The Effects of Discounting on Online Shopping

W241 Field Experiments and Causality

Jeffrey Hsu, Ramya Balasubramaniam and Roiana Reid
April 2017

Abstract

This paper presents the findings from a field experiment investigating the effects of discounting on online shopping behavior. The experiment is carried out in an Etsy online shop with discounts offered over a one week period for selected products as the treatment. Comparing demographic characteristics of the site visitors and other web session data between the treatment week, pre-treatment week and post-treatment week, we proved covariate balance and tried to evaluate that the change in sales performance and customer purchasing behavior in the treatment week is a causal effect from the treatment offered, rather than sales fluctuations with time. Sales per unique Item per order (SIO), measuring the purchasing tendency of customers toward higher or lower value items, was lower during the treatment week with differences that were statistically significant compared to control weeks. Combine this observation with the fact that total sales and sold item quantity are higher during the treatment week suggests that the discount increased customers' tendency toward buying lower priced items. Other sales performance metrics, measured by Sales per Order (SO), Order Value per Buyer (OVB), Order Quantity per Buyer (OQB) and Conversion Rate, do not show a statistically significant difference throughout the experiment. This limits our ability to interpret the relationship between the observed customer behavioral change and overall sales. Finally, Sales per Order (SO) data shows a

strong positive heterogeneity effect on returning customers. This implies the treatment effect is stronger among returning customers. However, given the small sample of returning customers and high fluctuation of sales level, further studies with randomization on individual web sessions are required to confirm the effect.

A. Introduction

In classical economics, the study of the demand curve and price elasticity is well established. Decrease in prices are expected to result in higher demand in the market. The demand curve is used to approximate behaviors in competitive markets, and is oftentimes combined with the supply curve to estimate the equilibrium price. Many studies have estimated price sensitivity and changes in demand (e.g. [1] Gerard J. Tellis, 1988 on the high negative price elasticity of selected market products), but most have taken place in physical retail markets. Price elasticity of products offered on E-commerce marketplaces or online shops are rarely studied. Many of the studies for online retail are focused on the effectiveness of online marketing (e.g. [2] Jansen and Molina, 2005 on the effectiveness of search engines on Ecommerce queries), the effect of lowering information barriers on competition (e.g. [3] Lynch and Ariely,2000 on the effect of lowering cost of search on wine price elasticity) and website features that seek to reduce checkout barriers, and increase sales and conversion rates. However, consumer behavior in response to price reduction may be markedly different in online shops than it is in physical shops. Moreover, understanding the impact of consumer behavior on sales is valuable for online retailers. As a shop owner, evidence-based knowledge of the causal effect of price variation on revenues, should affect sales strategy.

This study aims to answer the research question: “Do price reductions for selected products boost sales for online shops?” The study investigates two things. Firstly, it examines consumer behavior in response to price reductions of selected products offered in online shops. Secondly, it probes the overall effect of price reduction on the performance of the online shop. This answers interesting questions such as “does the price promotion cause a sales increase during the discount period but a decrease below baseline after discount has ended?” (intertemporal substitution).

The research approach controls for marketing factors and observes sales outcome coming from price reduction of selected products. Throughout the experiment, no additional advertisement were initiated. By controlling marketing factors and checking the balance of incoming shop visits, we ensure that changes in sales observed resulted from the treatment, rather than factors that can induce fluctuations in shop exposure. The study is carried out on an Etsy online shop. In terms of treatment effect on consumer behaviors, the change in the following factors are observed: likelihood of purchase for visitors to the website, order value of purchases, number of products purchased and sales per buyer, sales of unique item per order and long lasting effect of change in consumer behaviors. In terms of overall sales performance effect for the online shop, we observe the following factors: overall revenue, total items purchased, profit and intertemporal substitution of sales.

B. Experimental Setup and Methodology

The retail experiment was carried out on the online marketplace Etsy. The third party sources, “etsyonsale.com” and Google Analytics, were used to implement the experiment and retrieve key data, and R and Excel were used to analyze and visualize the data. The Etsy shop “3XUdesign” which designs and sells small, artsy, unique craft items such as matchboxes, passport covers and bookmarks, was used in the experiment. The prices of most items at “3XUdesign” are in an affordable \$6.99 to \$11.99 range and there are bundled item sets offered at discounted rates. The store is currently based in Vietnam and delivers worldwide, with a large share of buyers located in the US. The site typically receives between 150 and 200 visitors per day, a relatively small number, but it provides a large enough sample for the purposes of this experiment. The experimental design was influenced by the shop-owner's preferences, knowledge of the Etsy marketplace and features, and nuanced understanding of the patterns in sales.

Etsy has a feature, "Shop Stats", that enables shop-owners to obtain shopping metrics such as the number and types of items purchased, value of total purchases, purchase price, date of purchase, transaction ID, and buyer name and location – all available for download in CSV form. Etsy also provides visitor count data, but unfortunately these are not available for download. Although Etsy provides important fundamental stats to track sales, it fails to provide other useful statistics, such as conversion rates and demographic information that would allow shop-owners to identify the profiles and habits of site visitors. In order to compensate for this missing vital data, the use of Google Analytics was employed.

