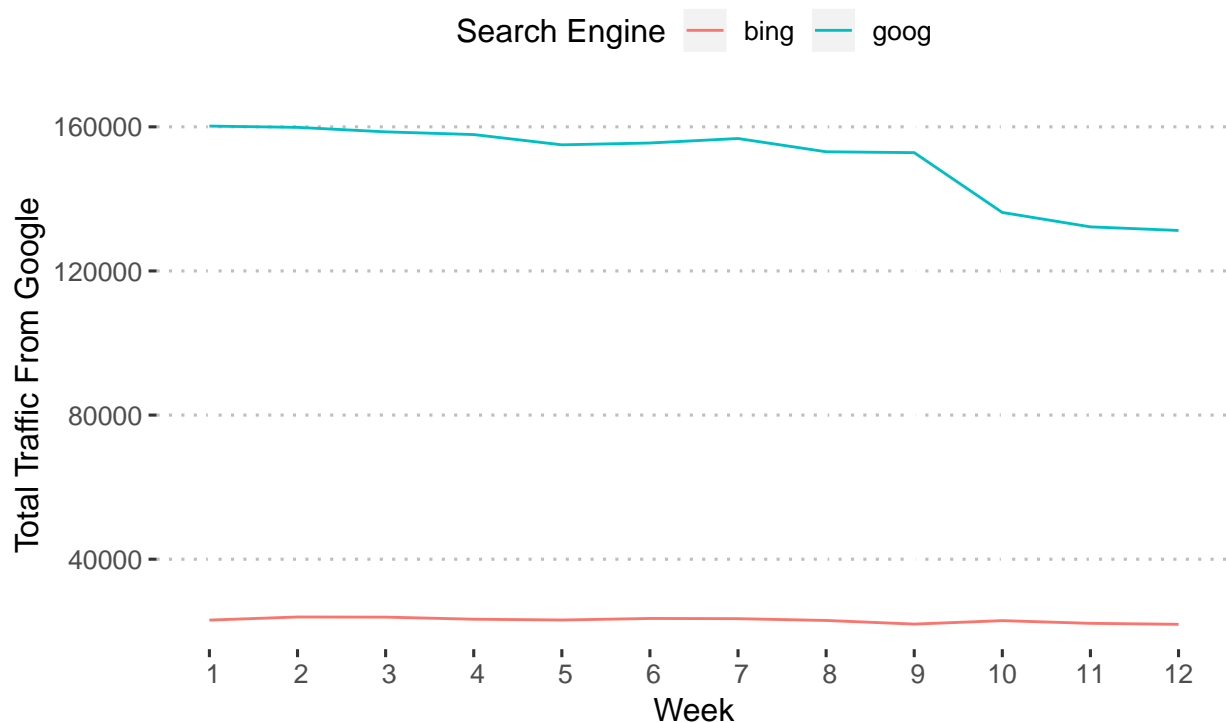
Clicks on Organic Ad Copies On Google Jumped After Week 9  
Figure 2



```
ggplot(google_bing) + geom_line(aes(x = week, y = tot_traffic, color = search_engine)) +
  labs(title = 'Total Traffic From Google Dropped After Week 9',
        subtitle = 'Figure 3', color = 'Search Engine') +
  ylab('Total Traffic From Google') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(1:12))
```

## Total Traffic From Google Dropped After Week 9

Figure 3



From the plots above, it was verified that the campaign was indeed cut short on Google after week 9. Not surprisingly, Bing provides substantially fewer customers than Google. There did seem to be a jump in organic clicks after week 9 from Google, but there was a sharp drop in total traffic from Google after week 9.

In order to develop a more robust metric to measure the impact of sponsored ads on click-throughs from Google, a **profit** metric was developed based on the assumptions below, irrespective of how consumers landed at the website.

