The Impact of Consumer Reviews on Product Sales: Evidence from Video Game Products on Amazon

By Ting Guo*

Consumer reviews are proven by most past literature to have an impact on product sales. This paper explores the impact of consumer reviews on product sales for video game products in particular, by using data extracted from Amazon and analyzing the relationship between review features and product Best Sellers Rank (BSR), an estimator of product sales. Several key findings emerged from analysis based on approximately 1.5 million reviews and 40,000 products. First, consumer reviews, especially ratings, have an impact on product sales: the higher the rating, the lower the product BSR, and the higher the number of sales. More recent ratings have a larger impact on older ratings. In addition, product features are also found to have an impact on product sales, although the effects are not as large as the ratings. The findings can help video game sellers improve their current business strategies to increase product sales.

These days, online retail is becoming more prevalent in people's lives and is making up a larger revenue generation part of a business. In the United States, sales from e-commerce were $871 billion in 2021, and have been growing at 16% on average since 2011 (Wang, 2022). Online retail sales share out of total retail sales increased from 16% to 19% in 2020 due to Covid-19, further signifying the increasing importance of online retail (UNCTAD, 2021). Therefore, it is vital for businesses to understand the mechanisms behind consumers' online purchasing decisions in order to boost their online sales. Implemented as a common feature on most, if not all, of the e-commerce platforms globally, the consumer reviews feature has played a key role in helping consumers make buying decisions and in helping sellers increase online sales (Chen et al., 2022).

Many pieces of literature have proven the impact of consumer reviews on product sales. Research conducted by Consumer Focus found out that consumers generally rely on others' reviews (OECD, 2019). An analysis by Ahsan (2017) shows the correlation between consumer reviews and product sales. Sun et al. (2021)'s duopoly model-based research find that online reviews have a stronger influence on high-quality sellers than on low-quality sellers. Chevalier and Mayzlin (2006) analyze the data of books sold on Amazon and discover that an increase in a book's reviews leads to an increase in the relative number of sales for the site, and that one-star reviews have a greater impact than five-star reviews in most

* Guo: University of British Columbia (UBC), 2329 West Mall, Vancouver, BC Canada V6T 1Z4, ting1107@student.ubc.ca

cases. For e-commerce in the fashion industry, Lohse, Kemper and Brettel (2017) find that positive online reviews have a greater impact on products with higher involvement, and that online reviews, both positive and negative, have a greater impact on weaker brands. In the video games industry, Zhu and Zhang (2010) find that online reviews have a larger impact on less popular games.

However, several pieces of literature show different findings. Duan, Gu and Whinston (2008)'s research shows that ratings do not have a significant impact on movies' box office revenues. Chen, Wu and Yoon (2004) do not find product sales to be related to consumer ratings; however, they do find a positive association between sales and the number of reviews. Kim and Kim (2022) find that star ratings have an inverted U-shaped relationship with product sales; their results also show a positive relationship between rating volume and product sales.

Amongst all of the previous literature, not much examined the impact of customer reviews on product sales for video game products specifically. By understanding the factors that impact video game product sales, businesses in the video games industry can improve their business strategies based on the influencing factors to increase their online sales. According to an online survey, 60% of the video gamers in the United States are under the age of 34 (Clement, 2022*b*). As video game users are mostly young people, the finding of this paper can also partially imply how consumer reviews play a role in young people's purchasing decisions. Hence, businesses in other industries whose main retail method is online and whose consumer age group are the young people can also make/adjust their business strategies based on the results of this paper. Therefore, this paper focuses on exploring the impact of consumer reviews on product sales for video game products.

To the author's knowledge, not much literature has used Amazon data to address the impact of consumer reviews on video game product sales. This paper will add value to the current literature by using video games data from Amazon. The hypothesis is that consumer reviews, especially ratings, will have an impact on product BSR: a higher overall rating of a product will lead to a smaller product BSR, or a higher product sales.

The data used in the analysis is obtained from the Amazon Review Data (2018) database which records product information and reviews information up until October 2018 (Ni, Li and McAuley, 2019). To explore the impact of consumer reviews on product sales, each product's number of sales is estimated by its Best Sellers Rank (BSR), a rank that indicates how well a product is selling in the most recent period of time. A lower BSR implies a greater number of sales.

Linear regression models are used to address the research question. Regression coefficients are analyzed to study the impact of review features such as mean rating, total number of reviews, total number of review images, etc. The model examines the relationship between review features and product BSR, and consequently find the impact on product sales. Multicollinearity tests are done before running the regression to ensure variables only have a limited multicollinearity.

All independent variables are standardized to the same mean value of 100 so that the impact strengths of variables can be ranked based on variable coefficients.

This paper applies natural language processing (NLP), a machine-learning technique which uses computers to study human language. A sentiment score is assigned to each review text using NLP, which measures how positive or how negative each review sounds to humans in a quantitative way. The empirical model evaluates the impact of each product's overall review sentiment on product sales. Not much literature in the research field uses this technique, and this paper adds value to the research area by using evidence from NLP.

In October 2016, Amazon banned incentivized reviews, which is an important milestone in its review system history. Before the ban, consumers were allowed to write reviews for products they purchased in exchange for a discounted or free product from product vendors. The original intention is to increase the number of "honest" feedback such that consumers can get more product information from the reviews. However, research shows that incentivized reviews give generally higher ratings, which distorts the original purpose of the program (ReviewMeta, 2016). Amazon banned this program to increase review fairness, and this event is taken into account in the analysis. The results indicate that review features indeed have an impact on product BSR: the mean rating of reviews after the ban of incentivized reviews, total number of reviews, percentage of verified reviews, percentage of reviews after the ban, and the total number of review images together have the power to explain 30.1% of the variance in product BSR. All variables, except the percentage of reviews after the ban, have negative effects on product BSR: higher values in these variables leads to a lower BSR and hence, a higher product sales. The mean rating of reviews after the ban has the largest impact among all of the review features, suggesting that the hypothesis is correct.

The empirical model is made more robust by conducting heterogeneity tests, where the impact of product features, such as brand popularity and number of product images, on product BSR are assessed in addition to review features. Product features are shown to have an impact on product sales, but the impact is smaller than the impact of consumer ratings. Higher product prices leads to lower sales, which partially explains vendors' motivation behind applying discounts on product prices.

The findings contribute to the existing literature by adding one more piece of evidence that consumer reviews indeed have an impact on product sales particularly for video game products. This suggests that businesses should pay attention to consumer reviews and react to consumer feedback promptly, since accumulating too many negative reviews in a recent period of time will likely affect the sales of a product negatively, which reduces business sales revenue for that period.

The rest of the paper is structured as follows. Section I reviews past literature and provides background on Amazon's ranking system and its reviews system. Section II describes the data. Section III describes the empirical methodologies. Section IV shows the results. Section V discusses the findings. Finally, section

VI concludes the entire paper and provides related implications.

I. Background

A. Amazon Ranking System

Every product sold on Amazon is assigned a Best Sellers Rank (BSR) after it achieves one sale, and a lower BSR indicates higher sales. Each product can have multiple BSRs under different category levels. For instance, a product can rank at 100 at video games level and at 50 at video games > PC games level simultaneously. BSR is updated every hour and is largely based on a product's current sales tendency and its historical sales volume (Connolly, 2022).

