NTNU – Trondheim
Norwegian University of
Science and Technology

Online Consumer Reviews

The Moderating Effect of Product Category

Einar Bjering
Lars Jaakko Havro

Industrial Economics and Technology Management
Submission date: June 2014
Supervisor: Øystein Moen, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

THESIS CONTRACTS

NTNU
Det skapende universitet

MASTERKONTRAKT
- uttak av masteroppgave

1. Studentens personalia

Etternavn, fornavn **Bjering, Einar**	Fødselsdato **11. aug 1988**
E-post **ebjering@gmail.com**	Telefon **45454350**

2. Studieopplysninger

Fakultet **Fakultet for samfunnsvitenskap og teknologiledelse**	
Institutt **Institutt for industriell økonomi og teknologiledelse**	
Studieprogram **Industriell økonomi og teknologiledelse**	Hovedprofil **Strategi og internasjonal forretningsutvikling**

3. Masteroppgave

Oppstartsdato **15. jan 2014**	Innleveringsfrist **11. jun 2014**
Oppgavens (foreløpige) tittel **Online Consumer Reviews** **The Moderating Effect of Product Category**	
Oppgavetekst/Problembeskrivelse A study on the moderating effect of product category on the relative impact of online consumer reviews on sales. The study includes an unprecedented number of product categories and reviews and also considers other aspects such as review helpfulness and certain review biases.	
Hovedveileder ved institutt **Professor Øystein Moen**	Medveileder(e) ved institutt
Merknader **1 uke ekstra p.g.a påske.**	

4. Underskrift

Student: Jeg erklærer herved at jeg har satt meg inn i gjeldende bestemmelser for mastergradsstudiet og at jeg oppfyller kravene for adgang til å påbegynne oppgaven, herunder eventuelle praksiskrav.

Partene er gjort kjent med avtalens vilkår, samt kapitlene i studiehåndboken om generelle regler og aktuell studieplan for masterstudiet.

Trondheim 22.05.2014

Sted og dato

Einar Bjøring

Student

Hovedveileder

Originalen lagres i NTNUs elektroniske arkiv. Kopi av avtalen sendes til instituttet og studenten.

NTNU
Det skapende universitet

MASTERKONTRAKT

- uttak av masteroppgave

1. Studentens personalia

Etternavn, fornavn **Havro, Lars Jaakko**	Fødselsdato **02. jul 1987**
E-post **havro@stud.ntnu.no**	Telefon **98487326**

2. Studieopplysninger

Fakultet **Fakultet for samfunnsvitenskap og teknologiledelse**	
Institutt **Institutt for industriell økonomi og teknologiledelse**	
Studieprogram **Industriell økonomi og teknologiledelse**	Hovedprofil **Strategi og internasjonal forretningsutvikling**

3. Masteroppgave

Oppstartsdato **15. jan 2014**	Innleveringsfrist **11. jun 2014**
Oppgavens (foreløpige) tittel **Online Consumer Reviews** **The Moderating Effect of Product Category**	
Oppgavetekst/Problembeskrivelse A study on the moderating effect of product category on the relative impact of online consumer reviews on sales. The study includes an unprecedented number of product categories and reviews and also considers other aspects such as review helpfulness and certain review biases.	
Hovedveileder ved institutt **Professor Øystein Moen**	Medveileder(e) ved institutt
Merknader **1 uke ekstra p.g.a påske.**	

4. Underskrift

Student: Jeg erklærer herved at jeg har satt meg inn i gjeldende bestemmelser for mastergradsstudiet og at jeg oppfyller kravene for adgang til å påbegynne oppgaven, herunder eventuelle praksiskrav.

Partene er gjort kjent med avtalens vilkår, samt kapitlene i studiehåndboken om generelle regler og aktuell studieplan for masterstudiet.

Trondheim 22/5.2014

Sted og dato

Student **Hovedveileder**

Originalen lagres i NTNUs elektroniske arkiv. Kopi av avtalen sendes til instituttet og studenten.

Side **2** av **2**

NTNU
Det skapende universitet

SAMARBEIDSKONTRAKT

1. Studenter i samarbeidsgruppen

Etternavn, fornavn **Bjering, Einar**	Fødselsdato **11. aug 1988**
Etternavn, fornavn **Havro, Lars Jaakko**	Fødselsdato **02. jul 1987**

2. Hovedveileder

Etternavn, fornavn **Moen, Øystein**	Institutt **Institutt for industriell økonomi og teknologiledelse**

3. Masteroppgave

Oppgavens (foreløpige) tittel **Online Consumer Reviews** **The Moderating Effect of Product Category**

4. Bedømmelse

Kandidatene skal ha *individuell* bedømmelse ☐
Kandidatene skal ha *felles* bedømmelse ☒

Trondheim 22.05.2014
Sted og dato

Einar Bjering

Hovedveileder

Lars Jaakko Havro

Originalen oppbevares på instituttet.

ABSTRACT

This paper tests a previously proposed model for assessing consumer generated online reviews effect on sales, the review impact continuum. Product category is found to play an important role as a moderating factor of several properties concerning user generated online reviews - including its impact on sales. The authors introduce a novel method for product category classification using natural language processing (NLP), and by applying this method show that reviews are more influential for subjectively evaluated products than objectively evaluated products. In addition, with a selection of hit products and random products for each category, it demonstrates that product popularity has a moderating impact on the effect of online reviews on sales, and that reviews tend to be more influential for the less popular products.

Additionally, the study mirrors previous research on the perceived helpfulness of reviews and finds that consumers seem to rate longer reviews as more helpful, and that the effect is stronger for reviews on objectively evaluated products. Information verifying the validity of the review is found to affect helpfulness positively as well.

The study is based on review data from Amazon, including an unprecedented 1.1 million unique reviews from more than 4,600 products in 30 different product categories, collected during the spring of 2014.

ABSTRACT (NO)

Denne studien tester en tidligere fremlagt modell for vurdering av digitale brukeranmeldelsers effekt på salg, ”the review impact continuum”. Produktkategori ser ut til å spille en viktig rolle som en modererende faktor på flere aspekter ved digitale brukergenererte anmeldelser - herunder innvirkningen de har på salg. Forfatterne innfører en ny metode for klassifisering av produktkategorier ved hjelp av ”natural languange processing” (NLP), og demonstrerer gjennom bruk av denne metoden at anmeldelsene er mer innflytelsesrike for subjektivt enn objektivt evaluerte produkter. I tillegg, med et utvalg av hit-produkter og tilfeldige produkter for hver kategori, viser studien at produktets popularitet har en modererende innflytelse på effekten av online vurderinger på salg, og at anmeldelser har en tendens til å være mer innflytelsesrike for mindre populære produkter.

I tillegg repliserer denne studien tidligere forskning på oppfattet hjelpsomhet av anmeldelser og finner at forbrukerne synes å rangere lengre anmeldelser som mer nyttig, og at effekten er sterkere for anmeldelser av objektivt evaluerte produkter. Informasjon som verifiserer påliteligheten av anmeldelsen blir også funnet å påvirke hjelpsomhet positivt.

Studien er basert på brukeranmeldelser fra Amazon, og inkluderer 1,1 millioner unike anmeldelser fra mer enn 4600 produkter i 30 ulike produktkategorier, samlet inn i løpet av våren 2014.

PREFACE

This master's thesis was written during the spring of 2014 at the Department of Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU).

The primary goal has been to shed light on the factor product category plays in the world of online consumer reviews. The interest in this particular topic arose during the writing of our pre-diploma thesis ”Word of Mouth in a Digital Age: The Case of Online Consumer Reviews” (Bjering and Havro, 2013). Some of the background theory is adapted from this paper.

We wish to thank our academical supervisor Øystein Moen at the Department of Industrial Economics and Technology Management, for his feedback and support during the writing of this thesis. We also wish to thank Frode Tobias Bjerke, MSc. student in computer science at the University of Oslo, for his valuable input on natural language processing.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

Part I

BACKGROUND AND METHODOLOGY

This part lays out the presuppositions and theoretical underpinnings for this study. We begin by giving a short introduction to the thesis. Further, we lay down some theoretical groundwork required for later parts, and present our set of hypotheses to be tested. Finally, we present our methodology used for collecting and processing data.

1 INTRODUCTION

User-generated online product reviews have become a natural part of the online marketplace experience for both retailers and consumers alike over the last few years. Some, like Yelp and Tripadvisor, have built their entire business model on such reviews, while others again like Amazon and Netflix use it to enhance their core model. The widespread use of these systems has also sparked interest from researchers. Several studies have been conducted aiming to understand different aspects of online consumer reviews. Research has already demonstrated an association between how positively a product is rated by consumers on a site and subsequent sales of the product on that site (Dellarocas et al., 2007; Chevalier and Mayzlin, 2006). In addition, a relationship between review volume and sales (Duan et al., 2008; Liu, 2006) has been established.

However, much of the important research only focus on a single category of products. For instance, highly cited work like Chevalier and Mayzlin (2006); Forman et al. (2008); Li and Hitt (2008) all include a large sample of products and reviews, but focus solely on books. Other studies like Ba and Pavlou (2002) and Mudambi and Schuff (2010) have included products from different categories, but these studies have a small sample of products (Mudambi and Schuff have 6), and do not use product category as a unit of analysis beyond grouping the products as search and experience goods. Several factors indicate, however, that product category properties moderate the influence reviews have on sales (Bjering and Havro, 2013). Therefore, it is in general difficult to assess the generalizability of prior research, and to determine whether different results stem from properties in product category or actual differences between review systems.

This paper aims to study the role product category plays as a moderating factor of properties concerning online reviews, by introducing a novel method for product category classification using natural language processing (NLP). With a selection of hit products and random products for each category, this paper also looks at the product popularity and its relation to the effect of online reviews on sales. Combining the impact of product category and popularity, the authors test a proposed model for a general understanding of reviews effect on sales, dubbed the review impact continuum. Finally, we take a look at review helpfulness and certain reporting biases concerning user generated online reviews.

The study includes a wide variety of categories, without compromising on either number of products or number of reviews. The dataset presented includes an unprecedented 1.1 million unique reviews from 4,600 products in 30 different product categories - spanning from the much studied movies and books to novelties like clothing, jewelry and hardware.

2 THEORETICAL BACKGROUND

This chapter introduces important literature as a foundation for understanding online product reviews and related concepts. We also introduce the concept of search and experience goods, before presenting the review impact continuum, a model introduced in Bjering and Havro (2013). We then outline natural language processing - needed as a foundation for an approach used to classify product categories in our study.

Sections 2.1-2.3 of this chapter are adapted from Bjering and Havro (2013).

2.1 WORD-OF-MOUTH AND ELECTRONIC WORD-OF-MOUTH

In marketing literature, an important concept is that of word-of-mouth communication (WOM). WOM can be defined as "all informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services or their sellers" (Hu et al., 2006, p. 324). WOM is a phenomenon that has been discussed for the better part of a century, with research attention gaining ground in the 1950s (see for instance Whyte Jr, 1954). The effect has since been thoroughly researched, and is now believed to be a strong influence on consumers' purchase intentions (Hennig-Thurau et al., 2004; Arndt, 1967).

The emergence of the internet has led to some interesting possibilities in the domain of WOM communication. Online communities where consumers share experiences have appeared in all shapes and sizes, bringing with them an entirely new way of conducting WOM communication. Messages can be stored indefinitely, in contrast to the "perishable" nature of traditional WOM. This allows online WOM communication to potentially saturate larger markets than before, as well as in quicker fashion. Indeed, the term viral is commonly used about the virus like spread of trends on the internet. Some even report that the life cycle of entertainment products has shrunk due to the speed with which messages, and with them, new trends, can spread (Dellarocas, 2006).

This new form of WOM communication has by much of the scientific community been dubbed electronic word-of-mouth, or eWOM (Hennig-Thurau et al., 2004; Racherla et al., 2012; Zhang et al., 2010). Hennig-Thurau et al. offer the following definition of eWOM: "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the internet" (Hennig-Thurau et al., 2004, p. 39). This is also the definition we employ throughout this thesis.

2.2 TYPES OF PRODUCTS

In economics and marketing, a search good is a product or service where the consumer is able to evaluate the quality, features and characteristics easily before purchase, such as a USB drive. According to Nelson (1970), to maximize expected utility, a consumer will search until the marginal expected cost of search becomes greater than its marginal expected return.

Contrasting search goods are experience goods. These are products or services where quality, features and characteristics are difficult to evaluate in advance of purchase, but can be obtained upon consumption. This can for instance be a hotel or a restaurant. Nelson (1970) asserts that marginal cost will be different in the experience case from that in the case of search goods. The expected cost of information in the experience case depends on the utility distribution. The marginal utility of an experiment is the potential loss in utility from consuming a brand at random rather than using the best brand that one has already discovered (Nelson, 1970). Nelson contends that this difference in dynamics has profound effects upon the market structure of consumer goods. He further predicts that the recommendations of others will be used more and have greater impact for purchases of experience goods than search goods.

This difference in market dynamics can also explain difference in behavior in advertising. Nelson (1970) goes on to show that for search qualities, advertising provides direct information about the characteristics of a brand. For experience qualities on the other hand, the most important information conveyed by advertising is simply that the brand advertises.

Some researchers contend that there is a third category as well, which is introduced somewhat later. This category, credence goods, is dominated by attributes that the consumer cannot verify confidently even after use. This is typically products or services such as for example dental services, where after being treated it is still difficult to assert its effect. The key feature of credence goods is that an expert knows more about the quality of the product than the consumer does himself (Darby and Karni, 1973).

