

INTRODUCTION

Online shopping has become more and more prevalent in average consumer's daily life. Electronic shopping brought many advantages to both sellers and buyers. One of the positive perks is Online Consumer Reviews (OCR). Online Consumer Reviews come in the form of electronic word-of-mouth (Dellarocas, 2003) and can be defined as "*peer-generated product evaluations posted on company or third party websites*" (Mudambi & Schuff, 2010). OCR were found to have an effect on product sales as well as consumer decision making process. (Dellarocas et al., 2007; Duan et al., 2008; Lee et al., 2011).

OCR help customers by decreasing their search costs because it is easier for the consumer to skim through many reviews on the internet than finding and asking people who actually bought the product in question. Consumers try to get information on the product they are going to buy so that they can decrease the uncertainty about that particular product. This information acquiring process (search) is costly. Furthermore, this information gathering process might be more or less costly depending on the product type, namely Search and Experience goods (Nelson, 1970). Search goods, by nature, have traits that can be observed objectively before buying (e.g., GHz of a CPU of a Laptop). On the other hand, the quality of Experience goods can not be evaluated before buying and evaluations could be very subjective (e.g., One needs to watch a movie in order to understand if it is a good movie or not. Also a movie being good or bad might depend on who watched it because people have different tastes.).

However, the increased rate of electronic commerce activity also created an abundance of product reviews on the internet. The vast amount of product reviews might increase the information overload on the customer (Fan, 2021; Sun et al., 2019). Unfortunately, this information overload can make the search process even costlier because it is really

time consuming to evaluate thousands of consumer reviews. Besides, not all reviews have the same informativeness level (Pan & Zhang, 2011). Therefore, answering the question of “what makes a helpful review?” is very important because reading more and more reviews without getting useful information can only increase the search costs. On top of this, the overload of useless information might impair the decision making ability of the potential buyer, leading to a sub-optimal choice. To make the understanding of a “helpful review” more clear it can be strictly defined as “*a peer-generated product evaluation that facilitates the consumer’s purchase decision process*” (Mudambi & Schuff, 2010).

Having a good understanding of how consumers handle information and how their behaviors change with respect to information they are exposed to might even unearth newer tools to increase the social welfare because it would lead to a decreased level of information asymmetry between consumers and producers. The decreased information asymmetry might produce two benefits. First, consumers can learn to evaluate the quality of the products before purchase much more efficiently. This leaves less room for any disutility that can arise from purchasing a sub-par quality product. The second, the increased knowledge of the consumers towards certain products may even facilitate competition among firms. This competition may lead increased quality of the products as well as decreased prices in a given market. This is why understanding the determinants of a helpful review is important because it might reveal new insights on consumer behavior and these new insights could be used in favor of both consumers and producers.

Since OCR do not share the same level of informativeness and usefulness (Pan & Zhang, 2011), most of the electronic retail websites implemented a very well known “helpfulness” feature to their customer review systems in order to help their customers navigate through thousands of reviews. On some websites, this feature is implemented in a way that *helpful / unhelpful* buttons are present and on some other websites, there is only a *helpful* button. In this study, the data was acquired from websites which support *helpful / unhelpful* system since it was found that helpfulness as a percentage (helpful votes over total votes) creates more powerful results than considering only helpful votes as a count variable (Ahmad & Laroche, 2015).

The research on perceived helpfulness is still growing. In Social and Behavioral Sciences,

it is common to have conflicting studies and results in the literature (Hunter & Schmidt, 1991). In this case, the study on perceived helpfulness is not an exception. Because of this reason, more empirical studies on perceived helpfulness are required. This study aims to add one more perspective to the growing literature of customer reviews and review helpfulness by employing a unique data set as well as questioning the subject with the intuition of economics. Furthermore, this study tries to show that unstructured data (e.g., text, image, etc.) is a viable option for economic research. With unstructured data, economics may be able to explore new venues.

The existence of the Negativity Bias and how the potential effects of Negativity Bias on Review Helpfulness could be moderated by either Search And Experience goods will be the main focus of this study. A lesser degree of Negativity Bias is expected for Experience Goods. Additionally, the amount of information that a review has and its connection to Review Helpfulness with respect to Review Ratings will be analyzed. Next, the diagnostic values of the negative and positive sentiments are going to be discussed. Because of the Negativity Bias, a positive relationship between Negative Sentiments and Review Helpfulness is expected. Lastly, the positive relationship between the Review Helpfulness and the timeline of a review will be inquired because an updated and new information might be more helpful than an old information.

This study uses secondary data. The reviews were collected from two websites. Namely, IMDB and BestBuy. The data was scrapped with Selenium Webdriver for Python (Van Rossum & Drake, 2009). Over 400,000 reviews and their meta data were gathered initially. After the data transformation process the final sample size was over 23,000. Then this data was subjected to empirical tests. It was found out that consumers usually find extreme reviews more helpful than neutral ones. Furthermore, consumers prefer longer reviews when reviews are rated negatively. On the contrast, positive reviews with fewer words were deemed more helpful by readers on the average. In addition, the sentiment of a review also has a significant effect on the helpfulness of a review. These effects were further moderated by product types, particularly, Search and Experience goods.

The plan of this study is outlined as the following. In the next section, the literature on the subject of Review Helpfulness is going to be presented. In the second section, the related

theories which would help with the understanding of the consumer behavior and review helpfulness will be laid out. Next, the data and the research methodology are going to be introduced. The results will be shared in the fourth section. In the fifth section the results are going to be discussed and then the study is going to be concluded in the last section.

CHAPTER 1

LITERATURE REVIEW

The research on Online Consumer Reviews can be considered to have two main branches. One branch is heavily marketing oriented and tries to understand the effects of OCR on sales (Ghose & Ipeirotis, 2011; Öğüt & Onur Taş, 2012). The other branch is concerned with rather behavioral or psychological outcomes and focuses on perceived helpfulness only. The boundary between these two branches are somewhat blurry and the same independent variables were used to explain both helpfulness and sales. Furthermore, Topaloglu & Dass (2019) found that review helpfulness have a moderating effect on sales. Many other papers used the perceived helpfulness to explain sales as well (Ghose & Ipeirotis, 2011; Lee & Choeh, 2018). Thus, these two branches are inter-connected with each other. Contrary to sales oriented papers which were mentioned above, this study tries to understand behavioral aspects of the review helpfulness rather than its effects on sales.

Many different independent variables and their relationship between perceived helpfulness were examined by researchers. Hong et al. (2017) groups these independent variables into two chunks. The first group is reviewer-related predictors. Variables within this group try to understand the relationship between reviewer status (reviewer expertise, number of past reviews etc.) and perceived helpfulness. (Ghose & Ipeirotis, 2011; Sun et al., 2019). The second group tries to understand the review-related relationships such as review length, review extremity, discrete review emotion etc. (Eslami et al., 2018; Ghose & Ipeirotis, 2011; Mudambi & Schuff, 2010; Ren & Hong, 2019). Ghose & Ipeirotis (2011) showed that review-related predictors have stronger prediction power

than reviewer-related ones. It should be noted that these two groups are not mutually exclusive and can be employed in the same study. The choice of the variables is heavily dependent on the data itself. In this particular study, only review-related predictors were examined because the data which is available for this study only allows for review-related variables. Reviewer-related variables could not be employed because IMDB and BestBuy use different conventions to identify their users (reviewers).

Review Length (Depth) could be considered as one of the most important independent variable explaining the perceived helpfulness. It might be implemented in different ways but for the majority of the studies it is the word count in a given review. For the most of the researchers, longer reviews contain more information about the product. As the potential buyer gets more information from a single review, he decreases his search costs with minimum effort so the amount of information (words) that a review contains becomes very important. (Choi & Leon, 2020; Hong et al., 2017; Korfiatis et al., 2012; Lee & Choeh, 2016, 2014; Liu et al., 2014; Mudambi & Schuff, 2010; Wang et al., 2018). However, other group of researchers found out that there was a threshold on review length, meaning unnecessarily long reviews might have useless information in them(Huang et al., 2015; Schindler & Bickart, 2012). Another group indicates the word count of the review shows diminishing returns to helpfulness, supporting the threshold idea (Baek et al., 2012; Es-lami et al., 2018; Kuan et al., 2015). Although very rare, a negative relationship between Review Length and Perceived Helpfulness was also reported (Chatterjee, 2020; Yang et al., 2019).

Review Rating (Review Extremity) is another important identifier of perceived helpfulness. This is usually the star rating on a 1-5 or 1-10 scale and resembles the evaluation of the reviewer for a given product. Its effect on perceived helpfulness is rather divergent in the literature. A number of researchers found out that there is a non-linear relationship between review extremity and perceived helpfulness. Specifically, a U-Shaped function which means that reviews that are on the extremes (either very negative or very positive) are more helpful than neutral reviews (Cao et al., 2011; Ghose & Ipeirotis, 2006; Liu et al., 2014; Mudambi & Schuff, 2010; Wang et al., 2018). In other studies, linear positive relationship was found. In other words, positively rated reviews are more useful (Choi &

Leon, 2020; Korfiatis et al., 2012; Lee & Choeh, 2016; Pan & Zhang, 2011; Schindler & Bickart, 2012). There are also results that support negative linear relationship (Chua & Banerjee, 2014) and no effect at all (Hong et al., 2017).

Review Age is an indicator of how long the time passed after a review published on a web site. In some papers the opposite of Review Age, Review Freshness (Recency) could be employed as well. Different effects on perceived helpfulness were also observed by researchers. Some researchers discussed that a timely review is very important at conveying the true information about a certain product, making older reviews more helpful (Li et al., 2019; Liu et al., 2014; Zhou & Yang, 2019). Others discussed that some web sites promote newer reviews and recent information is much more valuable than old information. Because of these more recent reviews can be considered as more helpful (Choi & Leon, 2020; Lee & Choeh, 2014; Wu et al., 2021).

The counts of negative or positive words out of total word count (the proportion of negative and positive words) were employed by a handful of studies (Baek et al., 2012; Kuan et al., 2015). These metrics can measure the sentiment of the review itself. They are usually different than Review Extremity (Review Rating). In most cases Review Extremity and Review Sentiment (Valence) do not match (Zhang et al., 2011). Baek et al. (2012) and Kuan et al. (2015) discovered a positive relationship between the proportion of negative words and perceived helpfulness. The relationship between the proportion of positive words and perceived helpfulness was found to be negative (Kuan et al., 2015).

The classification of Experience and Search goods plays a moderating role in the literature (Baek et al., 2012; Lee & Choeh, 2016; Mudambi & Schuff, 2010; Sun et al., 2019). Some of these studies showed that Review Length has a greater effect for search products than experience products (Baek et al., 2012; Lee & Choeh, 2016; Mudambi & Schuff, 2010). These findings suggest that search goods can be evaluated in a more objective way than experience goods. There is also moderating effect of product type on Review Extremity. Mudambi & Schuff (2010) showed that moderate reviews are more helpful for experience products than search products. Baek et al. (2012) and Lee & Choeh (2016) had the results which indicated that more extreme reviews (negative or positive) is much more helpful for search goods than experience goods.

Many different types of research methodologies were employed in the literature of the Review Helpfulness. To my knowledge the most common empirical analysis method that was used in the literature was Tobit Regression because of the limited and the censored nature of the Reviews Helpfulness (Choi & Leon, 2020; Chua & Banerjee, 2014; Huang et al., 2015; Korfiatis et al., 2012; Mudambi & Schuff, 2010; Ren & Hong, 2019). Some of these studies also shared Ordinary Least Squares results alongside with the Tobit Regression results for the purpose of benchmarking (Choi & Leon, 2020; Kuan et al., 2015; Mudambi & Schuff, 2010). Ordinary Least Squares were also used as a main empirical methodology in some papers (Ghose & Ipeirotis, 2006; Lee & Choeh, 2016; Liu et al., 2014). Furthermore, Ordinary Least Squares was employed in a hierarchical manner as well (Baek et al., 2012). Some other studies employed Multi-Level Regression model in order to deal with the potential selection bias which is present in the data (Ghose & Ipeirotis, 2011; Kuan et al., 2015; Wu et al., 2021). In a rare case Ordinal Logistic Regression was one of the methods that was used to explain the determinants of the review helpfulness (Cao et al., 2011). Predictive methods such as Neural Networks were used to understand the components of a helpful review as well (Lee & Choeh, 2014). One paper attacked the problem of the review helpfulness with three methods, specifically, Partial Least Squares - Structural Equation Modelling, ANOVA Analysis and Artificial Neural Networks (Eslami et al., 2018).

Any website which offer online reviews for products could be used as a data source (Yelp, Booking, BestBuy, eBay) in the literature of review helpfulness but the majority of the aforementioned studies used the data either from websites Amazon or Internet Movie Database (IMDB) because these websites offer relatively easier ways to get the needed data. They have their respective APIs (Application Programming Interface) and on top of that, it is relatively easier to scrape both websites. Additionally, both websites offer product reviews for a numerous different types of products. Amazon is a huge e-retailer and offers a wide range of items. IMDB is a home for both critic and user reviews of Movies and TV Shows. In some other studies local equivalents of Amazon were used as well. (Sun et al., 2019). Cao et al. (2011) used data from CNET Download.com which is a software distributor.

There might be variety reasons why there are some conflicting results in the literature. First of all, many different data sources were used in the studies. Even in the same data source, for instance Amazon, there exist a wide range of products with different brands. Different sub-groups of products can receive different feedbacks. This might increase the potential noise in the data, leading conflicting results. Secondly, many different research designs and research methodologies were employed. In these different research designs different independent and control variables were used because the data at hand may allow for different sets of variables. Lastly, there is no established theory of review helpfulness. This means that studies are rather unorganized and researchers are usually inclining towards their own domain of expertise when it comes to explaining review helpfulness.

Since there are many conflicting results in the literature of review helpfulness additional empirical contributions are important. As well as adding new empirical findings to the literature, this study tries to extend the potential predictors with the continuous interaction term between review length and review rating to test for the existence of the Negativity Bias. Furthermore, some under-utilized predictors from the literature such as the proportion of negative words and the proportion of positive words are also used to test Negativity Bias and Loss Aversion.