Each product is also assigned an organic ranking, which is the rank it gets on Amazon's search engine results pages (SERP). A product's organic ranking does not necessarily reflect its BSR, so does vice versa (Connolly, 2022).

B. Amazon Reviews System

Amazon has implemented a consumer reviews system since the very beginning. A few major changes happened to the reviews system before the latest date of the datasets used in this paper (October 2018), which includes the prohibition of incentivized reviews, the launch of the Early Reviewer Program, and the deactivation of a large number of accounts that are suspected members of underground review communities.

Although Amazon has long been prohibiting businesses to pay for fake reviews, it allowed incentivized reviews, where businesses offer free or discounted products to consumers to get "honest" feedback, until this was banned in October 2016 (Perez, 2016). For incentivized reviews, reviewers write reviews of a product and in turn receive a large discount or free product from the vendor; all they need to do is to disclose their affiliation with the vendors when writing the reviews. Although the original purpose of incentivized reviews is to encourage consumers to write truthful reviews that help other consumers make their purchasing decisions, in reality, incentivized reviews have significantly biased toward higher ratings. An analysis conducted by ReviewMeta (2016) based on 7 million reviews observes that incentivized review texts are less critical than non-incentivized review texts, and incentivized ratings on average are 0.38 stars (out of 5 stars) higher than non-incentivized ratings, which strongly favours the vendors. By removing incentivized reviews, the reviews will likely reflect more honest opinions.

In February 2017, Amazon introduced the Early Reviewer Program, where shoppers are encouraged to write product reviews by receiving low-value gift cards in exchange (Masters, 2021). Unlike the incentivized review feature that was banned in October 2016, this program is only applicable to products with less than 5 reviews. This helps products with low sales to attract more customers by providing minimal consumer reviews, which mitigates the crackdown that October 2016 policy change brought to the Amazon reviews system.

In April 2018, Amazon deactivated a large number of customer accounts that use the marketplace for commercial purposes. This is an effort to reduce underground reviews after the ban on incentivized reviews and make consumer reviews more credible.

It is important to account for the above changes in the analysis, as the removal of incentivized reviews could make reviews written after October 2016 more trustworthy to some consumers. Yet it is also important to keep in mind that incentivized reviews still exist after the ban under the table.

In addition, it is necessary to understand how product star rating is calculated. Amazon uses machine learning algorithms to detect fake reviews, and only authentic and verified (if the purchase is unverified, one needs to include more details in review texts for the review rating to get included in the star rating calculation) review ratings are used in calculating the star rating. Furthermore, more recent reviews are given more weight (Amazon, n.d.). Hence, the product star rating is different from the simple mean of all consumer ratings.

II. Data

A. Data Sources

As an e-commerce giant, Amazon earned almost $470 billion USD in its net sales revenue in 2021, which makes it a good e-commerce platform to study (Coppola, 2022). This paper use the Amazon Review Data (2018) provided by Ni, Li and McAuley (2019), which records reviews and product data on Amazon from May 1996 to October 2018. Amongst the 29 categories in the entire dataset, the video games category comprises 1.12% of all reviews and 0.56% of all products. Although it is not one of the largest categories in terms of review and product volumes on Amazon, Amazon is the most popular platform to purchase video games. According to the 2022 Statista Global Consumer Survey (multi-selection survey), 44% of all U.S. respondents have purchased online video games on Amazon in the past 12 months (Kunst, 2022). Furthermore, there are approximately 1.5 million reviews and 40,000 products available after data cleaning, which ensures that the sample size is large enough for analysis. Hence, it is convincing to analyze video game products and review data using data from Amazon even though the category does not make up a large portion of Amazon's overall product and review database. Lastly, Amazon has been an online seller from the very beginning and has always been online only up til today, which eliminates the impact of store factors on product sales as well as the substitution effect between online sales and in-store sales.

Using data from Amazon platform only removes e-commerce platform heterogeneity, since the user interface varies between different platforms, which may highlight different aspects of consumer reviews and result in different consumer purchasing decisions. In addition, different platforms also calculate product star ratings differently, which makes it hard to analyze the impact of star ratings on product sales. Thus, it is the best choice to analyze data from the same platform.

Conducting analysis on video game data exclusively removes product category heterogeneity, since people's purchasing decisions can vary amongst different categories of products. For instance, consumer reviews may be less important to some consumers for necessities and/or very cheap products. Video games are not necessities, and the price range is also relatively small, which reduces heterogeneity.

This paper uses two datasets from the Amazon Review Data (2018). One dataset contains consumer reviews information and the other contains product information; both datasets focus specifically on the category of video games and record information up to October 2018 (Ni, Li and McAuley, 2019). The two datasets are linked together by ASIN, Amazon Standard Identification Number which is assigned uniquely to each Amazon product.

The product data specifies the characteristics of each Amazon product under the video games category, and includes information on product ASIN, name, price, brand, categories, BSR at different category levels, details, descriptions, technical specifications, images, high-resolution images, and relevant products.

The reviews data specifies the details of each review of the products under the video games category, and includes information on product ASIN, rating out of 5 stars, reviewer ID, review time, review summary, review text, whether the review is verified, number of votes, style of the product purchased, and images attached to the review.

B. Data Construction

Data cleaning is performed to identify and remove irrelevant data. Data manipulation is also conducted to extract more useful information.

15.22% of the product dataset are duplicate data, meaning that some products are recorded multiple times; the repetitive rows are unnecessary, therefore removed. Products without a single review are excluded. Products under accessories or console subcategories are excluded as the sales of accessories and consoles can be different from video games. Only the products that have a BSR at video games level are retained to ensure that the sales of all products in the analysis are compared at the same level. The above actions (including the removal of duplicates) remove 44,829 of 84,819 (52.85%) data points from the original product dataset, leaving approximately 40,000 data points remaining for analysis. Rank at video games level will be the dependent variable for the models in Section III.

For the reviews data, 1,079,485 of 2,565,349 (42.08%) data points are excluded after removing reviews of products that are under accessories or console subcategories or do not have a rank in the video games level, leaving approximately 1.5 million data points remaining for analysis.

Table 1—: Data Cleaning Summary

Duplicate rows	12,908	15.22%
Products without BSR at video games level	10,296	12.14%
Products under accessories or console subcategories	21,623	25.49%
Products with no reviews	2	0.00%
Products remaining	**39,990**	**47.16%**
Total Products	84,819	100%
Reviews whose products do not have BSR at video games level, or are under accessories or console subcategories	1,079,485	42.08%
Reviews remaining	**1,485,864**	**57.92%**
Total Reviews	2,565,349	100%

Note: The table summarizes the data cleaning process for the product data and review data respectively. After data cleaning, around 40,000 products and 1.5 million reviews remain for analysis in Section III.

As summarized in Table 1, although the data cleaning process has removed a significant amount of data points from the original datasets, the process ensures that the final data to be used in the empirical model is clean, high-quality and more homogeneous. Moreover, the final data is still large in size despite the large exclusion of data points in percentage, hence the analysis based on the final data is still credible.