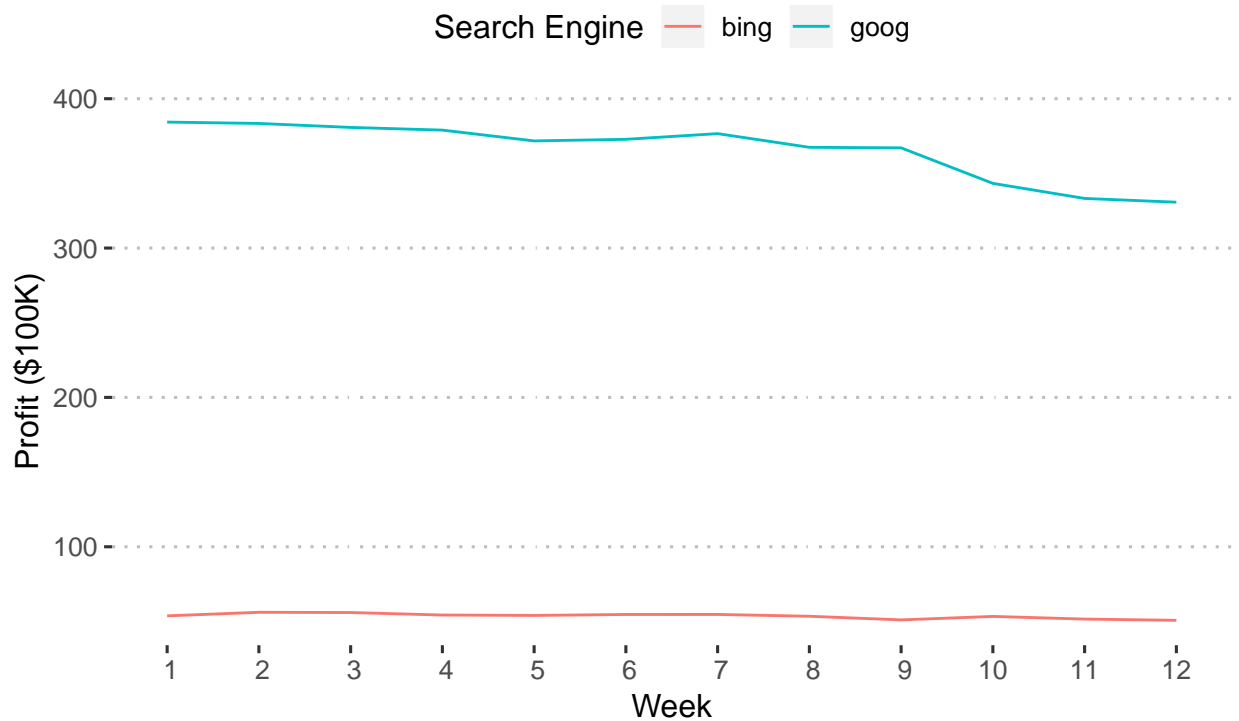
- Average cost per sponsored click = \$0.60
- Average margin per conversion = \$21
- Average probability of a click-through making a purchase = 0.12
- Avg revenue per click-through:  $21 * 0.12 = \$2.52$

```
google_bing <- google_bing %>% mutate(profit = ((2.52 * tot_traffic) - (0.6 * avg_spons)))
google <- google %>% mutate(profit = ((2.52 * tot_traffic) - (0.6 * avg_spons)))

ggplot(google_bing) + geom_line(aes(x = week, y = profit / 1000, color = search_engine)) +
  labs(title = 'Average Profit From Google Dropped After Week 9',
        subtitle = 'Figure 4', color = 'Search Engine') +
  ylab('Profit ($100K)') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(1:12))
```

## Average Profit From Google Dropped After Week 9

Figure 4



A more gradual decline in **profit**, compared to the decline in total traffic, was observed after week 9 once revenue and cost assumptions were taken into consideration. The fact that total traffic and expected profits dropped significantly after week 9 is contradictory regarding Myra's hypothesis that customer behavior, when a branded keyword search is used, will be unaffected if sponsored ad copies are not used by Bazaar.

With the exploratory analysis complete and a better sense of the problem at hand, the quantifiable impact of sponsored ad copies targeted at branded keyword searches on click-through-rate from Google was ready to be analyzed via two econometric methodologies - Difference-in-Differences & Regression Discontinuity.