2.3 REVIEW IMPACT CONTINUUM

In this section, we will present the review impact continuum, a hypothesized model for explaining differences in the effect of online consumer reviews on different types of products. The model was first presented by Bjering and Havro (2013).

2.3.1 *Model background*

Researchers seem to agree to a large degree that online product reviews affect product sales. The effect has also been seen for several different types of products. It has been found to hold for beer (Clemons et al., 2006), video games (Zhu and Zhang, 2010), books (Chevalier and Mayzlin, 2006; Li and Hitt, 2008; Hu et al., 2008), movies (Dellarocas et al., 2007) as well as restaurants (Luca, 2011). However, the reported effects vary in magnitude, with some studies even finding that review valence or ratings hold no explanatory power for sales (Duan et al., 2008). In order to explain these differences, several moderating effects have been researched and found to be influential. For products, two particular effects have been found; namely those of product type, and product popularity (Bjering and Havro, 2013).

Product type refers to the distinction between search and experience products. Nelson (1970) predicted that recommendations between consumers would be more important for experience products. This effect is found to be accurate for eWOM; online reviews seemingly have a larger impact on experience products than for search products (Senecal and Nantel, 2004; Park and Lee, 2009).

The second effect is related to the product popularity – whether the product appeals to niche groups, or to a larger, mass market. This is less studied, and usually only distinguishing between the extreme ends of the scale. However, the research seems to agree that the informative value of online reviews is larger for niche products, which in turn lends them a relatively larger degree of influence on sales (Dellarocas et al., 2007; Luca, 2011; Zhu and Zhang, 2010).

The referenced articles look at several moderating effects of review impact, but fail to put their findings into a bigger picture. Usually they stick to one particular effect, or at most, compare results along one of the dimensions. This has led to sometimes contrasting conclusions between researchers (cf. differences between Duan et al., 2008; Dellarocas et al., 2007). A new conceptual model was therefore proposed, that encompasses both axes. This would potentially be able to explain the observed differences, as well as more accurately predict the review impact on specific products.

2.3.2 *Review impact matrix*

A 2×2 matrix is constructed, encompassing both effects.

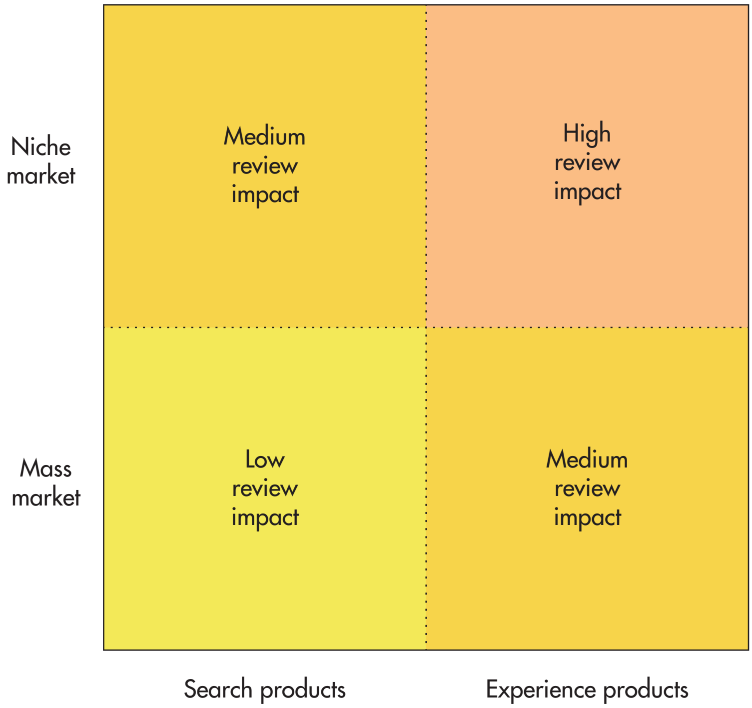

Figure 2.1: The 2×2 Review Impact Matrix.

Products placed within the top-right quadrant will see the largest effect of online reviews. An example could be an independent restaurant or a movie with a limited release. The theory contends that other sources of product information are particularly lacking for these products, and as such, WOM becomes an important channel of product information. In contrast is the argument for the lower left quadrant. We would argue that a common USB stick is a mass-market product, mostly evaluated on the available storage space. WOM will therefore be of less informational value – review impact is therefore relatively lower.

Products in the upper left quadrant are harder to evaluate based on objective information alone. We could imagine a USB stick with wireless capabilities. Objective information would exist, but many consumers may be confused as to how one would install and use it, since this is not a run-of-the-mill product. We note that most studies does not specifically measure the difference in effect between niche and mass products in the search category. It has, however, been conceptually posited (Chen and Xie, 2008).

The experience products in the lower right are harder to evaluate beforehand than their counterparts in the lower left corner, but they are easier to evaluate than the niche products in the upper right. Big production movies have famous actors and directors prominently displayed on marketing material. Media builds hype and expectations several months beforehand, and trailers go viral on the internet. There is an abundance of information which means less influence is given to online consumer reviews.

2.3.3 *Review impact continuum*

Although the above framework already seems to better explain the impact of reviews in a holistic sense, there are some weaknesses that should be addressed. Most importantly, products can very well fall between categories. Let us for instance consider a smartphone. Computers typically feature some quantifiable aspects, such as screen resolution, storage space, and battery life. However, many consumers are more concerned with ease of use and a solid user experience (UX) design. These two aspects pull in different directions concerning the classification of the product as a search or experience good. This discrepancy is true for many products, as they can often have different sets of features. In order to account for these cases, we increase the complexity of our model. By introducing continuums along both axes, we produce a diagram where products can be plotted on variable points. To describe the varying degrees of search or experience product features in a product, the horizontal axis is modified to indicate the degree of subjectivity with which one evaluates the product. Pure search products are expected to be evaluated based on largely objective criteria, while pure experience products are expected to be evaluated more with subjective experiences.

Further, products may be aiming for something in between the mass and niche markets. It is also not necessarily true that mass-market experience products always see strong or medium effects of reviews, some reporting very low effects for highly popular products (cf. Luca, 2011). The vertical axis is dubbed product popularity, by which we mean the relative amount of consumers that use or purchase the product. With these definitions in place, we can construct a review impact continuum for products, shown in Figure 2.2.

The review impact continuum immediately reconciles some of the differences in research findings. For instance, it offers one possible explanation for why Duan et al. (2008) do not find any direct online consumer review effect on sales, even when studying an experience product like movies, while most other researchers do (cf. Section 2.3.1). The

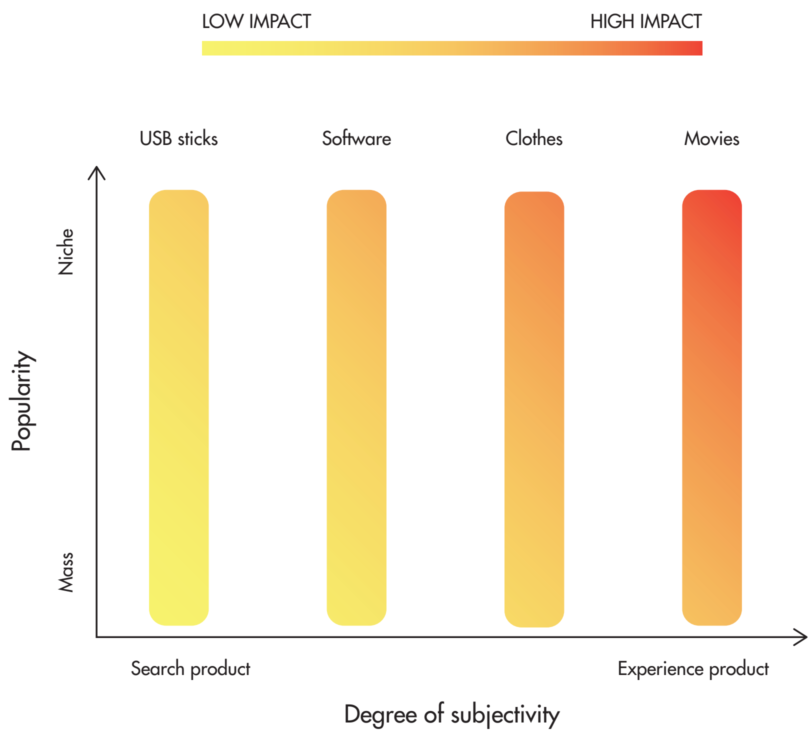

Figure 2.2: The Review Impact Continuum.

movies reviewed in Duan et al. are a selection of the absolute highest grossing movies in the market. This is the extreme end of the popularity dimension, considering that most movies do not even make it to the box office, and the lack of effect might be explained by the "hit" nature of these products.

Similary, Luca (2011) finds that the effect of reviews on sales are non-existing for chain restaurants. Again, our model suggests that this is because of the popularity of these restaurants, and thus not conflicting with the claim that online consumer reviews impact sales for experience products.

2.4 NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS

2.4.1 *Introduction to NLP*

Natural Language Processing (NLP) is a branch within language research that aims to provide means of computerized textual analysis that approaches human levels of contextual apprehension. Liddy (2001) offers the following definition:

> *"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications."*

In short, NLP provides researchers tools with which to quickly analyze large amounts of textual content, often aided by machine learning. Machine learning is a technique that allows the language processing to be adaptive, i.e. that the algorithms improve with extended use (Wikipedia, 2014c). Specifically, the more it analyzes, the more accurate it becomes.

2.4.2 *Sentiment analysis and subjectivity classifiers*

A specific application of NLP is known as sentiment analysis or opinion mining. The terms, which often denote the same field of study, started appearing in the beginning of the 2000s, and are closely related to the emergence of the internet and eWOM (Pang and Lee, 2008). Sentiment analysis is often employed to analyze the opinions of the masses, either from blogs, social media or through consumer reviews on online retailers. In short, sentiment analysis systems find, store and analyze opinions from textual sources. Sentiment analysis can also be an important tool in marketing, allowing brands or marketers to quickly and efficiently "test the waters" by programmatically analyzing the opinions of consumers on different aspects of products or services (Pang and Lee, 2008).

A subjectivity classifier is a common component in sentiment analysis systems. Its purpose is to analyze the contextual and literal meaning of any sentence and decide whether the sentence holds a subjective or objective sentiment (Pang and Lee, 2008). For instance, the sentence *"I love the look of the product."* is subjective, whereas *"The ladder has 12 steps."* is an objective one.

3 HYPOTHESES

This section introduces our 6 hypotheses, split into three groups: effect of reviews on sales, rating distributions and biases, and finally, review helpfulness. The overall goal of the thesis is to test the the review impact continuum model as presented in the previous chapter, as well as our new approach to classifying product categories. This is reflected in particular in hypothesis 1, regarding the effects of reviews on sales, where we specifically test both axes in the model.

We extend this line of inquiry to hypotheses regarding reporting biases as well as helpfulness of reviews. First, we aim to verify findings from prior research across different categories to control for any differences. Second, we look to strengthen our novel category classification by replicating previous results using our variables.

Parts of this section have been adapted from Bjering and Havro (2013).

3.1 EFFECT OF REVIEWS ON SALES

The most obvious question when discussing online consumer reviews is of their efficacy, or rather, whether a positive review leads to more sales. Perhaps the most cited study looked at the online book market. Using data from Amazon and BarnesAndNoble.com, Chevalier and Mayzlin (2006) found that improvement in a book's reviews lead to an increase in relative sales at the respective site. These findings were later corroborated by Hu et al. (2008), who also analyzed data for books sold on Amazon.com. By adopting a transaction cost and investment portfolio framework, effectively treating books as financial assets and reviews as favourable or unfavourable news, Hu et al. (2008) found that consumers responded positively to positive reviews, and negatively to negative reviews. Both of these studies also find that the negative impact of one-star reviews is greater than the positive impact of five-star reviews. The effect of reviews has also been researched outside the realm of Amazon. By combining reviews and ratings from Yelp.com for roughly 70% of all the restaurants in Seattle and quarterly revenue data over several years from Washington State Department of Revenue, Luca (2011) finds that a one-star increase in average rating leads to a 5-9% increase in revenue.

There are also studies that do not find a link between sales and ratings. Duan et al. (2008), using data from Yahoo! Movies and `boxofficemojo.com` find that the ratings of online user reviews have no significant impact on movies' box office revenues. However, assessing the available literature it can be expected that our data should show some effect on sales from reviews, and it will more likely follow the valence of the reviews than not. We formulate our first hypothesis:

HYPOTHESIS 1A: An increase in average rating on a site is associated with increased sales on that site.

Some research indicates that these effects on sales are moderated by the popularity of the product or service. Luca (2011) shows that while ratings on Yelp of independent restaurants in Seattle are affecting their revenue, ratings do not affect restaurants with chain affiliation. In fact, Luca finds that chains have become less popular after the introduction of Yelp, losing market share as Yelp has gained traction. He suggests that this is because the increased information about independent restaurants through online reviews is replacing more traditional sources of information like marketing.

Zhu and Zhang (2010) similarly find that online reviews are more influential for less popular games, where players need to rely more on sentiments from other consumers to assess game quality. Dellarocas et al. (2007) agree, finding that forecasting sales for niche movies can to a larger extent be done on the basis of reviews.