Natural language processing, a field of machine learning which focuses on linguistics, is used to analyze the sentiment of each review text in a quantitative way. This paper uses VADER (Valence Aware Dictionary for Sentiment Reasoning), a text sentiment analysis model that breaks down sentences into individual words, and analyze the lexical features of each word based on their semantics and suppositions in the sentence (Bajaj, 2021). Lexical features include sentiment polarity (positive, negative, or neutral) and sentiment intensity. The lexical features of individual words are combined together to produce an overall sentiment indicator of the texts. Each text is assigned a sentiment intensity score that ranges from -1 to 1, where a positive score indicates a positive valence, a negative score indicates a negative valence, and a score of 0 indicates neutral sentiment. The larger the magnitude of the sentiment score, the stronger the text is toward that valence; a sentiment score of 1 indicates absolutely positive sentiment and a sentiment score of -1 indicates absolutely negative sentiment. This paper uses the Natural Language Processing Tool Kit (NLTK) library in Python to perform sentiment analysis to each review text. The sentiment score not only quantifies the emotion of each review text, but also serves as an alternative measurement to review rating when analyzing the impact of reviews on product sales.

Table 2—: Summary Statistics

		Unscaled				Scaled			
		mean	std	min	max	mean	std	min	max
Dependent Variable	RankVideoGames[1]	66102.66	49716.75	2.00	241788.00	66102.66	49716.75	2.00	241788.00
Review Feature Variables	MeanRating	3.85	1.01	1.00	5.00	100.00	26.24	25.98	129.88
	MeanVerifiedRating	3.61	1.48	0.00	5.00	100.00	41.01	0.00	138.68
	MeanAfterBanRating	2.06	2.16	0.00	5.00	100.00	104.80	0.00	242.30
	MeanVerifiedAfterBan-Rating	2.03	2.17	0.00	5.00	100.00	107.09	0.00	246.83
	MeanSentiment	0.48	0.33	-1.00	1.00	100.00	67.26	-206.20	206.25
	MeanVerifiedSentiment	0.43	0.34	-1.00	1.00	100.00	77.37	-229.88	230.25
	MeanAfterBanSentiment	0.21	0.33	-1.00	1.00	100.00	152.24	-466.04	465.81
	MeanVerifiedAfterBan-Sentiment	0.21	0.32	-1.00	1.00	100.00	153.91	-475.20	475.72
	NumReviews	37.16	132.28	1.00	7630.00	100.00	356.02	2.69	20535.10
	VerifiedPercentage	0.67	0.33	0.00	1.00	100.00	48.61	0.00	148.87
	AfterBanPercentage	0.15	0.27	0.00	1.00	100.00	174.66	0.00	650.97
	TotalNumReviewImages	0.38	2.29	0.00	100.00	100.00	598.13	0.00	26138.96
Product Feature Variables	BrandPopularity	632.34	791.13	0.00	2387.00	100.00	125.11	0.00	377.48
	NumImages	3.04	3.09	0.00	14.00	100.00	101.62	0.00	460.95
	NumHighResImages	3.04	3.09	0.00	14.00	100.00	101.62	0.00	460.95
	NumDetails	0.01	0.21	0.00	7.00	100.00	2628.47	0.00	86665.63
	FeatureLength	264.13	356.85	0.00	7993.00	100.00	135.10	0.00	3026.18
	DescriptionLength	878.02	1242.89	0.00	18788.00	100.00	141.56	0.00	2139.81

Note: The table describes the summary statistics for all variables used in the empirical analysis, both before and after variable standardization. All review feature variables ($Feature_{rp}$'s in Equation 1) and product feature variables ($Feature_{ap}$'s in Equation 2) are standardized to the same mean value of 100 so that one can compare their magnitudes of coefficients to identify the variable that has the largest impact on product BSR.
1: $RankVideoGames$ represents the Best Sellers Rank (BSR) at video games level. It is the dependent variable in the empirical models. Unlike the independent variables, the BSR values are not scaled.

The Unscaled Section of Table 2 shows the summary statistics of the product features, aggregate review features, and product Best Sellers Rank at video games level. Reviews features are aggregated by product ASIN; features include mean rating ($MeanRating$), mean sentiment intensity score ($MeanSentiment$), number of reviews ($NumReviews$), percentage of verified reviews ($VerifiedPercentage$), percentage of reviews after the ban of incentivized reviews ($AfterBanPercentage$), and total number of review images ($TotalNumReviewImages$). In addition, the meaning rating/sentiment score of verified reviews, of reviews after the ban of incentivized reviews, and of verified reviews after the ban, are also computed to serve as alternative mean rating/sentiment values; they may have a stronger impact on BSR than the rating/sentiment of all reviews. If, for instance, a product does not have any reviews that are written after the ban of incentivized reviews (October 1, 2016), then that product has a mean after-the-ban rating ($MeanAfterBanRating$) of 0. Product features include brand popularity ($BrandPopularity$), number of images ($NumImages$), number of high-resolution images ($NumHighResImages$), number of details ($NumDetails$), length of product description ($DescriptionLength$), and length of product feature description ($FeatureLength$). The popularity of product brand is calculated as the number of occurrences that each brand appears in the entire product dataset. For instance, if 20 out of the 39,990 products belong to the brand ABC, then each product under that brand is assigned a brand popularity of 20.

Product BSR at video games level is the dependent variable for the empirical model in Section III, and the aggregate review features are the independent variables. Product features such as brand popularity are used in heterogeneity tests to increase model robustness.

C. Sample Description

Table 3 describes the rating distribution of the entire sample, verified vs. unverified purchases, as well as before vs. after the ban of incentivized reviews (October 1, 2016). The first four groups, as shown in columns (1) to (4) in the table, all follow the same rating distribution where 5-star rating reviews comprise the largest portion of the sample, followed by 4-star, 1-star, 3-star, and last 2-star ratings. The only exception is the sample of reviews after the ban of incentivized reviews, as indicated by column (5) in the table, whose 1-star ratings make up a larger portion than 4-star ratings. Rating distribution has become more extreme after the ban on incentivized reviews, where the proportion of 5-star and 1-star ratings both increase compared to before the ban. Verified reviews comprise of a larger percentage of 5-star ratings and a smaller percentage of 1-star ratings than unverified reviews.

Figure 1 shows the relationship between the mean rating and number of reviews of each product after dividing products based on their verification status. For products with less than ~ 300 number of reviews, review verification status

Table 3—: Rating Distribution

	(1) Full Sample	(2) Verification Status	(3)	(4) Before or After Ban	(5)
		Verified	Unverified	Before	After
1 star	10.99%	9.37%	14.48%	10.63%	13.08%
2 stars	5.12%	4.04%	7.45%	5.31%	4.00%
3 stars	8.41%	7.54%	10.28%	8.77%	6.28%
4 stars	16.73%	15.12%	20.20%	17.68%	11.10%
5 stars	58.76%	63.93%	47.60%	57.61%	65.53%
Sample Size	1,485,864	1,015,507	470,357	1,269,926	215,938

Note: This table reports the rating distribution of all reviews in which their corresponding product is not under accessories or console subcategories and has a rank at video games level. Columns (2) and (3) restrict the sample based on review verification status, and columns (4) and (5) restrict the sample based on whether the review is written before or after the ban of incentivized reviews (October 1, 2016).