The Google Analytics (GA) platform connects seamlessly to Etsy stores, tracking the number of site visits, number of sessions, the time, date and duration of sessions, location of users, language the website is viewed on, new versus returning users and, most importantly, the key demographics, age and gender. GA also tracks other characteristics of users such as interests, buying preferences, the technology used to interact with the shop page, including operating system and type of device – mobile, laptop, tablet – and even the model/make of the device. Unfortunately, GA does not provide details on these categories for all sessions, but it provides information for over 50% of sessions, which is sufficient for this analysis.

It was important to ensure that the treatment was very conspicuous to the shop “3XUdesign” visitors. However, the Etsy platform does not provide a systematic process for creating sales events, so we used the website “etsyonsale.com.” This site enabled us to advertise the sale on the shop’s website banner - the first thing visitors see when they enter the shop page – and calculated and displayed the final sales price, based on the discount rate placed on each item selected for the sales event. It also allowed us to set the length of time that the sales event should be displayed. Importantly, for each item on sale, “etsyonsale.com” enabled us to put a line through the original price, display the final discounted price and display “ON SALE” on the product page. As a result, there were many opportunities for visitors to see that a major sales event was taking place.

The experiment took place over a continuous three-week period, from March 13^{th} to April 2^{rd}. There were two control groups and one treatment group, which were determined by time -- seven-day periods for each. The first control group consisted of all site visitors between March 13^{th} and March 19^{th}, the treatment group included visitors between March 20^{th} and March 26^{th} - with the treatment delivered during that time - and the second control group included all visitors between March 27^{th} and April 2^{rd}. It was instructive to include a second control group in order to provide additional evidence that any changes in quantity of items purchased and value of sales in the treatment group were independent of both the timing of the first control measurement and the timing of the treatment and were only due to the treatment intervention. The store owner was uncomfortable with the idea of changing prices too often – on a daily basis for example – and informed us of the maximum amount of time for which the sales event could be conducted. Taking these factors into account, we decided to deliver the treatment for a week and only deliver it once. It would have been ideal to randomly assign treatment to visitors by identifying unique visitor ID or cookies, but this was not a viable solution given the limitations of the Etsy platform.

This experiment was designed to test the hypothesis that a price incentive would boost the quantity of purchases and total sales at the online retailer. The treatment offered was an across-the-board 20% discount at the 3XUdesign shop for the entire treatment week. It was decided that a 20% discount was a large enough incentive to boost sales and was unlikely to cause a significant loss to the shop-owner. The same sized percentage discount was applied on all items in order to ensure consistency and precision of comparisons. The original and discounted prices are as follows: $6.99 to $5.59, $9.99 to $7.99, $19.99 to $15.99, $11.99 to $9.59, $8.99 to $7.19. The bundled items that were already discounted were excluded from this treatment. For the control groups, only original prices were offered with no displays of a sales event anywhere on the website. Total sales value, total quantity sold, sales per order (SO), order value per buyer (OVB), order quantity sold per buyer (OQB), sales conversion rate and sales per unique item per order (SIO) were analyzed to determine if there was any significant impact from the treatment intervention. The shop owner was enthusiastic about this experiment because March sales are usually slow, so any strategy to boost sales was highly valued.

In proving the validity of an experiment, one of the main challenges is showing that the selection of members of the treatment and control groups is done randomly. We were unable to directly select subjects of the treatment and control groups, so we are tasked with proving that the persons who visited the website during the control weeks were of similar composition, whether demographic or other covariates, to those who visited during the treatment week. The assumption going into this experiment was that there was no reason to believe that the control groups and treatment group should be fundamentally different and thus website visit patterns are assumed to be consistent throughout the experiment. This assumption proved true with the actual experiment and is discussed in further detail in the "Discussion" section below.

As we enter the "Discussion" section, it is instructive to discuss how the various data sources were used. The GA data were used to perform covariate balance checks to prove randomness. As mentioned above, for some of the covariates, GA only obtained data for a certain portion of sessions - usually over 50% - so we just assume that patterns observed in these proportions are roughly consistent across the entire sample. The Etsy data were used to determine the magnitude and significance of the treatment effects. We also obtained historical Etsy sales data from the shop-owner in order to best determine how long the experiment would have to be undertaken to provide any sort of useful results and to develop a better feel for the data as we designed the experiment.

C. Results and Discussions

The outcome parameters were evaluated for three consecutive weeks between March 13th and April 2nd, as described in the section above. Analysis of the week following treatment, the second control group, was included to test for any persistent effects of applying the discounts and to ensure that any changes seen in the outcome parameters during the treatment week could be attributed to the treatment, thus ruling out any temporal/seasonal trends which could have naturally boosted sales for instance effects of regional festivities, payday effects. Additionally, by analysing outcome parameters in the week after the treatment, it allows us to estimate the direction of the long-term effect of the treatment and if there is any intertemporal substitution effect present.

C.1. Covariate Balance

The most important assumption of this experiment is based on the notion of covariate balance among the visitors of the site during the control and treatment weeks. Table 1 below displays the results of the covariate balance analysis. Various demographic and sociological features of the visitor population are compared and checked for statistically significant differences. The similarities in the statistical distributions are checked using t-tests and Wilcoxon tests (rank sum test for independent samples and signed rank test for matched pairs), but in absence of evidence for making normal approximations, the results of Wilcoxon tests are considered more reliable and are displayed here. The respective p-values from t-tests are tabulated for reference and are placed in square brackets to the right of the Wilcoxon test statistic. The hypothesis tests here are calculated paired wise, this allows us to compare demographics data of each day within the treatment and control weeks. The values have been analyzed from data collected for the respective weeks: Control Week 1 (C1), Treatment Week (TR) and Control Week 2 (C2).