CHAPTER 2

THEORETICAL FRAMEWORK

2.1 SEARCH AND EXPERIENCE GOODS CLASSIFICATION

Relaxing the assumption that consumers have perfect information about the products on the market raises the problem of information acquisition for consumers. Getting the complete information about products is important because economic agents want to maximize their utility by finding the best good with respect to their budget constraints on the market. The complete information is the information about both the prices and the quality levels of the goods which are available on the market. Acquiring the information about the quality of a product is harder than getting the information about its price. To explain, a consumer can check the price list of laptops and have an idea about the maximum and the minimum prices but getting the information about a laptop's technical components or a brand's technical support for its products is a lot more effort taking and costly.

In order to solve this problem Nelson (1970) differentiated goods based on their search and experience attributes (Nelson, 1974a, 1974b; Nelson, 1981). Search Attributes are the qualities that can be assessed before purchase. On the contrary, Experience Attributes can only be assessed by using the product itself, in other words, evaluation is only possible after purchase. A single good can have both search and experience attributes. Depending on its dominant attributes the good can be classified as Search or Experience. Thus, search and experience classification is not usually a clear cut one, rather it could be considered as a spectrum. Darby & Karni (1973) extended Search and Experience goods framework with Credence goods where consumers can not evaluate the good even after

the consumption. Though, Credence goods will not be discussed any further in this study.

Consumers can choose from two different strategies when they want to acquire more information about products. Consumers “search” in order to get information about a product assuming that they perfectly know where they can obtain the knowledge. Nelson (1970) defines the search process as “*The most obvious procedure available to the consumer in obtaining information about price or quality is search*”. The search process is complete when consumers inspect the option and that inspection happens before the purchase. Contrary to the search process, consumer may choose to “experience” the product. The consumer follows this strategy when the search for information is either inappropriate, inefficient or costly compared to just experiencing the good. This is true for cheap goods especially. For instance, in order to find out the best chocolate in the market it is best for the consumer to buy every other brand available and taste them. Thus, consumers will always prefer sampling or trial use (experiencing the product) given the fact that the cost of the sampling or trial is lower than the cost of the search. This is why search is costlier for Experience goods. Therefore, the kind of strategy that a consumer chooses depends on the product’s Search or Experience attributes and how dominant these attributes are.

Nelson (1970) also states that consumers can use other consumers’ experiences in their search processes. The information about a certain good is usually acquired in the form of “word-of-mouth” with this kind of search process. This type of search is viable when getting the information about a certain product is much more costly by “experience” compared to by “search”, i.e., sampling the product is very costly. For instance buying a laptop and trying to figure out if it is the best laptop for a certain task may even require buying additional laptops because of the need for comparison. This is the way of “experiencing” the good and this would be very costly. On the other hand, the potential buyer may want to hear from people who do the similar tasks with their laptops and try to sort out the best possible laptop model for her own task. This way she gets information about the product without having to “experience” it and this approach is way cheaper.

Klein (1998) discusses that experience sharing becomes easier with the increasing usage of the internet. Because experiences about products can be shared much more easily on the internet, consumers’ pre-purchase behavior might change accordingly. The reason

of this is that internet decreases search costs considerably. Furthermore, the potential consumer may engage in “vicarious learning” where he/she gets indirect usage experience through the experiences of other consumers in the search process (Murray, 1991).

Ultimately, to put the theory in a more generalized way, a product’s Total Cost consists of both Search Cost and Product Cost. This Search Cost is higher for Experience goods, in other words, information search is costlier for Experience goods than Search goods. In the consumer’s perspective, to make the search viable, the marginal benefit of an additional search must be greater than the marginal cost of the additional search. The increasing use of internet decreases the search costs. Because of the decreased search costs, the consumer may undertake more search processes and this might change how people perceive the good types, in other words, internet and virtual experience sharing can convert some Experience goods to Search goods (Klein, 1998).

2.2 NEGATIVITY BIAS AND LOSS AVERSION

Negativity Bias argues that humans or even animals have tendencies to be more aware of negative entities than positive ones. In Rozin & Royzman (2001)’s words Negativity Bias can be best exemplified such as the following:

“Brief contact with a cockroach will usually render a delicious meal inedible. The inverse phenomenon rendering a pile of cockroaches on a platter edible by contact with one’s favorite food is unheard of. More modestly, consider a dish of a food that you are inclined to dislike: lima beans, fish, or whatever. What could you touch to that food to make it desirable to eat that is, what is the anti-cockroach? Nothing! . . . One of the best generic descriptions of this relative power of negative contamination is embedded in an age old Russian adage: ‘A spoonful of tar can spoil a barrel of honey, but a spoonful of honey does nothing for a barrel of tar.’ ”

Rozin & Royzman (2001) explains the Negativity Bias by describing it with four main aspects. Namely, Negative Potency, Steeper Negative Gradients, Negative Dominance and Negative Differentiation (Eslami et al., 2018; Rozin & Royzman, 2001; Sen & Ler-

man, 2007). Loss Aversion has a strong connection to Negativity Bias since negative cues might be used as an early warning against potential losses (Herr et al., 1991).

Negative Potency defends that the exact opposite of something positive carries heavier weight. In other words, negative events are much more impactful than their exact positive counterparts. For instance, losing 300\$ could have a harder psychological effect on a person than that particular person is gaining 300\$. Loss Aversion (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991) is the phenomenon where the Negative Potency can be observed the best. People risk more (less) when they are losing (gaining) something. This behavior can be explained by the Endowment Effect (Kahneman et al., 1990). Endowment effect occurs because people usually over-estimate the value of their belongings, for example a pencil. However, they would have under-valued the same pencil if the item did not belong to them. The reason to this behavior is that the dis-utility of losing that pencil is greater than the utility of acquiring the same pencil.

Greater Steepness of Negative Gradients simply implies that the marginal psychological effect of a negative entity is much more higher than a positive one. In other words, negative things cause more negativity as the time goes on. At the very least, the effects of the negativeness are a lot more long-standing than effects of the positiveness.

Negativity Dominance asserts that a balanced mixture of positive and negative evaluations creates a perception of negativity. This means that negative entities have higher weight than positive entities. Rozin & Royzman (2001) also puts forwards as an example:

“The image that comes to mind from the physical domain is that of a single cancerous growth or germ that radiates itself through and ultimately consumes a perfectly healthy body. The image that comes to mind from the moral domain is that of a single vice corrupting and perverting and bringing the moral downfall of an otherwise perfectly good person.”

Greater Negative Differentiation discusses the possibility that human cognition is fine-tuned to negative entities because being aware of negative stimuli might correlate with survival in the process of the evolution. As a consequence, the vocabulary that reflects and describes the state of the negativeness is much more wider than its counterpart the

positiveness. (Peeters, 1971).

Prospect Theory (Loss Aversion) criticizes Expected Utility Theory by introducing the thought that Expected Utility Theory is insufficient explaining the decision making behavior of economic agents who are under risk. In Expected Utility Theory outcomes are weighted by their probabilities but Prospect Theory discusses that people dynamically give more or less weight to the outcomes depending on if it is a loss or a gain. This is called *certainty effect* (Kahneman & Tversky, 1979). The Certainty Effect causes people to behave differently whether they are going to lose or gain something. In the positive domain, people will prefer a small amount of a certain gain over a big gain which is merely probable (risk averse). On the contrary, in the negative domain, people will have a preference towards a big loss which is merely probable over a small certain loss (risk seeking).

Since the cognitive abilities of humans are limited they tend to focus on what is crucial to them. Therefore, in the information processing context and because of the existence of the negativity bias, it could be expected that people process negative cues more thoroughly than positive ones. Furthermore, people may develop much more advanced memory about negative entities. This can also cause high awareness to negative information (Baumeister et al., 2001; Taylor, 1991).

CHAPTER 3

THE DATA AND RESEARCH METHODOLOGY

3.1 DATA COLLECTION

The data was scraped with a custom web scraper which was constructed with Selenium Framework for Python 3 (Van Rossum & Drake, 2009). The scraper acquired the customer reviews from IMDB and BestBuy. The former is the biggest platform for Movies and TV Shows, the latter is one of the biggest electronic retailers in United States of America. Both websites host a great amount of online reviews. After the data was collected it was further transformed and analyzed with the help of R Programming Language (R Core Team, 2020; Silge & Robinson, 2016; Wickham et al., 2019).

Online Consumer Reviews from IMDB were collected as an experience goods data since movies can be considered as pure-experience goods and their search attributes are very limited. It is nearly impossible to evaluate a movie before watching it. Furthermore, evaluations for a movie can both have subjective (e.g., personal attachment to the protagonist, atmosphere of the movie etc.) and objective aspects (e.g., writing, production quality etc.). There are also numerous studies that analyzed movies as experience goods (Baek et al., 2012; Ghose & Ipeirotis, 2006; Lee & Choeh, 2016; Liu et al., 2014; Nelson, 1970).

BestBuy is a very big electronics retailer and this is why the website offers a sizable number of OCR for search goods. Contrary to experience goods, search goods usually have more objectively quantifiable traits. For example, resolution or screen size of a TV, RAM or GPU of a Laptop, storage or battery lifetime of a Smart Phone. Additionally, in the literature, TVs, Laptops and Phones were used as search goods (Baek et al., 2012;

Mudambi & Schuff, 2010; Ren & Hong, 2019; Sun et al., 2019). Therefore, search goods data was collected from BestBuy as it was a very suitable candidate for this research.

From IMDB reviews of 22 movies were collected as experience goods. In total, 22 movies account for 31,863 reviews. From BestBuy, total of 383,357 reviews were collected as search goods. Specifically, 201,113 reviews for TVs, 130,270 for Laptops and 51,974 for Phones. However, in the data there are many reviews that were never got voted and probably they were never read because they got lost in the crowd of other reviews. To make sure that the analysis sample contains only the reviews which were actually read by potential buyers, a threshold of 5 was set for Total Votes. In other words, reviews that had less than 5 Total Votes were removed from the analysis data (Baek et al., 2012; Choi & Leon, 2020). At the very end, the analysis data was consisted of 15,024 observations for experience goods and 8,042 observations for search goods. It should be noted that 47 percent of the reviews had total votes of 5 or greater for experience goods, whereas only 2 percent of the search goods were voted at least 5 times. Most of the search product reviews are not getting voted and even possible that they are not getting read at all. This might support the idea that people form more homogeneous opinions about search goods (Hong & Pavlou, 2014). Once a helpful review is written most of the other reviews become redundant since other reviews most likely convey similar information. On the contrary, more attendance to the reviews can be observed for experience goods (movies) since nearly half of them were voted at least five times. This might be because people form heterogeneous opinions about experience goods (Hong & Pavlou, 2014).

As a data cleaning note, in IMDB data there were cases (20 observations to be exact) where the same reviews were written by different user names. All duplicates along with original reviews were removed because the reviews had different helpfulness ratios.

3.2 EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) has no set of procedures to be followed. The way EDA is conducted almost always depends on the context of the research. Jebb et al. (2017) discusses that EDA encourages scientific replication if done properly. This section is going to be constructed with two notions in mind:

1. Transparent data sharing for the sake of increased reproducibility value
2. Helping readers build a better intuition about the model and its components through transparency

The goal of this section is not “confirmatory” whatsoever. Both dependent and independent variables are going to be introduced with an emphasis on their quantitative features such as their distributions and potential transformations which can be applied to the variable in question. In addition to this, bivariate relations between independent and dependent variables will be plotted in order to get an initial idea about the potential effects along with similar findings from the literature. Thus, in a sense, EDA in this section does not carry out the purpose of discovery, rather it merely puts the data on the spot for the sake of transparency.

3.2.1 Dependent Variable

$$\text{Perceived Helpfulness} = \frac{\text{Helpful Votes}}{\text{Total Votes}}$$

The ratio of helpfulness is derived from the formula above. Below on Figure 3.1 the distribution of the helpfulness for search goods can be observed along with the sample mean value which is represented as a vertical line.

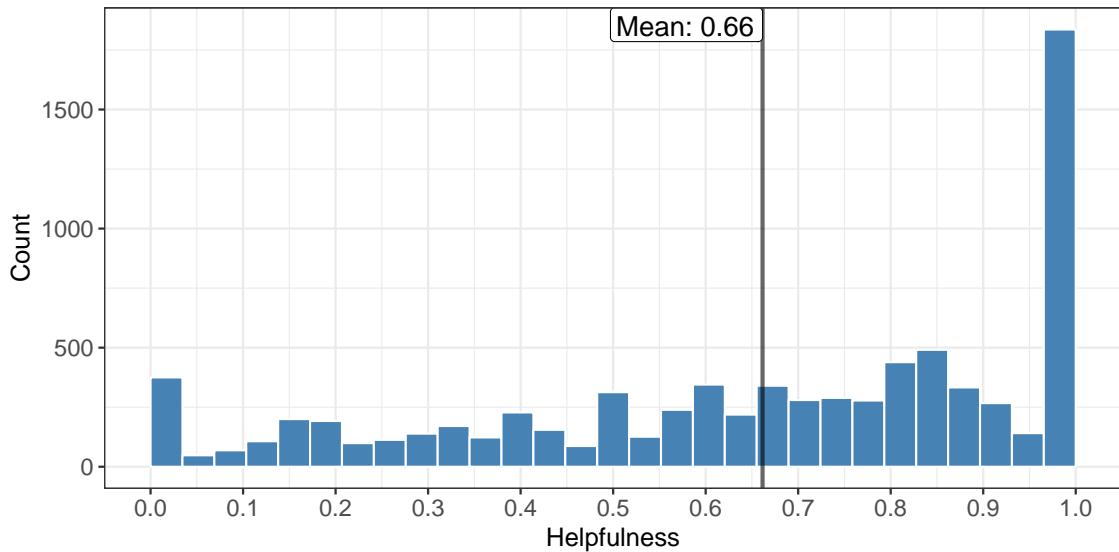


Figure 3.1: Distribution of the Helpfulness of Search Goods

As it can be seen the distribution is very skewed. Looks like in the electronics case, readers find some reviews very helpful. In fact, in the data there is one review which had over 170 helpful votes with zero unhelpful votes. On the other hand, distribution for experience goods on Figure 3.2 presents a different picture. This time helpfulness ratio is sort of normally distributed. There are not very helpful reviews. The most helpful review interval seems to be lying between 70% to 80%.

On a small note, the way the dependent variable was distributed for search goods may raise some questions about whether the standard errors will be calculated correctly or not in a linear regression method. Li et al. (2012) found out that given a relatively high sample size ($N > 3000$) linear regression techniques produces sound estimates of coefficients. Since the sample size for search goods is over 8000 it might be assumed that the use of linear regression methods is acceptable.

