Exploring the Causal Effect of Product Images on Online Sales

Nathan Nusaputra, Ankitkumar Patel, Chris. Wen

W241 - Final project report

Abstract

In online marketplaces, where physical examinations of products are infeasible, product images and descriptions play key roles in improving consumer confidence. visual images are the powerful tools to convey crucial information about a product and its quality, and thus, they highly influence consumer choices. In this project, we explore how image quantity impacts online product sales. Our hypothesis is, product sales increase with the number of posted images in online marketplace since images provide a channel to build trust between buyers and sellers. In this study, we uncover the causal effects of image quantities on the online product sales through a controlled-experiment. We apply the applications of data mining and knowledge discovery approaches on the information collected from the experiment to test the hypothesis and identify the causal effects. Our causal analysis demonstrates that product sales volume and conversion rates can be increased by including either 1 or 5 photos of the product in the product listing. Including 1 photo increases sales volume by 2.1 units (4.4%) on average relative to 0 photos, while including 5 photos increases by sales volume 2.8 units (5%) on average relative to 0 photos. We find no statistical difference in the treatment effect of 5 photos relative to 1 photo. The outcome of the study can benefit not only to sellers to strategically increase online sales on highly competitive marketplace, but also to buyers to avoid adverse product selections.

Introduction

Background

Online shopping has exploded since the advent of the internet, enabling large multinational e-Retailers such as Amazon, Alibaba, and Shopify to accrue a collective multi-trillion dollar market capitalization.[1] Indeed, the already-mature business of online shopping received a further boost in the midst of global lockdown due to the Coronavirus Pandemic; both Amazon and Shopify posted earnings that vastly outperformed expected estimates in spite of global economic downturn.[2,3] Other online retailers have done similarly well in the face of forced closures of physical retail space. Despite the convenience and lower pricing that online shopping enables, the lack of physical interaction between buyers and sellers for physical products has long posed a barrier to purchase, particularly among secondhand or independent markets. Information on product quality and physical dimensions communicated only through text descriptions create a sense of anxiety and risk for consumers, due to the subjective nature of such descriptions, as well as the prevalence of fraud. Consumers do not

want to receive products that do not match their expectations, especially with the transaction cost of returns and disputes being nontrivial.

The use of product images may help to shore this trust gap by conveying more objective visual information of physical products to buyers -- ultimately enabling greater sales conversion rates. Indeed, studies have shown that product images are among the most influential risk-reducing factors in online shopping, which are driven by trust level, perceived risk, consumer attitude, and social influence.[4] Data also supports this by suggesting that clear and detailed pictures of products help improve online purchasing power.[5] Observational studies highlighted in literature have analyzed this hypothesis; in particular, studies have analyzed the correlative relationship between the number of images and sales rates on the Ebay platform, establishing a clear positive relationship.[6,7]

Research Question

While correlative behavior suggests that pictures do indeed help improve sales power of a listed physical product, a formal causal relationship cannot yet be established on this data. Therefore, the broad research question still remains: Do online sales increase with the number of posted images in a product listing?

This field experiment explores a subset of this broad question, examining the causal relationship between product sales and the presence of 0, 1, or 5 images for physical products sold on e-commerce platforms that best represent the “individual / re-sell / third-party” category.

Hypothesis

Based on prior literature and prevailing reasoning above, the null hypothesis posits that increasing the number of images increases the volume (or likelihood) of online sales in this e-commerce market category. Formally stated, in the equation:

$$y = \beta^T X + \varepsilon = \sum \beta_i x_i + \varepsilon$$

where y denotes the number sales for a product in the “individual / re-sell / third-party” market category, x_i denotes a sales determinant, and β_i represents the causal impact of that determinant, the null hypothesis posits that the coefficients $\beta_{j=0} < \beta_{j=1} < \beta_{j=5}$ for the sales determinants $x_{j=0}, x_{j=1}, x_{j=5}$ where $j \in \{0, 1, 5\}$ refers to the number of images attached to the product posting.

Stated another way, holding all other things constant, we expect the number of sales to increase with the number of images attached to a posting, from 0, to 1, to 5 images.