All of the p-values calculated for the covariate balance checks for C1 and TR are not statistically significant at the 10% significance level, so we fail to reject the null hypothesis that the distributions are the same. Thus, the distributions could be assumed to be similar in the two weeks (C1 and TR) - one important step towards our randomization claim. Covariate balance check on language between C2 and TR show p-values slightly below the 5% significance level. However, by performing covariate checks on all available factors, we need to consider correction of multiple comparison. The correction would yield a significance level well below 0.05, thus relaxes concerns for covariate imbalance. Thus, the distributions could be assumed similar in the two weeks for most of the demographic features (C2 and TR).

Table 1: Covariate Balance Analysis

Covariates	Total Number of Web Sessions			Wilcoxon Test p-value (t test p-value)	
	C1	TR	C2	C1 vs TR	C2 vs TR
Age (18-24)	237	210	209	0.469 [0.424]	1.000 [0.972]
Age (25-34)	216	254	217	0.172 [0.117]	0.269 [0.251]
Age (35-44)	82	98	43	0.527 [0.536]	0.142 [0.065]
Age (45-54)	12	40	39	0.361 [0.321]	1.000 [0.972]
Age (all groups)	547	602	508	0.611 [0.487]	0.397 [0.317]
Females	574	608	559	0.612 [0.562]	0.812 [0.556]
Males	125	94	81	0.352 [0.236]	0.588 [0.538]
Language	1136	1215,1214*	1099	0.924 [0.371]	0.049 [0.088]
Country	1133	1215,1207**	1096	0.907 [0.281]	0.261 [0.134]

Note: * and ** signify that out of two values mentioned in the cell, the first value represents the value used while comparing C1 and TR weeks and the second value represents the value used while comparing C2 and TR weeks. Additionally, * refers to total number of sessions in the common group of languages and ** refers to total number of sessions in the common group of countries amongst the treatment and control weeks. Source: Etsy and Google Analytics

C.2. Sales Summary

Table 2 below displays the observations for the Total Sales and Total units sold in control and treatment weeks. Number of observation represents the total web sessions in the given week and estimate represents the sum of sales (qty or revenue) during the week.

Table 2: Estimates of Outcome Parameters in Control and Treatment Weeks

Outcome Parameter	No. of web session observations			Estimates		
	C1	TR	C2	C1	TR	C2
Total Sales Value (in US $)	1161	1239	1129	325.71	452.95	162.81
Total Number of Orders				22	19	14
Total Quantity Sold				32	67	19

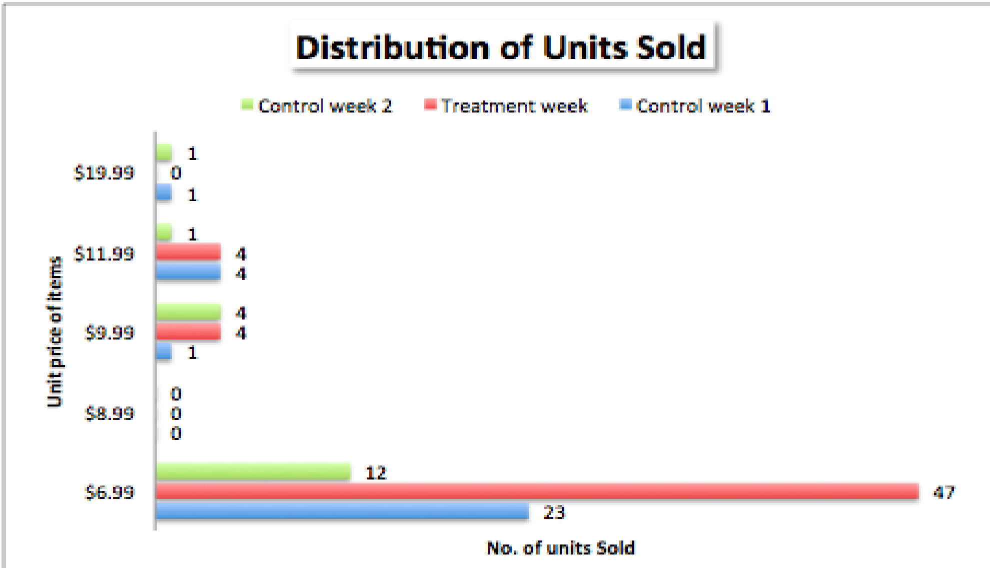
Source: Etsy and Google Analytics

1. There is a 39% rise in the value of sales and a 94% rise in the quantity of items sold from C1 to TR. Similarly, for the week immediately after the treatment week, C2, there is a 64% fall in the value of sales and 69% fall in the quantity relative to the treatment week. **This is a crucial result as it suggests that the discounts were able to successfully boost sales. These are probably the results that the store owner cares about the most. Also, the drop in C2 indicates an intertemporal substitution effect.** This effect, however, will require comparison of sales in the past to confirm. We will discuss that later.