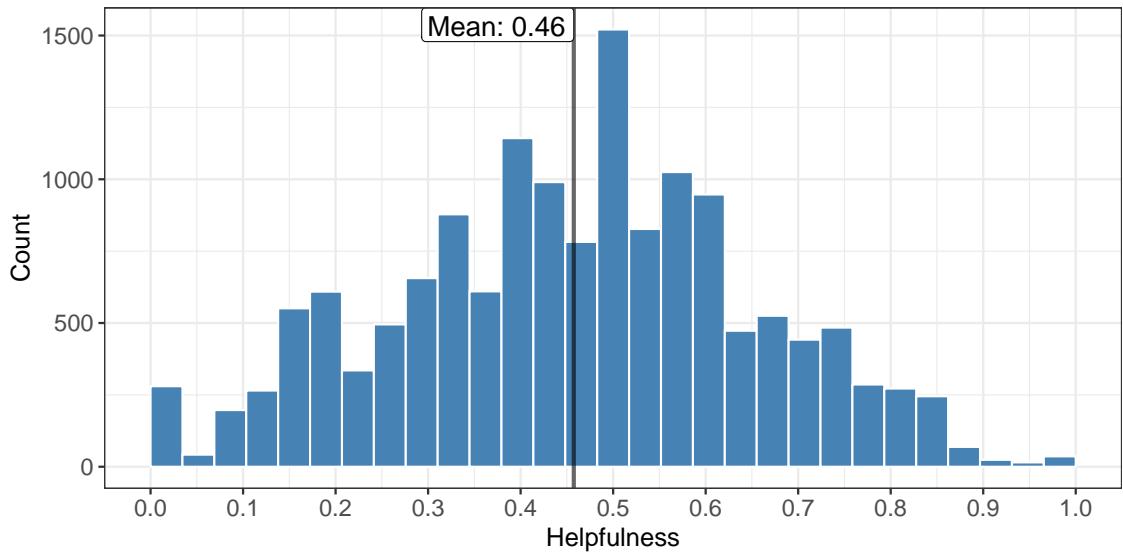


Figure 3.2: Distribution of the Helpfulness of Experience Goods

Hong & Pavlou (2014) states that consumers likely to form uniform opinions about search goods (Sun et al., 2019). On the contrary, the same paper suggests that consumers' opinions are more diversified on experience goods. Intuitively, what these histograms show is the same. For instance, positive or negative attributes of search goods can be expressed in a limited number of ways. The storage space of a phone might be small or CPU of a laptop might not be fast as expected. Thus, one good review is enough to reveal the quality of a product. This might be the reason why such a big spike at 100% was observed on Figure 3.1. In the movies (experience good) case on Figure 3.2, it can be seen that readers usually disagree with each other because the observations of the perceived helpfulness distributed around 50%. This might be because evaluation process of movies is rather subjective. For instance, one person might not like Spider-Man: Homecoming and another would love it. This might depend on how a person idealizes the Spider-Man or Peter Parker image in her mind which is completely subjective.

3.2.2 Independent Variables

3.2.2.1 Review Rating (Review Extremity)

Review Rating is the star rating that was given by the reviewer as an evaluation of a product.

For both search and experience goods the bi-modality of the distributions can be observed on Figure 3.3 (A) and Figure 3.4 (A) respectively. Hu et al. (2006) also pointed out a similar distribution concerning Online Product Reviews. This might mean that consumers usually tend to write product reviews when they are extremely happy or unhappy with the product.

Black vertical lines show the mean ratings for both product types. It is nice to see that averages are close to the neutral rating of 3. With this it can be made sure that the sample reviews do not address to positive or negative side only.

Because the data was taken from two different websites (BestBuy and IMDB), there are two different scales. For IMDB that is 1 to 10 and for BestBuy the scale is between 1 and 5. To make the interpretation of the regression coefficients uniform, IMDB's scale was divided by 2.

On Figure 3.3 (B) and Figure 3.4 (B) the bivariate relationship between perceived helpfulness and review extremity can be seen. Once again, it must be emphasized that this analysis is not confirmatory. On figures, we can observe "U" type of relationship for both Search and Experience goods, though the non-linearity seems to be stronger with Search goods. This might suggest that on the average extreme reviews might be more helpful than moderate reviews (Cao et al., 2011; Ghose & Ipeirotis, 2006; Liu et al., 2014; Mudambi & Schuff, 2010; Wang et al., 2018). On the other hand, it is observable that positive reviews on the average are more helpful than the rest. This might indicate that there is a linear positive relationship (Choi & Leon, 2020; Korfiatis et al., 2012; Lee & Choeh, 2016; Pan & Zhang, 2011; Schindler & Bickart, 2012).

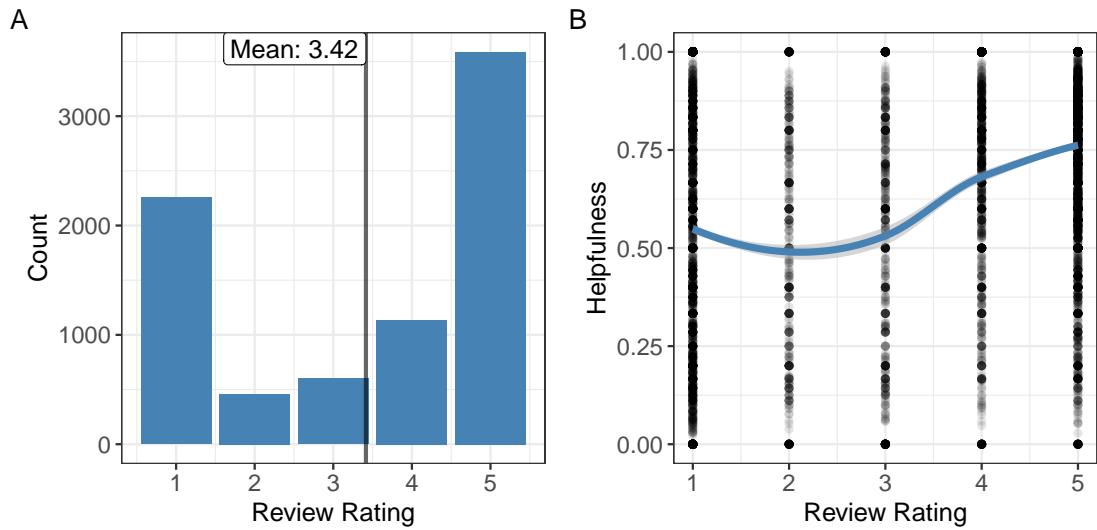


Figure 3.3: The Distribution(A) and the Scatter Plot(B) of Review Rating for Search Goods.

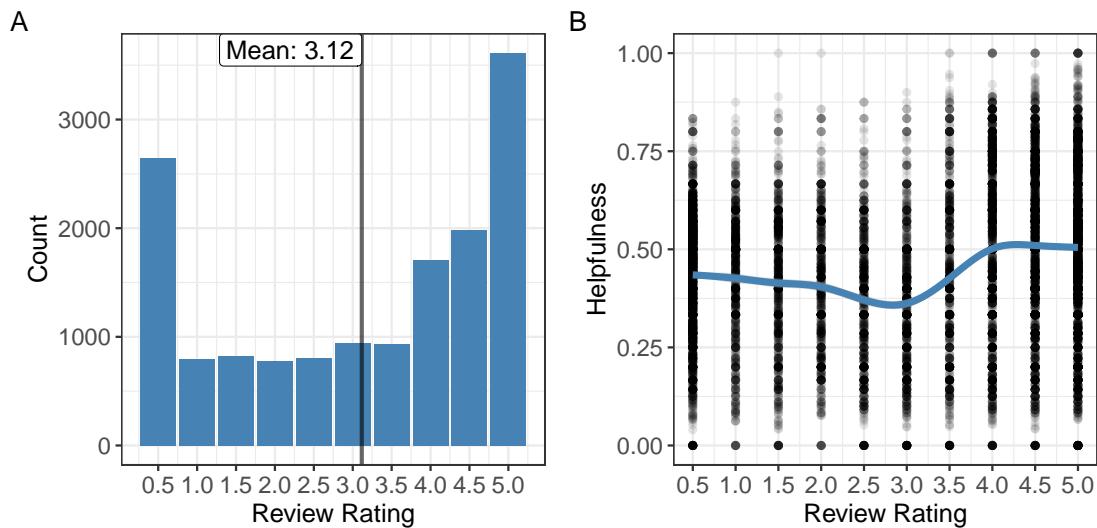


Figure 3.4: The Distribution(A) and the Scatter Plot(B) of Review Rating for Experience Goods.

3.2.2.2 Review Length (Review Depth)

Review Length is the number of the words in a given review.

The exponential-like distribution of the number of words can be observed for both good

types on Figure 3.5 (A) and Figure 3.6 (A). This means that reviews become less prevalent as they get longer. This might suggest that logarithmic transformation can increase the model fit. For the transformed distributions see Figure 3.5 (B) and Figure 3.6 (B). As it can be observed from the vertical mean lines the reviews of the experience goods tend to be longer than the reviews of the search goods.

The bivariate relationship between the number of words and Helpfulness seems to be varying for different product types. For Search goods on Figure 3.5 (C) there is a clear positive relationship (Choi & Leon, 2020; Hong et al., 2017; Korfiatis et al., 2012; Lee & Choeh, 2016, 2014; Liu et al., 2014; Mudambi & Schuff, 2010; Wang et al., 2018). However, this positive relationship appears to stop at a threshold (Baek et al., 2012; Es-lami et al., 2018; Huang et al., 2015; Kuan et al., 2015; Schindler & Bickart, 2012). With this in mind, logarithmic transformation might make more sense (see Figure 3.5 (D)) for modelling purposes. For Experience goods, on the other hand, there seems to be a weak negative relationship or no relationship at all (Chatterjee, 2020; Yang et al., 2019) (see Figure 3.6 (C) and (D)).

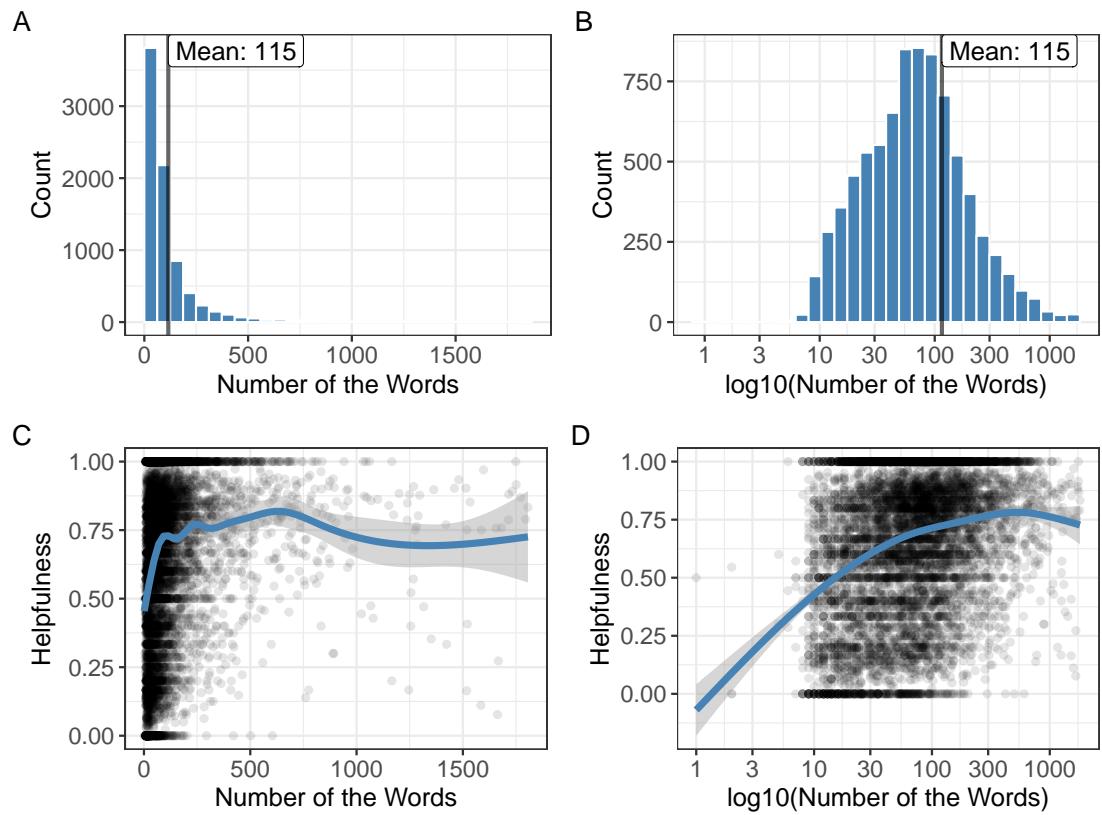


Figure 3.5: The Distribution(A) and the Scatter Plot(C) of Review Length for Search Goods and respective logarithmic transformations (B), (D)

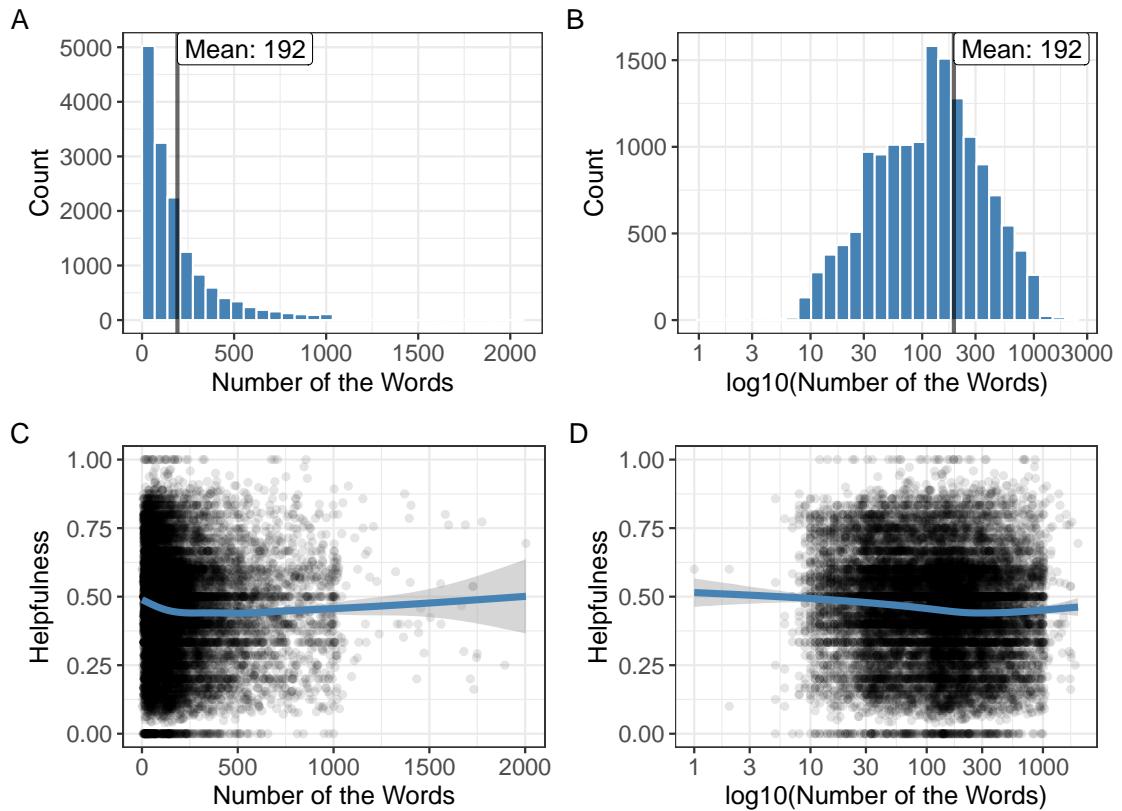


Figure 3.6: The Distribution(A) and the Scatter Plot(C) of Review Length for Experience Goods and respective logarithmic transformations (B), (D)

3.2.2.3 Freshness of A Review

Review Freshness indicates how recent a review is. To engineer this variable following steps were taken:

1. The first review date for a particular brand and model of a product (Movie, Phone, Laptop, TV) was taken.
2. The first dates for particular products then subtracted from other reviews' dates under the respective brand and the model resulting an integer variable representing days.
3. Bigger numbers represent more recent reviews.