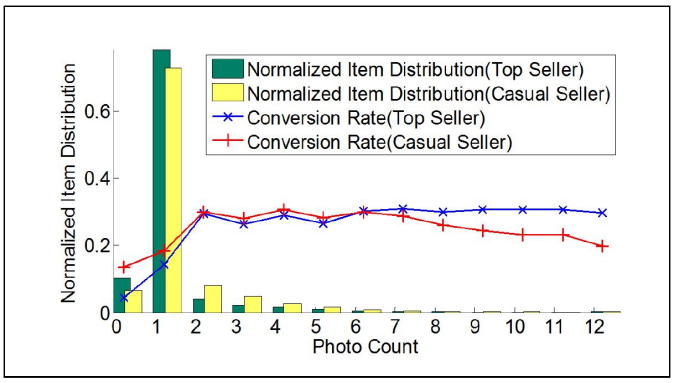
Experimental Design

A randomized controlled experiment was proposed to investigate the null hypothesis, which would entail posting products with varying numbers of images to record the number of sales.

Justification/Power Calculation

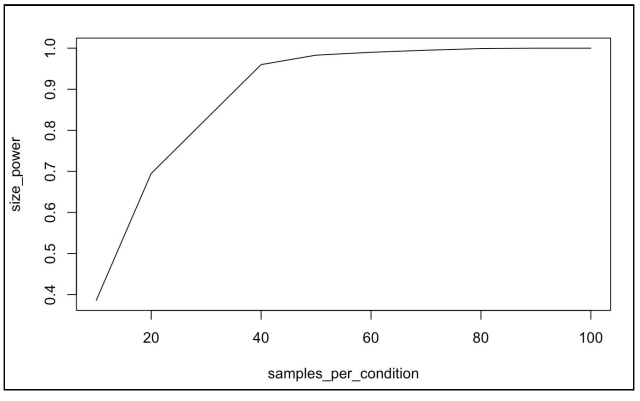
Pre-experimental power analysis enabled accurate experiment scale. A previous study exploring conversion rate of product sales to product photo count was used to estimate the expected treatment effect.[5] Results from the experiment, which was conducted on the eBay platform, are summarized in Figure 1. Results suggest that for casual sellers, the conversion rate of shoppers to buyers for a product increases from 14% to 18% when the number of photos increases from 0 to 1. Similarly, the conversion rate increases from 18% to ~30% when the number of product photos increases beyond 1, to as many as 7, before it begins to decrease.

Figure 1: Conversion Rate vs. Product Image Count, eBay Platform



For considerations of labor and manual effort, the proposed experiment explores the incremental impact of listing only 0, 1, or 5 product images. Using the observed effect sizes from the experiment seen in Figure 1, a power analysis was conducted to determine the number of samples needed for each treatment group through power analysis. Results can be seen in Figure 2.

Figure 2: Statistical Power Analysis; Power vs. # of Samples



For the mean conversion rate of 14% with standard deviation of 5% for the Treatment Group 0 (control) and the mean conversion rate of 18% with standard deviation of 5% for the Treatment Group 1 (products with 1 image), Figure 2 shows the probability of rejecting the null hypothesis at the 95% confidence-level as the sample size increases. The results from Figure 2 indicate that we need at least 30 samples in each treatment group to reject the null hypothesis with probability 0.80. Treatment was ultimately assigned by geography at the

zip-code level, implying that the experiment required at least 30 zip codes for each treatment.

Timeline

Experiment timeline proceeded as follows:

July 1	July 6	July 12	July 18	July 25
Sign up on the Applications	Pilot Study	End Pilot Study and Review Experiment	Start experiment by posting pictures	End Experiment and record results

A pilot study was conducted with a broad selection of products used across 99 completely randomly selected zip codes. Information gleaned from the pilot study was used to fine-tune experimental specifications, including product and category selection, zip code population thresholds, and experimental runtime. The pilot experiment also enabled understanding of platform security measures to avoid, including how to avoid getting an account banned or a product listing removed. This was a significant barrier faced in the pilot study.

Individual postings were initiated on a Saturday evening between the hours of 9 PM and 3 AM Pacific Daylight Time, and were allowed to run for a total of 7 days, in order to eliminate heterogeneous and cyclical day-of-week effects.