### 0.0.3 Difference-in-Differences

This approach was naturally suited to isolate the effect of interest since the data was panel data, the mix of customers on Google and Bing is assumed to be identical, and one platform received treatment while the other did not. The Difference-in-Differences (DiD) methodology exploits this experimental design by canceling out the time-invariant confounds specific to each platform via Fixed Effect Regression and it controls for the time-variant confounds common to both platforms which are of particular interest in this case. There is a drawback that comes with these powerful benefits and that is there needs to be statistical evidence that there are not any time-variant confounds that are unique to either the control or test group throughout the experiment. This statistical evidence can be gathered by analyzing the parallel trend of the control and treatment group to see if anything varied between these two entities before treatment. It is also assumed that the treated subjects are not influencing the control subjects which there was no reason to believe this was apparent within the study.

```
google_bing <- google_bing %>% mutate(test = ifelse(week > 9, 1, 0))
google <- google %>% mutate(test = ifelse(week > 9, 1, 0))
google_bing$test <- as.factor(google_bing$test)
google$test <- as.factor(google$test)
```

```
summary(lm(profit ~ test + search_engine + test * search_engine, data = google_bing))
```

```
##
## Call:
## lm(formula = profit ~ test + search_engine + test * search_engine,
##     data = google_bing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8815.1 -2667.0  -69.7  1784.7  8425.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      54210      1578  34.364 < 2e-16 ***
## test1           -2324       3155  -0.736    0.47
## search_enginogoog 321704       2231 144.200 < 2e-16 ***
## test1:search_enginogoog -37841      4462  -8.481 4.68e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4733 on 20 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.9991
## F-statistic: 8760 on 3 and 20 DF,  p-value: < 2.2e-16
```

Interpeting the DiD regression, the impact of going from Bing to Google had a significant impact on the relationship of going from weeks 1 - 9 to weeks 10 - 12 on the expected profits for Bazaar. This second-order effect was estimated to be -\$37,841.

Before moving forward with a conclusion, the parallel trend between Google and Bing had to be analyzed to ensure the results above were reliable.

```
did_dyn <- lm(profit ~ search_engine + as.factor(week) + search_engine * as.factor(week),
              data = google_bing)
```

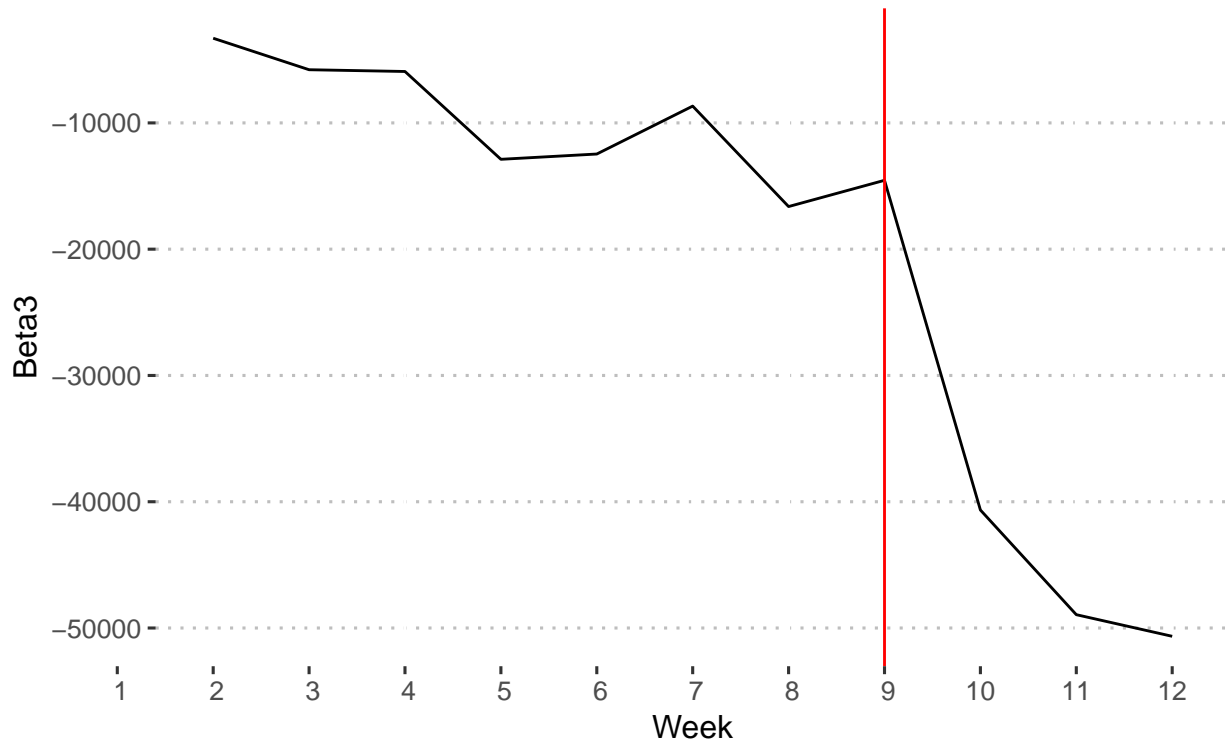
```
summary(did_dyn)
```

```
model <- summary(lm(profit ~ search_engine + as.factor(week) +
                    search_engine * as.factor(week), data = google_bing))
```

```
coefs_ses <- as.data.frame(model$coefficients[14:24, c("Estimate", "Std. Error")])
colnames(coefs_ses) <- c("beta", "se")
coefs_ses <- coefs_ses %>%
  mutate(ub95 = beta + 1.96 * se,
         lb95 = beta - 1.96 * se,
         week = 2:12)
```

```
ggplot(coefs_ses) + geom_line(aes(x = week, y = beta)) +
  geom_ribbon(aes(x = week, ymax = ub95, ymin = lb95, alpha = 0.3), show.legend = FALSE) +
  geom_vline(xintercept = 9, color = 'red') +
  labs(title = 'Unable to Confirm Parallel Trend Between Google & Bing',
       subtitle = 'Figure 5') +
  ylab('Beta3') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(1:12))
```