Distributions of both product types resemble exponential distribution (see Figure 3.7 (A) and Figure 3.8 (A)). Again a logarithmic transformation might be useful here (see Figure

3.7 (B) and Figure 3.8 (B)) . Interestingly, even after 3 years from the release date some products are still getting reviewed from customers. Vertical mean lines indicate that the products in the experience goods sample are older than the products in the search goods sample because the margin between the first review for a given product and last review for a given product is wider.

On the scatter plots a weak relationship between perceived helpfulness and review freshness could be observed. The smoother line suggests both positive (Choi & Leon, 2020; Lee & Choeh, 2014; Wu et al., 2021) and negative (Li et al., 2019; Liu et al., 2014; Zhou & Yang, 2019) linear relationship between two variables. This non-linearity may explain different findings in the literature.

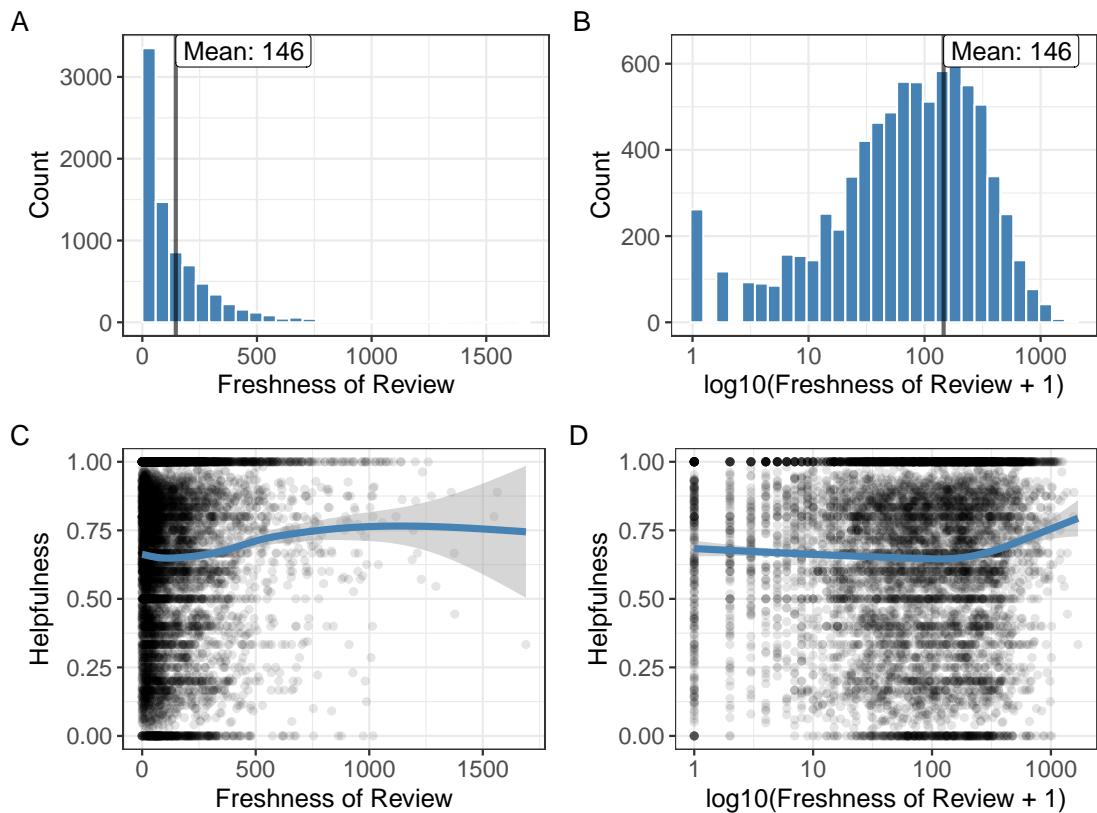


Figure 3.7: The Distribution(A) and the Scatter Plot(C) of Review Freshness for Search Goods and respective logarithmic transformations (B), (D)

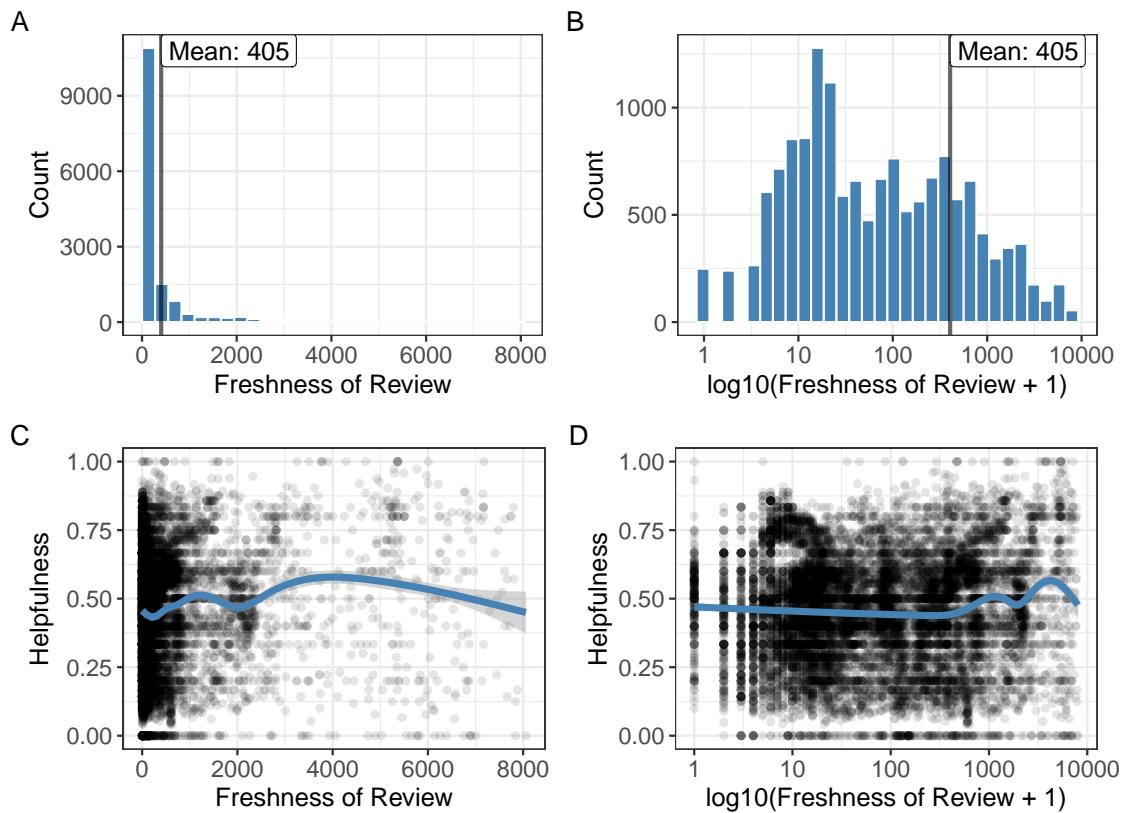


Figure 3.8: The Distribution(A) and the Scatter Plot(C) of Review Freshness for Experience Goods and respective logarithmic transformations (B), (D)

3.2.2.4 Proportion of Negative Words

Negative words were derived by using NRC Word-Emotion Association Lexicon (Mohammad & Turney, 2013). A lexicon is basically a dictionary of words. These words might have similar emotions or sentiments. In this case, negative words lexicon was used in order to engineer this variable. A sample of negative words from the lexicon can be observed below.

“caricature, superstition, compost, misconception, grief, violation, scrub, omen, erratic, porcupine”

$$\text{Negative Word\%} = \frac{\text{Number of Matched Negative Word}}{\text{Total Words in the Review}}$$

These steps were taken in order to create this variable:

1. Words in a review were tokenized.
2. Tokenized words were matched with the words in the lexicon.
3. Matches were counted and divided by the total number of words with the respective review then multiplied by 100 in order to make regression results more interpretable.
4. Logarithmic transformation was applied in order to scale the variable in a more normalized way.

On Figure 3.9 (A) and Figure 3.10 (A) the distributions of the variable for both good types can be seen. On logarithmic transformed plots, X-axis labels were given in the logarithm of base 10 so 1.00 on the X-axis implies that 10% of the review was negative words. For example in this case the mean negative proportion rate is 2,3% ($10^{0,37}$) for search goods and 3,7% ($10^{0,57}$) for experience goods. It should be noted that 0.33 percent of the reviews for search goods do not have any negative words in them. For experience goods this number is 0.14 percent.

On Figure 3.9 (D), non-linear relationship between perceived helpfulness and the logarithm of the proportion of negative words can be observed for search goods. This might suggest that too much negative information might fall into slander rather than informing the customer about the potential risks of the product. On the other hand, somewhat weak negative linear relationship between perceived helpfulness and proportion of negative words can be seen for experience goods on Figure 3.10 (D).

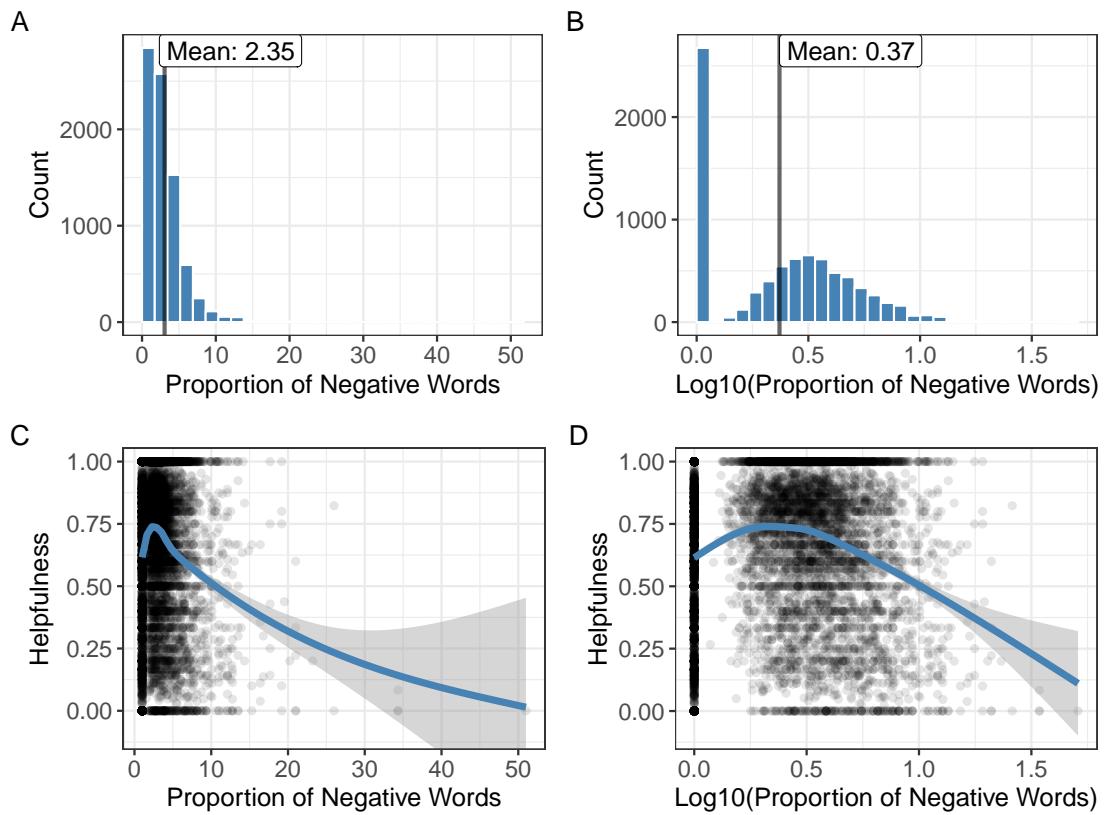


Figure 3.9: The Distribution(A) and the Scatter Plot(C) of Proportion of Negative Words for Search Goods and their respective logarithmic transformations (B), (D)