2. The number of sessions during the two control weeks C1 and C2 are similar at 1161 and 1129, respectively. The number of sessions during the treatment week is 1239. Treatment week also shows a higher incoming shop visit compared to control weeks.

Figure 1 displays the distribution of items sold (grouped according to the original prices). Note that only the item groups on which the 20% discount were applied are represented here.

Figure 1: Number of units sold by the original price of the item



Source: Etsy

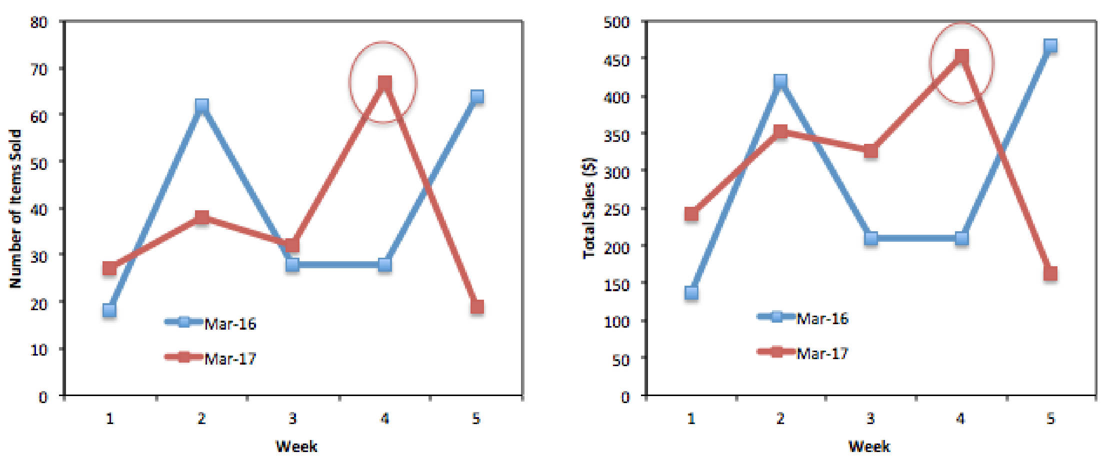
1. In order to test if the distributions of the number of items sold is the same across the three time intervals, the Wilcoxon's Signed rank test (for matched pairs) was used. The p-values for C1 with respect to TR and for C2 with respect to TR, are 0.423 and 0.586 respectively. So, the null hypothesis that the distributions of number of items sold are similar across treatment and control cannot be rejected.

2. There is a considerable increase in the number of units with the price tag of $6.99 that were sold. There is a rise of 104.3% from C1 to TR and a fall of 74.5% from TR to C2.

Sales in March 2017 were compared to sales in March 2016 to determine if any seasonality affected results (see Figure 2). Monthly sales level fluctuation is high (e.g. December revenue is 4 times of March). To eliminate monthly sales pattern affecting the interpretation of sales baseline, we compare previous sales outcome of the same month. Number of items sold in March 2016 were slightly higher, but total sales in March 2017 were higher. There appeared to be no seasonal connections in sales across the weeks in March in both years. While activity in the aggregate appeared to be similar, the variations in sales within the month (from week to week) seemed different. In March 2016, sales were more volatile, but in 2017 they were more stable. In fact, for Week 4, which was the week in which the treatment was applied in 2017, sales were actually quite low in 2016. It is hard to draw any definitive conclusion given the fluctuation in sales level. Nonetheless, the higher revenue and quantity sold from the treatment week compared to 2016 March sales and other control weeks in 2017 March suggest that treatment has a short term boost on sales in the fourth week of March relatively to norm.

Comparing Week 5, which is the post-treatment week of March 2017, to that of March 2016 shows a drastic lower revenue and quantity sold in 2017. This drastic decrease of sales from treatment week to post-treatment week, which is not observed in 2016 March, implies an intertemporal substitution effect from the discount.

Figure 2: Items and Sales in March 2016 and 2017



Source: Etsy

In light of the observation that along with total sales and quantity of items sold, there is a concomitant increase in the number of sessions, the average effects of applying 20% discounts need to be evaluated and the statistical and practical significance of the differences observed, need to be assessed.

C.3. Sales Treatment Effect Estimation

For assessing the effects of applying the 20% discount on 15 items over a span of a week, five outcome measurements were used: Sales per Order (SO), Sales per Unique Item per Order (SIO), Order Value per Buyer (OVB), Order Quantity per Buyer (OQB) and Conversion Rate. The definition of each outcome measurements is listed below in Table3. The treatment effect estimate is calculated by the difference in means. To test statistical significance of the differences, randomization inference under sharp null hypothesis was used. The resulting p-value obtained by sharp null hypothesis were compared with p-values from t-test given in square brackets. All the tests conducted during this experiment are two-tailed. The standard error of the effect estimate is calculated based on the ATE distribution under sharp null hypothesis and is provided in parenthesis under the the effect estimate. Additionally, to interpret the meaning of practical significance of the treatment effect, it is discussed in its original units, rather than transforming to other statistics (e.g. Cohen's d).