Figure 1. : Mean Product Rating vs. Number of Reviews

Note: This scatterplot shows the relationship between the mean rating and number of reviews of each product after dividing products based on their verification status.

does not play an important role in product's mean rating. For products with a large number of reviews, verified reviews tend to lead to a higher mean rating. This suggests a possibility to incorporate the combined effects of verification status and the number of reviews into the empirical analysis.

Figure 2 displays the distribution of sentiment intensity score of each review, grouped by review ratings. In general, sentiment score and rating show a positive correlation with each other. The positive correlation is more obvious for lower ratings. On the other hand, sentiment scores show a similar distribution for 4-star and 5-star ratings, and it does not necessarily reflect the review rating. For instance, some reviewers show very negative sentiments in the review texts, yet give a 4- or 5-star rating at the same time. This can be explained by the inaccuracy of the sentiment analyzer (e.g.: the analyzer fails to detect sarcasm in the reviews) and/or the personality of reviewers (e.g.: some people tend to not give a low rating despite dissatisfaction with products). The above observations suggest that sentiment scores can be considered as an alternative variable for product rating in the empirical model in Section III: although it has a positive correlation with review rating, the correlation is far from perfect; it might have a larger impact on BSR than product rating, which makes it is worthy to compare the impact of these two metrics, keeping all other variables the same.

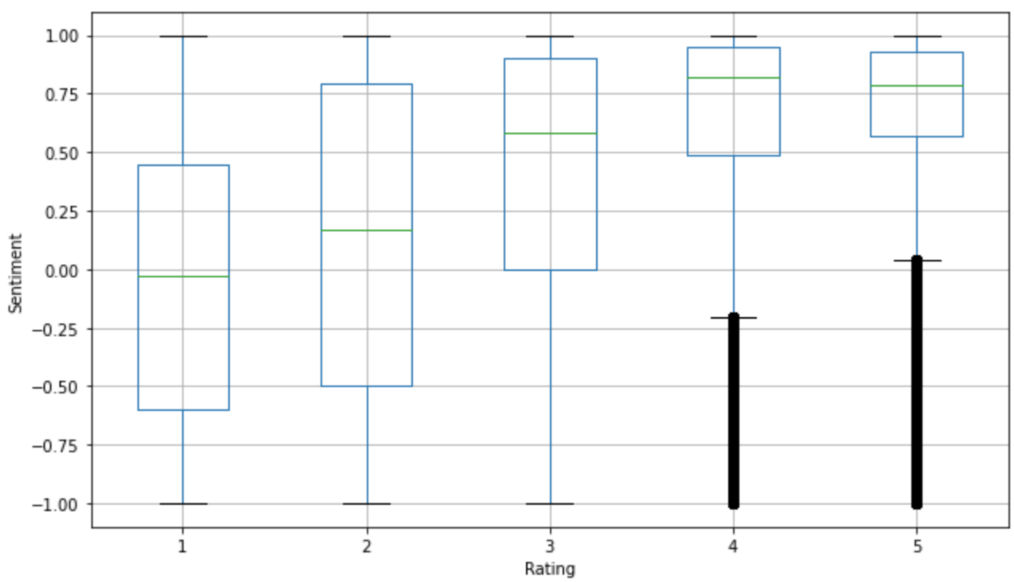

Figure 2. : Boxplot of Sentiment Score by Review Rating

Note: The boxplots show the distribution of the sentiment intensity score of each review, grouped by review ratings.

Figure 3 shows the distribution of mean rating of all reviews, mean rating of all verified reviews, mean rating of all reviews after the ban of incentivized reviews, and mean rating of all verified reviews after the ban of incentivized reviews, all aggregated by product. Compared to the mean rating of all reviews, the mean rating of verified reviews shows a similar distribution, but the mean rating of reviews after the ban on incentivized reviews and that of verified reviews after the ban show a very different distribution, where the 75th percentile values are similar,

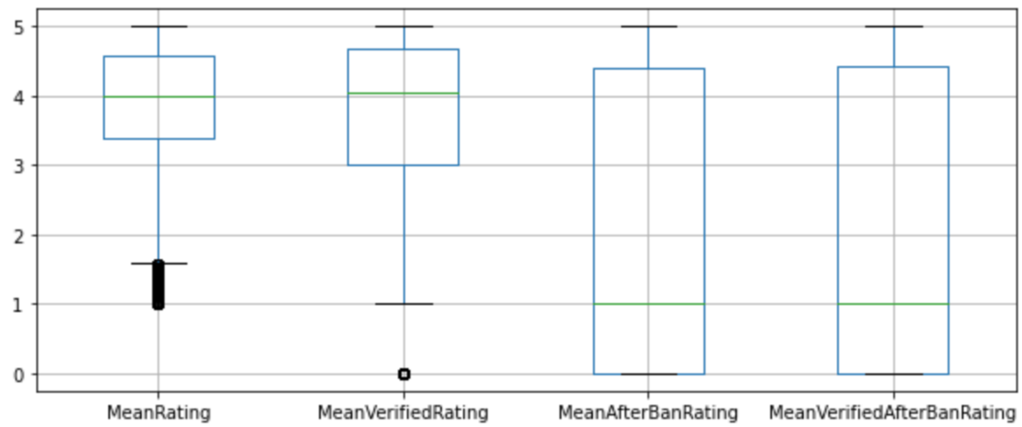

Figure 3. : Boxplots of Different Mean Ratings

Note: The boxplots show the distribution of mean rating of all reviews, mean rating of all verified reviews, mean rating of all reviews after the ban of incentivized reviews, and mean rating of all verified reviews after the ban of incentivized reviews, respectively. All means are aggregated by product.

but the median and the 25th percentile values are very low. This is because a large number of products do not have any reviews written after the ban of incentivized reviews, which results in their mean rating being 0 when conditioned on the reviews after the ban. As alternative means show different distributions from the simple mean of all reviews, it is worth testing in the empirical model to find out the mean value that impacts BSR the most.

III. Empirical Methodology

This paper uses linear regression models to estimate the impact of a product's aggregate consumer review features, such as mean rating and total number of reviews, and its Best Sellers Rank (BSR) at video games level.

Regression model is chosen to address the research question since the dependent and independent variables are all numeric, which makes the regression model a perfect choice. A linear relationship is assumed since non-linear functions are much more complex and would not likely result in a better fit than linear regression. Furthermore, the ordinary least squares (OLS) technique is applied to solve the linear regression model by minimizing the variance in the results.

The exact number of sales of a product is only available to the product vendor and is unavailable to the public. Hence, a proxy is needed for the product number of sales and it needs to be publicly available. Product Best Sellers Rank (BSR) serves as a good proxy, since BSR is publicly available under each product's Product Details section, is recorded in the datasets, and is determined largely by a product's current sales trend (Connolly, 2022). Several websites provide calculators that estimate the number of sales of an Amazon product based on its BSR at different category levels (Jungle Scout, n.d.). Unfortunately, the services are not free, and the fees required to obtain the estimated number of sales for all $\sim 40,000$ products in the datasets are out of the research budget. Therefore, product sales are estimated using product BSR, and in particular, using product BSR at video games level. This ensures that the products are compared at the same level.