Setup

3 different e-commerce platforms were chosen to represent the “individual / re-sell / third-party” market category: Craigslist, LetGo, and Offerup. These platforms were chosen due to availability, trends, and feature consistency -- in particular, the ability to select and specify individual geographies.

Table 1: List of Products and Respective Prices

ID	Product	Price
1	Love Letter Board Game	$ 5
2	Logitech G Pro Gaming Mouse	$ 10
3	Bicycle from Rocket Next	$ 35
4	Apple Keyboard	$ 65
5	VTECH Sit to Stand Learning Walker	$ 5
6	Kid's Study Table	$ 35
7	Apple Mouse	$ 35
8	Patio Chairs	$ 20
9	Retrospec Longboard	$ 40
10	Speaker	$ 5
11	Keurig	$ 50

11 distinct physical products were randomly selected from heterogeneous categories including toys, home decor, electronics, kitchen appliances, furniture, and games. Products and categories were selected based on ease of acquisition, in order to ensure the ability to generate standard product images, as well as viability informed from the Pilot Study. Certain product categories with extremely high price variance and / or no consumer interest were excluded. Product prices were selected based on the median price observed across all platforms enrolled in the experiment, and were determined via key-word search for the

particular product for postings in the last 12 months. Prices ultimately ranged from $5.00 to $70.00. Table 1 shows the listing of all final products and their respective prices.

Each product was exposed to all three treatment groups (11 products * 3 treatment groups = 33 total postings), which included:

(I) Control: Pure Text Description

(II) Treatment 1: Text Description + 1 Product Image

(III) Treatment 2: Text Description + 5 Product Images.

Text descriptions were structured to be of passive tone, contain 150 total words, and 3 descriptive bullet points. Products used all satisfied product condition categories for “Used - Good” or “Used - Very Good”, as laid out by Amazon.[8] This criteria was chosen due to consensus use as the standard of excellence in the retail industry. Images were taken in standardized fashion. For the single-image treatment, a profile of the product encompassing the entire product and as much surface area coverage as possible was taken. Products were taken against standard wood-grain or concrete backgrounds. For the five-image treatment, two additional profile images were added at 120 degree and 240 degree rotations from the first profile image, and an above / below image was also provided.

Each posting was also exposed to each e-commerce platform (33 listings * 3 platforms = 99 total final postings -- 33 for each treatment group as specified by the 30 minimum in the pre-experiment power analysis). Each individual final posting was randomly assigned to one of 99 zip codes, which were each randomly selected from a list of U.S. zip codes with a population of more than 10,000 people, according to the 2016 Census. Zip codes with fewer than 10,000 were excluded to increase likelihood of response and data collection.

Randomization Engineering

Subjects in this experiment are consumers from the United States who are seeking for new and used products on well-known peer-to-peer marketplaces like LetGo, OfferUp, and Craigslist. We decided to use both clustered random assignment with clustering on the zip code level and standard random assignment with each zipcode as a data point. We performed all our randomizations for this experiment in R. We randomly chose 99 zip codes with populations of over 10,000. To avoid product-specific biases, 11 products are randomly selected from heterogeneous categories, such as toys, home decor, electronics, kitchen appliances, furnitures, and games with prices ranging from $5.00 to $70.00. A product is listed over a randomly selected geographic US market targeting subjects with the varieties of demographics backgrounds. We show the randomization graphically below.

As shown in Figure 1, the leftmost column shows each of the products that is posted for sale which are applied to different types of treatments(control, treatment 1, treatment 2) over different geographical markets. The multiple treatments are then applied to consumers through one or more of the applications(OfferUp, LetGo, Craigslist). Finally, each of these products are posted on a single geographical market using one of the treatments to avoid any biases.