The literature seems to agree that for popular, what we can call hit products, a greater array of information channels exist. Large studio movies have immense marketing budgets and famous actors that contribute to the sales of the movies. Chain restaurants have a recognizable and trusted brand, consumers expecting the same service and product regardless of location. Smaller, independent producers can to a lesser extent afford expensive marketing campaigns, leaving consumer opinions a greater share of the available product information. Thus, we formulate the second part of our first hypothesis, which corresponds to the Y-axis in the review impact continuum:

HYPOTHESIS 1B: The association between average ratings and sales is stronger for non-hit products than for hit products.

Further, the impact of online consumer reviews seems to vary with the product type. Nelson (1970) predicted that recommendations between consumers would be more important for experience products than for search products. Since experience products pose a greater challenge to evaluate before trying or consuming, consumers will likely find greater utility of opinions from others for such products. For example, a simple mailing envelope needs only to match a few objective measurements; the dimensions and perhaps the inclusion of a plastic window for displaying the address. A consumer will know if the product is a match withouth resorting to experiences by peers. Contrasting this, a restaurant can inform potential patrons of their menu, any awards or accolades, but no objective info can tell him how it ultimately tastes. Thus, reviews are expected to be more persuasive for experience products.

Indeed, some research has been done on the subject, and gives support to the supposition. Senecal and Nantel (2004), constructing an experiment allowing subjects to make purchasing decisions after receiving product recommendations, find that recommendations for experience products are significantly more influential than recommendations for search products. Park and Lee (2009) also show that the eWOM effect is greater for experience goods than for search goods by having test subjects rate their perceived influence of reviews for a set of search and experience products. Considering all this, the third part of hypothesis 1 becomes:

HYPOTHESIS 1C: The association between average ratings and sales is stronger for experience products than for search products.

A problem with the original classification of search and experience goods by Nelson (1970) is that it is binary in nature; later literature tends to treat the distinction as somewhat less discrete. Mudambi and Schuff (2010) argued that some products would hold qualities from both categories, and be difficult to classify as either search or experience products. Hpwever, they do not have a proper way of classifying the categories along a scale.

In order to treat product categories in a more nuanced fashion and having the ability to classify products that seemingly belong somewhere in the middle, a new variable is needed to determine the relative position of a product between pure search goods and pure experience goods. Looking at previous research, a common denominator in the evaluation process seems to be the degree of subjectivity that is used to assess product quality (Mudambi and Schuff, 2010; Bjering and Havro, 2013). Since search products are to a larger extent defined by objective facts, it seems plausible that reviews for search products contain more objective statements. For instance, a USB stick review would likely contain information about it's storage capacity. Conversely, since experience products cannot as readily be evaluated on objective facts alone, one would expect the ratio of subjective statements in the reviews to be larger.

By introducing a new variable, using natural language processing analysis of review texts to quantify the degree of subjectivity with which a product is evaluated, we expect to see differing impacts of reviews, depending on the product's position on the subjectivity axis. This corresponds to the X-axis in the proposed review impact continuum.

HYPOTHESIS 1D: The association between average ratings and sales is stronger for products that tend to be subjectively evaluated than for products that tend to be objectively evaluated.

Finally, some of the reviewed literature also sees an effect from the volume of reviews. Duan et al. (2008) did not find a relationship between the rating of movies and box office sales, however they show that sales are significantly influenced by the volume of online posting. Duan et al. attribute the effect of online user reviews to it being an indicator of the underlying word-of-mouth that plays a dominant role in driving box office revenues. This theory is supported by Zhang et al. (2010) and Dellarocas et al. (2007). Liu (2006) also finds that WOM information offers significant explanatory power for both aggregate and weekly box office revenue, especially in the early weeks after a movie opens. Most of this explanatory power, Liu argues, comes from the volume of WOM. Our second hypothesis is:

HYPOTHESIS 2: A comparatively high number of reviews on a site is associated with comparatively higher sales on that site.

3.2 RATING DISTRIBUTIONS AND REVIEW BIASES

For a consumer conducting online research, it is beneficial that the opinions posted are trustworthy and present a credible picture of the marketplace. If the available reviews are for some reason skewed towards one end of the scale, the consumer may be enticed to purchase a product that does not represent the optimal choice. This phenomenon is often called *review bias*.

A commonly cited shortcoming of online reviews is *under-reporting bias*. Under-reporting refers to the notion that the reviews posted for a product are not accurately describing the whole of consumers' opinions - the population of reviewers may be biased or lacking in magnitude, reaching a verdict that does not reflect the objective quality or value of a product. Under-reporting bias is likely primarily a consequence of the motivations for posting reviews, in which extremely satisfied or extremely dissatisfied consumers are more likely to post reviews. Consumers with mediocre or average experiences simply don't find the same utility in expressing their views (Anderson, 1998; Hennig-Thurau et al., 2004; Hu et al., 2006). As such, the rating distributions approach U-shaped curves, where the average values are underrepresented. In fact, Hu et al. (2006) found that about 53% of products reviewed on Amazon.com have bimodal rating distributions, showing signs of the U-shape. We expect that our data should reflect previous findings, giving us:

HYPOTHESIS 3: The distribution of ratings for a product tends to be bimodal, with the low and high end of the scale as local modes.

Another described bias is the self-selection bias. This is a phenomenon that occurs when products have a subset of consumers that are especially invested in that product, its producer or category (Li and Hitt, 2008). For instance, if an author has a loyal following, it is likely that this subset of consumers will be strongly represented in the early adopters of new books. The first reviews may therefore be positively biased, misrepresenting the true quality of the product until a sufficiently large amount of unbiased reviewers pitch in. Further evidence of the bias has been reported in several instances (Dellarocas et al., 2007; Hu et al., 2011; Zhu and Zhang, 2010), leading us to expect similar effects in our dataset. We formulate:

HYPOTHESIS 4A: The average rating of a product tends to decrease over time before stabilizing at a long term value lower than the initial value.

The self-selection of early reviewers may also cause secondary problems. When disappointed consumers, having bought into the biased early reviews, post their experiences, they over-compensate and post reviews that are more negative than the average long term value (Li and Hitt, 2008). These negative reviews then affect the average rating below the long-term average. This causes a tell-tale dip in the average ratings. Li and Hitt (2008) find that products that are affected by undershooting on average see the dip in ratings between the 6th and 19th weeks after release. This is known as the undershooting period, and leads us to the next part of our hypothesis:

HYPOTHESIS 4B Some products with a difference in initial average rating and long term average rating go through an ”undershooting” period after the initial period where the rating is lower than the long term average.

Li and Hitt (2008) go on to show that the undershooting effect is stronger for products with more heterogenous consumer preferences. As previously argued, experience products are believed to be evaluated more on subjective preferences than search products. For products being evaluated purely on objective criteria, it is easy to imagine consumer preferences to converge; either the product matches the objectively defined need or usage, or it doesn't. However, products being evaluated on a basis of subjective taste will likely see more diverging preferences. Li and Hitt (2008) found support for their hypothesis based solely on data from books, looking at differences in consumer preferences for products within that category. However, considering the assumed dynamic between search and experience products, it seems plausible that differences in the magnitude of any undershooting can also be observed between product types. Specifically:

HYPOTHESIS 4C: The undershooting effect is stronger for experience products than for search products.

3.3 REVIEW HELPFULNESS

The final area of focus for this study is based on the content and other accompanying information about specific reviews. Specifically, we will look at a particular metric provided by several online retailers with reviews, namely that of review helpfulness. As will be described in Section 4.3.2, consumers are often asked if they find a particular review to be helpful or not. The number of helpful votes are then displayed right next to or below the review, giving consumers some added information of the quality of specific reviews. The review helpfulness has been shown to be influential for the impact of reviews on several occations (see for instance Hu et al., 2008; Forman et al., 2008)

A previous analysis of reviews from Amazon.com across six products indicated that review extremity, review depth, and product type affect the perceived helpfulness of the review (Mudambi and Schuff, 2010). Independently of product type, review depth, or rather, the length of the review measured in word count, was found to have a positive effect on the helpfulness of the review. This is simply believed to stem from the increased information provided in the review. Longer reviews are assumed to include more nuanced descriptions of the products, considering both positive as well as negative sides (Chevalier and Mayzlin, 2006). However, Mudambi and Schuff found that the effect was greater on the helpfulness of the review for search goods than for experience goods. Since reviews for search goods are often presented in a fact-based, sometimes bulleted format, reviews can be relatively short. The factual nature of search reviews implies that additional content in those reviews is more likely to contain important information about how the product is used and how it compares to alternatives. While additional review content is helpful for all reviews, Mudambi and Schuff (2010) found that the incremental value of additional content in a search review was more likely to be helpful to the purchase decision than the incremental value of additional content for experience reviews.

Expanding on previous findings, we expect to see similar effects, but also with our new subjectivity classification of product categories. Our fifth hypothesis, in three parts:

HYPOTHESIS 5A: Reviews that are perceived as helpful tend to be longer than other reviews.

HYPOTHESIS 5B: The association between helpfulness and the length of the review is stronger for search goods than for experience goods.

HYPOTHESIS 5C: The association between helpfulness and the length of the review is stronger for products which tend to be objectively evaluated than for those which tend to be subjectively evaluated.

One trend that has been seen amongst online retailers, is the growing inclusion of social network functionality (Bjering and Havro, 2013). This development finds support in research. Forman et al. (2008) found that online community members rate reviews containing identity-descriptive information more positively, and that the prevalence of reviewer disclosure of identity information is associated with increases in subsequent online product sales. Reviewers who disclosed real name or location had 12.2 percentage points more helpful votes than otherwise identical reviewers. Wang (2010) found that there were an order of magnitude more *prolific* reviewers on Yelp than for two competing sites, prolific referring to the productivity and perceived helpfulness. This is believed to be a consequence of increased trust in the reviewers stemming from Yelp's encouragement of creating social profiles on their review system. Wang contends that this trust is critical for consumers when assessing reviews. We formulate our sixth hypothesis.

HYPOTHESIS 6A: Reviews written by reviewers that use their real name are perceived as more helpful than other reviews.

Another element that may increase trust in the reviews is a verification that a transaction has taken place. If consumers can be certain that the reviewer actually has used the product, a greater amount of credibility can be attributed to the opinions in the review. Many online retailers display such information (e.g. Amazon.com), whereas some require a purchase for the consumer to be able to review the item (e.g. Hotels.com). Consumers are increasingly wary of review manipulation and fake or shill reviews (Bambauer-Sachse and Mangold, 2013), so it should be conceivable that reviews with verified purchases should see larger amounts of trusts, and thus be perceived as more helpful. We postulate:

HYPOTHESIS 6B: Reviews written by reviewers with verified purchases are perceived as more helpful than other reviews.

3.4 OVERVIEW OF HYPOTHESES

Table 3.1 provides an overview of the hypotheses as described in sections 3.1 to 3.3

#	HYPOTHESIS
1a	An increase in average rating on a site is associated with increased sales on that site.
b	The association between average ratings and sales is stronger for non-hit products than for hit products.
c	The association between average ratings and sales is stronger for experience products than for search products.
d	The association between average ratings and sales is stronger for products that tend to be subjectively evaluated products than for those who tend to be objectively evaluated.
2	A comparatively high number of reviews on a site is associated with comparatively higher sales on that site.
3	The distribution of ratings for a product tends to be bimodal, with the low and high end of the scale as local modes.
4a	The average rating of a product tends to decrease over time before stabilizing at a long term value lower than the initial value.
b	Some products with a difference in initial average rating and long term average rating go through a ”undershooting” period after the initial period where the rating is lower than the long term average.
c	The undershooting effect is stronger for experience products than for search products.
5a	Reviews that are perceived as helpful tend to be longer than other reviews.
b	The association between helpfulness and the length of the review is stronger for search goods than for experience goods.
c	The association between helpfulness and the length of the review is stronger for products which tend to be objectively evaluated than for those which tend to be subjectively evaluated.
6a	Reviews written by reviewers that use their real name are perceived as more helpful than other reviews.
b	Reviews written by reviewers with verified purchases are perceived as more helpful than other reviews.

Table 3.1: Overview of hypotheses.

4 METHODOLOGY

4.1 RESEARCH DESIGN AND VALIDITY

An important concept in research methodology and design is the notion of research validity and reliability. As ours is an explorative study, aiming to test certain hypotheses about online product reviews, two of our main concerns in the research design will be to secure Construct validity as well as reliability (Yin, 2009). Construct validity refers to the degree to which the study measures what it claims to be measuring. In other words, whether the research is constructed in such a way that it can make any inferences on the variables it purports to research. Reliability describes the extent to which other researchers will be able to replicate the research that has been conducted. We will give brief descriptions of how these concepts are handled for this thesis, before detailing every step of the research design in the following sections.

CONSTRUCT VALIDITY The study aims to measure the effect of reviews on sales. As such, there is arguably no more sound data than reviews themselves, along with connected sales points. However, high resolution sales data is difficult to obtain, so our study uses a proxy called the Amazon sales rank. Much of the previous research focusing on Amazon also uses the sales rank as a proxy for sales, amongst others Schnapp and Allwine (2001); Chevalier and Goolsbee (2003); Chevalier and Mayzlin (2006). These demonstrate that the relation between $\ln salesrank$ and $\ln sales$ is approximately linear.

Our statistical models build on methods presented in peer-reviewed and highly cited articles (for instance Chevalier and Mayzlin, 2006; Li and Hitt, 2008; Mudambi and Schuff, 2010), which should ensure that our tests are valid.

RELIABILITY In order to allow for other researchers to replicate our research, we provide detailed descriptions of the data collection, data structuring, statistical models and variable construction. The relevant source code used to collect data is provided in Appendix B, and the employed dataset will also be provided by the authors upon request.