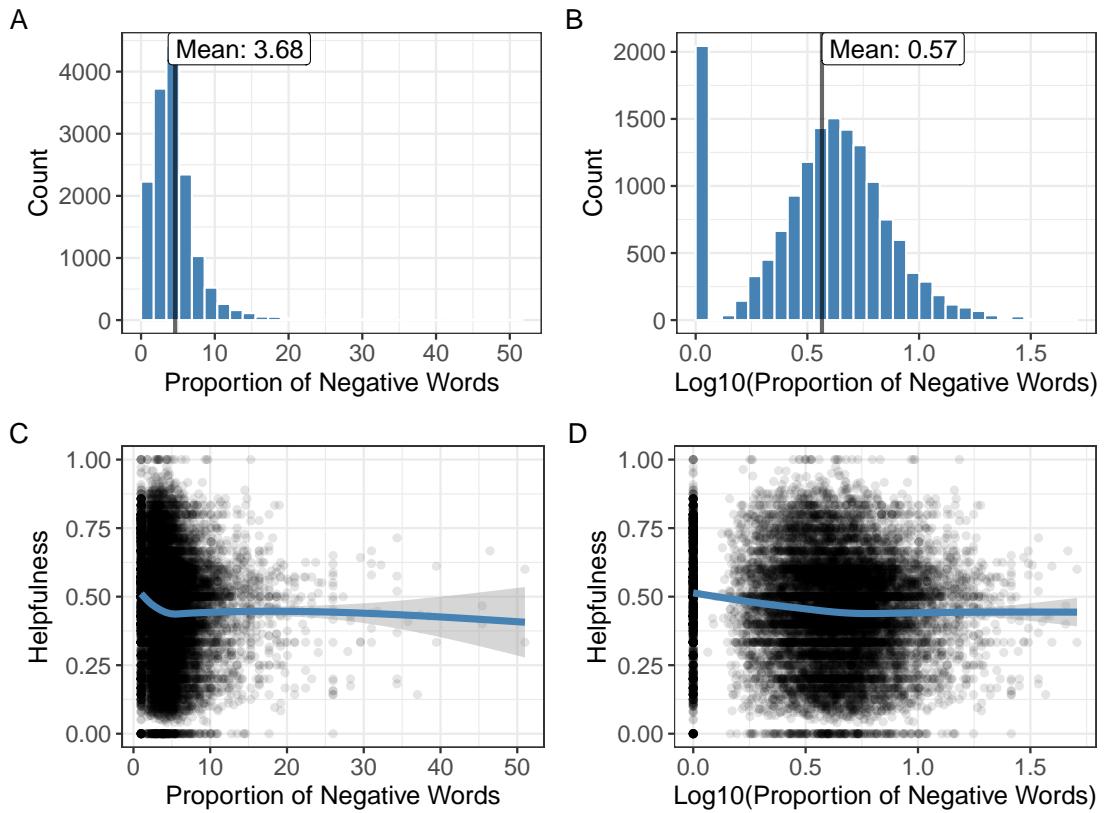


Figure 3.10: The Distribution(A) and the Scatter Plot(C) of Proportion of Negative Words for Experience Goods and their respective logarithmic transformations (B), (D)

3.2.2.5 Proportion of Positive Words

$$\text{Positive Word\%} = \frac{\text{Number of Matched Positive Word}}{\text{Total Words in the Review}}$$

This variable was engineered in the same manner with its negative words counterpart. In this case, however, reviews of both good types have more positive words in them proportionally when compared to negative words. 0.14 percent of the reviews of search goods do not have any positive words in them and this number is only 0.08 for experience goods. Mean proportion values are 4.4% ($10^{0.64}$) and 5.2% ($10^{0.72}$) for search and experience goods respectively. Except for zero values both distributions are looking quite normal (see Figure 3.11 (B) and Figure 3.12 (B)).

A sample of positive words from EmoLex:

“covenant, jovial, score, foundation, affirmation, reliable, mending, garden, luxurious, wit”

On scatter plots (see Figure 3.11 (D) and Figure 3.12 (D)), relationships between perceived helpfulness and proportion of positive words are looking very similar to their negative counterparts. Furthermore, these plots are kind of similar to perceived helpfulness versus number of words plots (see Figure 3.5 (B) and Figure 3.6 (B)). In a sense, both negative and positive proportions imply the number of words. As the proportion increases so does the number of words. That is why aforementioned figures show similar relationships between variables. On higher dimensions these relationships might be different. That is why Exploratory Data Analysis should never be conducted in a confirmatory sense.

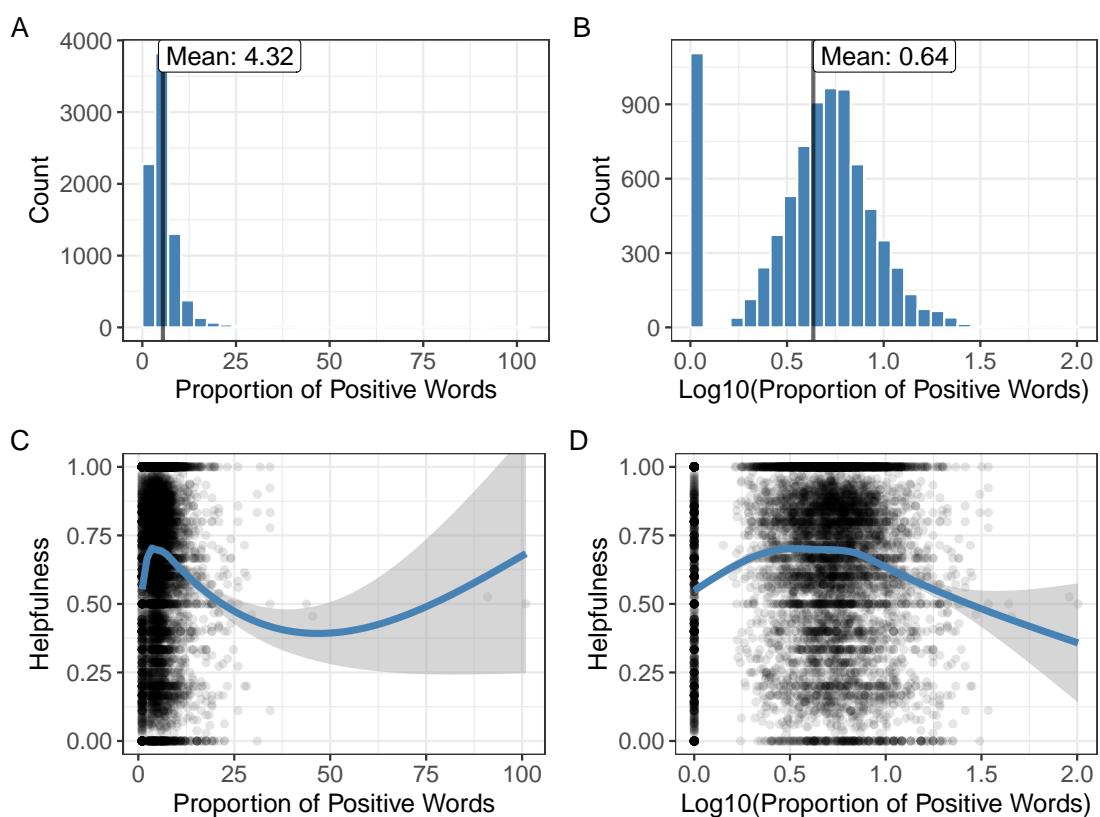


Figure 3.11: The Distribution(A) and the Scatter Plot(C) of Proportion of Positive Words for Search Goods and their respective logarithmic transformations(B), (D)

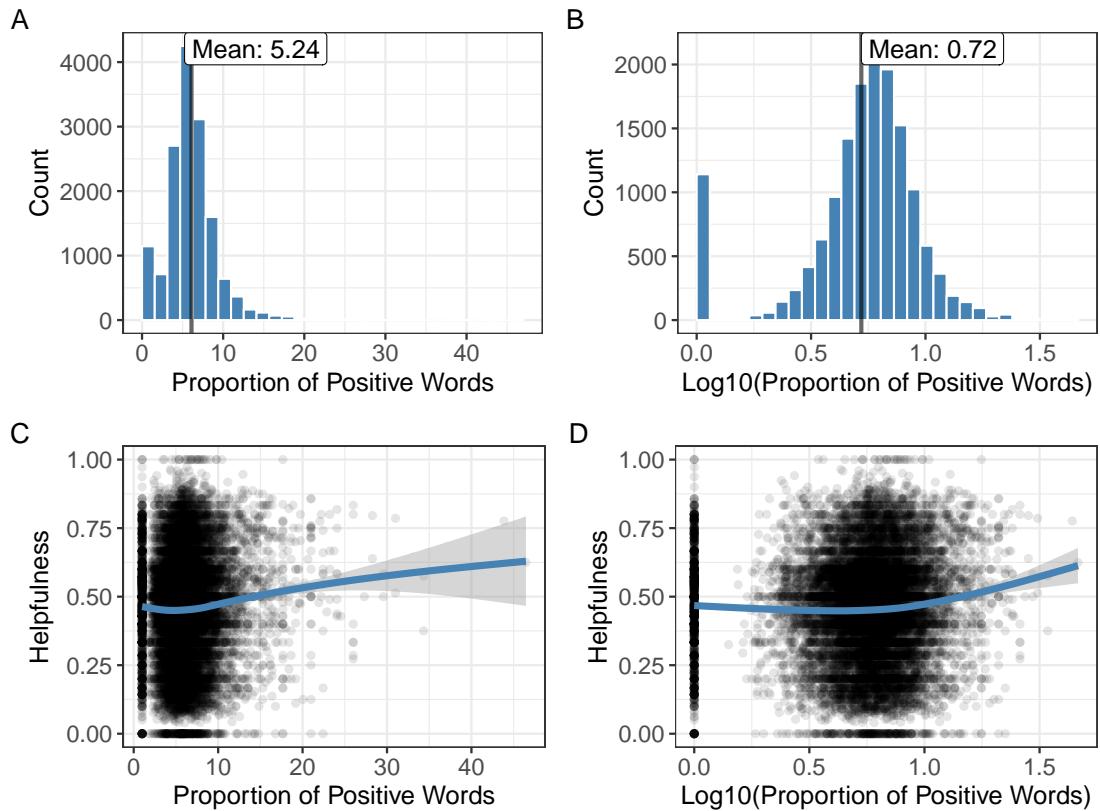


Figure 3.12: The Distribution(A) and the Scatter Plot(C) of Proportion of Positive Words for Experience Goods and their respective logarithmic transformations (B),(D)

Readers may be curious about the existence of negation words such as “not”, “no”, etc.. Additionally, reviewers might also be sarcastic when writing their reviews and they might say thing like “This product was so good that I could not believe it!”. Silge & Robinson (2016) discusses that both negation and sarcasm were used in a lesser degree in the English language. Therefore, given the enough sample size non-negated words and non-sarcastic use of the language get dominated by the normal use of the words.

3.2.3 Control Variables

3.2.3.1 Total Votes

The number of Total Votes is going to be used as a control variable since 80% Perceived Helpfulness can arise from both 4 Helpful votes over 5 Total votes or 80 Helpful votes

over 100 Total votes (Choi & Leon, 2020; Liu et al., 2014; Mudambi & Schuff, 2010).

From Figures 3.13 (B) and 3.14 (B) it can be observed that logarithmic transformation scales the variable in a better way for both good types. This is also good for linear models. Distributions for both good types are looking similar. There is a little difference though. Movies got voted many more times than search goods. This makes sense because relatively older movies are present in the data. In the case of electronics, it is hard to find reviews for older devices because they are not sold anymore so it is impossible to vote on them. This explains why there are so few amount of highly voted reviews for search goods.

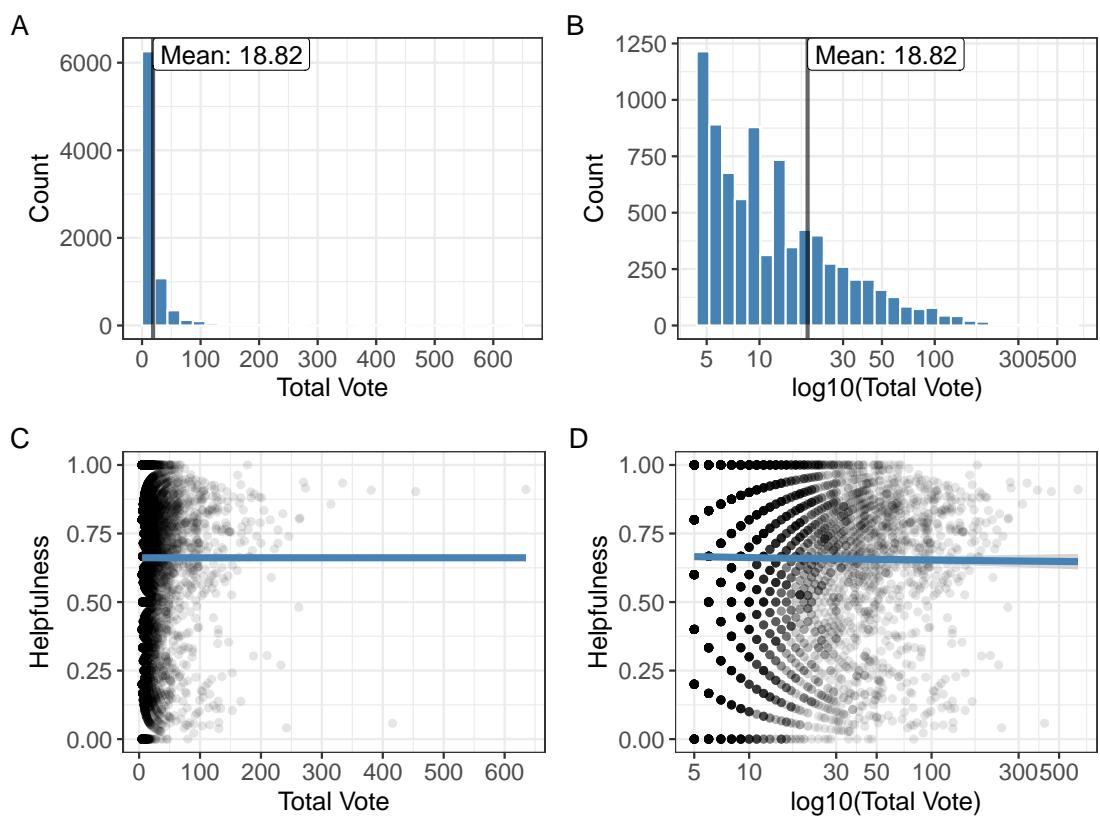


Figure 3.13: The Distribution(A) and the Scatter Plot(C) of Total Votes for Search Goods and respective logarithmic transformations (B), (D)

On Figures 3.13 (C) and (D) the smoother line may look like a straight horizontal line because the plot is small but there exist a weak negative slope. This might indicate two things. There is no association between these two variables or there is a negative relation-

ship.

Figure 3.14 (C) and (D) show a different story. Figures might indicate that there is a positive relationship between these two variables.

Nonetheless, it should be noted that the variable Total Votes is the denominator part of the dependent variable Perceived Helpfulness so any interpretations to these plots should be done with care.