Figure 3: Experimental Design Diagram

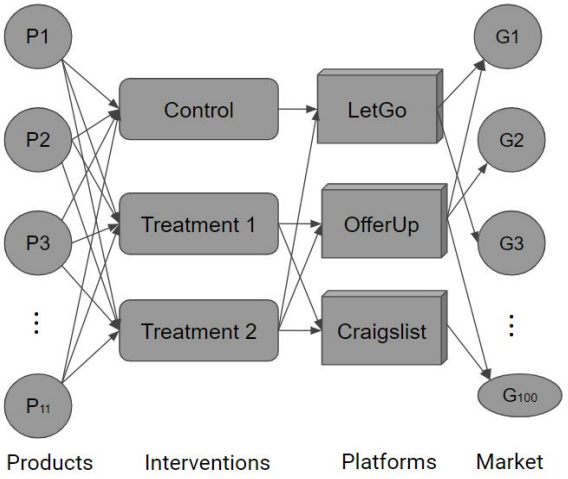


Table 2: Experimental Design Results Post-Randomization

	Application	Treatment	Count
1	letgo	placebo	10
2	letgo	t1	12
3	letgo	t2	11
4	cragslist	placebo	12
5	cragslist	t1	9
6	cragslist	t2	12
7	offerup	placebo	11
8	offerup	t1	12
9	offerup	t2	10

Outcomes Collection

The number of unique individual product inquiries were counted for each posting to represent a “sale”. The applications LetGo and OfferUp also supplied unique total product view counts; however, Craigslist did not. These view counts were used to represent the number of individual experiment enrollees (compliers) for listings within those two applications, which were clustered at the zip code level. With this in mind, outcomes and treatment effects were measured in two similar but different analyses.

In the first analysis, each individual zip code was defined as an experimental enrollee, and the response was defined as the cumulative number of responses to the product listing. This integer outcome was compared across treatment groups to determine the estimated Average Treatment Effect (ATE) when moving from Control to Treatment 1 to Treatment 2.

In the second analysis, individual consumers are treated as experimental enrollees. Each individual view, then, is treated as a compliant experimental enrollee. Those individuals within the zip code who did not view the product are seen as non-compliant; they are exposed to the treatment (product is listed openly on each platform), but do not take (they did not view the actual listing). Non-compliance is not visible to the experiment, as no data

exists that reveals the total number of unique users on the application during the experimental time frame. Individuals are therefore clustered at the zip code level, as treatment assignment was performed at the zip code level. In this analysis, the outcome is binary (viewed & responded, or viewed & did not respond), and outcomes are compared across treatment groups to determine the estimated Complier Average Causal Effect (CACE). This view offers more granular data and more data points, suggesting potentially more robust results. However, due to the inability of the Craigslist platform to provide this information, this analysis is done with only 66% of the data afforded by the experimental design.

Results, Causal Analysis & Discussion

Data Cleansing

Data was first cleansed and transformed into usable format. Only unique inquiries were counted; follow-up and duplicate inquiries were excluded. Automated / robotic inquiries were also removed; determination was made based on email address or profile of responder, as well as Google-search results of response texts (standard scripts are used in spam inquiries, and other sellers have posted warnings on the internet). Responses that did not explicitly inquire about purchasing the product were also excluded (offers to sell product, view images of product, etc).

For the first analysis, counts of all product responses were tallied at the zip code level. Results can be seen in Table 3.

Table 3: Experimental Results

	Application	Treatment	Count	Total_Responses	Total_Views
1	letgo	placebo	10	5	176
2	letgo	t1	12	60	567
3	letgo	t2	11	80	695
4	cragslist	placebo	12	2	-
5	cragslist	t1	9	11	-
6	cragslist	t2	12	15	-
7	offerup	placebo	11	4	200
8	offerup	t1	12	17	498
9	offerup	t2	10	13	395

For the second analysis, data subsetted to include only the platforms that offer views data. Each observation was then flattened into n records, in which n refers to the number of views received. Of the n records, r were flagged with outcome 1, where r refers to the number of responses. The remaining $n - r$ records were flagged with outcome 0. Outcome 1 indicates that this individual responded, while outcome 0 indicates that this individual did not respond. This process can be seen in Figure 4