Unable to Confirm Parallel Trend Between Google & Bing  
Figure 5



Even though the above Beta3 estimates take a substantial dip after treatment, the above plot hinders the results found in the previous DiD Regression since there was not enough data to estimate the 95% confidence interval and the coefficients are trending decreasingly below zero before treatment. Thus, it cannot be confirmed whether the DiD Regression results found above are reasonable.

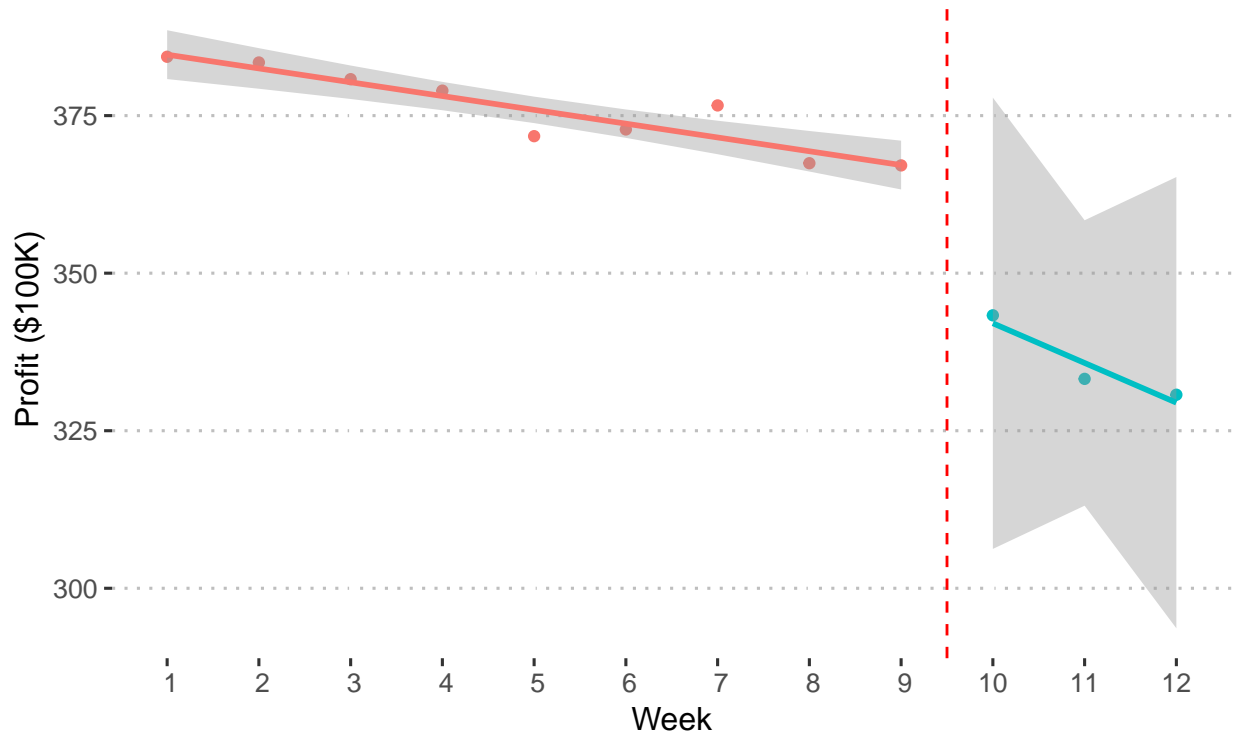
#### 0.0.4 Regression Discontinuity

Regression Discontinuity (RDD) was applicable since the treatment assignment of whether sponsored ads were used on Google was entirely arbitrary due to a technical glitch. Also, figure 4 gives some evidence that the model specification should not be difficult to accommodate and it is likely to be linear. Note, data collected on Bing has been filtered out for the following analysis.

```
