ECON 492: Research Methods in Economics

Daniel Ortega

April 7th, 2020

An Empirical Analysis: Do Online Amazon Reviews Have an Impact on Price?

1. Introduction

With the advent of technology, more and more people are purchasing items online. One of the biggest online retailers is Amazon, which dominates the market with almost half of the U.S market share (Yahoo Finance, 2020). Compared to shopping in traditional retail stores, where a commodity's features could be analyzed and seen in person, with online shopping one does not have this luxury. Instead, many online retailers have implemented a review system in which purchasers can leave a review and rate the satisfaction that they have received from the item. For potential buyers of a smartphone, ratings can be a proxy for quality. If a smartphone has a high rating, then that signifies that previous purchasers were happy with the quality of the product.

One of the problems with purchasing online is the asymmetry of information, potential buyers do not know the product that they are going to get until it is received in person. Potential buyers use many online tools such as online ratings to gain more information about the product. The more information that is gained from these tools (ratings, reviews, photos, YouTube reviews, etc), the more asymmetric information is reduced. Amazon specifically, has a five-star rating system in which users can rate a product, leave feedback, and even take photos of the product. With all this data available on Amazon, and many other online retailers, many economists have examined some of these variables and specifically the relationship between online rating and price.

My paper will focus primarily on Amazon's smartphone market in which the relationship between online reviews and prices will be analyzed. My hypothesis is that average ratings will have a negative relationship with prices, all else equal. I am primarily focusing on the online smartphone market to precisely extract the relationship between online reviews and prices from my data, avoiding any unnecessary noise from trying to focus on multiple products. My research is important because it gives more insight on how online retail prices can be potentially impacted by online ratings.

2. Literature Review

In one study, researchers found that Airbnb prices are strategic complements and are influenced by the review score (rating), characteristics of room, and features of neighborhood (Lawani, Reed, Mark, & Zheng, 2019). This means that ratings and prices mutually reinforce each other. They also found that not only do good ratings cause prices to increase, but it also causes rival Airbnb rentals to increase their prices. This study looked at many different variables that could affect the price of an Airbnb rental such as the distance of other Airbnb competition, location, and number of reviews. They conducted multiple models, including an Ordinary Least Squared one. The OLS regression reported that rating was a significant determinant of price at the 1% level. They also did a log-level model to be able to easily interpret the effect of a variable on the price in terms of percentage.

Another study analyzed the relationship between professional wine ratings and reviews comparing many different past studies conducted on this topic using a technique called Akaike Information Criteria (Snipes & Taylor, 2014). In these past studies, they found that wine reviews affected prices of many different wines while holding other variables such as type of wine constant. Snipes and Taylor further reinforced these findings and concluded that there was a causal relationship between ratings and price by comparing several previous models. Higher professional ratings for a wine meant an increase in price for said wine. It should be noted that compared to these professional ratings that are published in wine journals, Amazon ratings are done by random people and are published in real time. People who buy wine or any product, might not respect the review of a random person over an expert..

One study conducted looked at how trending status and online ratings affect the price of homogeneous products (Cenk Kocas & Can Akkan, 2016). The homogeneous products were books, and they conducted an ordinary least squared model as well as a two staged least squared model. They controlled twenty-four book genres and their model suggested lower prices for books that were high rated and trending. This is contrary to my hypothesis; however, I believe phones are better homogeneous products since the factors that affect their price are easily controllable compared to a book. Since this study also is using data from Amazon, it's very possible that I will see similar results, but it is important to mention that I am not also controlling for trending status.

In a lot of these studies, ratings are treated as given and none really try to assess the validity of these ratings. One study analyzes the validity of Amazon ratings and assesses the motivation behind why people rate and what kind of distortion might take place (Lafky, 2014). In his research, he discovered that people are motivated by the desire to reward or punish sellers and to inform future buyers. He showed that any cost of rating, no matter how small, can create a "blind spot" in the ratings distribution and cause inaccurate ratings. The higher the cost of ratings the more polarizing and inaccurate ratings become. He also talks about the possibility of there being additional motivations such as not wanting to review because there are already a lot of reviews for a product or wanting to be the first review. It is important to mention that Amazon has recently changed their rating policy since this study, and consumers can now leave a star rating without having to write a review. This lowers the cost of rating, which according to this study's results, will cause an increase in the accuracy of the ratings.