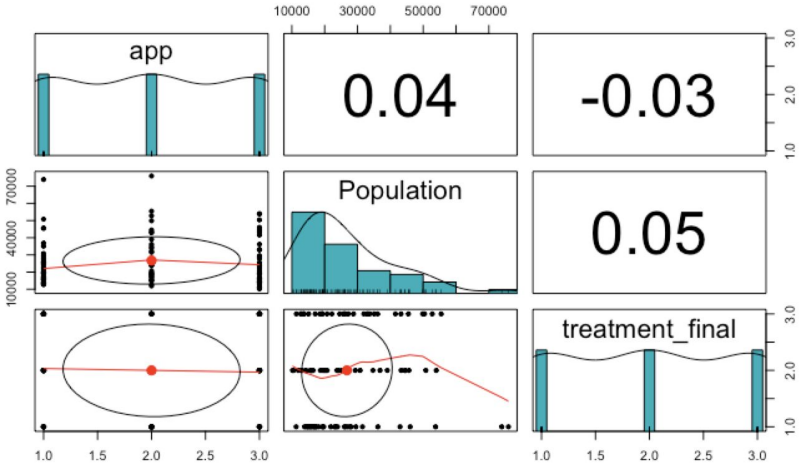
Figure 4: Data Flattening Process for Analysis II

Application	Treatment	Count	Total_Responses	Total_Views
letgo	placebo	10	5	176

Application	Treatment	Count	Total_Responses	Total_Views	Outcome	ID
letgo	placebo	10	5	176	1	1
letgo	placebo	10	5	176	1	2
letgo	placebo	10	5	176	1	3
letgo	placebo	10	5	176	1	4
letgo	placebo	10	5	176	1	5
letgo	placebo	10	5	176	0	6
letgo	placebo	10	5	176	0	7
letgo	placebo	10	5	176	0	8
letgo	placebo	10	5	176	0	9
⋮	⋮	⋮	⋮	⋮	⋮	⋮
letgo	placebo	10	5	176	0	176

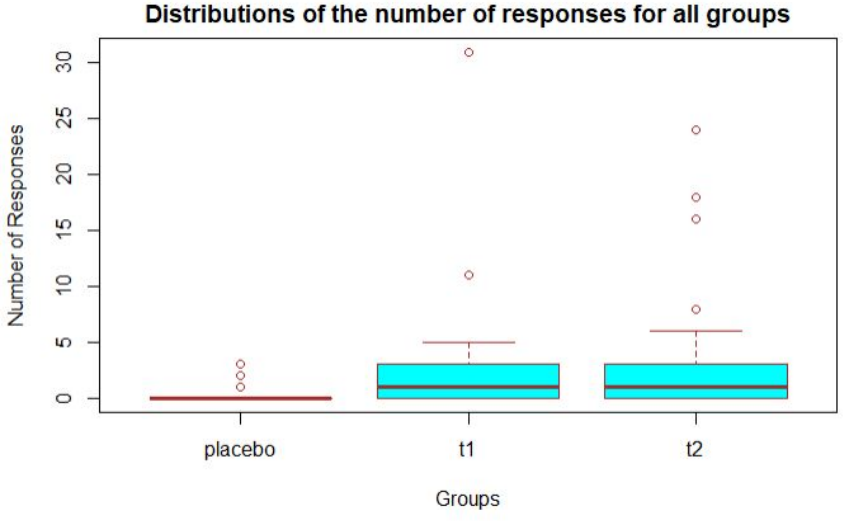
Experimental Data Analysis

Figure 5: Randomization Checks, Correlation Plots & Histograms



The first EDA we ran was confirming that each treatment and application had even amounts of data. We see on the top right that each application had similar amounts of treatment. The middle graph shows a right skewed distribution for population which follows our original population distribution for zip codes. Lastly in the bottom right graph we see that the treatments are evenly distributed amongst applications.

Figure 6: Exploratory Boxplot of Treatment Effects



The second EDA we did was on the outcome variable, responses. This showed that the placebo(control) group had the least number of views with a very small range. The treatment groups, on the other hand, had a larger range and were similar to each other.

Analysis I & Discussion

To analyze the average treatment effect (ATE), we perform linear regression analysis with heteroskedastic-robust standard errors, The results are as shown in Table 4.