An advantage of using data from video game products is that it has a small number of returning consumers. Unlike grocery products such as tissue paper, video games are likely to be purchased once only for each consumer, which reduces the probability that a product's sales are constantly from the same group of repeat consumers without new consumers. Having repeat consumers can underestimate the impact of reviews on product sales, since repeat consumers will likely purchase products based on their user experience rather than other people's reviews. As video game products have a lower probability of getting repeat consumers, the impact of consumer reviews is likely not decreased by it.

A. Basic Model Setup

The general linear regression model is as follows:

$$BSR_p = \sum_{rp} \beta_{rp} Feature_{rp} + \epsilon_p \tag{1}$$

where BSR_p is the dependent variable which indicates the BSR of product p at video games level, $Feature_{rp}$'s are the aggregate values for product p for review feature r (e.g.: mean rating, mean sentiment of verified reviews, number of reviews, percentage of verified reviews, etc.), β_{rp}'s are the coefficients of $Feature_{rp}$'s, and ϵ_p is the error term.

A lower BSR represents a higher number of sales, hence a very negative coefficient would indicate a large positive impact on the number of sales of a product. The hypothesis is that rating or sentiment will have a positive impact on product sales, meaning that their coefficients are hypothesized to be negative.

$Feature_{rp}$'s are standardized so that all of them share the same mean. This allows one to identify the aggregate review feature that has the largest impact on BSR_p, which is the $Feature_{rp}$ with the largest magnitude of the coefficients. In this model, all independent variables are standardized so that they share the same mean value of 100. Table 2 shows description of $Feature_{rp}$ variables and their summary statistics before and after the standardization; summary statistics before the standardization are under the Unscaled columns, and summary statistics after the standardization are under the Scaled columns.

B. Heterogeneity Model Setup

This section explores heterogeneous factors that might have an impact on BSR by testing the effects of product attributes.

The regression equation after adding heterogeneous factors is as follows:

$$BSR_p = \sum_{rp} \beta_{rp} Feature_{rp} + \sum_{ap} \beta_{ap} Feature_{ap} + \epsilon_p \tag{2}$$

where $Feature_{ap}$'s are the values of product p's product feature a (e.g.: number of images, brand popularity, etc.), and β_{ap}'s are the coefficients of $Feature_{ap}$'s.

Again, all of $Feature_{rp}$'s and $Feature_{ap}$'s are standardized so that they all share the same mean (100 for this model), to assess the variable with the largest impact on BSR_p. The summary statistics of the variables are again shown in Table 2.

C. Multicollinearity and Variable Selection

OLS technique requires variables to have no perfect multicollinearity. Although a strong multicollinearity does not violet OLS assumptions, it is not recommended since it will result in a large variance in the results. Hence, correlation and multicollinearity tests are conducted before determining the independent variables used in both the basic model and heterogeneity model.

First, the correlations between different variables are computed to detect pairwise collinearity. The correlation matrix in absolute values is shown in Figure 4. A larger absolute correlation coefficient (cells with darker backgrounds) indicates a stronger pairwise correlation, and a smaller absolute correlation coefficient (cells with lighter backgrounds) indicates a weaker pairwise correlation.

Most of the pairs are weakly correlated to each other, and most variables do not show a strong correlation to BSR ($RankVideoGames$ in the heatmap). The mean rating of reviews after the ban of incentivized reviews ($MeanAfterBanRating$ in the heatmap) is moderately correlated to BSR with an absolute correlation of 0.5, holds a stronger correlation strength compared to the mean of all ratings ($MeanRating$ in the heatmap), and is in fact the variable with the strongest correlation to BSR out of all the variables. This suggests that the mean rating for reviews after the ban will likely fit the model better than the mean rating for all reviews.

Some pairs show a very strong correlation. For instance, a product's number of images ($NumImages$) and number of high-resolution images ($NumHighRes-Images$ have an absolute correlation of 1. This suggests that the two variables should not be included in the same model simultaneously.

Multicollinearity checks are also performed to ensure the independent variables selected for the model show limited multicollinearity. Variance inflation factors (VIF) are used to measure multicollinearity. A VIF greater than 10 indicates

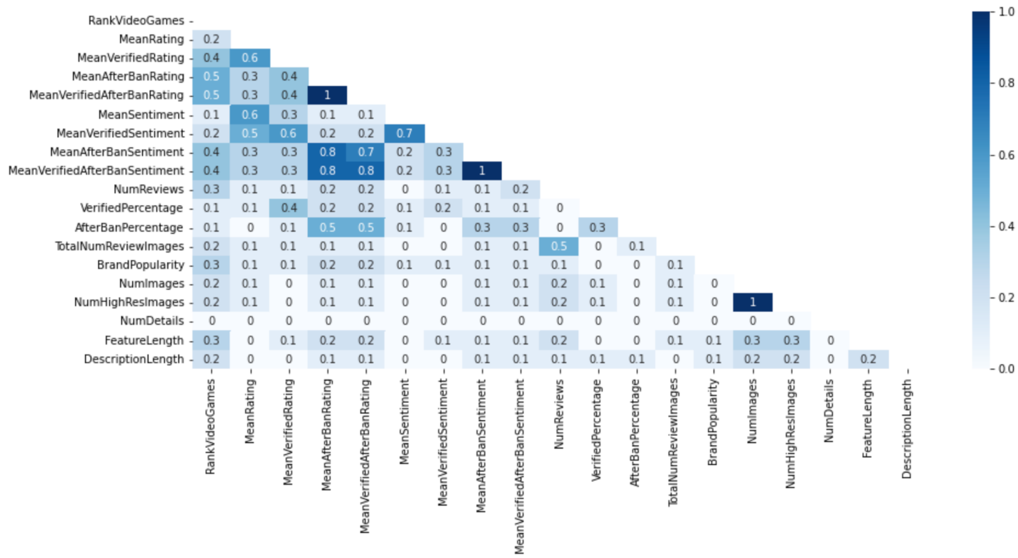

Figure 4. : Correlation Heatmap of Aggregate Review Features

Note: The heatmap shows pairwise correlations between different pairs of product features. All correlation coefficients are in absolute values. A larger absolute correlation coefficient (cells with darker backgrounds) indicates a stronger pairwise correlation, and a smaller absolute correlation coefficient (cells with lighter backgrounds) indicates a weaker pairwise correlation.

strong multicollinearity, and at least one variable needs to be removed from the list of independent variables to ensure that the OLS assumption is met.

The results of multicollinearity test of sample two sets of variables are shown in Table 4. In Table (a), MeanRating has a VIF larger than 10, which indicates strong multicollinearity between the set of variables. After removing MeanSentiment, the VIF of the rest of the variables are shown in Table (b); no VIF exceeds 10, which suggests limited multicollinearity between the variables, hence these variables can be used in the same regression model simultaneously.