4.2 RATIONALE FOR CHOOSING SOURCE OF DATA

When deciding on the source of data for this research, several possibilities were considered, among them TripAdvisor, Yelp, the Norwegian tender portal MittAnbud as well as Amazon. In the end, the choice hinged on three criteria:

- Volume of reviews
- Availability of sales data
- Variety of products

VOLUME OF REVIEWS In order to make meaningful conclusions about reviews, their content and ratings, a sufficiently large volume of reviews is necessary. In this regard, Amazon is arguably the largest player on the internet, in 2010 being cited as the most important single source of product reviews by consumers in a study by PowerReviews (BusinessWire, 2010). In addition, Amazon is unique in that reviews are being written in significant numbers on several categories usually not exposed to online consumer reviews, like can openers (more than 22,000 reviews) or ladders (more than 12,000 reviews).

AVAILABILITY OF SALES DATA An imperative for determining the effect of reviews on sales is having actual indications of how sales are faring. In this regard, there are mainly two types of data sources. The first is official revenue data from businesses being reviewed at online review portals such as Yelp or TripAdvisor, the second is direct sales through an online store (that also has reviews).

In order to connect businesses' revenue data with reviews, one needs access to revenue filings or tax records. Access to Norwegian records is readily available, but data is released with yearly intervals, meaning short-term rating changes could not be analyzed properly. In addition, review density is lackluster. The US has a very high review density, and businesses file revenues on a quarterly basis. Access to this data, however, is difficult to obtain.

The second alternative, sales data from web merchants, is not readily available either. Few web merchants, if any, allow access to detailed sales information of products they offer. One notable half-exception, however, is Amazon. Sales data is not offered directly, but rather a relative sales indicator known as the Amazon Sales Rank (Wikipedia, 2014a). The sales rank has been used in previous research (Schnapp and Allwine, 2001; Chevalier and Goolsbee, 2003; Chevalier and Mayzlin, 2006) and has been found to be a good proxy for sales figures, lending a strong argument for Amazon as our data source.

VARIETY OF PRODUCTS One especially lacking area in existing research is a comparison of the effects of reviews and ratings on different product categories in a holistic way. With this in mind, finding a data source with a wide spectrum of products and product categories was important. Several online stores offer a wide selection of products, amongst them eBay, Amazon.com and WalMart.com. Common for most online retailers with a wide product selection, however, is that they are all limited in at least one of our two other criteria. Some have limited reviewing functions (e.g. eBay), and few, if any, give out any sales information.

CONCLUSION Amazon is known as "the world's largest online retailer" (Wikipedia, 2014b) and carries millions of products over hundreds of different categories. Combining in the fact that they have arguably the deepest set of product reviews, as well as a method to determine the magnitude of sales, Amazon seems like a reasonable choice for our data source.

4.3 THE AMAZON WEB PAGE

We will here elaborate some aspects of the Amazon web page, more specifically how reviews are displayed to consumers, important terminology and the Amazon Product Advertising API.

4.3.1 *Layout*

The Amazon web page has a number of different product views depending on where you are on the page. The first view for most consumers is usually the product listing view. This is shown when a product is displayed on the front page, or in a list of results for a search. An example of a product listing view is shown in Figure 4.1.

Figure 4.1: Amazon product listing.

This view aggregates the products review information to an average rating, rounded to the nearest half point and displayed as filled up stars. In addition, it shows the number of ratings and reviews written for the product.

The next view, shown in Figure 4.2, is the product page view. This is the page shown to users when they click on a specific product.

Figure 4.2: Amazon product page.

The product page includes all information shown in the listing view, in addition to product variation selections and a larger photo. Here, the consumer may also hover the mouse over the rating to bring up the detailed rating view, as shown in Figure 4.3.

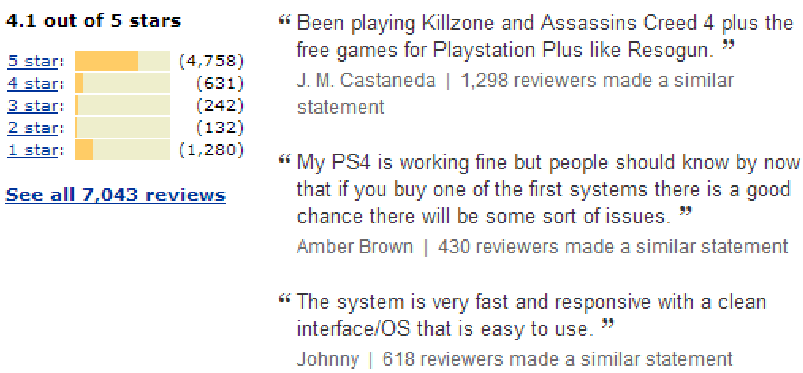

Figure 4.3: Detailed rating view.

At the bottom of the product page is the actual review listings, sorted on "most helpful" as default. A sample of a review is given in Figure 4.4.

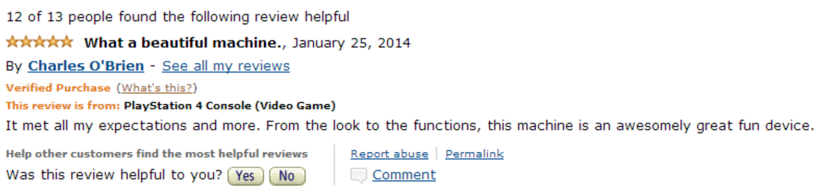

Figure 4.4: Sample review.

4.3.2 *Terms and concepts*

Here we introduce some terms and concepts that the reader should be familiar with before reading later parts of the thesis.

ASIN An abbreviation of Amazon Standard Identification Number, which is a unique 10 character alphanumeric product identifier assigned by Amazon.

REVIEW HELPFULNESS As shown in Figure 4.4, users are asked "Was this review helpful to you?" on each review listing. This can be answered with a "Yes" or a "No". We refer to the share of people who answer "Yes" to this question as the "perceived helpfulness". The total number of votes (either "Yes" or "No") is referred to as the "total helpful votes".

This is displayed at the top of each review (cf. Figure 4.4) in the sentence "X of Y people found the following review helpful".

SALES RANK One of the datapoints Amazon offers through the Product Advertising API is the sales rank of any given product. The sales rank is a numerical value that denotes the relative sales of a product within a product category. A low sales rank thus means high sales, and vice versa.

BROWSENODES Product categories in Amazon's systems are sorted within so-called browsenodes, each with aunique numerical ID. Browsenodes can have both parent browsenodes as well as child browsenodes, forming trees of product types. For example, the browsenode for Books (#1000) has several children, among them History (#9) and Law (#10777). The topmost browsenode in any category is also known as a search index. The sales rank value for a product is relative within the product's associated search index, not the immediate parent browsenode.

4.3.3 *The Product Advertising API*

One of the services offered by Amazon Web Services is dubbed the Product Advertising API (Amazon.com, 2014c) which gives programmatical access to the product offerings and discovery methods on Amazon.com. The API (Application Programming Interface), is a ready made set of code libraries and functions that developers can use to access different services. Third-party web site owners and bloggers can through the Product Advertising API offer their users a convenient way to browse Amazon's products without ever leaving their site – in effect offering a mini Amazon storefront. For sales generated through these storefronts, Amazon remunerates the third-party sites with a percentage of the transactions.

Access to the Product Advertising API is entirely free of charge, and requires only an active Amazon account, with which one needs to sign up for the Amazon Associates program. This in turn gives an associate tag that is needed to access the API. In addition, one needs to obtain a unique Access Key ID as well as a Secret Access Key, both available through the AWS websites. These keys are needed in order to sign requests to Amazon's servers.

In order to access the Product Advertising API a set of methods were written in Java. With Java, Amazon allows developers access to the API using the SOAP request protocol (Wikipedia, 2014d). The code written for this research is included in Section B.2.

4.4 SELECTION OF CATEGORIES

In selecting categories, our main goal was to have categories that varied on a number of parameters:

- *Search/Experience categorization.* We wanted to include categories that are undoubtedly representatives for each of these classifications (such as USB drives (search) and books (experience), and in addition some that are more ambiguous in their categorization (e.g. clothing).
- *Market size.* Both high-volume product categories and low-volume

In addition, the category should:

- *Provide sales ranks on its products.* Some categories do not, for unknown reasons, return sales ranks through the API. These categories were not considered.
- *Have products that are reviewed.* The volume of reviews varies greatly from product to product, but also from category to category. The included categories should at least have some reviews

and ratings for most its products, or else they will be difficult to study.

- *Be specific enough to primarily include one type of product.* For instance, the category "Electronics" is regarded as too general since it contains a large variety of electronics products with very different properties.
- *Be easily understandable* in regards to which types of products it includes.
- *Mutually exclusive.* Meaning that one particular product can only appear in one of the selected categories.

Using these criteria a total of 30 categories were selected. These categories are listed in Table 4.1.

CATEGORIES	
Board Games	Hobby Fabric
Books	Ink and Toner
Bowls	Jewelry
Can Openers	Ladders
Candy	Movies
Car Electronics	Perfumes
Clothing	Restroom Fixtures
Copy Paper	Screws
Desktop Computers	Shoes
Digital Cameras	Software
Dog Food	Test & Measure
Envelopes	USB Drives
Guitars	Video Games
Hard Drives	Vitamins
Hardware	Watches

Table 4.1: Selected product categories.

4.5 SELECTION OF PRODUCTS

To make it possible to compare hit products with non-hit products, we decided to have two different selection methods.

4.5.1 *100 best selling products*

Amazon provides lists with the 100 best selling products within a large selection of categories. These lists are updated hourly (Amazon.com, 2014b), and display the products within a category with the highest sales rank. This does not mean that the product in first place has the sales rank 1, because as explained in 4.3.2, the ranks are connected to the search index, not the necessarily the immediate category (browsenode). However, it does mean that the sales rank for the product at the top of the list has a lower rank than the one at number 2, number 2 on the list has a lower rank than number 3, and so on. It also means that these products are still to be viewed as the hit products within their category.

The products collected as top 100 products are the ones that topped the Amazon Best Sellers list as of March 19, 2014. The exact time of collection varies from category to category within that day. It is worth noting that the product ASINs are collected directly from the Best Sellers listing on the web page, while the sales ranks are collected through the Product Advertising API.

4.5.2 *100 random products*

In addition to the top selling 100 products, an additional 100 random products were selected for each product category. This was done for several reasons. First of all, simply looking at the top 100 products would not allow many of the products room to climb the sales rankings, which would make it harder to measure the effects on sales from positive reviews. Second, a randomized selection offers a way to compare the effects of reviews for products with varying degrees of popularity. Third, the top 100 products may see a large degree of biases from different types of exposure on Amazon.com that lower ranked products do not. As such, the random sets may serve as a control group, should the top 100 products be too affected by forces other than ratings and reviews.

As there is no ostensibly reliable way to retrieve sets of random products from either Amazon's web pages or the Product Advertising API, a script was written in java to access the API and manually make lists of randomized products. In order to make the searches as random as possible, the search terms were pulled from Wordnik (2014), an online dictionary and language resource. Wordnik has its own API that allows developers to include dictionary functionality in their applications or web sites, with one of the available functions (getRandomWords) returning a list of random words. The employed script fetched 1000 random words at a time, and using a predefined browsenode ID searched the Amazon API for products matching the random words, one word at a time. In order to avoid duplicates, any matches were controlled against the top 100 lists. In addition, matches had to return a valid sales rank between 100 and 40,000. This was implemented to segregate the random selections from the top 100 products, while still not being so low as not to see any sales or review activity. The 40,000 limit was deduced after sampling products at different levels of sales rank. A complete list of the required criteria for random products is displayed in Table 4.2. The code written to fetch random products is included in Appendix B.2.2.

TYPE	CRITERION
Search term	Randomly selected noun, adjective, verb or adverb
Sales rank	Valid sales rank between 100 and 40,000 returned
ASIN	Not already in retrieved top 100 list
Availability	Product is available for sale as new (not used)

Table 4.2: Criteria for selection of random products.

4.6 COLLECTING REVIEWS

After selecting categories and products, the next task was collecting the reviews. Since the Product Advertising API of Amazon does not support the collection of individual reviews, a custom VBA script was created to download the required data. The code is included in Section B.1. The script operates in two separated phases to increase stability and reduce the possibility of errors. The data is stored after each step, making the process traceable and individual steps repeatable.

The first step is the actual data gathering, accessing the reviews using the unique ASIN-number. The script accesses the Amazon web page, iterates through the lists of ASINs until the last page number is reached. The section marked by the unique document object model (DOM) identifier “productReviews” is collected. This section contains all reviews on that web page (10 for all pages, between 1 and 10 for the last page), but excludes everything else (headers, footers, comments and more). Further, the script splits the collected section into individual reviews using a unique string as separator. This approach is required since Amazon does not provide a unique DOM-id for each individual review, making string-based separation the only option. However, the number of reviews collected are controlled against the known total number of reviews for that product, ensuring that no reviews are lost in this process. The result of phase 1 is stored as a long string in an excel sheet.

The second step is structuring the data. Since Amazon does not provide any DOM-id for the individual data-fields either, the same string-based separator approach is required. Using a variety of textual identifiers and pattern-recognition, the data for each individual review is split into several fields, as listed in Table 4.3.