Table 4: Analysis I - Raw Response Volumes

	Dependent Variable: Responses				
	Responses				
	(1)	(2)	(3)	(4)	(5)
treatment_finalt1	2.333**	2.015	2.120**	2.129**	4.471
	(0.984)	(1.269)	(0.948)	(0.949)	(2.730)
treatment_finalt2	2.939***	2.787**	2.840***	2.820***	6.822**
	(0.992)	(1.082)	(0.917)	(0.902)	(2.681)
appcragslist			−3.439***	−3.500***	−0.399
			(1.275)	(1.287)	(0.343)
appofferup			−3.278**	−3.295**	−0.159
			(1.290)	(1.298)	(0.388)
Population				0.00001	0.00002
				(0.00003)	(0.00004)
treatment_finalt1:appcragslist					−3.416
					(2.789)
treatment_finalt2:appcragslist					−5.805**
					(2.775)
treatment_finalt1:appofferup					−3.344
					(2.850)
treatment_finalt2:appofferup					−5.963**
					(2.796)
Constant	0.333**	−0.819	2.676***	2.365*	0.016
	(0.130)	(0.565)	(0.966)	(1.294)	(0.940)
Product Fixed Effects?	No	Yes	No	No	No
Observations	99	99	99	99	99
R^2	0.074	0.261	0.190	0.191	0.249
Adjusted R^2	0.055	0.158	0.156	0.148	0.173
Residual Std. Error	4.543 (df = 96)	4.289 (df = 86)	4.294 (df = 94)	4.314 (df = 93)	4.249 (df = 89)
Note:					*p<0.1; **p<0.05; ***p<0.01

Analysis I included 5 individual successive regressions, each adding additional covariates. Regression (1) estimates the ATE purely using the treatment status, and finds statistically significant treatment effects of both the 1 photo (*p < 0.05, 95%CI = [0.38,4.29]*) and 5 photo (*p < 0.05, 95%CI = [0.97,4.9]*) treatments relative to the text-only control. Both 95% confidence intervals are also entirely in the positive region, reinforcing the positive impact of photos in both cases. We note, however, that the estimates are similar and their confidence intervals overlap substantially. Thus, while this suggests statistically significant treatment effect relative to the text-only control, it does not suggest an improved treatment effect of the 5 photo treatment relative to the 1 photo treatment.

Product fixed effects were included in Regression (2) on the idea that individual products might have specific tendencies to elicit consumer responses, and the need to control for this variance. Inclusion of this, however, results in an increase in standard errors of estimates, an unexpected result. Removal of these fixed effects and inclusion of application-specific fixed effects can be seen in Regression (3). Cursory view of the data seemed to imply that LetGo

performed vastly superior relative to either OfferUp or Craigslist. Regression (3) appears to confirm this discrepancy, as both Craigslist and OfferUp yield negative coefficients, indicating that LetGo inherently produces more responses than either of the other two applications. Indeed, both 95% confidence intervals are entirely in the negative value region (Craigslist: *p < 0.01, 95%CI = [-6.0,-0.90]; OfferUp: p<0.05, 95%CI = [-5.8,-0.72]*). Both estimates are also statistically significant relative to LetGo at a 5% critical value; however, we note again that the two estimates are similar with overlapping confidence intervals. The estimates for Craigstlist and OfferUp do not appear to be statistically significant relative to each other. With the inclusion of these application fixed effects, however, treatment effect estimates do not change substantially (Regression (3) vs. Regression (1)), and standard errors improve marginally.

Population is included in Regression (4) as a randomization check; population as an attribute for Zip Code does not affect any of the treatment effect estimates; the estimate is both practically insignificant (even with zip code populations in excess of 100k; median is more in the 10k order of magnitude) and statistically insignificant (SE = 3*Estimate Size). This result is expected for a well-randomized experiment, as we would not expect any zip code attributes to correlate with treatment.

Finally, heterogeneous treatment effects for each application are explored in Regression (5). Here, results suggest that there are statistically significant heterogeneous effects for the 5 photo treatment for LetGo relative to OfferUp and Craigslist, but none for the 1 photo treatment. The 5 photo treatment effect is estimated at approximately 6.8 responses relative to the text-only treatment, and this is statistically significant. Together, this would imply that a product listed on OfferUp or Craigslist with 5 photos will see an increase in responses relative to baseline by +6.8 responses, but an incremental decrease by approximately 6 responses for not being listed on LetGo. Overall, the same pattern: while significance exists relative to LetGo as the baseline, significance is not supported between OfferUp and Craigslist.

Ultimately, use of an F test to compare model explanatory power can be used to identify the best model. While Regression (5) appears to include many more covariates and explores heterogeneous treatment effects, a comparison with Model 3 yields an F Statistic of 1.4, for a p-value of 0.23 (Figure 7). We see that the incremental explanatory power of variance in Model 5 relative to Model 3 is not statistically significant.