Both the pairwise correlation and multicollinearity tests are performed based on non-standardized data, and only the regression model uses standardized data (i.e.: all independent variables are standardized to a mean value of 100 and the dependent variable stays unchanged).

IV. Results

Multicollinearity tests are performed over different combinations of variables, and models that regress on the combinations that show insignificant multicollinearity are compared against each other based on R^2.

Table 4—: VIF of Sample Sets of Variables

(a) Significant Multicollinearity

Variables	VIF
RankVideoGames	2.42
MeanRating	10.88
MeanSentiment	4.76
NumReviews	1.45
VerifiedPercentage	5.04
AfterBanPercentage	1.50
TotalNumReviewImages	1.34

(b) Limited Multicollinearity

Variables	VIF
RankVideoGames	2.41
MeanRating	5.29
NumReviews	1.45
VerifiedPercentage	4.93
AfterBanPercentage	1.49
TotalNumReviewImages	1.34

Note: The tables show the variance inflation factor (VIF) values of two sample sets of variables to check multicollinearity of each set of variables. The variables in Table (a) are strongly multicollinear since *MeanRating* has a VIF larger than 10. The variables in Table (b) are limited in multicollinearity since all variables have a VIF less than 10.

A. *Basic Model*

The review features of a product include its total number of reviews, percentage of verified reviews, percentage of reviews after the ban on incentivized reviews, total number of review images, as well as an overall rating/sentiment score. The choice of the overall rating/sentiment score is one of the mean value (rating or sentiment) of all reviews, mean value of all verified reviews, mean value of all reviews after the ban, and mean value of all verified reviews after the ban.

The regression model that achieves the highest R^2 consists of the following independent variables: total number of reviews, percentage of verified reviews, percentage of reviews after the ban of incentivized reviews, total number of review images, and mean rating of all reviews after the ban. Compared to the most ordinary model that uses the mean rating of all reviews and has an R^2 of 0.122, the best-performing model gives an R^2 of 0.301, which significantly increases the power of the independent variables in explaining the variance in the dependent variable, BSR at Video Games level.

The coefficients of the best-performing basic model are summarized in Table 5. All coefficients, except the percentage of reviews after the ban ($AfterBan-Percentage$), have a negative coefficient, which indicates a negative relationship with product ranking and hence, a positive relationship with product sales. The mean of all reviews after the ban, or $MeanAfterBanRating$, displays the strongest magnitude. As all of the independent variables are standardized to the same mean value, this suggests that $MeanAfterBanRating$ has the largest impact on product sales. On the other side, the total number of review images, or $TotalNumReviewImages$ exhibits a very small, although non-zero, impact, on

BSR, which implies that it does not impact product sales much.

Table 5—: Basic Model Coefficient Summary

	Coefficient	Standard Error	2.5% CI	97.5% CI
MeanAfterBanRating	-258.06	2.35	-262.66	-253.46
NumReviews	-18.26	0.68	-19.59	-16.92
VerifiedPercentage	-54.69	4.50	-63.50	-45.87
AfterBanPercentage	45.84	1.43	43.04	48.64
TotalNumReviewImages	-2.39	0.40	-3.17	-1.61

Note: This table summarizes the coefficients of the independent variables of the best-performing basic model, as well as coefficient standard errors and 2.5% and 97.5% confidence interval values. All of the independent variables are standardized to the same mean value of 100 before conducting linear regression using the OLS method to ensure that the magnitudes of coefficients are unbiased for comparison.

An R^2 of 0.301 implies that the review features together play a rather important role in explaining product sales. The result is aligned with the hypothesis, which predicts a positive impact of consumer reviews, especially consumer ratings, on product sales, or a negative impact on product ranking.

B. Heterogeneity Model

The heterogeneity model adds heterogeneous factors, the product features of each product, to the basic model. Heterogeneous factors include brand popularity (measured by the number of times the brand appears out of all products), number of high-resolution images, number of details, length of product features, and length of product description. VIF checks are done and no significant multicollinearity is found.

The coefficient summary of the heterogeneity model is shown in Table 6. The explanatory power of review features is diluted as additional variables with nonzero coefficients are added in. Despite that, the mean rating of all reviews after the ban of incentivized reviews still appears to be the variable with the largest impact on product ranking. Brand popularity, or $BrandCount$, has the strongest impact out of the five product feature factors, and all product features except number of details ($NumDetails$) show a negative impact on product BSR, or a positive impact on product sales; number of details of a product exhibits a negligible impact.

The R^2 of the heterogeneity model is 0.358, a small increase from the basic model. A regression model that explains BSR using only the product features is developed as a supplement, where the model formula is displayed in Equation 3. The R^2 of the product features-only model is 0.165, a much smaller R^2 compared to the heterogeneity model.

Table 6—: Heterogeneity Model Coefficient Summary

	Coefficient	Standard Error	2.5% CI	97.5% CI
MeanAfterBanRating	-219.34	2.34	-223.93	-214.75
NumReviews	-11.51	0.67	-12.81	-10.20
VerifiedPercentage	-65.30	4.36	-73.84	-56.77
AfterBanPercentage	32.71	1.39	29.98	35.43
TotalNumReviewImages	-2.62	0.38	-3.37	-1.88
BrandCount	-61.84	1.65	-65.06	-58.61
NumHighResImages	-23.11	2.07	-27.18	-19.05
NumDetails	-0.05	0.08	-0.20	0.10
LenFeature	-49.38	1.59	-52.50	-46.26
LenDescription	-24.76	1.50	-27.69	-21.83

Note: This table summarizes the coefficients of the independent variables of the heterogeneity model, as well as coefficient standard errors and 2.5% and 97.5% confidence interval values. All of the independent variables are standardized to the same mean value of 100 before conducting linear regression using the OLS method to ensure that the magnitudes of coefficients are unbiased for comparison.

$$BSR_p = \sum_{ap} \beta_{ap} Feature_{ap} + \epsilon_p \tag{3}$$

The comparisons between R^2's of different models, along with the analysis of coefficients, indicate that although product features do have some impact on product sales, the impact overall is not large compared to review features.

The impact of the product price on product sales is also analyzed. Only 6,632 (16.58%) products have a valid price in the products dataset, so it is inappropriate to include it in the heterogeneity model. To test the effect of price, the subset of products that have a valid price is selected, and a comparison between the heterogeneity model without the presence of price and the heterogeneity model with the presence of price is made. The heterogeneity model without the presence of price has an R^2 of 0.281, and the heterogeneity model with the presence of price has an R^2 of 0.286. The coefficient of the price variable is 19.99, which indicates a positive impact on product BSR and thus a negative impact on product sales. This is expected since consumers are expected to be more willing to purchase cheaper products.

V. Discussions

A. Explaining Mechanisms Behind the Results

Linear regression models are used in the empirical methodology since it is easy to implement and interpret. However, some drawbacks include that the relationships may not be linear, and the model is sensitive to outliers. It is also important to ensure no significant multicollinearity between the variables, since multicollinearity dampens the coefficients and reduces the explanatory power of the regression model.