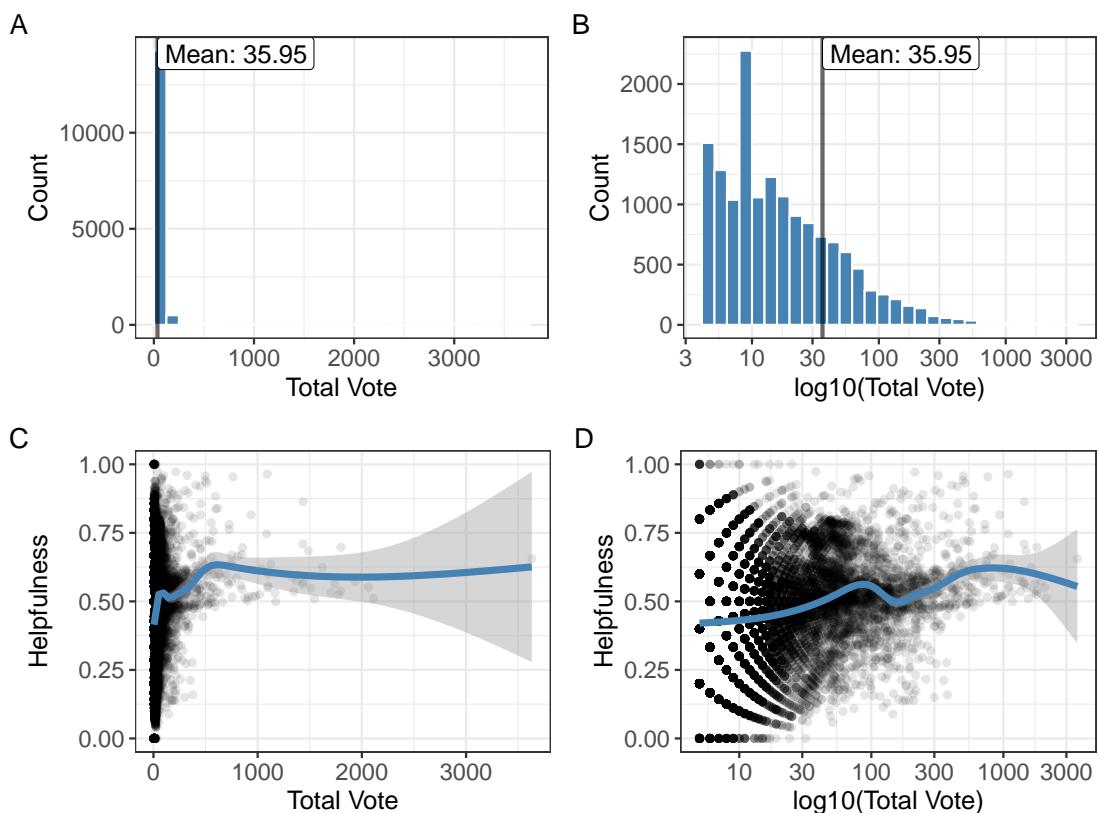


Figure 3.14: The Distribution(A) and the Scatter Plot(C) of Total Votes for Experience Goods and respective logarithmic transformations (B), (D)

3.2.4 Correlation Table

Before progressing to the modelling section, it would be a good practice to check the correlations between independent and control variables in order not to cause any multicollinearity problems. As it can be seen on Figure 3.15 there are not any major correlations between independent and control variables for both good types.

The relatively high positive / negative correlation between the Review Rating and Positive / Negative word proportions might give a hint about the usefulness of the lexicon. As the Review Rating gets higher the review contains more positive words and, on the other hand, low rated reviews have more negative words in them. This makes sense intuitively.

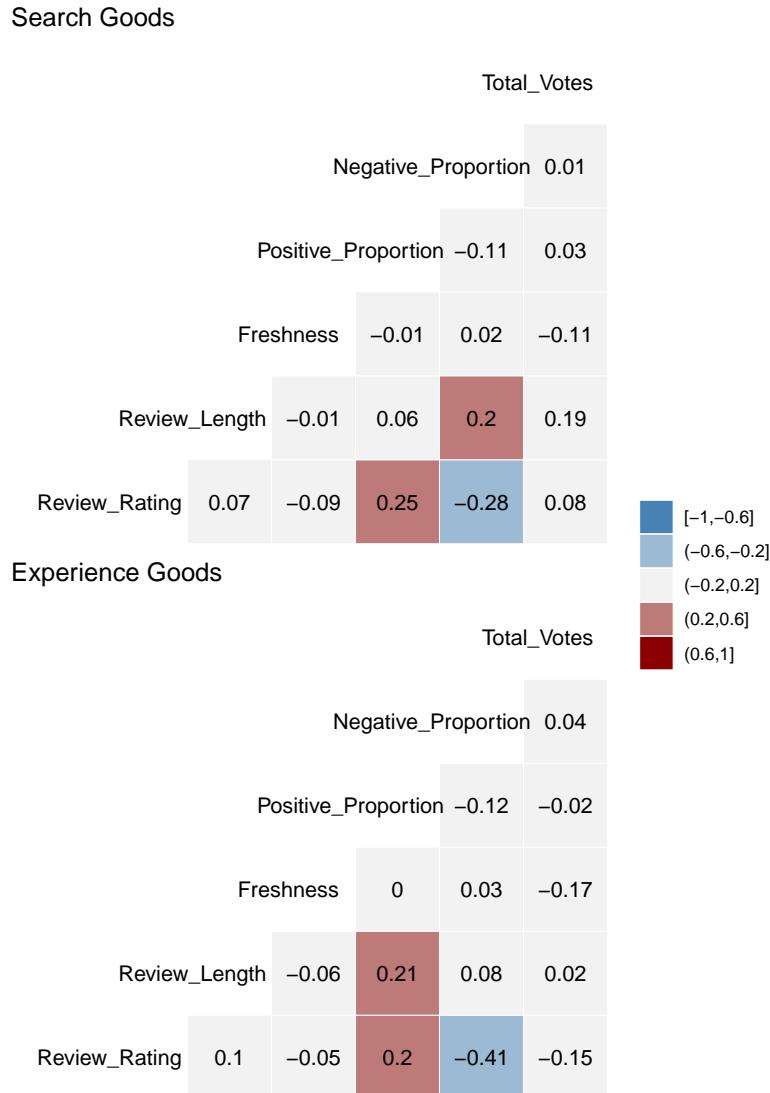


Figure 3.15: Correlations between Independent Variables.

3.3 HYPOTHESIS DEVELOPMENT

Korfiatis et al. (2012) found that neutral rated reviews (3 out of 5 stars) imply indifference to readers. This means that prior information of the potential buyer can not be influenced by these neutral reviews and likely to be found less helpful compared to extreme reviews (1 or 5 stars) (Cao et al., 2011; Ghose & Ipeirotis, 2006; Liu et al., 2014; Wang et al., 2018). As the reviews get more extremely rated they might strengthen or sway prior information that a potential consumer has, making them more helpful. This extremeness has more objective nature for Search Goods since there are fewer ways to express pleasantness or unpleasantness for a particular product. For instance a bad design in a laptop might cause fans to make much more noise than they should normally do. This information along with a very low rating may signal that there are some inherent problems with the particular product, making the feedback more helpful. On the other hand, for Experience Goods and in the specific case of movies, opinions for a given product might be much more subjective. For example, a reviewer might find a movie “childish” and give it one star, another reviewer might find the movie “light-hearted and colorful” and give the movie 5 stars, making the helpfulness subjective and decreasing its usefulness. Furthermore, for experience goods, it was found that opposite extremes to a potential buyer’s prior expectations are not impactful on the decision making process (Hoch & Deighton, 1989). In other words, prior expectations play a bigger role in decision making process for experience goods. Therefore it is hypothesized that:

H1: *There is a U-Shaped relationship between Perceived Helpfulness and Review Rating. This U-Shape is stronger/weaker for Search/Experience Goods.*

In order to decrease the uncertainty in the most optimal way, search costs must be reduced. In Section 2.1 it was mentioned that the action of search was costlier for Experience Goods than Search Goods. Furthermore, people tend to have similar opinions about Search Goods (Hong & Pavlou, 2014). This might mean that information cues are much more uniform for Search Goods (Baek et al., 2012; Huang et al., 2015). In other words, there are limited ways to construct a review because search attributes are more objective. Therefore, there might be a positive linear relationship between Perceived Helpful-

ness and Review Length because longer reviews can help the consumers in their decision making process by increasing the amount of information they have. However, since the attributes of electronics are objectively identifiable and limited (e.g., storage space of a phone, RAM of a laptop, color depth of a TV etc.) there are not much to review for a reviewer. Because of this, there might be diminishing marginal returns to Review Length. After some point more information can cause confusion and may render the review unfocused (Huang et al., 2015; Schindler & Bickart, 2012). On the contrary, for Experience Goods, people have tendency to form heterogeneous opinions (Hong & Pavlou, 2014). This means that movie reviews are much more subjective by nature. Since there are less objectivity, a helpful review to one person might not be helpful to another. This might decrease the average responsiveness of the Perceived Helpfulness to the Review Length. However, since there are other ways to objectively evaluate a movie such as the quality of its effects, movie's budget and how well this budget was utilized, a weaker positive linear relationship is expected. Therefore it is hypothesized that:

H2: There is a positive relationship between the logarithm of the Review Length and the Perceived Helpfulness. This effect is stronger for Search Goods than Experience Goods.

As it was mentioned in Section 2.2, people tend to try to minimize their losses. Because of this, negative cues draw more attention than positive ones since negative cues may imply potential losses. Guo et al. (2019) found out that people spent more time when evaluating negative reviews. This may suggest a linear interaction between Review Extremity and Review Length. Since people take more time to evaluate negative reviews, this may indicate that the effect of Review Length is stronger when Review Rating is low. And when a review is longer it is perceived as more helpful for negatively rated reviews than positively rated reviews. In this case, it is hypothesized that:

H3: There is a negative linear interaction effect between Review Rating and the logarithm of the Review Length for both Search and Experience Goods.

Potential buyers may find more recent reviews helpful because they may want to get updated information on a product. This might especially be true for products which are on the market for more than a year. For instance, a laptop which was produced in 2018 could be a good choice for that year but for year 2021 it is probable that its features

and specifications would be below average. A similar behavior might be possible for movies as well. However, this time reviewers might need to take their time to digest and evaluate the movie. A review which is carefully constructed may be posted late, making it more recent compared to a review which was hastily written. Furthermore, some popular movies which are considered as classics may attract reviews even way later than they were first released. Since some websites may employ algorithms that may push newer reviews to the top in order to be able to increase the interaction on their platforms, these kind of reviews may attract more helpful votes. Thus, it is hypothesized that:

H4: *There is a positive relationship between Perceived Helpfulness and Freshness of a Review for both Search and Experience Goods.*

So far it should be understood that people try to cut down their losses. That is why negative cues draw more attention than positive ones (Kahneman & Tversky, 1979; Schindler & Pfattheicher, 2017). Additionally, people find negative information more credible and persuasive than positive information (Ito et al., 1998; Kanouse, 1984). It is known that Review Rating is the satisfaction level of a customer towards a product. However, Review Rating itself is not enough to convey the information about how a customer expresses her satisfaction / dissatisfaction towards a product. The way a customer expresses her opinion about a product might have an effect on the perceived helpfulness of a review. Reviews with negative sentiments may bring out the risks associated with the purchase better than reviews with positive sentiments. Since in this study movies and electronics are compared, it is costlier to make a wrong decision on Search Goods than Experience Goods. Therefore, it is hypothesized that:

H5: *There is a positive relationship between the Proportion of Negative Words and the Perceived Helpfulness. This effect is stronger / weaker for Search/Experience goods.*

Kuan et al. (2015) discussed that since the majority of the reviews are positive these positive reviews lose their diagnostic values for a given product. Compared to negative information which may lay out potential risks for the product in a much more objective sense, positive information might be found more subjective and lay out a more ambiguous information about the product which may confuse potential consumers. Therefore, it is hypothesized that:

H6: *There is a negative relationship between the Proportion of Positive Words and the Perceived Helpfulness for both product types.*

3.4 RESEARCH METHODOLOGY

Tobit model is also named as Censored Regression model. Censored data arises when there exist lower (censored from below) or upper (censored from above) or both lower and upper boundaries for the dependent variable. This type of data is fairly prevalent in economics. For instance, in consumer economics there is no such thing as “negative consumption” or in labor economics there is no such thing as “negative work hours”. Furthermore, there might be cases where some really high values could be recorded as “50,000 or more” because of the data collection design. Ordinary Least Squares can not take account of the censored observations in the data because it extrapolates lower and upper ends when it fits the model and this leads biased estimations.

Tobit regression approaches censored dependent variable problem from two angles. In a sense Tobit regression could be considered as the summation of Probit regression and Truncated regression. The probit side considers if a particular observation is above or below the given limit. If the observation, for example, is below the lower limit then it must be equal to the lower limit. If it is higher than the lower limit the problem turns to a normal regression. In this case, Perceived Helpfulness can be expressed such as this:

$$y = \begin{cases} 0 & \text{if } y^* \leq 0 \\ y^* & \text{if } 0 < y^* < 1 \\ 1 & \text{if } y^* \geq 1 \end{cases}$$

Tobit Regression is selected as the analysis method because of three reasons. First of all, the dependent variable, perceived helpfulness, is defined in a continuous range from 0 to 1 since perceived helpfulness is a percentage as mentioned above. The second reason as stated by Mudambi & Schuff (2010) is the potential selection bias in the sample. Unfor-

tunately, it is not possible to know how many people read a particular review. Neither BestBuy nor IMDB disclose this information. Because of this, being included in a sample could be correlated with independent variables. For instance, potentially more voted reviews could attract more readers or readers could be drawn to reviews which have certain ratings. Lastly, there are many studies from the literature that used Tobit Regression as their research methods (Choi & Leon, 2020; Huang et al., 2015; Korfiatis et al., 2012; Mudambi & Schuff, 2010; Ren & Hong, 2019).

CHAPTER 4

RESULTS

Tobit regression results are presented at Table 4.1 alongside with OLS results as a benchmark at Table 4.2 (Choi & Leon, 2020; Kuan et al., 2015; Mudambi & Schuff, 2010). Both Tobit regression results and OLS results have the same signs and significances which might indicate that the results are somewhat robust. For Search goods, the magnitudes of the coefficients are usually higher . This is quite expected since Search good data has more censored observations compared to Experience goods data.