Figure 7: F-Test of Regression (5) vs. Regression (3)

```
## Analysis of Variance Table
##
## Model 1: Responses ~ treatment_final * (app) + Population
## Model 2: Responses ~ treatment_final + app
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     89 1607.1
## 2     94 1733.4 -5   -126.36 1.3996 0.2321
```

Taking Regression (3) as the primary model reflective of treatment outcomes, the ATE of 1 photo relative to text-only is approximately 2.1 additional sales, while the ATE of 5 photos relative to text-only is approximately 2.8 additional sales. The incremental 0.7 sales seen in

the 5 photo treatment vs. the 1 photo treatment is low relative to the literature upon which this experiment's power analysis was conducted. This incremental difference is not statistically significant; while the conclusion that photos help sales rate relative to no photos, no conclusion can be drawn about larger numbers of photos relative to small numbers of photos (at least not 5 vs. 1).

Analysis II

As information and behavior on non-compliers is not available, Analysis II deals purely with the Complier Average Causal Effect (CACE). The CACE was estimated via Linear Regression with cluster-robust standard errors (clustering dimension = Zip Code); results can be seen in Table 5. The outcome is measured in fraction of views converted to sales, or the sales conversion rate.

Table 5: Analysis II - Individual Enrollee Outcomes

	Dependent Variable: Response for each View Observation			
	outcome			
	(1)	(2)	(3)	(4)
treatment_finalt1	0.048**	0.050**	0.044**	0.077**
	(0.024)	(0.025)	(0.021)	(0.037)
treatment_finalt2	0.061***	0.071***	0.050***	0.087***
	(0.019)	(0.020)	(0.016)	(0.022)
appofferup		−0.067***	−0.066***	−0.008
		(0.016)	(0.019)	(0.014)
treatment_finalt1:appofferup				−0.063
				(0.039)
treatment_finalt2:appofferup				−0.074***
				(0.025)
Constant	0.024***	0.033	0.059***	0.028***
	(0.007)	(0.047)	(0.014)	(0.009)
Product Fixed Effects:	No	Yes	No	No
df:	2528	2527	2517	2515
Observations	2,531	2,531	2,531	2,531
R^2	0.006	0.041	0.023	0.025
Adjusted R^2	0.006	0.036	0.021	0.023
Residual Std. Error	0.256	0.252	0.254	0.253
Note:				*p<0.1; **p<0.05; ***p<0.01

Analysis II included 4 individual successive regressions, each adding additional covariates. Regression (1) estimates the CACE purely using the treatment status, and finds statistically significant treatment effects of both the 1 photo (*p < 0.05, 95%CI = [0.00036,0.096]*) and 5 photo (*p < 0.05, 95%CI = [0.024,0.098]*) treatments relative to the text-only control. Both 95% confidence intervals are also entirely in the positive region, reinforcing the positive impact of photos in both cases (though the lower end of the 1-photo confidence interval is very close to 0). We note again, as in the prior analysis, that the estimates are similar and their confidence intervals overlap substantially. Thus, while this suggests statistically significant treatment effect relative to the text-only control, it does not suggest an improved treatment effect of the 5 photo treatment relative to the 1 photo treatment.

Product fixed effects were included in Regression (2), again on the idea that individual products might have specific tendencies to elicit consumer responses, and the need to control for this variance. Inclusion of this, like in the prior analysis, results in an increase in standard errors of estimates, an unexpected result. Removal of these fixed effects and inclusion of application-specific fixed effects can be seen in Regression (3). In this case, note that Craigslist information was subsetted out of the data due to availability. Regression (3) again confirms the existence of a fixed effect for the OfferUp relative to LetGo. On average, it appears that OfferUp sees an absolute decrease in sales by 6.6% relative to LetGo; this result is statistically significant to the 1% critical value (*p < 0.01, 95%CI = [-0.10,-0.030]*). The confidence interval is entirely negative, suggesting that OfferUp sees between 3% and 10% fewer sales relative to LetGo.