Table 3: Definitions of the Outcome Measures

Measures	Definition
Sales per Order (SO)	Sales value (in US $) of each order
Sales per Unique Item per Order (SIO)	Sales (in US $) of unique item per order
Sales Value per Buyer (OVB)	Total sales per unique buyer
Order Quantity per Buyer (OQB)	Total items purchased per unique buyer

Conversion Rate	Percentage of web sessions that turns into purchases

Sales per Order (SO)

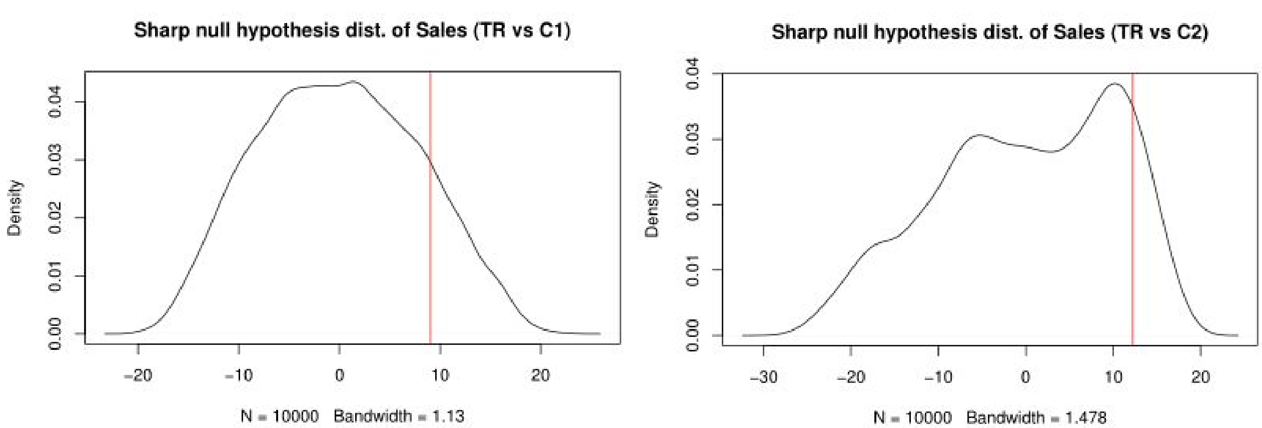
For analysing SO, each order is one observation. The ATE estimate of SO evaluated by randomization inference is shown in Table 4. Figure 3 shows the ATE distribution under sharp null hypothesis.

Table 4: Testing the Differences of Sales per Order (SO) Across Weeks

Outcome Measures	No. of Observations			Outcome Estimate			Effect Estimate		p-value	
	C1	TR	C2	C1	TR	C2	TR-C1	TR-C2	TR-C1	TR-C2
SO	22	19	14	14.805	23.839	11.629	9.034 (7.972)	12.210 (10.363)	0.278 [0.289]	0.275 [0.141]

Source: Etsy and Google Analytics

Figure 3: Differences in Sales per Order under Sharp Null Hypothesis



Source: Etsy and Google Analytics

We find that the Sales per Order has increased from $14.80 to $23.84 during control week C1 and treatment week TR respectively. Also, there is a fall from $23.84 to $11.63 during treatment TR and control week C2. However, the increase in Sales per Order during the treatment week is not statistically significant. Looking into order data, it is likely that the increase in SO for treatment week was due to the orders made by two returning customers. These buyers purchased a large number of items while placing returning coupons on top of the treatment offered. The distribution of SO for other orders during control and treatment weeks remained roughly the same.

Sales per Unique Item per Order (SIO)

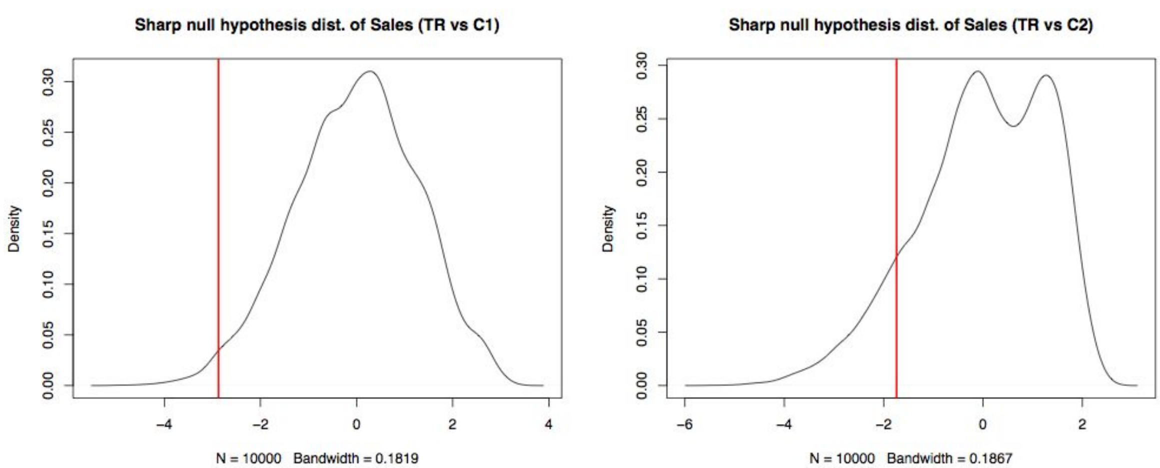
To assess the purchasing behavioral change, we estimated the difference in Sales per Unique Item per Order (SIO) across treatment and control weeks.