ggplot(google, aes(y = profit / 1000, x = week)) +
  geom_point(aes(color = test), show.legend = FALSE) +
  geom_vline(xintercept = 9.5, linetype="dashed", color="red") +
  geom_smooth(aes(group = factor(test), color = test), method = "lm", show.legend=FALSE) +
  labs(title = 'RDD Looks Promising', subtitle = 'Figure 6') +
  ylab('Profit ($100K)') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(1:12))
```

## RDD Looks Promising

Figure 6

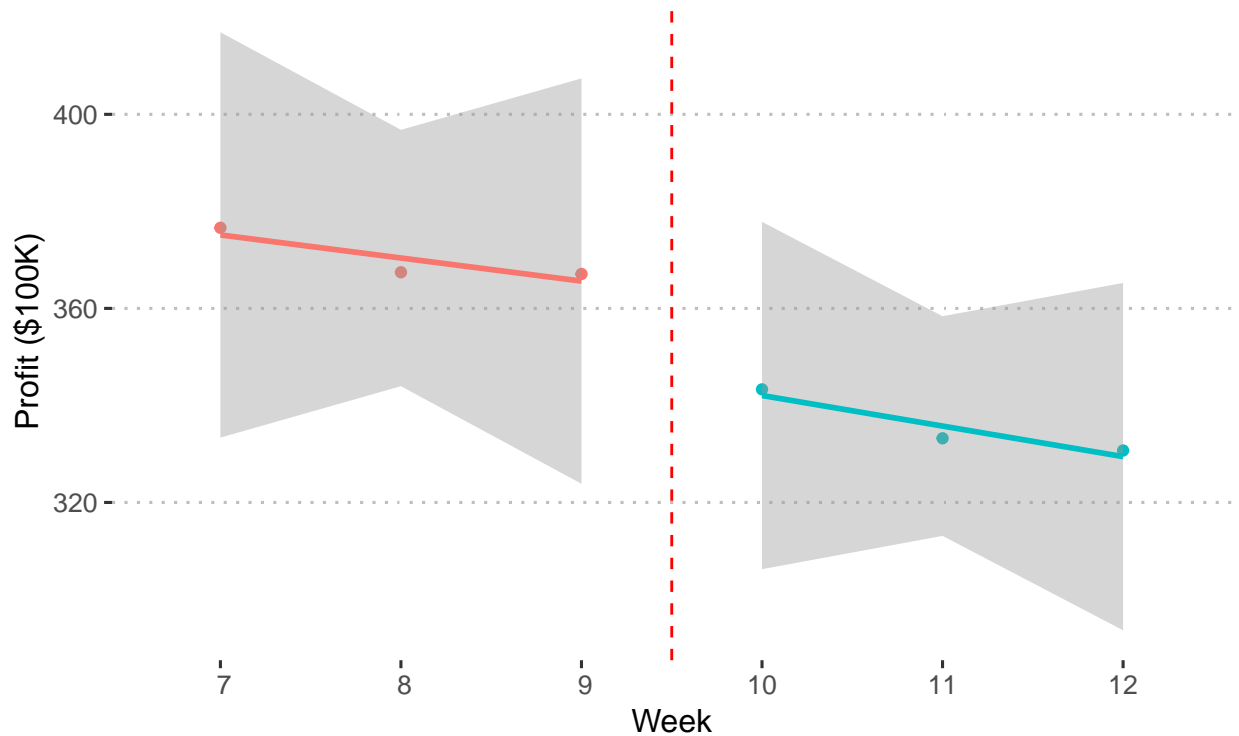


```
google2 = google %>% filter(week >= 7)
ggplot(google2, aes(y = profit / 1000, x = week, color = test)) +
  geom_point(show.legend = FALSE) +
  geom_vline(xintercept = 9.5, linetype="dashed",color="red") +
  geom_smooth(aes(group = factor(test), color = test), method = "lm", show.legend=FALSE) +
  labs(title = 'RDD Isolated on Weeks 7 - 12', subtitle = 'Figure 7') +
  ylab('Profit ($100K)') + xlab('Week') + theme_pubclean() +
  scale_x_discrete(limits=c(7:12))
```