Finally, in Regression (4), heterogeneous treatment effects were explored again. Results are similar to the prior analysis, suggesting a statistically significant heterogeneous treatment effect for the 5-photo treatment based on the application used. Use of OfferUp relative to LetGo when applying a 5-photo treatment appears to have an incremental decrease in response rate of 7% (*p < 0.01, 95%CI = [-0.12,-0.025]).* The 5 photo treatment effect is estimated at approximately 8.7% additional responses relative to the text-only treatment, and this is statistically significant. Together, this would imply that a product listed on OfferUp with 5 photos will see an increase in responses relative to baseline by +8.7%, but an incremental decrease by approximately 7% for not being listed on LetGo.

An F test is again used to compare model explanatory power can be used to identify the best model. While Regression (4) appears to include many more covariates and explores heterogeneous treatment effects, a comparison with model (3) yields an F Statistic of 2.997, for a p-value of 0.05, just outside the 5% critical value (Figure 8). We see that the incremental explanatory power of variance in Model 5 relative to Model 3 is not compelling enough from a statistical significant standpoint to justify its use.

Figure 8: F-Test of Regression (5) vs. Regression (3)

```
## Analysis of Variance Table
##
## Model 1: outcome ~ treatment_final * app
## Model 2: outcome ~ treatment_final + app
##   Res.Df    RSS Df Sum of Sq     F  Pr(>F)
## 1   2525 162.21
## 2   2527 162.60 -2  -0.38507 2.997 0.05011 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Taking Regression (3) as the primary model reflective of treatment outcomes, the CACE of 1 photo relative to text-only is approximately 4.4% increase in absolute sales conversion rate, while the CACE of 5 photos relative to text-only is approximately 5% increase in absolute sales conversion rate. The incremental 0.6% seen in the 5 photo treatment vs. the 1 photo treatment is again low relative to the literature upon which this experiment's power analysis was conducted. This incremental difference is not statistically significant in this experiment; while the conclusion that photos help sales rate relative to no photos, no conclusion can be

drawn about larger numbers of photos relative to small numbers of photos (at least not 5 vs. 1).

Conclusion

As the world continues to move and gravitate away from physical shopping, the need to understand retail consumer needs in the market for physical products will only increase. Prior literature has investigated the determinants of consumer online purchases of physical goods, and validates intuition that product images can convey valuable information and foster trust between buyer and seller.

This experiment seems to corroborate this, finding that inclusion of product images in product listings produces more sales at a greater rate when compared to the absence of images in product listings. This experiment, however, does not corroborate results in prior literature that suggest that more is better (to a point). Treatment estimates for the use of 5 photos vs. 1 photo had very small practical significance, and no statistical significance. Still, this experiment suggests that, for a normal product in the “individual / re-sell / third-party” markets, including 1 or 5 photos can increase sales volume by up to 5 units (upper bound of 95% confidence interval), or increase sales conversion rates by up to 8% (upper bound of 9% confidence interval). Data also suggests that, if deciding between the marketplace platforms of OfferUp, LetGo, or Craigslist, LetGo will generally have the largest sales volume and sales conversion rate -- and may even provide incremental boosts for postings with images. Thus, using the findings of this study, sellers can strategically increase their online sale and buyers can avoid the selection of adverse products.

References

1. https://www.fool.com/investing/2020/08/02/the-3-largest-e-commerce-stocks-in-2020.aspx
2. https://variety.com/2020/tv/news/amazon-earnings-q2-2020-coronavirus-1234721317/
3. https://www.investors.com/news/technology/shopify-earnings-shopify-stock-q2-2020/
4. E. M. Bland, G. S. Black, and K. Lawrimore, Risk-reducing and risk-enhancing factors impacting online auction outcomes: empirical evidence from ebay auctions, Journal of Electronic Commerce Research, 8(4):236–243, 2007.
5. D. Koehn, The nature of and conditions for online trust, Journal of Business Ethics, 43(1):3–19, 2003.
6. G. Lewis, Asymmetric information, adverse selection and online disclosure: The case of ebay motors, The American Economic Review, 2009.
7. Wei Di, Neel Sundaresan, Robinson Piramuthu, Anurag Bhardwaj, Is a picture really worth a thousand words? - On the role of images in e-commerce, WSDM 2014 - Proceedings of the 7th ACM International Conference on Web Search and Data Mining. 633-642, 2014
8. https://sellercentral.amazon.com/gp/help/external/200339950