Table 5: Testing the Differences in Sales per Unique Item per Order Across Weeks

Outcome Measures	No. of observations			Outcome Estimate			Effect Estimate		p-value	
	C1	TR	C2	C1	TR	C2	TR-C1	TR-C2	TR-C1	TR-C2
SIO	32	62	18	10.178	7.306	9.045	-2.873 (1.283)	-1.739 (1.316)	0.018** [0.040]	0.178 [0.104]

Source: Etsy and Google Analytics

Figure 4: Differences in Sales per Unique Item per Order under Sharp Null Hypothesis



Source: Etsy and Google Analytics

The distribution of the ATEs for TR with respect to C2 does not approach normal even after 10000 randomizations. So, the p-value calculated using randomization inference is more reliable than p-value calculated by t-test for the TR v.s. C2 estimate.

The difference in the estimates of SIO for pre-treatment week and the treatment week is $2.873 and the difference in the estimates of SIO for post-treatment week and the treatment week TR is $1.739. There is a drop in SIO estimate for the treatment week, TR, in both the cases. This drop in SIO could be interpreted as the number of units sold increasing particularly for the low price items in an order (e.g. items that originally cost $6.99). The differences in the estimates show statistical significance for TR v.s. C1 with the p-value of 0.018 but no statistical significance for TR v.s. C2 with the p-value of 0.178 under sharp null hypothesis. The differences in mean of $2.87 (TR v.s. C1) and $1.74 (TR v.s. C2) can be seen as practically significant given that the magnitude of the differences approach half and one-third of the value of the lowest priced item (starting from $5.59) during the treatment week.

Order Value per Buyer (OVB)

For analyzing the differences in Order Value per Buyer, the sales data were grouped by each unique buyer in the respective weeks. The results are shown in Table 6:

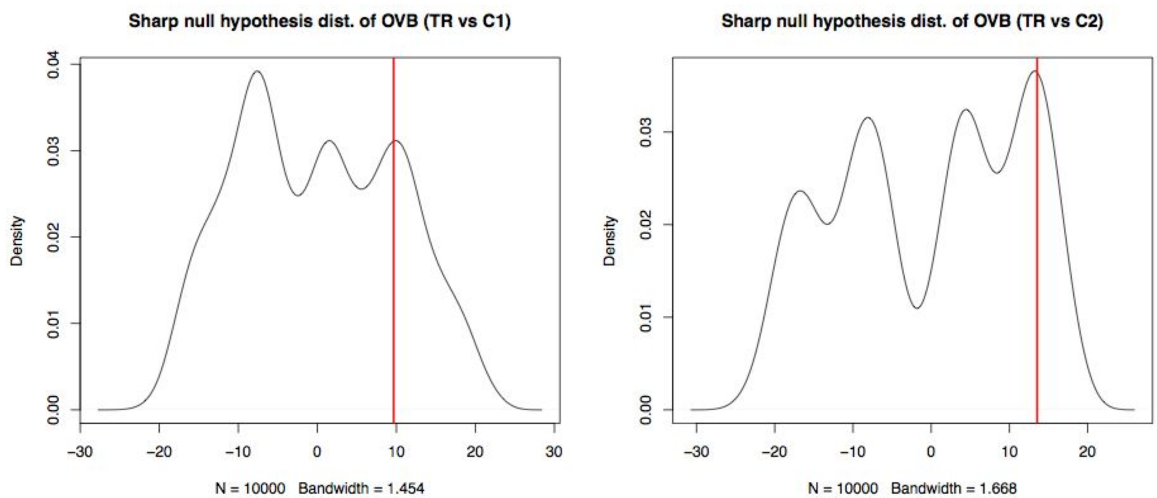
Table 6: Testing the Differences in Order Value per Buyer Across Weeks

Outcome Measures	No. of observations			Outcome Estimate			Effect Estimate		p-value	
	C1	TR	C2	C1	TR	C2	TR-C1	TR-C2	TR-C1	TR-C2
OVB	21	18	14	15.510	25.164	11.629	9.654 (10.194)	13.535 (11.768)	0.416 [0.382]	0.324 [0.208]

Source: Etsy and Google Analytics

The estimates show that there are practically significant differences in the Order Value per Buyer estimates for TR and C1 and for TR and C2. The average differences of $9.65 and $13.53 are almost 1.5 and 2 times (respectively) of the price of the cheapest item during the control weeks ($6.99). But the p-values of the estimates show no statistical significance. The high differences in the estimate could be due to the presence of two large value orders during the treatment week. The presence of these outliers may have influenced the overall estimate of OVB for the treatment week.

Figure 5: Sharp Null Hypothesis of Differences in Order Value per Buyer Across Weeks



Source: Etsy and Google Analytics

From the distributions, it is clear that the t-test will not give any reliable p-values as even 10,000 simulations cannot yield approximately normal distributions for the ATEs. The distribution shows the sign of the outlier high value orders which drives the non-normality of the distribution.