While the previous study examines the validity of Amazon reviews, another study examines the self-selection and information role of online product reviews. They created a model that examines how idiosyncratic preferences of early buyers can affect long term consumer purchase behavior. They concluded that self-selection bias, if not corrected, decreases consumer surplus. This occurs when early buyers hold different preferences than do later consumers about the quality of a given product (Li & Hitt, 2008). In this study, they specifically looked at book ratings on Amazon. They found that products have consistent increases and decreases in ratings, in which can be explained by early buyers being avid fans of the product and later buyers being average consumers who give a more unbiased rating.

These studies have illuminated some of the future problems I might encounter such as self-selection bias and whether my ratings are even valid. Considering Lafky's assessment on the validity of Amazon reviews, I am using smartphones that have a decent amount of reviews, above thirty, to show that there is a low cost of reviewing as opposed to a phone that has a low amount of reviews. The results of Li and Hitt are interesting, and with my own research, I am also looking at the distribution of ratings except I am not looking at how they change over time because my data is cross sectional. The studies conducted on the prices of wine, Airbnb's, and books are the most similar to mine research since they are directly looking at how rating affects price. My research will further contribute to their findings by looking at the same question in the context of smart phones.

3. Theory

Prices are determined by both the demand and supply of a given good. Supply is determined by several determinants which includes price of good, number of sellers, price of inputs, price of related goods, technology, and suppliers' expectation. While demand is determined by price of good, income of buyers, price of related goods, expectations, and consumer preferences. Ratings represent not only quality but the preferences of other consumers. If a good has a high rating then that signifies that many consumers have a strong preference for the good, and vice versa, if the good has a low rating then that signifies many consumers have a low preference for the good. If there is an increase in the preference for a good, we know from the law of demand that the demand will increase for the good and so too will the price due to a right shift in the demand curve (See Figure 1.) Additionally, I believe ratings might affect a supplier's expectation. If a supplier sees that, let's say their smart phone, has a high rating they might be compelled to raise their price, shifting supply to the left (See Figure 2). My hypothesis is that as ratings increase for smartphones, the price of the smartphone will also increase, holding other things equal.

4. Empirical Methods

A. Description of Data

My data is from Amazon.com because I am limiting the smartphones to only ones sold and fulfilled by Amazon. That way I am controlling for shipping companies who might have an impact on the rating, and only allowing for one shipper. I control for common features of a phone such as storage and megapixels of camera. I am using phones whose storages are 32 and 64 gigabytes. 46% of my phones are 64 gigabytes while the remaining 54% are 32 gigabytes. Megapixels varies highly, in which I have collected data on the rear camera. Rear camera usually ranges from 8-20 MP and the average is 14.3. Battery is an important factor that I am controlling for which is measured in milli ampere hour(mAh). I am also controlling for whether the phone is unlocked or not because phones that are tied to contracts tend be cheaper than those that are not. I am also controlling for competition in which I am controlled for, the number of new and used substitutes. This competition is between sellers who sell the same model phone, both used and new. Brands is another variable I have collected and controlling for, in which I will mainly be looking at Apple, LG, Motorola, Samsung, and Google. I have another variable called "Other" which is acting as my reference group so it will not be in my model, but it's comprised of 26.02% of my phone sample and is composed of lesser known brands. Lastly, I have collected data on the average rating for each individual smartphone. As seen in Table 1, the average rating given price is around the mean which is 3.76. There are two phones in my samples that have no rating and are outliers. My sample size is 74.