INFO	DESCRIPTION
Review date	Accurate to the date, exact time is not available.
Rating	Ranges from 1 to 5 stars, 5 being the best.
Reviewer name	This can be a user name or a real name.
Review title	The title of the review.
Review text	The actual review text.
Helpfulness	People who found the review helpful (cf Section 4.3.2)
Total helpful votes	People who voted on whether they found the review helpful (cf. Section 4.3.2)

Table 4.3: Basic information available on reviews.

On Amazon, users can earn badges by creating what users and Amazon regards as good content. Some badges are temporary, and some are permanent. In addition there are some tags giving extra information about the reviewer or the purchase, aiming to help the reader assess the quality and credibility of the review. An overview of important badges and tags is given in Table 4.4. A full overview of available badges can be found on Amazon.com (2014a).

BADGE	DESCRIPTION
Verified purchase	This means that the customer who wrote the review actually purchased the item at Amazon.com, and Amazon was able to verify it.
Real Name	Awarded to users use their real names. Names are verified using a credit card issued to the user.
Top reviewer	This includes ”#1 reviewer”, “Top 10 reviewer”, “Top 50 reviewer”, “Top 500 reviewer”, “Top 1000 reviewer” and “Hall of fame reviewer”. These badges are given to whom Amazon view as their best reviewers.
Vine voice	Amazon Vine is a program that enables a select group of Amazon customers to post opinions about new and pre-release items to help their fellow customers make educated purchasing decisions.

Table 4.4: Collected review badges and tags.

In addition to the basic information and badges, the script calculates two new data-points when processing the reviews, cf. Table 4.5.

METRIC	DESCRIPTION
Typed characters	The number of characters in the review text field.
Distinct words	The number of distinct words in the collection of reviews for a product.

Table 4.5: Calculated information about reviews.

4.7 CLASSIFICATION OF PRODUCTS

4.7.1 *Search and experience products*

Remembering traits identified as most often found in either search or experience, we here do a manual classification of categories. Mudambi and Schuff (2010, page 191) describe a search product as ”one for which it is relatively easy to obtain information on product quality prior to interaction with the product; key attributes are objective and easy to compare, and there is no strong need to use one’s senses to evaluate quality.” Contrasting this, an experience product is identified as ”one in which it is relatively difficult and costly to obtain information on product quality prior to interaction with the product; key attributes are subjective and difficult to compare" (Mudambi and Schuff, 2010, page 191).

Using this understanding, we identify a subset of product categories in our data as experience products in Table 4.6 and a subset of product categories as search goods, shown in Table 4.7.

CATEGORY	EVALUATION CRITERIA
Board games	How engaging and fun it is
Books	"Soft" qualities like story and characters
Candy	Taste, smell and texture
Movies	"Soft" attributes like story and characters
Perfumes	Subjective preferences for smells
Video games	How engaging and fun it is, as well as graphics, story, music and characters.
Vitamins	Taste and smell, as well as post-consumption effects.

Table 4.6: Product categories identified as experience products.

CATEGORY	EVALUATION CRITERIA
Bowls	Size, volume or type of material
Can openers	Type of metal, warranties of durability, electrical motor
Ink and toner	Which printing or copying system it is made for
Paper	Type of paper, size and weight.
Hard drives	Storage capacity and transfer speeds
Envelopes	Dimensions, padding and similar.
Screws	Size, diameter, type of metal
USB drives	Storage capacity and transfer speed.

Table 4.7: Product categories identified as search products.

4.7.2 *Degree of subjective evaluations*

In order to measure the degree of subjectivity used for evaluating a certain product category, we need some content to measure. In this regard, there is arguably no better content than the reviews themselves; in effect documentations of how each consumer evaluated the product. Our chosen method for measuring the subjectivity of reviews is through computerized sentiment analysis with subjectivity classification, as outlined in Section 2.4.

For analyzing the sentiments of reviews, we employ the freely available *OpinionFinder* library, developed by researchers at the University of Pittsburgh, Cornell University, and the University of Utah (MPQA, 2011). The set of OpinionFinder classifiers have been widely used in previous research, reporting good results in classifying subjectivity (see for instance He et al., 2008). The subjectivity classifiers included in the toolkit are based on work by Riloff and Wiebe (2003); Wiebe and Riloff (2005) from the universities of Utah and Pittsburgh. The OpinionFinder toolkit includes two separate subjectivity classifiers.

The first classifier is a model-based classifier, meaning it is based on a model that can be trained through machine learning. This classifier has a reported accuracy of 76%, subjective precision of 79% and subjective recall of 76% (MPQA, 2011). The precision denotes how many of the reported subjective sentences in fact are deemed subjective manually, the recall represents the percentage of manually tagged subjective sentences that are classified as such by OpinionFinder. This method classifies all sentences as either objective or subjective.

The second is rule-based, working by applying pre-defined rules to determine whether a sentence is subjective or objective. The rule-based classifier is reported to have a higher accuracy (91,7% for subjective sentences, 83% for objective sentences), but with lower recall (30.9% subjective recall, 32.8% objective recall), since it will only classify a sentence as subjective or objective if it can do so with confidence (MPQA, 2011). The result is therefore 3 classifications, objective, subjective or unknown. When calculating fraction of subjectivity with these result, we disregard those classified as unknown, and employ the number of subjective sentences divided by those classified as either subjective or objective as our fraction.

As input for the OpinionFinder tool, we pulled 1000 random product reviews from each category. The reviews were then run through both subjectivity classifiers to determine the fractions of subjective sentences for each category, which forms the basis of our subjectivity variable. The results are presented in Section 5.5, and in Table A.11 and Table A.12 in the appendix.

4.8 MODEL SPECIFICATION

We will here discuss the different statistical methods employed to test the hypotheses outlined in Chapter 3.

4.8.1 *Effect of reviews on sales*

For hypotheses 1 and 2 we define two multiple regression models that aim to predict the effect of several different variables on the natural logarithm of the sales rank. Ideally, our dependent variable would be the natural logarithm of sales, but as outlined in Section 4.3.2, the relationship between $\ln sales$ and $\ln salesrank$ is approximately linear, making $\ln salesrank$ a adequate substitute.

STATIC MODEL A product's sales rank on Amazon is likely affected by several variables. Building on work by Chevalier and Mayzlin (2006); Luca (2011); Forman et al. (2008), our regression model assumes that the salesrank is mainly a function of a product's average rating, price and volume of reviews. In addition, we will be testing for whether or not product type or the degree of subjectivity in evaluation can contribute to the effect of ratings on sales. Finally, it is assumed that certain products experience certain fixed effects. These fixed effects can be the relative popularity of an author or producer, offline promotions, or simply the quality of the product. These fixed effects, however, are difficult to observe and quantify across such a large and varied set, and will not be treated in this model.

The specification becomes:

$$\begin{aligned}
&\ln \mathsf{SALESRANK}_{pt} = \\
&\alpha + \beta_1 \ln \mathsf{PRICE}_{pt} + \beta_2 \mathsf{AVGRATING}_{pt} \\
&+ \beta_3 \ln \mathsf{NUMREVIEWS}_{pt} + \beta_4 \mathsf{PRODUCTTYPE}_{p} \\
&\qquad + \beta_2 \mathsf{AVGRATING}_{pt} \times \mathsf{PRODUCTTYPE}_{p} + \epsilon
\end{aligned} \tag{4.1}$$

With subjectivity variables:

$$\begin{aligned}
&\ln \mathsf{SALESRANK}_{pt} = \\
&\alpha + \beta_1 \ln \mathsf{PRICE}_{pt} + \beta_2 \mathsf{AVGRATING}_{pt} \\
&+ \beta_3 \ln \mathsf{NUMREVIEWS}_{pt} + \beta_4 \mathsf{SUBJECTIVITY}_{p} \\
&\qquad + \beta_2 \mathsf{AVGRATING}_{pt} \times \mathsf{SUBJECTIVITY}_{p} + \epsilon
\end{aligned} \tag{4.2}$$

where the subscript p denotes *product* and t denotes *time*. PRICE_{pt} thus denotes the price for product p at time t. The coefficient β_1 may therefore be seen as a measure of the effect of the product price on $\ln \mathsf{SALESRANK}$, or in effect, a proxy for the price elasticity of the product.

$\mathsf{AVGRATING}_{pt}$ represents the average star rating for a product p at time t. Since the sales rank data has been extracted at daily intervals, the average rating for any specific day includes all reviews submitted before or at that specific date. The coefficient β_2 thus represents the effect of the average star rating on the sales rank.

The variable $\mathsf{NUMREVIEWS}_{pt}$ denotes the number of reviews submitted for a product p before or at time t. This is in line with Duan et al. (2008), who suggest that the most important review variable when looking at sales is the volume of reviews, rather than their valence. As per Chevalier and Mayzlin (2006), we use the logarithms of price and number of reviews so that we can compare the effect of percentage change in either variable on the percantage change in sales rank.

Further, we include the dummy variable $\mathsf{PRODUCTTYPE}_p$ to control for any effects on sales rank that stem from the product being classified as either a search or experience product. Since our product selection does not guarantee that categories have similar levels of sales ranks, categories may have significantly different mean sales ranks, which could bias the regression. Since not all product categories have been classified as either search or experience goods (c.f. Section 4.7.1), this variable will only be used with those products. In similar fashion, in Equation 4.2 we substitute in the variable $\mathsf{SUBJECTIVITY}_p$ to control for any differences in mean sales ranks for the different levels of subjectivity.

To test for the interaction between product type and the average rating, we include the compound variable $\mathsf{AVGRATING}_{pt} \times \mathsf{PRODUCTTYPE}_p$. This interaction term is meant to pick up if the effect of the average rating on sales is larger for any of the product types. Similarly, the $\mathsf{AVGRATING}_{pt} \times \mathsf{SUBJECTIVITY}_p$ variable in Equation 4.2 is included to test if the degree of subjectivity really moderates the effect of ratings on sales.

ALTERNATIVE RATING VARIABLES An alternative approach used by Chevalier and Mayzlin (2006) involves substituting the average star rating with variables denoting the fractions of five-star and one-star reviews. This method allows for a more nuanced view of the impact of review valence on salesrank. Substituting these two new variables into Equation 4.1 we get:

$$\begin{aligned} &\ln \mathsf{SALESRANK}_{pt} = \\ &\alpha + \beta_1 \ln \mathsf{PRICE}_{pt} + \beta_2 \ln \mathsf{NUMREVIEWS}_{pt} \\ &+ \beta_3 \mathsf{PRODUCTTYPE}_p + \gamma_1 \mathsf{ONESTAR}_{pt} \\ &+ \gamma_2 \mathsf{FIVESTAR}_{pt} + \gamma_3 \mathsf{ONESTAR}_{pt} \times \mathsf{PRODUCTTYPE}_p \\ &\qquad + \gamma_4 \mathsf{FIVESTAR}_{pt} \times \mathsf{PRODUCTTYPE}_p + \epsilon \end{aligned} \tag{4.3}$$

Where $\mathsf{ONESTAR}_{pt}$ now denotes the fraction of reviews with a rating of one star, and $\mathsf{FIVESTAR}_{pt}$ denotes the fraction of reviews with a rating of five stars. The coefficients γ_1 and γ_2 represent the effects of the respective fractional variables, or specifically, to what degree the one-star and five-star reviews affect sales.

DIFFERENCE MODEL In order to not only measure the correlation between snapshots of sales rank and average rating, we also include a regression model that concerns the total change in a product's sales rank throughout the recorded period. By subtracting the starting point $t = 0$ from any arbitrary time t in Equation 4.1 (disregarding coefficients) the following relation appears:

$$\begin{aligned}
&\ln \mathsf{SALESRANK}_{pt} - \ln \mathsf{SALESRANK}_{p0} = \\
&\ln \mathsf{PRICE}_{pt} - \ln \mathsf{PRICE}_{p0} \\
&+ \ln \mathsf{NUMREVIEWS}_{pt} - \ln \mathsf{NUMREVIEWS}_{p0} \\
&+ \mathsf{PRODUCTTYPE}_{p} - \mathsf{PRODUCTTYPE}_{p} \\
&+ \mathsf{AVGRATING}_{pt} - \mathsf{AVGRATING}_{p0} \\
&+ \mathsf{AVGRATING}_{pt} \times \mathsf{PRODUCTTYPE}_{p} \\
&\qquad - \mathsf{AVGRATING}_{p0} \times \mathsf{PRODUCTTYPE}_{p} + \epsilon
\end{aligned} \tag{4.4}$$

Performing all operations, this gives us:

$$\begin{aligned}
&\Delta \ln \mathsf{SALESRANK}_{pt} = \\
&\Delta \ln \mathsf{PRICE}_{pt} + \Delta \ln \mathsf{NUMREVIEWS}_{pt} \\
&+ \Delta \mathsf{AVGRATING}_{pt} + \Delta \mathsf{AVGRATING}_{pt} \times \mathsf{PRODUCTTYPE}_{p} + \epsilon
\end{aligned} \tag{4.5}$$

We see that the variable $\mathsf{PRODUCTTYPE}_p$ has been cancelled, which means we do not need to control for differences in mean sales rank for the different categories. This also extends to any unobserved fixed effects that were not included in the first model; as long as the fixed effects are assumed constant through time they will be cancelled through the transformation to a difference regression model.

It should also be noted that a more traditional first-difference model measuring daily differences was formulated for this purpose, but it proved to limit the available sales data too much, as daily changes in average rating can be very minute. Sales ranks extracted for longer periods of time are likely necessary for such a model to return any significant results.

We will in Chapter 6 employ Equations 4.1, 4.2, 4.3 and 4.4 to test hypotheses 1 and 2.