The dataset does not record product star ratings, and due to the black box algorithm of Amazon star ratings, the star ratings that actually appear on each product page cannot be fully replicated. Different measures of mean rating and mean sentiment are estimates only, which may overstate or understate the actual product star rating. The lack of information on product star ratings makes it harder to analyze, which may result in either overestimation or underestimation of the impact of consumer ratings on product sales.

Keeping all other independent variables the same, the basic model that includes $MeanAfterBanRating$ shows a greater explanatory power on product ranking than the basic model that includes $MeanRating$. This may be related to review recentness: all reviews prior to the ban of incentivized reviews (October 2016) are relatively old compared to when the data is extracted (October 2018). If a product has a large number of reviews after the ban, consumers may not scroll far enough to see the reviews before the ban. If a product does not have a large number of reviews after the ban, consumers may consider the product not as popular or of high quality, which may hinder their decision to purchase. Either way suggests that the reviews written before the ban are not as important.

This result has implications for vendors: vendors should pay attention to consumer reviews, especially recent reviews, since they can influence product sales in the short run. For example, if many recent reviews of a product complain about a product's packaging, the vendor should check out the problem with the packaging of the product and develop solutions to fix the problem before the product's overall rating starts declining, which will likely lead to a decrease in product sales. The importance of customer reviews also suggests that businesses could increase the number of employees in the customer aftercare team to make sure customer feedback are received promptly and the business can adjust strategies to better suit consumers' needs.

The basic model which includes $MeanVerifiedAfterBanRating$, the mean rating of all verified reviews after the ban, shows a slightly smaller R^2 of 0.300 than the basic model that includes $MeanAfterBanRating$, which has an R^2 of 0.301. It can be due to the following reasons: (1) Consumers do not care much about whether a review is verified or not; (2) There is only a very small amount of unverified reviews after the ban; (3) The unverified reviews show a similar rating distribution as verified reviews.

For the basic model, different means of sentiment scores are not as good an indicator of the product BSR as means of ratings, suggesting that consumers care more about the overall rating than reading the actual review texts. Indeed, a quantified number is often more representative than texts, and easier to understand for people who are lazy to read long lines of reviews. It is also possible that reviews are written in a language that one does not understand, which makes it impossible for consumers to make purchase decisions based on review texts.

Both the basic model and the heterogeneity model show that $MeanAfterBan-Rating$ has the largest impact on product BSR, hence it has a greater effect than the total number of reviews. This indicates that the overall rating is more important than the number of reviews. Both models also show that the total number of review images, or $TotalNumReviewImages$, has a very small impact on product sales. This is slightly unexpected since the author expects a larger impact. This may be explained by the category of the products being analyzed in the models, which is the video games category. Consumers can access game demonstrations and gameplay videos from the internet, which reduces the strength of consumer review images in impacting product sales. However, the impact of the number of review images can be very different for products under other categories. Further research can compare the impact of the number of review images for various categories of products.

In the heterogeneity model, a subset of the data that contains a valid price is used to test the impact of price on product sales. Although the subset sample does not make up a large percentage of the entire dataset, the sample size (6632) is still relatively large, which makes the result credible. The result shows that price does have a negative impact, although not too large, on product number of sales. This can be explained by the observation that the price range of video game products, which is shown in Figure 5, is not large. Most (over 75%) of the products have a price below \$200, and half of the products have a price below \$100. Nevertheless, the negative impact of price on product sales suggests that it is worthwhile for merchants to consider applying price discounts to their products if they want to increase the number of sales, depending on the current sales and price of the product. An important consideration is the trade-off between the increase in the number of sales and the decrease in the profit earned per product sold.

The findings can be used to explain why businesses are paying expensive advertisement fees to platforms and/or influencers to promote their products: advertising increases product exposure and encourages product reviews amongst consumers; by asking influencers to give positive feedback on their product, they further increase the positive reviews for the product, which leads to an increase in product sales. The cost of advertisement is lower than the revenue earned from the advertisements, which explains the motivation behind this action.

The findings also explain the reasons behind Amazon's decision to ban incen-

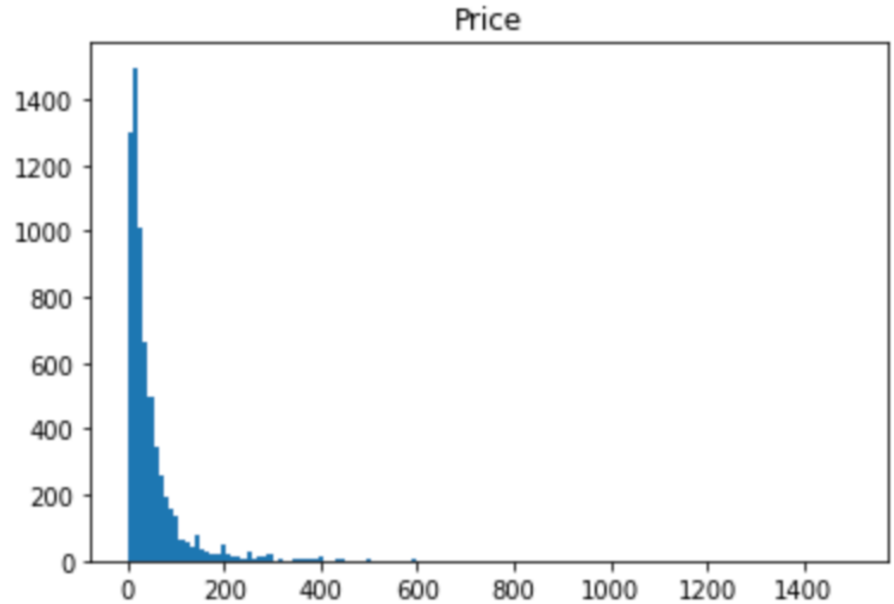

Figure 5. : Boxplot of Product Prices

Note: The histogram shows the distribution of product prices.

tivized reviews and the continuing existence of underground review communities after the ban. Before the ban, sellers well understand the importance of consumer reviews, hence they are willing to offer discounted or free products in exchange for likely higher reviews, which will lead to an increase in sales. Understanding the unfairness created by this program, Amazon banned it to promote more honest feedback. After the ban, some sellers are willing to illegally pay for positive reviews to boost their sales, which is why underground review communities still exist. In response to this, Amazon keeps on detecting possibly fake reviews and hence in April 2018, it deactivated a large number of customer accounts suspicious of using the marketplace for commercial purposes.

B. The Impact of Other Variables on Product Sales

The heterogeneity model result indicates that only 35.8% of the variance in product BSR is explained by review features and product features. This is not surprising since other, more external factors are expected to influence consumer purchasing decisions. The information on reviews and products is also not complete.

The datasets do not record whether the product prices are at discount or not, and the most recent price and the date the product is sold at that price. The price of the product that each review is based on is also unknown, hence the prices recorded in the datasets are only an estimate of the true prices.

The shipping fee for each product is also unknown. The hypothesis is that the presence of shipping will lower product sales. Further research can investigate the impact of (the presence of) shipping fees on product sales.