It is important to mention how the average ratings are determined. Amazon uses a machine learning model to determine a product's average rating. The proprietary model is not public, and specifics are not known. What is known is that only verified purchases contribute to the average rating and the model considers factors such as how the recent the review was. However, I believe this is advantageous because by only allowing verified purchases this prevents any distortion of the rating by people who did not buy the product and may have gotten it for free. The fact that it gives more weight to recent reviews is also good, since it will allow the review to reflect the current price better since Amazon is known to marginally shift prices of their products. With this being said, I am using phones that are newer since newer phones have stable prices.

2. Description of the Empirical Model

I am using the same methods as prior studies, but in a new context. There have been a few studies done in the recent years on the relationship between rating and price. Notably, one done in the context on how Airbnb reviews affect price done by Lawani, Reed, Mark, and Zheng (2019). The other one that was conducted was on how Amazon book ratings affect their price by Kocas and Akkan (2016). In both studies, OLS regressions were conducted as well as several other statistical techniques. I am doing an OLS regression for my research. My key differentiation is that I am focusing on the smartphone market. The reason I chose smartphones over other products is that I believe that the features of a smartphone are easily controllable, and smartphones do not differ much except in features such as storage, operating system (only two: iOS and Android), and megapixels for the camera. The regression model I am doing is:

(1) $$\text{AvgRating}_i = \beta_0 + \beta_1\text{Price}_i + \beta_2\text{ScreenSize}_i + \beta_3\text{Battery}_i + \beta_4\text{Unlocked}_i + \beta_5\text{Totalnumratings}_i + \beta_6\text{Storage}_i + \beta_7\text{Apple}_i + \beta_8\text{Google}_i + \beta_9\text{LG}_i + \beta_{10}\text{Motorola}_i + \beta_{11}\text{Samsung}_i + \beta_{12}\text{MP}_i + \beta_{13}\text{Age}_i + \epsilon$$

The Unlocked variable will be a dummy variable in which it will take on values 1 if the phone is unlocked, and 0 if the phone is locked. Amazon have majority of their ratings centered around the mean 3.76, with very few ratings below 3 (See Table1). Each brand will be represented by a dummy variable. Again, I am testing the hypothesis that average rating of a product has a positive effect on the price of said product. This can be done with a p-value to see if there is significance: I expect to observe b1>0 and statistically significant.

5. Results

I estimated four models to illustrate how the significance changes as I add more important variables. My first model includes just four explanatory variables: average rating, battery, screen size, and the dummy variable unlocked. In this model, screen size and unlocked are significant at the 0% level. In the second model, I add megapixels of rear camera and the number of months the phone has been released. Age and screen size are significant at 0% level, while unlocked loses its significance. Battery also becomes significant at the 5% level. Adding this variable increased the Adjusted R-Squared from .25 to .41. In the third model, I factor in total number of ratings and the dummy variable storage. Total number of ratings had no significance and storage had significance at the 0% level. In the fourth and final model, I added the dummy variables for brands into the model. Apple is significant at the 5% level, Google is significant at the 1% level, LG is significant at the 5% level, and Samsung is significant at the 1% level. Screen size and storage keep their significance. Adding brands in the model modestly increased Adjusted R-Squared from .57 to .60.

Based on the p value test for all four regressions, I conclude that average rating does not have a significant impact on price. The p-values had no significance in every model, and in the first model had a negative coefficient estimate. In the second, third, and fourth model the estimate was positive. My hypothesis was that having a higher rating would cause the price to increase. This is consistent with previous studies, however that it is only true in the first model. In the other three models $b1<0$.

The most significant determinants of price are the screen size, storage, brand and age of the phone. Screen size, Storage, Apple, Google, and LG have a positive impact on price and surprisingly, so does Age. The brands all have high estimates for their effects on price. For example, An Apple phone compared to Other phones is associated with an additional 42.2% increase in price. This increase in price represents what consumers are willing to pay for extra given a certain brand. Google had the highest mark up at 48%. One of the reasons Age could be positive is that older phones are no longer manufactured and therefore there is a decrease in supply as time goes on. This decrease in supply results in an increase in price of the phone. Megapixels, battery, and whether the phone was unlocked also had no significance. This is surprising since these features are essentially inputs that have cost and therefore should affect the overall price of the phone. One of the reasons why they could have no significance is that consumers are not paying the price for a phone depending on these features, but rather they look at other superficial features of a phone such as screen size and brand.