## RDD Isolated on Weeks 7 – 12

Figure 7



```
summary(lm(profit ~ test + week2,
  data = google2 %>% mutate(week2 = week - 9)))
```

```
##
## Call:
## lm(formula = profit ~ test + week2, data = google2 %>% mutate(week2 = week -
## 9))
##
## Residuals:
```

	1	2	3	4	5	6
	700.1	-2940.8	2240.7	2030.4	-2520.3	489.8

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	364858	2199	165.905	4.83e-07 ***
test1	-18048	4918	-3.670	0.0350 *
week2	-5530	1440	-3.841	0.0311 *

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2879 on 3 degrees of freedom
## Multiple R-squared:  0.9872, Adjusted R-squared:  0.9787
## F-statistic: 115.9 on 2 and 3 DF,  p-value: 0.001444
```

```
summary(lm(profit ~ test + week2 + test * week2,
  data = google2 %>% mutate(week2 = week - 9)))
```

```
##
```

```
## Call:
## lm(formula = profit ~ test + week2 + test * week2, data = google2 %>%
##   mutate(week2 = week - 9))
##
## Residuals:
##      1      2      3      4      5      6
## 1470 -2941  1470  1260 -2520  1260
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   365629      3062 119.414 7.01e-05 ***
## test1         -17278       5969  -2.895   0.101
## week2          -4760       2372  -2.007   0.183
## test1:week2    -1541       3354  -0.459   0.691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3354 on 2 degrees of freedom
## Multiple R-squared:  0.9884, Adjusted R-squared:  0.9711
## F-statistic: 57.02 on 3 and 2 DF,  p-value: 0.01728
```

When the interaction term was included in the second regression it resulted in insignificant results. This provided evidence that the impact of going from weeks 1 - 9 to weeks 10 - 12 did not impact the effect one unit change in week would have on expected profits for Bazaar. Thus, based upon the first regression without the interaction term, there is a 3.5% chance results as extreme as an \$18,048 drop-in expected profits for Bazaar when sponsored ads on keyword searches are not utilized on Google would be observed during repeated analyses. This drop-in expected profits is likely due to customers clicking on competitor ad copies since there was no sponsored ad copy placed by Bazaar. Giving consumers the chance to test out other options on the market could potentially cause a portion of those customers to leave Bazaar forever, compounding the short-term negative impact of expected profits dropping.

### 0.0.5 Final Recommendations

Myra's hypothesis that sponsored ad copies are insignificant in regard to swaying customers to come to Bazaar's website compared to organic ad copies has been invalidated; Bazaar should continue to run sponsored ads on keyword searches on Google.