4.8.2 *Rating distributions*

To test for the existence of an under-reporting bias and thus an U-shaped rating distribution in hypothesis 3, we have developed a simple logic test to run on the overall sample as well as on the individual products in the different categories. To have the hypothesised U-shape, showing a tendency towards bimodality, we acknowledge that the number of 1 star ratings need to be larger than 2 star rating, and the number of 5 stars need to be larger than 4 stars. Lastly, we want to exclude those with at spike of ratings in the middle of the distribution. Our test is as follows:

$$\text{Bimodality} = (f1 > f2) \wedge ((f2 >= f3) \vee (f3 <= f4)) \wedge (f4 < f5) \quad (4.6)$$

In Equation 4.6, f1, f2, f3, f4 and f5 represents the frequency of the 1, 2, 3, 4, and 5 star rating, respectively. This test does not perform any check on how "deep" the U-shape is if it exists, it simply shows a tendency towards bimodality in the rating distribution. The set of statistical bimodal distributions will be a subset of the one identified by our test. Nevertheless, we see this test as sufficient for our use, as we only wish to demonstrate the tendency towards this type of distribution and the possible differences between the categories.

4.8.3 *Review biases*

In hypothesis 4, we look at the expected self-selection bias as well as over- and undershooting periods, which requires a more mathematical approach. In order to assess the existence of self-selection bias and review undershooting, Li and Hitt (2008, see pages 14-19) developed an enhanced negative exponential model to fit the trend in book reviews over time. Their model is formulated as follows:

$$AVGRATING_{pt} = f_0 + f_1 \exp(-f_2 \cdot T_{pt}) \cos(f_3 \cdot T_{pt}) + u_p + e_{pt} \quad (4.7)$$

In Equation 4.7, $AVGRATING$ again represents the average rating of all reviews posted for product p between the time it was released and time t. T denotes the amount of three-day intervals that have passed since release at time t. u_p represents a similar fixed effect as described in Equation 4.1 and Equation 4.2. e_{pt}, again, denotes a random error. Both these variables are included for symbolic representation, because as we are not able to observe or measure them, they will not be included in the actual calculations.

The coefficients, f_1, f_2 and f_3 illustrate the trend in the average rating over time. Depending on the signs of f_1 and f_2, the model can display an increasing ($f_1 f_2 < 0$), decreasing ($f_1 f_2 > 0$) or no trend ($f_1 f_2 = 0$) over time (Li and Hitt, 2008). The test for self-selection is thus simply an assessment over the signs of f_1 and f_2. A positive sign will imply the existence of self-selection.

The model also includes a cosine term. If the final coefficient, f_3, is zero, we see that the cosine term equals 1, and the model becomes a standard negative exponential model (Li and Hitt, 2008). However, if f_3 is non-zero, the cosine term will produce the revealing ”dip” characteristic to undershooting.

4.8.4 *Review helpfulness*

For the last set of hypotheses, nos. 5 and 6, the dependent variable becomes helpfulness of the review. Previous literature (see e.g. Mudambi and Schuff, 2010; Forman et al., 2008) suggests that helpfulness is moderated by review depth, product type as well as disclosure of reviewer identity and purchase verification. We formulate a regression model to incorporate these variables:

$$\begin{aligned}\mathsf{PCTHELPFUL} &= \alpha + \beta_1\mathsf{RATING} + \beta_2\mathsf{CHARCOUNT}\\ &+ \beta_3\mathsf{PRODUCTTYPE} + \beta_4\mathsf{CHARCOUNT} \times \mathsf{PRODUCTTYPE}\\ &+ \beta_5\mathsf{IDENTITY} + \beta_6\mathsf{VERIFIED} + \epsilon \end{aligned} \tag{4.8}$$

PCTHELPFUL denotes the percentage of votes awarded to the review deeming it helpful. RATING of course refers to the rating awarded by the speficic review, and CHARCOUNT represents the total number of characters typed in for the review. PRODUCTTYPE is a binary (dummy) variable coded to 1 or 0, with 1 indicating an experience product and 0 indicating a search product. The CHARCOUNT × PRODUCTTYPE aims to catch the effect sought in hypothesis 7b: that the review depth is more important for helpfulness of reviews for search products than for experience products. IDENTITY is a binary variable indicating whether or not the reviewer has chosen to disclose his or her identity. The final variable, the dummy VERIFIED, indicates whether or not the review is connected to a verfied purchase.

To test if the hypotheses also hold for our constructed subjectivity variable, a second model, with PRODUCTTYPE replaced by SUBJECTIVITY is formulated:

$$\begin{aligned}\mathsf{PCTHELPFUL} &= \alpha + \beta_1\mathsf{RATING} + \beta_2\mathsf{CHARCOUNT}\\ &+ \beta_3\mathsf{SUBJECTIVITY} + \beta_4\mathsf{CHARCOUNT} \times \mathsf{SUBJECTIVITY}\\ &+ \beta_5\mathsf{IDENTITY} + \beta_6\mathsf{VERIFIED} + \epsilon \end{aligned} \tag{4.9}$$

Both these models use traditional multiple linear regression as the method of analysis, where the basis from Mudambi and Schuff (2010) used a Tobit regression. This hinged on the notion that the review helpfulness is bounded in its extremes, and the sample is censored in nature. Consumers can only vote helpful or not helpful, and there may be self-selection issues in who actually decides to vote. However, for robustness, they also test their model as a multiple linear regression. This gave qualitatively the same findings and significances. As our statistical software does not feature Tobit regression, we will assume that the linear approach will suffice.

Part II

DATA AND RESULTS

In this part we elaborate the data collection and give an overview of our dataset. We present the results of our statistical modeling and analysis, and discuss the findings. Lastly, we offer our concluding remarks, before presenting implications for both managers and researchers.

5 DATA COLLECTION AND DESCRIPTIVES

This chapter outlines the data collection period and presents important aspects and characteristics of our dataset not necessarily related to the hypotheses. The data was collected from Amazon between April 1st and May 8th 2014 using the techniques described in Chapter 4. The reviews were collected at the end of this period, which means that the data set includes all reviews written about the products before the 8th of May. Sales ranks were logged every 24 hours, with two exceptions due to technical problems.

5.1 PRODUCTS

In the product selection phase, we focused on creating a set of unique product ASINs and thus unique products. However, during the data collection period, it became obvious that even with unique ASINs, some products on Amazon share review sets. This phenomenon occurs for instance in the "movies" category, because different releases of the same movie on different mediums (DVD, blu-ray, Amazon Instant) are assigned unique ASINs, but their reviews are shared. Similarly, versions of a USB drive with different storage capacities are given unique ASINs and sales rank, but share reviews with other versions. Therefore, we have divided our products into two main groups: the primary and the secondary group. The primary group consists of products with unique ASINs and unique review sets. The secondary group has unique ASINs and therefore a unique associated sales rank, but share review sets with products in the primary group.

There are no formal connections between the products in these groups, and the phenomenon is not trivial to detect automatically. We used a combination of number of reviews, sum of number of words and sum of number of characters in all reviews for a product as a unique identifier for a review set. This method has excellent precision as the number of reviews increases, but is still fairly accurate even at a small number of reviews. This method should detect all duplicates, and worst case is that is classifies something as duplicate that is not. To reduce this risk, the review sets with less than 10 reviews that were identified as not unique were checked manually for uniqueness. Using this method a total of 312 duplicate review sets were identified, and the products classified as secondary. In most cases, only the primary set is used. Table 5.1 shows the number of products within each selection and classification.

SELECTION	PRIMARY	SECONDARY	ALL
Top 100	2,653	299	2,952
Random	1,958	13	1,971
Total	4,611	312	4,923

Table 5.1: Number of primary and secondary products in both selections.

Our ambition was to collect data for the 100 best selling products in addition to 100 randomly selected products within each of our 30 categories, as described in Chapter 4. However, in three of the categories (hard drives, ink and toner, USB drives) it proved difficult to obtain search hits with our random words. This could be because of a smaller selection of products, or that there are particularly few words that are in some way associated with these categories. Regardless the reason, after running our random product generator for more than a full day with only a few hits, we chose not to pursue random products within these categories any further.

At the end of our data collection phase, we discovered an error in our code used for checking for duplicate products in the generated lists of random products. The consequence of this error was that the lists contained several duplicates that needed to be removed from our final set of products, making them shorter than our target of 100. After removing these, as well as products without reviews at the end of the data collection period, we arrive at our final set of 1,971 randomly selected products, where 1,958 are primary. More detailed information about the excluded random products is given in Table A.2 in the appendix.

The selection of top 100 products is the ”best selling” list in each of the respective categories as it appeared on Amazon.com on 19th of March 2014. We chose to include all categories regardless of whether we were able to generate random lists or not. 7 products were removed by Amazon during our data collection period and were thus not included in our final set. After also excluding products without reviews and duplicates, our final set includes 2,952 unique ASINs where 2,653 had unique review set in the top 100 selection. Please refer to Table A.3 in the appendix for detailed information about the excluded top 100 products.

The number of primary and secondary products for each category is given in Table 5.2.

CATEGORY	PRIMARY	SECONDARY	ALL
Board Games	186	1	187
Books	199	0	199
Bowls	136	20	156
Can Openers	139	11	150
Candy	180	5	185
Car Electronics	167	5	172
Clothing	190	0	190
Copy Paper	121	6	127
Desktop Computers	147	4	151
Digital Cameras	163	20	183
Dog Food	181	6	187
Envelopes	157	2	159
Guitars	146	21	167
Hard Drives*	69	30	99
Hardware	188	1	189
Hobby Fabric	145	5	150
Ink and Toner*	91	9	100
Jewelry	178	0	178
Ladders	142	11	153
Movies	148	49	197
Perfumes	190	1	191
Restroom Fixtures	111	18	129
Screws	126	9	135
Shoes	189	0	189
Software	157	11	168
Test & Measure	176	1	177
USB Drives*	52	48	100
Video Games	171	15	186
Vitamins	189	4	193
Watches	169	0	169
Total	4,611	312	4,923

Table 5.2: Number of primary and secondary products for all categories.
**No random product list generated.*

5.2 REVIEWS

Following the distinction between primary and secondary data sets in products, our set of reviews split in the same way. The primary data set is unique reviews, while the secondary are duplicates of the ones in the primary set, but attached to a non-duplicate ASIN with a distinct sales rank. Table 5.3 gives an overview of the number of reviews included in each set, split on top 100 and random products.

SELECTION	PRIMARY	SECONDARY	ALL
Top 100	986,344	340,655	1,326,999
Random	161,144	1,051	162,195
Total	1,147,488	341,706	1,489,194

Table 5.3: Number of primary and secondary reviews for top 100 and random selection.

Our data set includes more than 6 times as many primary reviews from the top 100 selection than the random selection. This is partly because of more products in our top 100 selection as shown in Table 5.3, but mainly because high sales are associated with more reviews - as we will demonstrate in Section 6.1.

Table 5.4 gives an overview of number of reviews in the primary and secondary sets for each category. We see that books has by far the most primary reviews constituting almost 19% of our review set. This is not surprising, as it is where Amazons started out and is still considered part of its core business. Amazon reportedly has a 65% of the US online (digital and print) books market (Kohn, 2014), indicating that lots of traffic and reviews is to be expected - especially for the best selling books. However, the table also shows that the dataset includes a considerable amount of reviews for less studied and less obvious categories like can openers (22,068), vitamins (73,816) and jewelry (21,530). For more detailed information about the number of reviews in the top 100 and random selection, please see Table A.4 and Table A.5 in the appendix.

CATEGORY	PRIMARY	SECONDARY	ALL
Board Games	34,407	395	34,802
Books	216,361	0	216,361
Bowls	4,116	2,832	6,948
Can Openers	22,068	4,984	27,052
Candy	13,793	1,973	15,766
Car Electronics	93,027	19,822	112,849
Clothing	68,466	0	68,466
Copy Paper	3,352	743	4,095
Desktop Computers	5,594	390	5,984
Digital Cameras	41,132	8,601	49,733
Dog Food	15,697	3,071	18,768
Envelopes	3,189	97	3,286
Guitars	9,239	5,293	14,532
Hardware	40,094	293	40,387
Hard Drives*	19,375	24,304	43,679
Hobby Fabric	1,838	265	2,103
Ink and Toner*	27,581	4,671	32,252
Jewelry	21,530	0	21,530
Ladders	12,960	3,390	16,350
Movies	103,586	127,393	230,979
Perfumes	21,724	358	22,082
Restroom Fixtures	2,735	968	3,703
Screws	1,919	120	2,039
Shoes	59,652	0	59,652
Software	41,169	10,400	51,569
Test & Measure	18,186	110	18,296
USB Drives*	48,272	97,432	145,704
Video Games	90,018	19,724	109,742
Vitamins	73,816	5,017	78,833
Watches	31,652	0	31,652
Total	1,147,488	341,706	1,489,194

Table 5.4: Number of primary and secondary reviews for all categories.
**No random product list generated.*

Table 5.4 shows that the number of reviews varies greatly across the different categories. However, it is difficult compare these numbers and get any insight to the typical review activity for product within the category since the number of products vary as well. To get a better sense of this, we have included the mean number of reviews per product and standard deviation for the random selection in Figure 5.1. We consider this to be the most comparable selection, as the top 100 selections across categories vary immensely in relative popularity.

We still observe that the variation between the categories is large, all the way from movies at the top (351.42 reviews per products), to envelopes at the bottom (4.54 reviews per product). We also see that the standard deviation in almost all cases are greater than the mean, indicating that the variation within the categories is vast.