6. Conclusion

This paper set out to test the hypothesis on whether Amazon ratings have an impact on price. The analysis that was conducted showed that this hypothesis is rejected, and average rating has no effect on the price of a phone. The significant determinants of a phone’s price were shown to be screen size, age, and brands such This means that when consumers purchase a phone, the price they’re willing to pay is not based on features embedded in the phone such as MP of camera and battery capacity, but mainly by the brands, storage, and size of the screen. However, since average rating had no effect in this is not consistent with the findings of previous studies. One possibility by Lafky (2014) states that there are many possible distortions in the accuracy of these ratings and with the lack of variation in the average rating this could produce an inaccurate result. Compared to the Airbnb and wine study, that showed a positive significant relationship between rating and price, mine did not. In the study done by Kocas and Akkan (2016), they found that ratings and trending status had a negative effect on the price of books. It could be that I omitted an important variable which is the trending status of the phone. Another possibility for why my results are inconsistent with other studies, is that for smartphones consumers rely less on reviews but rather word-of-mouth from friends and so on. This would be a logical conclusion since everyone has a smartphone these days, so information on what phones are preferred by consumers is abundant.

My model estimates could also be biased. I believe a reason for this is that there is not enough variation in the ratings I have collected. Amazon seems to have majority of their ratings centered around the mean 3.76, with very few ratings below 3 (See Graph 3). It is also important to mention that there might be reverse causality between ratings and price. If a product has a high price that might result in a lower rating because the customer might deem it not worth of such a price. If that is the case, there is endogeneity from simultaneity bias and my ratings coefficients estimate is biased. Overall, the weakness in my research comes from the fact that I possibly omitted an important variable such as trending status and my model has endogeneity. The strengths of my research compared to others, is that my research gives key insights on what consumers prefer from their smartphones. Policy implications of this could be that smartphone companies use this information to better price their smartphones.

References

Duggan, W. (2020, February 4). Latest e-commerce market share numbers highlight Amazon's dominance. Retrieved from https://finance.yahoo.com/news/latest-e-commerce-market-share-185120510.html

Kocas, C., & Akkan, C. (2016). How Trending Status and Online Ratings Affect Prices of Homogeneous Products. *International Journal of Electronic Commerce, 20*(3), 384-407. doi:10.1080/10864415.2016.1121765

Lafky, J. (2014). Why do people rate? Theory and evidence on online ratings. *Games and Economic Behavior, 87*, 554-570. doi:10.1016/j.geb.2014.02.008

Lawani, A., Reed, M. R., Mark, T., & Zheng, Y. (2019). Reviews and price on online platforms: evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, *75*, 22–34.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product Reviews. *Information Systems Research*, *19*(4), 456–474.

Snipes, M., & Taylor, D. C. (2014). Model selection and akaike information criteria: an example from wine ratings and prices. *Wine Economics and Policy*, *3*(1), 3–9.