Order Quantity per Buyer (OQB)

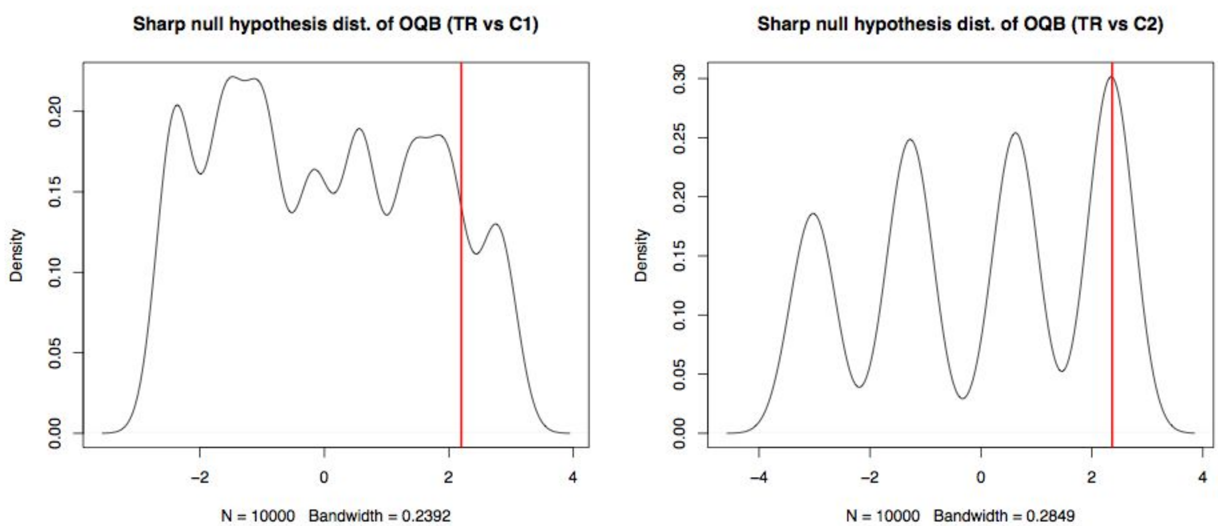
For analyzing the differences in Order Quantity per Buyer (OQB), the sales data was grouped by each unique buyer in the respective weeks. The ATE estimate of OQB tested by randomization inference are shown below in Table 7 and Figure 6.

Table 7: Testing the Differences in Order Quantity per Buyer Across Weeks

Outcome Measures	No. of observations			Outcome Estimate			Effect Estimate		p-value	
	C1	TR	C2	C1	TR	C2	TR-C1	TR-C2	TR-C1	TR-C2
OQB	21	18	14	1.524	3.722	1.357	2.198 (1.665)	2.365 (1.992)	0.239 [0.230]	0.358 [0.193]

Source: Etsy and Google Analytics

Figure 6: Sharp Null Hypothesis of Differences in Order Quantity per Buyer Across Weeks



Source: Etsy and Google Analytics

From the distributions again it is clear that the t-tests will not provide any reliable estimates. Again the results show practically significant differences in the OQB. Approximately 4 units (3.72) were bought on an average during the treatment week TR, compared to approximately 2 units (1.53) and 1 unit (1.35) bought respectively during control weeks C1 and C2. These results are however not statistically significant because the two high value orders have large amount of quantities. These two orders are consist of large amount of lower price items. The non-normality (spikes) of the sharp null hypothesis distribution confirmed the presence of these high quantity orders. These outliers possibly influenced the estimate for the treatment week.

Conversion Rate

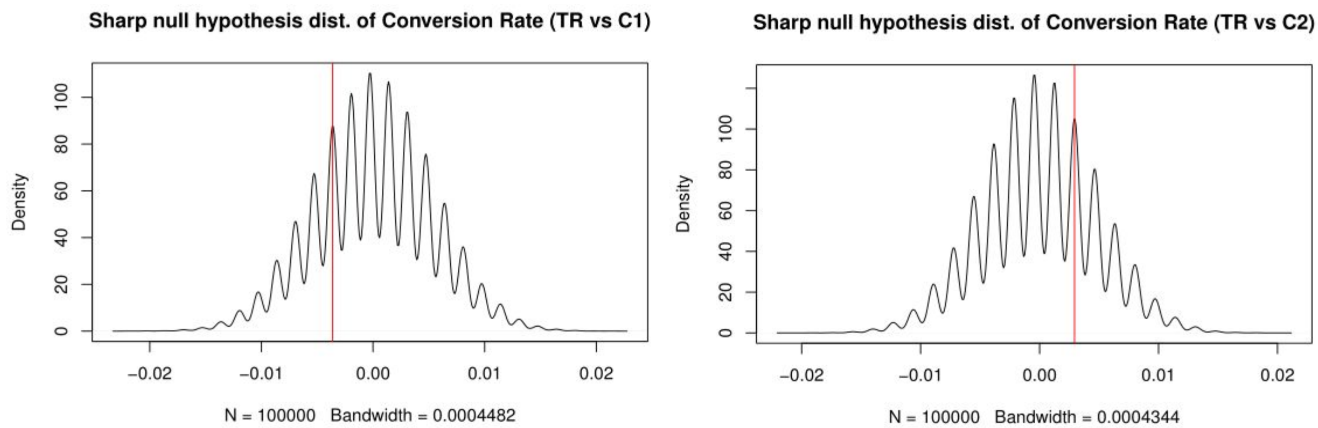
The treatment effect on conversion rates - percentage of web sessions which turned into purchases - and their corresponding p-values estimated by randomization inference are shown below in Table. Note that the number of sessions increased during the treatment week so the conversion rate provides a robust way to normalize number of purchases. The conversion rate for the treatment week was lower than that of the first control group, as a result of the increase in sessions, but was higher than the rate in the second control group. Overall, the differences were not statistically significant and the differences in means are small as well.