Hypothesis 1 proposed a U-Shape for the relationship between Review Extremity and Perceived Helpfulness. For Search Goods this is true because the linear Rating Extremity coefficient (-0.061, $p < 0.01$) has a negative sign and the quadratic Rating Extremity coefficient (0.036, $p < 0.001$) has a positive sign. The presence of the significant linear interaction effect between Review Extremity and Review Length accentuates the linear Rating Extremity coefficient (-0.043, $p < 0.001$). This means that as reviews get longer negative extreme reviews become more useful. For Experience Goods, on the other hand, both linear (0.026, $p < 0.001$) and quadratic (0.006, $p < 0.001$) Review Extremity terms are positive. A U-Shaped function can not be observed. However, the significant interaction effect between Review Length and Review Extremity (-0.015, $p < 0.001$) suggests that for longer reviews the linear term turns to negative. This means that the function takes a U-Shape after a certain review length. It can also be observed that Search Goods have stronger coefficients than Experience Goods meaning the expected U-Shape is more prominent for Search Goods. Therefore, it can be said that for shorter reviews Hypothesis 1 is partially supported and for longer reviews it is fully supported.

A positive linear relationship between the logarithm of Review Length and Perceived Helpfulness was proposed with Hypothesis 2. A strong relationship between Review Length and Perceived Helpfulness is observed (0.404, $p < 0.001$) for Search Goods. This effect is moderated by Review Extremity since the interaction effect is significant with a negative sign (-0.043, $p < 0.001$). As review rating increases the responsiveness of the Perceived Helpfulness to Review Length decreases. This means that for positive extreme reviews relatively shorter texts are more helpful. Logarithmic form also proves that an additional word in a review shows diminishing returns to helpfulness. Experience Goods somewhat deviates from the results of Search Goods. There is a weak positive relationship between Review Length and Perceived Helpfulness (0.021, $p < 0.01$). However, this positive relationship quickly turns to negative for higher ratings because of the presence of the interaction term between Review Length and Review Extremity (-0.015, $p < 0.001$). For lower ratings Hypothesis 2 is fully supported but as the Review Extremity increases this support weakens.

For both good types the interaction effects are significantly negative (-0.043, $p < 0.001$; -0.015, $p < 0.001$). This means that as Review Rating increases longer reviews become less helpful, supporting the idea of people spending more time evaluating negative information because it takes more time to read a longer review. However, there are nuances too. For Search Goods, the effect of the Review Length is always positive whether or not Review Rating is low or high. On the contrary, for Experience Goods, after some point of Review Rating longer reviews become “unhelpful” because the sign of the Review Length turns to negative. In this case Hypothesis 3 is supported wholly for Search Goods and somewhat supported for Experience Goods.

Hypothesis 4 is fully supported. Recent reviews seem to be more helpful for both good types (0.029, $p < 0.001$; 0.038, $p < 0.001$).

The results for Negative Word Proportion is quite interesting because it shows different characteristics for different product types. As expected, for Search Goods, Negative Word Proportion has a significant positive coefficient (0.061, $p < 0.001$). The presence of negative words indeed increases the perceived helpfulness. Contrary to this, there is a negative relationship between negative words and perceived helpfulness for Experience Goods (-

0.021, $p < 0.001$). Hypothesis 5 is only supported for Search Goods and not supported for Experience Goods.

Positive Word Proportion is only significant for Search Goods (-0.056, $p < 0.001$). The presence of positive words has a negative effect on perceived helpfulness just as expected. The coefficient for Experience Goods show the expected sign too but it is statistically insignificant (-0.007, $p > 0.5$). With this result, Hypothesis 6 is again only supported for Search Goods and not supported for Experience Goods.

The result of the only control variable should also be reported. It seems that total amount of votes that a review gets has a negative relationship (-0.157, $p < 0.001$) with perceived helpfulness for Search Goods. The opposite is true for Experience Goods (0.123, $p < 0.001$).

Table 4.1: Tobit Regression Results

	<i>Dependent variable:</i>	
	Helpfulness	
	Search	Experience
Rating	-0.061** (0.019)	0.026*** (0.006)
Rating Squared	0.036*** (0.003)	0.006*** (0.001)
log10(Review Length)	0.404*** (0.023)	0.021** (0.007)
log10(Total Vote)	-0.157*** (0.012)	0.123*** (0.004)
log10(Freshness + 1)	0.029*** (0.006)	0.038*** (0.002)
Positive Proportion	-0.056*** (0.013)	-0.007 (0.006)
Negative Proportion	0.061*** (0.014)	-0.021*** (0.006)
Rating x Review Length	-0.043*** (0.006)	-0.015*** (0.002)
Constant	0.059 (0.047)	0.152*** (0.017)
Observations	8,042	15,024
Log Likelihood	-4,294.455	3,492.175
Wald Test (df = 8)	1,852.288***	2,240.081***

Note:

*p<0.05; **p<0.01; ***p<0.001

Table 4.2: OLS Regression Results

	<i>Dependent variable:</i>	
	Helpfulness	
	Search	Experience
Rating	-0.030* (0.014)	0.025*** (0.006)
Rating Squared	0.026*** (0.002)	0.006*** (0.001)
log10(Review Length)	0.332*** (0.017)	0.020** (0.007)
log10(Total Vote)	-0.077*** (0.009)	0.121*** (0.004)
log10(Freshness + 1)	0.024*** (0.005)	0.038*** (0.002)
Positive Proportion	-0.041*** (0.010)	-0.007 (0.006)
Negative Proportion	0.044*** (0.010)	-0.021*** (0.006)
Rating x Review Length	-0.038*** (0.004)	-0.015*** (0.002)
Constant	0.064 (0.035)	0.159*** (0.017)
Observations	8,042	15,024
R ²	0.206	0.132
Adjusted R ²	0.205	0.131
Residual Std. Error	0.270 (df = 8033)	0.181 (df = 15015)
F Statistic	260.229*** (df = 8; 8033)	284.367*** (df = 8; 15015)

Note:

*p<0.05; **p<0.01; ***p<0.001

CHAPTER 5

DISCUSSION

The results show a clear contrast between the objective nature of search goods versus the subjective nature of experience goods which once again validates the notion of Search and Experience goods classification that was developed by Nelson (1970). It should not be ignored that this contrast is only relevant for the case of electronics versus movies since the available data only included these products. Electronics such as laptops or phones carry very strong search attributes whereas movies can be classified as “pure experience” goods which can be considered as a special case (Liu et al., 2014).

Negativity Bias and Prospect Theory had a strong support in the case of search goods. The regression results showed that the reviews which had low ratings and/or reviews which had more negative words in them were found to be more helpful on the average. Additionally, Perceived Helpfulness has higher responsiveness to Review Length when a review has a low rating. This means that people focus more on negative cues than positive cues because the marginal value of an additional negative information is higher. This way consumers are able to cut down potential sure losses. A similar result was reflected by the negative coefficient of the proportion of positive words in a review. This shows that reviews with positive sentiments do not offer diagnostic value for consumers. Nonetheless, the U-Shape of the Review Rating variable also suggests positivity bias (Hoorens, 2014). An explanation to this could be that certain brands or products might be favored by some customers and these customers might selectively tend towards positively rated reviews and find them helpful just because they love the brand (Zhu et al., 2021). The reason why there is a divergence between the “positively rated review” and “a review with positive

sentiment” might be explained by this. A consumer who has brand or product inclinations wants to be persuaded to have the product so this particular consumer does not really care about what is written in the review, he might skim through positive reviews until he feels persuaded. Since a preference towards a brand is subjective it might be speculated that a strong brand image could embed experience attributes into search goods. A similar notion was discussed by Mudambi & Schuff (2010) where they classified Apple iPod as an experience good. That is although an MP3 player such as iPod has many objective search attributes such as storage capacity and battery life, what makes this particular product preferable most of the time is the product image that was created by Apple.

The rationality why the statistical results for search goods came out just as expected can be accredited to their objective features. Properties of electronics are easily evaluated prior purchase. This means that it is comparatively easier to collect information on search goods than experience goods. Other than online consumer reviews on a retailer’s website, a potential buyer has an arsenal of product reviews which are available on tech enthusiastic websites where electronics are reviewed by professionals. Furthermore, some channels on YouTube offer product reviews where the reviewer inspects the product in question on a camera. Thus, the potential buyer can gather a great amount of information even before accessing to the online retailer’s website.

A similar strong support for Negativity Bias can not be observed in the case of movies. Only evidence for Negativity Bias was the significant linear interaction term between Review Length and Review Rating. According to this result the U-Shaped function of Review Rating is only observable when Review Length is high enough. A similar situation is the case for search goods as well but the interaction term does not interfere with the signs of the other coefficients so no matter the Review Length is a U-Shaped function is observed for search goods. However, this is certain that as the review rating gets higher the review loses its diagnostic value for both for experience and search goods because the coefficient of the Review Length weakens. This means that potential buyers require more elaboration when a review is rated low. For experience goods, high rated reviews become more helpful on the average when reviews are short. This also might be a sign for positivity bias where people pay more attention to reviews which were aligned with

their own feelings (Zhu et al., 2021). Since experience attributes are naturally subjective it is logical to think that readers look for reviews that they agree.

A number of reasons could be speculated about why the Hypothesis 5 and Hypothesis 6 did not hold for experience goods. The expectation was observing a negative significant coefficient for proportion of positive words and a positive significant coefficient for proportion of negative words. At the very least it can be said that the existence of positive words in a movie review does not offer a significant information value according to regression results. Thus, it is hard to say there is a strong deviation from Hypothesis 6. On the contrast, Hypothesis 5 for experience goods was not supported at all.

One reason of this might be that the NRC Word-Emotion Association Lexicon which was used to mine the negative words in the reviews was not suitable for the movie reviews. For instance, the word “*cage*” is a negative word in the lexicon and “*marvel*” is a positive word. One or more movies in the sample include movies where “Nicholas *Cage*” had the starring role. There are also “*Marvel Cinematic Universe*” movies present in the data. Cases like these may create noise for the variable and the model because aforementioned words were used frequently and outside of the context of the lexicon. To test this problem, negative and positive words such as “*cage, john, climax, force, king, mad, fury, hardy, margin, harry, marvel, stark, iron, vulture, green, goblin, bale*” were filtered out of the lexicon for both good types. Neither coefficient signs nor coefficient significances were changed. However, it should be pointed out that the magnitude of the coefficient of negative words for experience goods was weakened slightly. Further filtering was not pursued because of potential dangers of hypothesis fishing. (The results in Tables 4.1 and 4.2 are the results with the filter.)

Another reason may arise from the subjective nature of the experience attributes of the movies. Negative words, in the context of electronics, may have a narrower range of meanings because what they usually try to convey directly relates to quality of the good itself. The quality of the good is easily accessible because of the fact that electronics have dominant search attributes. Thus, negative words probably translates as potential monetary risks rather than affecting emotions. A risk might be exemplified as purchasing a sub-par performing laptop with a given budget, in other words getting sub-optimal

utility. Furthermore, electronics are usually bought to be used for a comparatively long period of time. A smartphone or a television can serve its owner for years so negative cues which relate to below average durability are very important. Buying a sub-durable product for a given budget can only increase the chances that the consumer will have to buy the same type of product prematurely in the future. On the other hand, the monetary risks associated with a bad movie are relatively lower. Consumers can go to cinema, buy a DVD, rent the movie on YouTube or access the movie as a part of their subscription plan on platforms such as Netflix. Subscription or ticket fees are not the only cost for the consumer. In order to consume the product the consumer must spend the time that is equal to the duration of the movie. Of course, the consumer can stop watching the movie anytime if it is too unbearable for her so she does not waste any more time. Because of these reasons the money spent on a movie might be irrelevant to an average consumer since accessing a movie is relatively cheaper than owning an electronic device. Therefore, negative words in a movie review may not alert the consumer for potential monetary losses. Instead, negative words could render a review in question unpleasant because emotions are priority in the case of consumption of movies.

There are still points to be made about the results of the movies. In the Section 2.1 it was mentioned that consumers choose from two different strategies when they want to acquire information about a particular product. They might choose to experience the good or they can undertake search in order to collect information. It was pointed out by Nelson (1970) that word-of-mouth type of information is viable when it is much more costlier to “experience” the good than undertaking a search process. It is probable that, in the case of movies, watching the movie (experiencing it) isn’t that much costlier or maybe less costlier than gathering information about that particular movie and then making a decision on watching it. Thus, “search” might not be very justifiable for movies. This might explain why electronic word-of-mouth (Review Length) showed rather weaker and erratic (positive for lower ratings, negative for higher ratings) results.

Unexpected and different results are common in the literature of Review Helpfulness (Fan, 2021). This work and the rest of the literature suffer from certain biases that the data imposes. Many other limitations may arise from potential omitted variables.

Firstly, as Mudambi & Schuff (2010) pointed out that potential selection bias exists in the data. It is not possible to know how many people actually read a particular review. Additionally, there is a strong possibility that as a review gets more votes this is going to draw more attention to the particular review. This can cause a positive feedback loop and this feedback loop can increase the chances that an observation making into the sample which leads to a selection bias. To fight with this, researchers usually resorted to setting a threshold of total votes where reviews with small amount of votes were filtered out of the sample. Furthermore, a multi-level regression model was developed by Kuan et al. (2015) in order to overcome the selection bias. In the first level of the regression the researchers modeled the binary case of whether a review gets voted or not. After what makes a review voted was established, they progressed to the next stage of modelling the review helpfulness. Along with the multi-level model results they also shared OLS results where non-voted reviews were filtered from the data. Most of the significance and coefficient signs were not different but coefficients differed in magnitude.

Secondly, as pointed out in Section 1, there were many different research designs present in the literature. Let alone different methodologies, even the utilizations of similar variables were contrasting. For instance, review extremity was employed in three different ways by researchers. This study and some others use the review ratings as it is. Another group of studies took the differences of review ratings from the mean rating of a given product. The resulting variable is usually called Review Inconsistency. The last group of researchers utilized a dummy variable for neutral reviews (3 out of 5) and extreme reviews (1,2,4,5 out of 5). In some studies these dummy variables were extended to negative and positive extreme dummies as well. The differences are not limited to Review Rating. The way the dependent variable was used, Perceived Helpfulness, diverged too. In some circumstances only absolute Helpful votes were used along with Negative Binomial Regression. Furthermore, some studies converted the percentage of the helpfulness into a binary problem where reviews with over 50% helpfulness score was considered as a helpful review. As a last example to different applications, scale transformations (logarithmic transformation, etc.) were applied liberally as well. All in all, it is clear that there is no standard approach in the literature which might be the reason why there are many conflicting results.