Figure 1

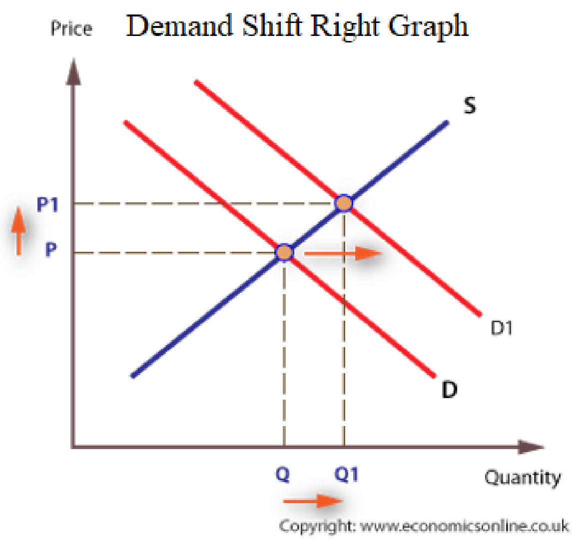
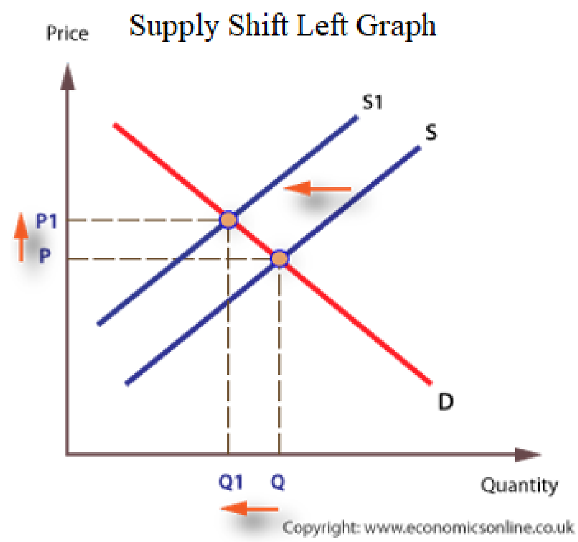