As mentioned in Section I, each product has a Best Sellers Rank (BSR) and organic ranking. The hypothesis is that organic ranking will have a positive impact on product sales. In other words, the more a product appears at the front line on Amazon's search engine results pages, the more likely it will get higher sales, when keeping all other variables constant. The dataset does not record product organic rankings, hence the impact of organic ranking is unknown. The exclusion of organic ranking in the dataset, however, is also understandable, since the organic ranking will change based on the keywords that consumer type in the search bar, which makes it hard to measure.

External factors such as consumer socioeconomic status can play a huge factor. The purchasing power of consumers at a higher socioeconomic status is higher, which allows them to buy more video games at higher prices with less hesitation.

The gender of the consumer can also impact their purchasing decisions. For instance, Chen et al. (2022) find that female consumers, compared to male consumers, are more affected by negative reviews than positive reviews. The reviews dataset does not record information on reviewers' genders, and using natural language processing to identify reviewer gender based on their name is not a good method, since people of different genders can share the same name, and some reviewers do not use their real names, making it impossible for computers to identity reviewer genders. A controlled experiment is suggested if one aims to explore the impact of consumer reviews between different genders. In addition, social media reviews and reviews by friends and family can also play a role in product sales, since they are also reviews just like the reviews written on Amazon.

Aside from the above factors, the country region of each product may also be a useful factor. People from different countries hold different review evaluation standards; Amazon also owns a different percentage of market share and faces different competitors. It would be valuable to analyze the degree of impact that reviews have on video game buying decisions in different countries, or even different regions of countries.

The analysis is based on the data from Amazon only, and results based on data from other platforms are unknown. Future research can focus on the analysis on another platform and compare the results from this paper, to find out the review impacts among various platforms.

As mentioned in Section I, product BSR is updated every hour to reflect its most recent sales. Hence, this paper analyzes the impact on product sales at a specific point in time only. Future research can consider keeping track of product BSR over time and performing a time series analysis to get a more holistic conclusion.

In 2020, people's lives were significantly impacted by the outbreak of the Covid-19 coronavirus. People were forced to stay home during the lockdown, which boosted video game sales as people try to find a way to entertain themselves without leaving home. In a survey, European gamers report that video games make them feel less detached and happier overall (Clement, 2022*a*). According to Statista, digital gaming sales on in-game content increased by 12% and the sales on paid downloads increased by 21% globally (Clement, 2022*a*). The statistics indicate that it is worthwhile to study the impact of Covid-19 on product sales. However, the dataset only records data up to October 2018, which is long before the onset of Covid-19, hence the impact of Covid-19 is untestable. Nonetheless, it is possible to analyze the Covid-19 factor if there is an updated dataset that contains more recent data.

Finally, it is important to recognize that people extremely satisfied or extremely dissatisfied are more likely to begin word-of-mouth reviews, which could influence the purchasing decisions of people close to them (Anderson, 1998). This also makes online reviews biased.

VI. Conclusion

The empirical model results show that the hypothesis is correct. Consumer reviews in general have a impact on product sales and explain approximately 30.1% of product Best Sellers Rank. Consumer ratings have the largest impact amongst all consumer review features: the higher the rating, the lower the product BSR, which indicates higher product sales. Ratings of the reviews after the ban of incentivized reviews have a larger impact than the ratings of all reviews; this can be partially. While product features also have an impact, the impact is not as large as the impact of ratings; review features and product features altogether explains approximately 35.8% of product BSR. Product brand has a positive impact on sales. On the other hand, product price can influence sales negatively: the higher the price, the lower the product sales.

The results suggest that sellers should consistently monitor consumer reviews and develop new strategies or alter existing strategies based on consumer feedback to keep up their product sales. Sellers whose products are sold at a higher price range may consider lowering product prices to bring up product number of sales as product price has a negative impact on sales.

REFERENCES

Ahsan, Ayesha. 2017. "Consumer ratings-reviews and its impact on consumer purchasing behavior."

Amazon. n.d.. "How Are Product Star Ratings Calculated?"

Anderson, Eugene W. 1998. "Customer Satisfaction and Word of Mouth." *Journal of Service Research*, 1(1): 5–17.

Bajaj, Aryan. 2021. "Can Python understand human feelings through words? – A brief intro to NLP and VADER Sentiment Analysis."

Chen, Pei-Yu, Shin-yi Wu, and Jungsun Yoon. 2004. "The Impact of Online Recommendations and Consumer Feedback on Sales." 58.

Chen, Tao, Premaratne Samaranayake, XiongYing Cen, Meng Qi, and Yi-Chen Lan. 2022. "The Impact of Online Reviews on Consumers' Purchasing Decisions: Evidence From an Eye-Tracking Study." *Frontiers in Psychology*, 13.

Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research*, 43(3): 345–354.

Clement, J. 2022*a*. "COVID-19 impact on the gaming industry worldwide - Statistics & Facts."

Clement, J. 2022*b*. "Distribution of video gamers in the United States in 2022, by age group."

Connolly, Brian. 2022. "Amazon Best Sellers Rank: Everything You Need to Know in 2022."

Coppola, Daniela. 2022. "Annual net sales revenue of Amazon from 2006 to 2021, by segment."

Duan, Wenjing, Bin Gu, and Andrew B. Whinston. 2008. "Do online reviews matter? — An empirical investigation of panel data." *Decision Support Systems*, 45(4): 1007–1016.

Jungle Scout. n.d.. "Amazon Sales Estimator."

Kim, Da Yeon, and Sang Yong Kim. 2022. "The impact of customer-generated evaluation information on sales in online platform-based markets." *Journal of Retailing and Consumer Services*, 68: 103016.

Kunst, Alexander. 2022. "Digital video game purchases by store brand in the U.S. in 2022."

Lohse, Tobias, Jan Kemper, and Malte Brettel. 2017. "HOW ONLINE CUSTOMER REVIEWS AFFECT SALES AND RETURN BEHAVIOR – AN EMPIRICAL ANALYSIS IN FASHION E-COMMERCE." *In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimarães, Portugal*, 2635–2644. ISBN: 978-0-9915567-0-0.

Masters, Kiri. 2021. "A Short History Of Amazon's Product Review Ecosystem, And Where We Are Today."

Ni, Jianmo, Jiacheng Li, and Julian McAuley. 2019. "Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects." 188–197. Hong Kong, China:Association for Computational Linguistics.

OECD. 2019. "Understanding online consumer ratings and reviews." *OECD Digital Economy Papers*, 289. Place: Paris.

Perez, Sarah. 2016. "Amazon bans incentivized reviews tied to free or discounted products."

ReviewMeta. 2016. "Analysis of 7 million Amazon reviews: customers who receive free or discounted item much more likely to write positive review."

Sun, Miao, Jing Chen, Ye Tian, and Yufei Yan. 2021. "The impact of online reviews in the presence of customer returns." *International Journal of Production Economics*, 232: 107929.

UNCTAD. 2021. "Global e-commerce jumps to $26.7 trillion, COVID-19 boosts online sales."

Wang, Alex. 2022. "What Is the Share of E-Commerce in Overall Retail Sales?"

Zhu, Feng, and Xiaoquan (Michael) Zhang. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics." *Journal of Marketing*, 74(2): 133–148.

Appendix

The datasets and code for analysis can be found on: `https://github.com/ting486/Consumer-Reviews-And-Purchasing-Decisions`.