Table 8: Testing the Differences in Conversion Rates Across Weeks

Outcome Measures	No. of Observations			Outcome Estimate			Effect Estimate		p-value	
	C1	**TR**	**C2**	**C1**	**TR**	**C2**	**TR-C1**	**TR-C2**	**TR-C1**	**TR-C2**
CR	1161	1239	1129	0.019	0.015	0.012	-0.004 (0.005)	0.003 (0.005)	0.530 [0.496]	0.602 [0.541]

Source: Etsy and Google Analytics

Figure 7: Sharp Null Hypothesis of Differences in Conversion Rate Across Weeks



C.4. Heterogeneous Treatment Effect on Returning Customers

The following section compares treatment effect on new customers and returning customers. The idea is to see if the discount has a larger effect on which group of buyers across pre-treatment (C1) and treatment week (TR). If an order is placed by customers who has order records during the past year, this record is marked as from a returning customer. Regression is performed on Sales per Order (SO) as the outcome variable. Model 3 estimates the heterogeneous treatment effect given the interaction term "Returning Customer:Treatment". Result shows a large and statistically significant higher SO value among returning customers. Comparing to model 1 which shows no statistical significance of treatment, this implies that while treatment has no overall boost on SO, it has a large effect no returning customers. However, the sales order data reveals that the large amount of returning customer sales in the treatment week were due to 2 customers placing coupon on top of the discount offered and purchased a large amount of items. In total across C1 and TR, there are 7 returning customer orders. Given the small sample base, the 2 large value orders in the treatment week can bias the interpretation. Further studies which spans over longer period can help observe more returning customers to confirm the result.

	Dependent variable:		
	OrderValue		
	(1)	(2)	(3)
treat	9.034	6.687	−5.058
	(7.905)	(7.053)	(6.292)
Returning_Customer		31.648***	−6.351
		(9.348)	(11.317)
treat:Returning_Customer			69.175***
			(15.269)
Constant	14.805***	10.489**	15.671***
	(5.381)	(4.945)	(4.179)
Observations	41	41	41
R^2	0.032	0.257	0.522
Adjusted R^2	0.008	0.218	0.483
Residual Std. Error	25.240 (df = 39)	22.412 (df = 38)	18.216 (df = 37)
F Statistic	1.306 (df = 1; 39)	6.560*** (df = 2; 38)	13.461*** (df = 3; 37)
Note:			*p<0.1; **p<0.05; ***p<0.01

D. Conclusion

Through this experiment, the following observations are made:

1. Treatment week shows an obvious boost in total sales and total quantity of units sold compared to the control weeks and March 2016. However, these quantities cannot be tested with statistical significance as there is no way that we could obtain the distribution of overall weekly sales using data collected for three weeks.
2. Post-treatment week shows an obvious decrease in both total sales and quantity sold. The fall in the corresponding values of these variables during the post-treatment week (C2) is noticeably greater than the change in these values between the pre-treatment week (C1) and

treatment week (TR). Given there is no such trend observed for March 2016, it implies an intertemporal substitution effect from the treatment.

3. There is a considerable increase in the number of units sold on lower price units.
4. Although the Order Value per Buyer (OVB) and the Order Quantity per Buyer (OQB) had noticeable and practically significant higher values, the differences were not statistically significant.
5. The Sales per unit Item per Order (SIO) shows statistical significance for TR v.s. C1 but not for TR v.s. C2. The differences are practically significant for both TR-C1 and TR-C2. During the treatment week the quantity of the lowest priced item is higher. This explains the lower SIO value for the treatment week as compared to the control weeks C1 and C2. The treatment resulted in a change of purchasing behavior on lower price items.
6. Returning customer shows a strong positive heterogeneous treatment effect on Sales per Order (SO). However, given the small sample and 2 high value outliers, further studies which gathers more returning customers orders is required to confirm the result.

Although treatment shows a boost in sales and quantity sold, given the observed intertemporal substitution post-treatment, the overall sales effect of full site discount is not positive. The treatment boosts sales mostly in lower price items, pointing to a possible sales strategy focused on low price item discounts. Given the high sales fluctuation across months and among different customers, more observations would need to formulate an accurate baseline sales level. Ideally, future study should allow randomization on individual web session level to give bigger sample base. This would also help confirm the heterogeneous treatment effect observed among returning customers.

Reference

[1] The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales, Gerard J. Tellis, 1988

[2] The effectiveness of Web search engines for retrieving relevant ecommerce links, Bernard J. Jansen and Paulo R. Molina, 2005

[3] Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution, John G. Lynch, Jr. and Dan Ariely, 2000

Appendix

[1] Project Github repository:

https://github.com/jeffrey-hsu/Statistics-and-Causality/tree/master/causality/discount_effect_etsy_shop

[2] 3XUdesign Etsy Shop: https://www.etsy.com/shop/3xudesign