Figure 2



Graph 1

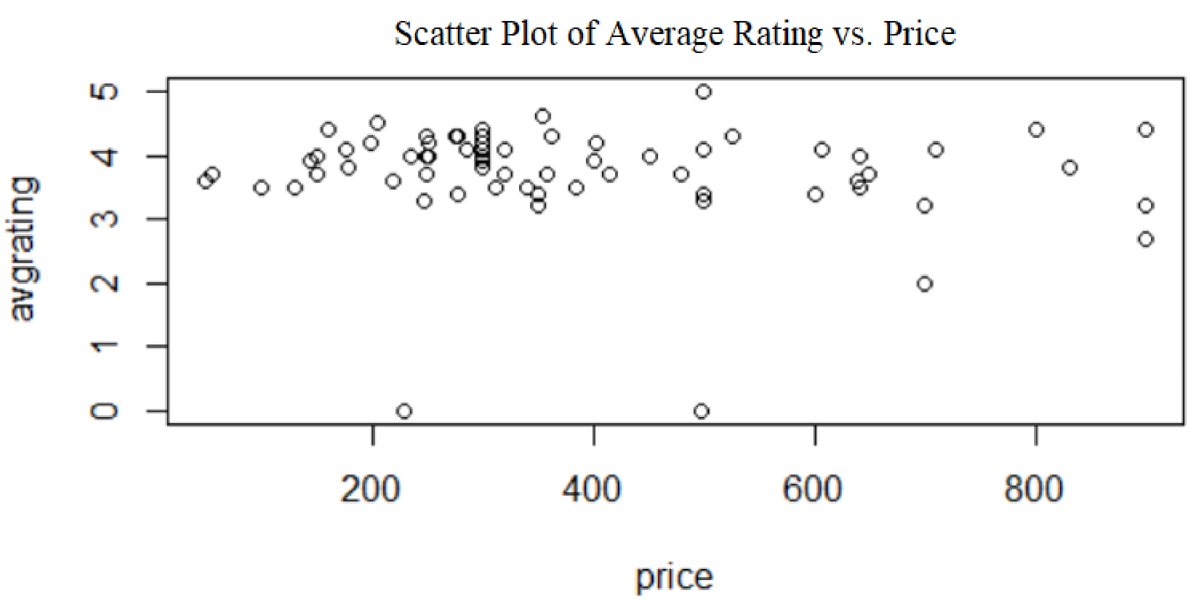


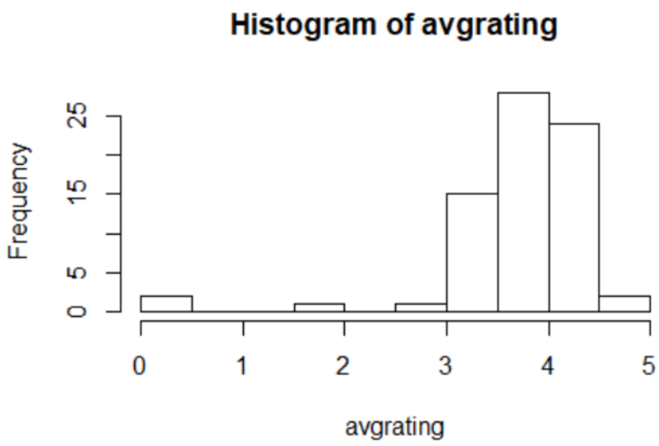
Table 1			
Definitions and Descriptive Statistics			
Variable Name	Definition	Mean	SD
Price	Amazon Phone Price	377.027	205.476
Average Rating	Average Numerical Rating (out of 5 stars)	3.761	.781
Screen Size	Size of phone measured in inches	5.676	.716
Battery	Battery capacity of phones measured in mAh	3127.835	665.24
Unlocked	=1 if the brand is Unlocked =0 If the brand is not Unlocked	.698	.462
Totalnumrating	Total number of reviews for a given phone	575.71	866.583
Storage	=1 if the storage is 64 gb =0 If the storage is 32 gb	46.027	15.987
Apple	=1 if the brand is Apple =0 If the brand is not Apple	.0821	.276
Google	=1 if the brand is Google =0 If the brand is not Google	.0547	.229
LG	=1 if the brand is LG =0 If the brand is not LG	.109	.314
Motorola	=1 if the brand is Motorola =0 If the brand is not Motorola	.328	.473
Samsung	=1 if the brand is Samsung =0 If the brand is not Samsung	.164	.373
MP rear	Number of mega pixels of the rear camera	14.301	6.376
Age	Number of months since the phone has been released	29.232	18.955

Table 2

OLS Regression Results

	Dependent variable:			
	log(price)			
	(1)	(2)	(3)	(4)
	Coefficient Estimate	Coefficient Estimate	Coefficient Estimate	Coefficient Estimate
Average Rating	-0.030	0.025	0.008	0.002
	(0.080)	(0.072)	(0.063)	(0.062)
Screen Size	0.510***	0.739***	0.578***	0.520***
	(0.146)	(0.140)	(0.123)	(0.123)
log(Battery(mAh)	-0.594	-0.782*	-0.480	-0.345
	(0.462)	(0.412)	(0.355)	(0.358)
Unlocked	-0.419***	-0.109	-0.131	-0.009
	(0.139)	(0.141)	(0.122)	(0.137)
Totalnumrating			0.00001	-0.00003
			(0.0001)	(0.0001)
Storage(64gb)			0.521***	0.487***
			(0.100)	(0.111)
Apple				0.422*
				(0.213)
Google				0.498**
				(0.215)
LG				0.290*
				(0.165)
Motorola				0.169
				(0.142)
Samsung				0.326**
				(0.147)
MP rear		-0.004	-0.009	-0.007
		(0.009)	(0.008)	(0.007)
Age		0.017***	0.020***	0.021***
		(0.004)	(0.003)	(0.004)
Constant	8.058***	7.405***	5.725**	4.686*
	(3.024)	(2.688)	(2.311)	(2.372)
Observations	73	73	73	73
R^2	0.291	0.459	0.621	0.677
Adjusted R^2	0.250	0.410	0.574	0.605
Residual Std. Error	0.506 (df = 68)	0.449 (df = 66)	0.381 (df = 64)	0.367 (df = 59)
F Statistic	6.988*** (df = 4; 68)	9.332*** (df = 6; 66)	13.132*** (df = 8; 64)	9.498*** (df = 13; 59)
Note:	Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Numbers in () under coefficient estimates are the standard errors			

Figure 5


Histogram of avgrating
Frequency
0
5
15
25
0
1
2
3
4
5
avgrating