Potential omitted variables such as consumer's buying motivation or his/her trust towards strangers may also explain the conflicting or unexpected results as pointed out by Zhu et al. (2021). A person who is more agreeable will likely find many more reviews helpful on the average. Additionally, buying motivation might play an important role in explaining the review helpfulness. Above, it was speculated that some consumers might have tendency towards certain brands or products. Favored brands might not be the only factor that affects the consumer motivation. For instance, a consumer with limited amount of budget may find positive reviews of products that have suitable prices helpful. Thus the product price relative to the consumer's budget can also be used as an indicator.

CHAPTER 6

CONCLUSION

Online Consumer Reviews data were collected from two different sources (IMDB and BestBuy) and analyzed within the context of consumer uncertainty in this study. Negativity Bias, Prospect Theory and Search / Experience goods paradigm from information economics were used as theoretical engines to explain the information gathering process of an average consumer. Results showed similarities from the existing Review Helpfulness literature. However, once again it should be noted that the literature has many conflicting results.

This study contributes to the literature as an additional empirical work with a unique Best-Buy review data and a unique combination of movie reviews from IMDB. In addition, to my knowledge, this is the first study which analyzed the linear continuous interaction between the Review Length and the Review Rating in terms of Negativity Bias. Additionally, there is a lack of emphasis on the sentiment analysis in the existing literature. By employing positive and negative sentiments in terms of Negativity Bias and Prospect Theory this study shows the viability of the sentiment analysis in this context.

Goods with very strong search or experience attributes were chosen in order to compare the information gathering process in the clearest way possible. The results prove the fact that information gathering is easier for search goods than experience goods. This means that consumers face less uncertainty against search goods. The results are only generalizable to laptops, phones, televisions and movies since different goods have different mix of search and experience attributes.

Findings show the importance of information sharing for an average online consumer at least for the search goods. Since helpful reviews aid consumers deal with their uncertainties by helping them acquire information concerning the quality of a given product, attendance to information sharing either by writing reviews or voting if the reviews are helpful or not is strongly advised. Search goods results at Table 4.1 can be used as a guide for writing a review. The guidelines can be listed such as this:

1. Reviewers should only write a review when their opinions of a given product is rather extreme.
2. Reviewers should share all the information and the experience they accumulated about the product without making it too long because it was observed that helpfulness shows diminishing returns to additional words. The review should be clear and to the point.
3. Reviewers should not refrain from using negative words as long as it objectively identifies something wrong or lacking with the product which is being reviewed.
4. When reviewers are reviewing a product which they are extremely unhappy about they should give more details and elaborate, on the opposite, when they are content they should write relatively shorter reviews.
5. Since recent reviews were found to be more helpful on the average, sharing updated information, i.e., writing new reviews is important and encouraged.

With this way consumers would be able to purchase the best good available in the market with respect to their budgets, i.e., maximizing their utilities. This collective behavior may even impact social welfare positively since economic agents are making much more informed decisions. Unfortunately, similar implications can not be made for experience goods, specifically for movies, since the realized results were mostly different than what was hypothesized. Thus, this study can not offer any suggestions for movie reviews.

There are some implications for the future studies. Firstly, aside from lexicon based sentiment analysis it is also possible to utilize deep learning methods to mine opinions from a text. One advantage of the deep learning method over a lexicon is its versatility. A negative word can change from one context to another and training a deep learning method is more robust to this type of contextual changes whereas with lexicon based methods a

negative word is a negative word no matter the context is. Next, future studies should consider collecting data from different regions and cultures. There might be some differences on how consumers process and react to the information in order to deal with uncertainty. Additionally, researchers can dig deeper into search and experience goods paradigm. It is well established in the literature that product type affects the information gathering process. However, there might be differences within product types depending on their search and experience attributes. An objective way to identify and index experience and search attributes may also shed some light to the conflicting results in the literature. Lastly, since Online Consumer Reviews data is ever growing more researchers should incorporate Exploratory Data Analysis into their works. Reviews and how helpful these reviews are perceived by potential buyers is relatively a new topic. In addition, new ways to analyze text data are being developed since the area of Natural Language Processing is quite hot so Online Consumer Reviews offer a great potential for new discoveries from the data for different domains. This increases the value of the Exploratory Data Analysis as well.

BIBLIOGRAPHY

- Ahmad, S. N., & Laroche, M. (2015). How Do Expressed Emotions Affect the Helpfulness of a Product Review? Evidence from Reviews Using Latent Semantic Analysis. *International Journal of Electronic Commerce*, 20(1), 76–111. <https://doi.org/10.1080/10864415.2016.1061471>
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*, 17(2), 99–126. <https://doi.org/10.2753/jec1086-4415170204>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521. <https://doi.org/10.1016/j.dss.2010.11.009>
- Chatterjee, S. (2020). Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. *International Journal of Hospitality Management*, 85, 102356. <https://doi.org/10.1016/j.ijhm.2019.102356>
- Choi, H. S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, 139, 113403. <https://doi.org/10.1016/j.dss.2020.113403>
- Chua, A. Y. K., & Banerjee, S. (2014). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, 66(2), 354–362. <https://doi.org/10.1002/asi.23180>
- Darby, M. R., & Karni, E. (1973). Free Competition and the Optimal Amount of Fraud.

- The Journal of Law and Economics*, 16(1), 67–88. <https://doi.org/10.1086/466756>
- Dellarocas, C. N. (2003). The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.393042>
- Dellarocas, C., Zhang, X. (Michael)., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45. <https://doi.org/10.1002/dir.20087>
- Duan, W., Gu, B., & Whinston, A. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233–242. <https://doi.org/10.1016/j.jretai.2008.04.005>
- Eslami, S. P., Ghasemaghaei, M., & Hassanein, K. (2018). Which online reviews do consumers find most helpful? A multi-method investigation. *Decision Support Systems*, 113, 32–42. <https://doi.org/10.1016/j.dss.2018.06.012>
- Fan, W. (2021). What makes consumer perception of online review helpfulness: Synthesizing the past to guide future research. *Proceedings of the Annual Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/hicss.2021.334>
- Ghose, A., & Ipeirotis, P. (2006). Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality. *Proceedings of the 16th Annual Workshop on Information Technology and Systems*.
- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512. <https://doi.org/10.1109/tkde.2010.188>
- Guo, D., Zhao, Y., Zhang, L., Wen, X., & Yin, C. (2019). Conformity feedback in an online review helpfulness evaluation task leads to less negative feedback-related negativity amplitudes and more positive P300 amplitudes. *Journal of Neuroscience, Psychology, and Economics*, 12(2), 73–87. <https://doi.org/10.1037/npe0000102>
- Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17(4), 454. <https://doi.org/10.1086/208570>
- Hoch, S. J., & Deighton, J. (1989). Managing What Consumers Learn from Experience.

- Journal of Marketing*, 53(2), 1–20. <https://doi.org/10.1177/002224298905300201>
- Hong, H., Xu, D., Wang, G. A., & Fan, W. (2017). Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, 102, 1–11. <https://doi.org/10.1016/j.dss.2017.06.007>
- Hong, Y. (Kevin), & Pavlou, P. A. (2014). Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents. *Information Systems Research*, 25(2), 328–344. <https://doi.org/10.1287/isre.2014.0520>
- Hoorens, V. (2014). *Positivity bias* (pp. 4938–4941). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_2219
- Hu, N., Pavlou, P. A., & Zhang, J. (2006). Can online reviews reveal a product's true quality? *Proceedings of the 7th ACM Conference on Electronic Commerce - EC '06*. <https://doi.org/10.1145/1134707.1134743>
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17–27. <https://doi.org/10.1016/j.chb.2015.01.010>
- Hunter, J. E., & Schmidt, F. L. (1991). *Meta-analysis* (pp. 157–183). Springer Netherlands. https://doi.org/10.1007/978-94-009-2195-5_6
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75(4), 887–900. <https://doi.org/10.1037/0022-3514.75.4.887>
- Jebb, A. T., Parrigon, S., & Woo, S. E. (2017). Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review*, 27(2), 265–276. <https://doi.org/10.1016/j.hrmr.2016.08.003>
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, 98(6), 1325–1348. <https://doi.org/10.1086/261737>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263. <https://doi.org/10.2307/1914185>
- Kanouse, D. E. (1984). Explaining negativity biases in evaluation and choice behavior: Theory and research. *ACR North American Advances*.

- Klein, L. R. (1998). Evaluating the Potential of Interactive Media through a New Lens: Search versus Experience Goods. *Journal of Business Research*, 41(3), 195–203. [https://doi.org/10.1016/s0148-2963\(97\)00062-3](https://doi.org/10.1016/s0148-2963(97)00062-3)
- Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3), 205–217. <https://doi.org/10.1016/j.elerap.2011.10.003>
- Kuan, K., Hui, K.-L., Prasarnphanich, P., & Lai, H.-Y. (2015). What makes a review voted? An empirical investigation of review voting in online review systems. *Journal of the Association for Information Systems*, 16(1), 48–71. <https://doi.org/10.17705/1jais.00386>
- Lee, J., Park, D.-H., & Han, I. (2011). The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls. *Internet Research*, 21(2), 187–206. <https://doi.org/10.1108/10662241111123766>
- Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 56(4), 849–866. <https://doi.org/10.1108/md-06-2017-0561>
- Lee, S., & Choeh, J. Y. (2016). The determinants of helpfulness of online reviews. *Behaviour & Information Technology*, 35(10), 853–863. <https://doi.org/10.1080/0144929x.2016.1173099>
- Lee, S., & Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, 41(6), 3041–3046. <https://doi.org/10.1016/j.eswa.2013.10.034>
- Li, H., Wang, C. (Renee)., Meng, F., & Zhang, Z. (2019). Making restaurant reviews useful and/or enjoyable? The impacts of temporal, explanatory, and sensory cues. *International Journal of Hospitality Management*, 83, 257–265. <https://doi.org/10.1016/j.ijhm.2018.11.002>
- Li, X., Wong, W., Lamoureux, E. L., & Wong, T. Y. (2012). Are Linear Regression Techniques Appropriate for Analysis When the Dependent (Outcome) Variable Is Not Normally Distributed? *Investigative Ophthalmology & Visual Science*, 53(6), 3082. <https://doi.org/10.1167/iovs.12-9967>

- Liu, Z., Hong, L., & Liu, L. (2014). An investigation of online review helpfulness based on movie reviews. *African Journal of Business Management*, 8(12), 441–450. <https://doi.org/10.5897/ajbm11.2628>
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
- Mudambi, & Schuff. (2010). Research note: What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly*, 34(1), 185. <https://doi.org/10.2307/20721420>
- Murray, K. B. (1991). A test of services marketing theory: Consumer information acquisition activities. *Journal of Marketing*, 55(1), 10–25. <http://www.jstor.org/stable/1252200>
- Nelson, P. (1974a). Advertising as Information. *Journal of Political Economy*, 82(4), 729–754. <https://doi.org/10.1086/260231>
- Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311–329. <http://www.jstor.org/stable/1830691>
- Nelson, P. (1974b). The economic value of advertising. *Advertising and Society*, 109–141.
- Nelson, P. J. (1981). *Consumer information and advertising* (pp. 42–77). Springer Netherlands. https://doi.org/10.1007/978-94-009-8168-3_5
- Öğüt, H., & Onur Taş, B. K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industries Journal*, 32(2), 197–214. <https://doi.org/10.1080/02642069.2010.529436>
- Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612. <https://doi.org/10.1016/j.jretai.2011.05.002>
- Peeters, G. (1971). The positive-negative asymmetry: On cognitive consistency and positivity bias. *European Journal of Social Psychology*, 1(4), 455–474. <https://doi.org/10.1002/ejsp.2420010405>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ren, G., & Hong, T. (2019). Examining the relationship between specific negative emo-

- tions and the perceived helpfulness of online reviews. *Information Processing & Management*, 56(4), 1425–1438. <https://doi.org/10.1016/j.ipm.2018.04.003>
- Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, 5(4), 296–320. https://doi.org/10.1207/s15327957pspr0504_2
- Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, 11(3), 234–243. <https://doi.org/10.1002/cb.1372>
- Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, 69, 172–177. <https://doi.org/10.1016/j.jesp.2016.09.009>
- Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing*, 21(4), 76–94. <https://doi.org/10.1002/dir.20090>
- Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in r. *JOSS*, 1(3). <https://doi.org/10.21105/joss.00037>
- Sun, X., Han, M., & Feng, J. (2019). Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. *Decision Support Systems*, 124, 113099. <https://doi.org/10.1016/j.dss.2019.113099>
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1), 67–85. <https://doi.org/10.1037/0033-2909.110.1.67>
- Topaloglu, O., & Dass, M. (2019). The Impact of Online Review Content and Linguistic Style Matching on New Product Sales: The Moderating Role of Review Helpfulness. *Decision Sciences*, 52(3), 749–775. <https://doi.org/10.1111/deci.12378>
- Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), 1039–1061. <https://doi.org/10.2307/2937956>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
- Wang, Y., Wang, J., & Yao, T. (2018). What makes a helpful online review? A meta-

- analysis of review characteristics. *Electronic Commerce Research*, 19(2), 257–284. <https://doi.org/10.1007/s10660-018-9310-2>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Gromelund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wu, C., Mai, F., & Li, X. (2021). The effect of content depth and deviation on online review helpfulness: Evidence from double-hurdle model. *Information & Management*, 58(2), 103408. <https://doi.org/10.1016/j.im.2020.103408>
- Yang, S., Zhou, Y., Yao, J., Chen, Y., & Wei, J. (2019). Understanding online review helpfulness in omnichannel retailing. *Industrial Management & Data Systems*, 119(8), 1565–1580. <https://doi.org/10.1108/imds-10-2018-0450>
- Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674–7682. <https://doi.org/10.1016/j.eswa.2010.12.147>
- Zhou, Y., & Yang, S. (2019). Roles of review numerical and textual characteristics on review helpfulness across three different types of reviews. *IEEE Access*, 7, 27769–27780. <https://doi.org/10.1109/access.2019.2901472>
- Zhu, Z., Liu, J., & Dong, W. (2021). Factors correlated with the perceived usefulness of online reviews for consumers: a meta-analysis of the moderating effects of product type. *Aslib Journal of Information Management*, 74(2), 265–288. <https://doi.org/10.1108/ajim-02-2021-0054>