For more details on the average number of reviews for both the top 100 and random selection, please refer to Table A.7 and Table A.6 in the appendix.

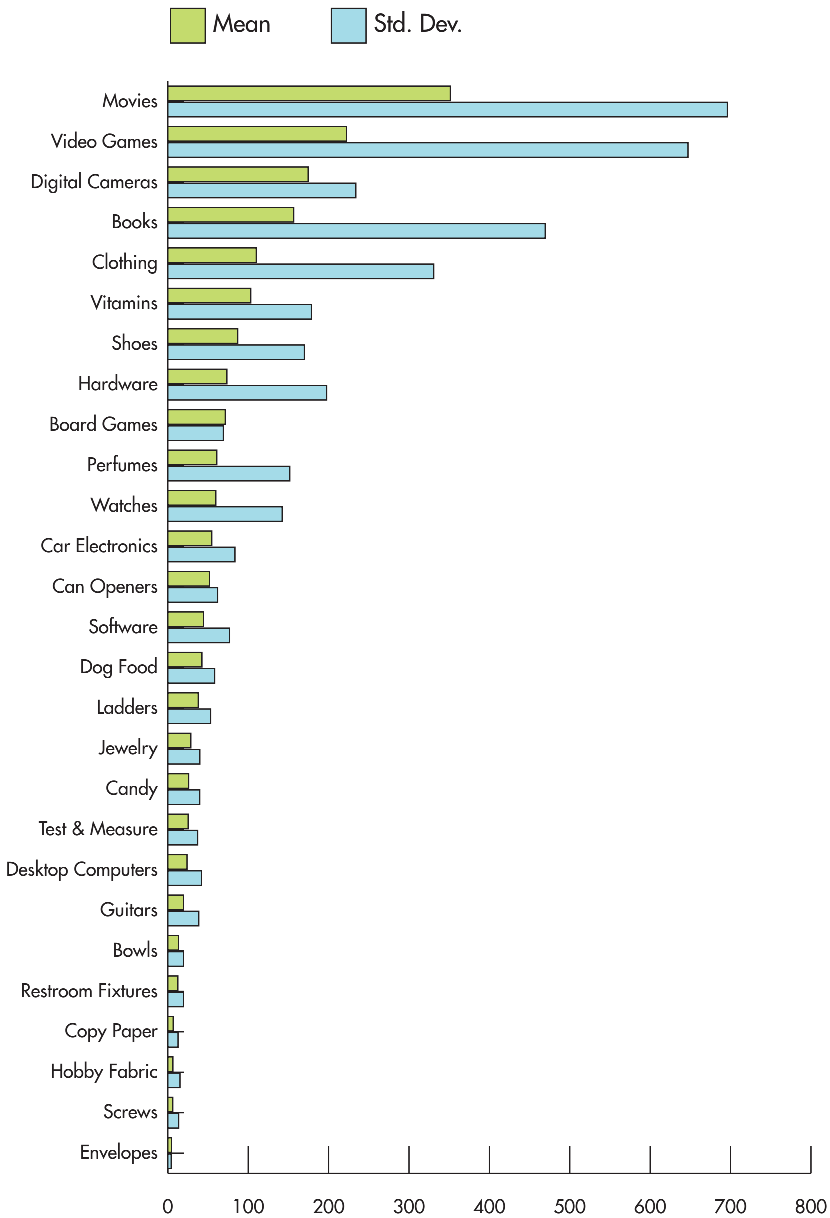

Figure 5.1: Mean number of reviews per product for random selection.

5.3 REVIEW LENGTH

To measure the length of the text in the reviews, we calculated the number of typed characters. Figure 5.2 shows the distribution of number of characters for all primary reviews. The longest 1% has been excluded for purposes of readability.

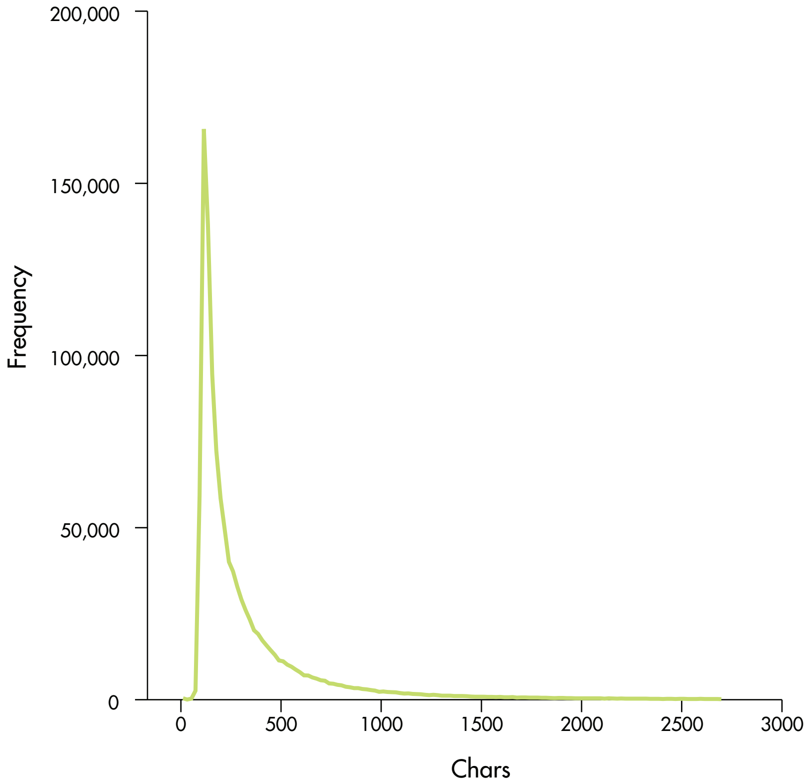

Figure 5.2: Distribution of number of characters for all reviews, longest 1% excluded.

Overall, 99% of the reviews have less than 2,705 characters, and 90% have less than 748. However, the longest have more than 32,000. We observe this type of distribution for all the categories individually as well, although the length of the tail varies. We therefore first take a look at the mode, the top point of the curve in Figure 5.3. The mode is found to be comparable for all categories and in most cases between 111-121 characters. Detailed numbers can also be found Table A.8 in the appendix.

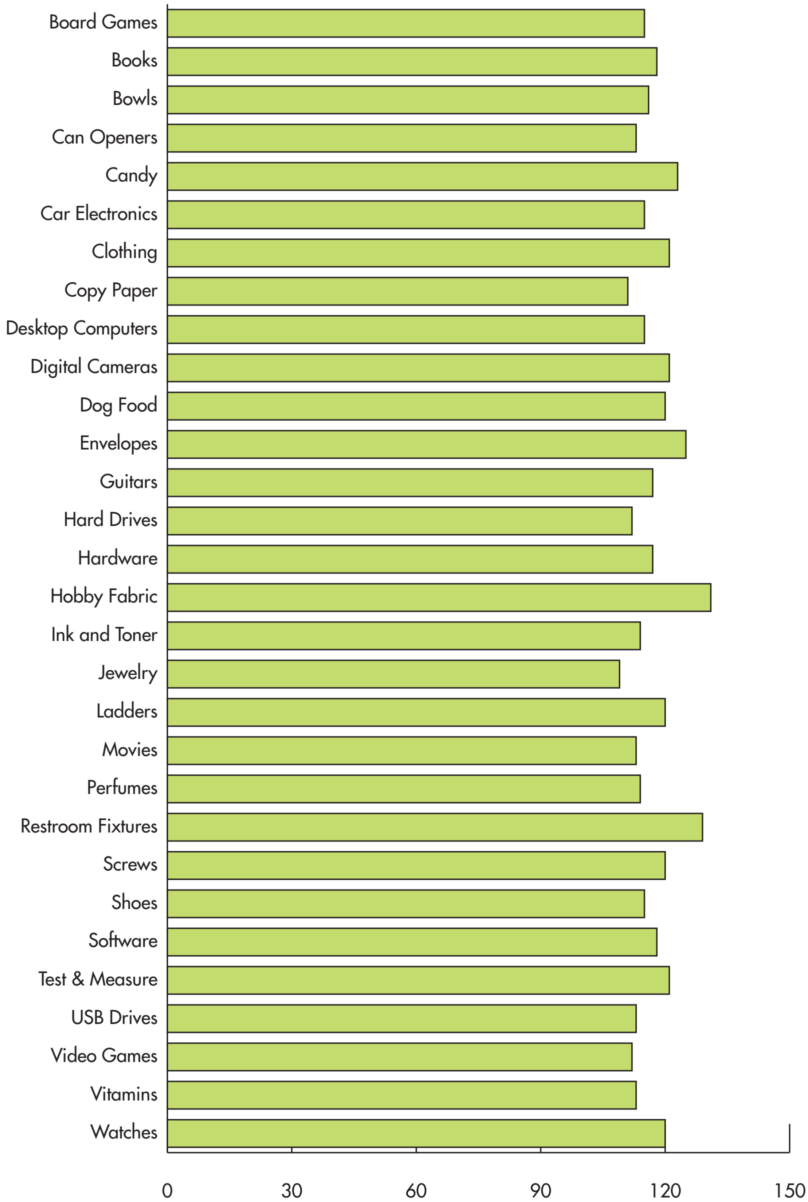

Figure 5.3: Mode number of characters for all categories.

The biggest difference in respect to review length between the categories seems to be the amount of presence of ultra-long reviews. Figure 5.4 shows that for the longest 1% in our set , more than 75% are found in the 4 categories books, digital cameras, movies and video games.

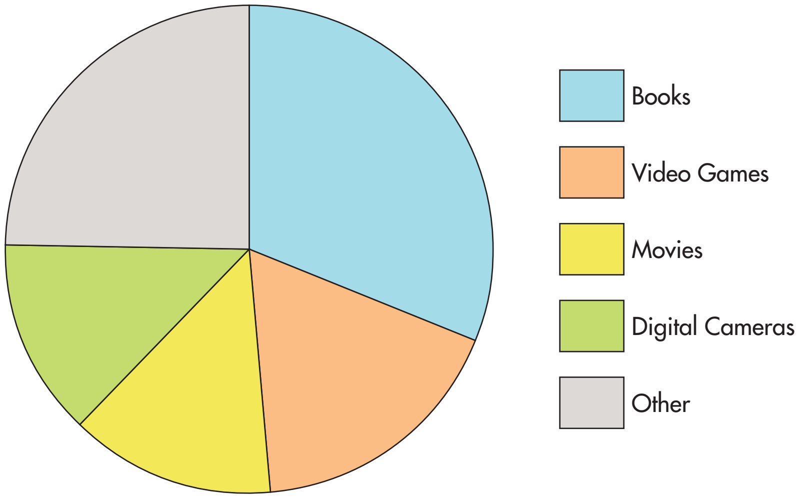

Figure 5.4: Category split for 1% longest reviews.

Even though the distribution shape and mode is similar for the categories, the mean length varies. Some of this is because of the difference in presence of ultra-long reviews, but there are differecenses in the slope as well. The mean length is given in Figure 5.5

The detailed number on mean length is included in the appendix in Table A.8.

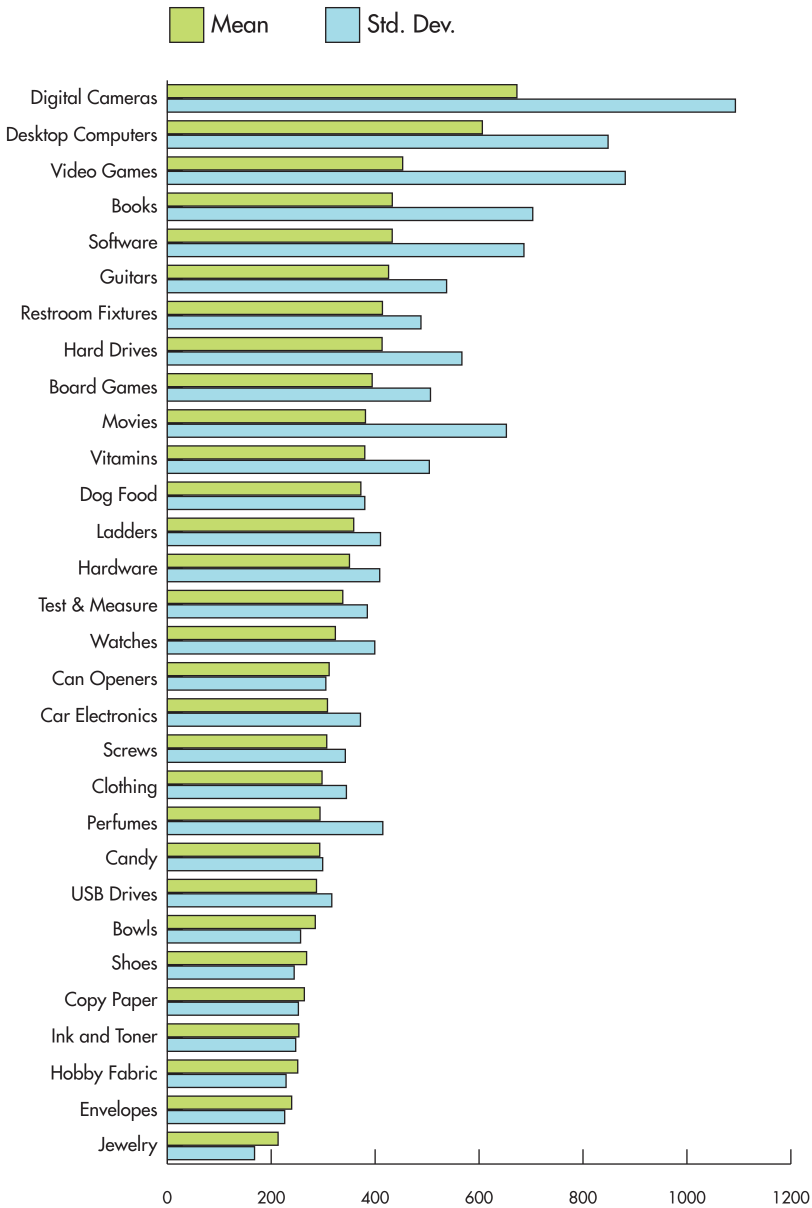

Figure 5.5: Mean length of reviews.

5.4 RATING

Looking at the overall distribution of ratings, the 5 star rating is by far the most used. 64.9% of all ratings are 5-star ratings, indicating that reviews in general are overwhelmingly positive. We also note that 2 is the least used rating, with only 4.3% of the overall ratings. The distribution is displayed in Figure 5.6.

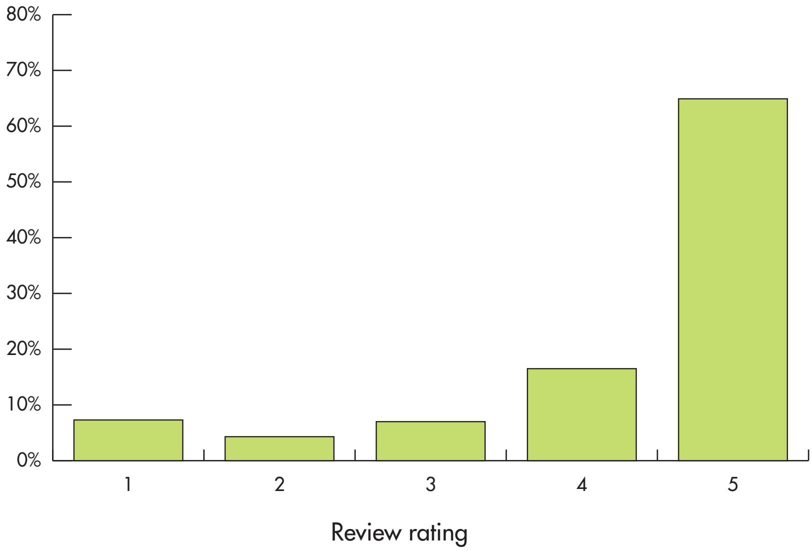

Figure 5.6: The distribution of ratings for all reveiws.

The mean rating for all reviews is 4.27, with a standard deviation of 1.215. However, the mean varies greatly from category to category, as demonstrated in Figure 5.7. Software seems to stand out as the only category with a mean below 4.0, whilst board games enjoys the highest mean rating. Most lie between 4.1 and 4.4 with a standard deviation between 1.05 and 1.30. Please see Table A.10 for detailed numbers with standard deviation for all categories.

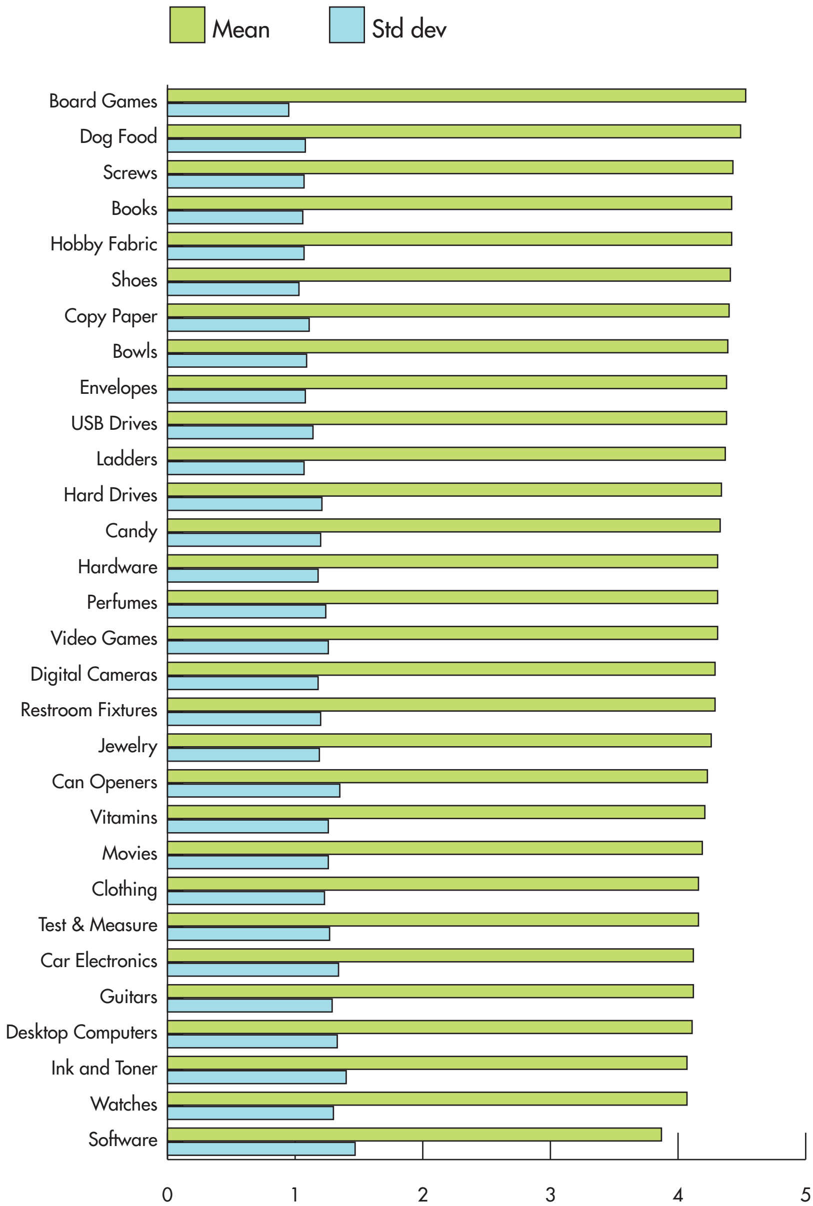

Figure 5.7: Mean rating and standard deviation for all categories.

5.5 SUBJECTIVITY SCALE

Using the method and the OpinionFinder tool described in Section 4.7, we classified all our categories according to the degree of subjectivity in the reviews. The results for both methods are found in Figure 5.8, sorted by the rule-based results. The number in the figure is the fraction of identified subjective sentences divided by the number of objective and subjective sentences (those classified as unknown in the rule-based approach are not included). The rule-based approach uses the top axis, while model-based uses the bottom.

We see that the result seems to fit fairly well with an intuitive evaluation. Books and movies are classified as the most subjective and internal hard drives and copy paper are the least subjective. The only fairly unexpected result for us was that ladders, which we would expect to be mostly objectively evaluated, are deemed one of the most subjective on both approaches. We emphasize that the score makes little sense on its own, but is to be regarded as a relative score. We also see that there are some differences in the results from the two methods, and that the model-based approach gives generally higher scores. It is the sorted order that is of greatest importance, and although there are some differences, the results are mostly comparable. However, some categories like USB drives and Can Openers are rated higher with the model-based than with the rule-based classifier.

For detailed results of the subjectivity analysis, please see Table A.11 and Table A.12 in the appendix.

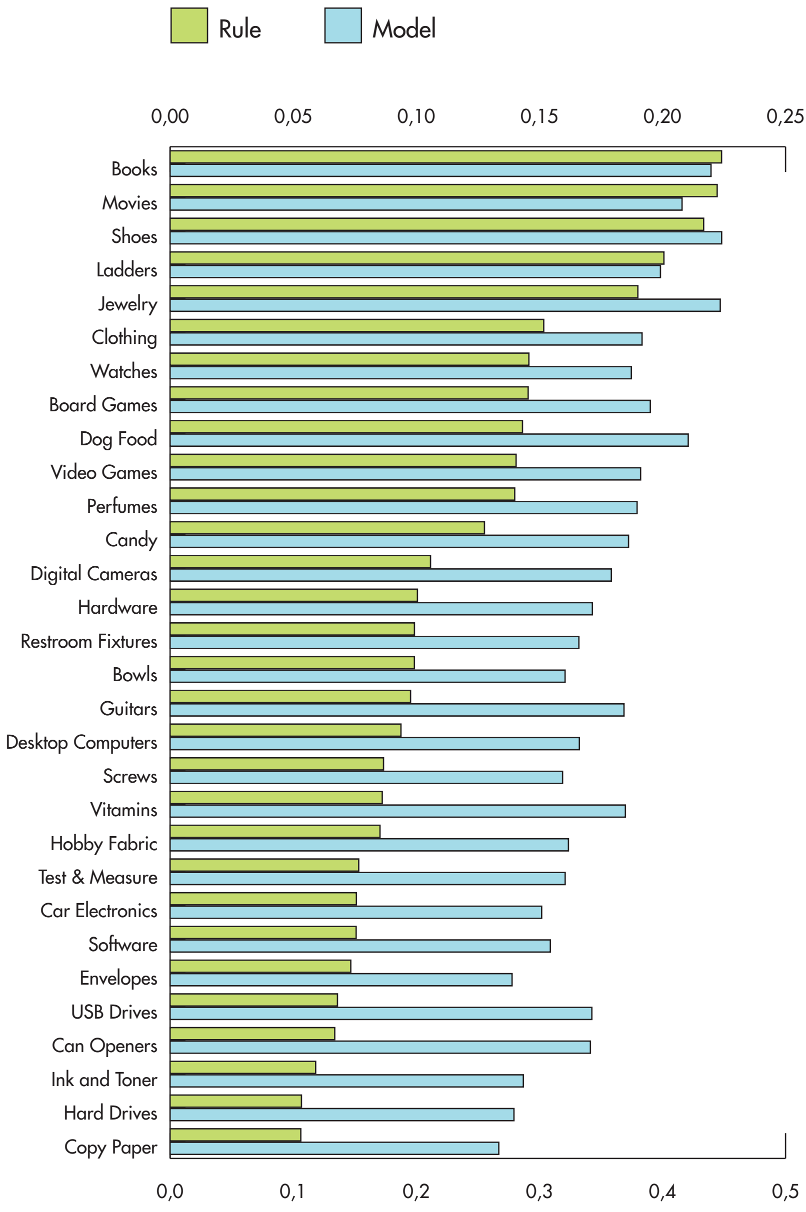

Figure 5.8: Categories by share of subjective review content. Rule-based on top axis and model-based on bottom. Entries are sorted in descending order by rule-based variable.

6 RESULTS AND FINDINGS

In this chapter present the results of our statistical analyses. The results are split into the same three sections as for the hypotheses. We begin by looking at the effects of reviews on sales, before analyzing the dataset for review biases. We then test our hypotheses regarding the helpfuness of reviews.

Throughout this chapter we use a set of abbreviations for some of our variables. An overview of the used abbreviations, as well as their meaning, is presented in Table 6.1. Further, R^2 values in the tables denote both R^2 and adjusted R^2, as these values have been the same for all our regressions.

ABBREVIATION	VARIABLE
AR	Average rating
SR	Sales rank
Type	Product type, search or experience
Subj.(M)	Subjectivity, model-based score
Subj.(R)	Subjectivity, rule-based score

Table 6.1: Variable abbreviations used in results.

6.1 EFFECT OF REVIEWS ON SALES

This section details the testing of hypotheses 1 and 2, as listed in Table 6.2. We begin by testing the suppositions with our static regression model, before delving deeper with the difference regression model.

Any missing values for these tests have not been included in the regression. This mainly concerns missing sales rank data, as well as products that have no reviews. We have chosen not to interpolate missing sales rank values, as the dataset was considered to contain a sufficient number of observations. For the same reason, we have not included products with missing reviews. All tests in this section use the secondary dataset, as defined in Chapter 5.

#	HYPOTHESIS
1a	An increase in average rating on a site is associated with increased sales on that site.
b	The association between average ratings and sales is stronger for non-hit products than for hit products.
c	The association between average ratings and sales is stronger for experience products than for search products.
d	The association between average ratings and sales is stronger for products that tend to be subjectively evaluated products than for those who tend to be objectively evaluated.
2	A comparatively high number of reviews on a site is associated with comparatively higher sales on that site.

Table 6.2: Hypotheses regarding the effects of reviews on sales.

6.1.1 *Static regression model*

The results of the regression using our static model are shown in Table 6.3 and Table 6.4. Limiting the variables to only price, number of reviews, as well as the average rating, we see clear and significant effects of all three. Since strong sales lead to a lower sales rank, variables connected to stronger sales will show a negative sign, whereas variables detrimental to sales will show a positive sign. As one would assume, we see that price is negatively associated with the sales rank, i.e. higher (lower) prices are correlated with lower (higher) sales. Further, the number of reviews shows a strong correlation with the sales rank, higher volumes of reviews being associated with higher sales. This suggests hypothesis 2 is correct. However, more detailed analyses would be necessary in order to determine the causality. From this simple regression, one cannot say whether a higher density of reviews per purchase leads to more sales, or if the larger number of reviews simply stem from more sales.

Finally, we see that the average rating, as well as the star fraction variables show the expected signs, and all are highly significant. In Table 6.3, the overall average rating shows correlation with higher sales (lower sales rank). This translates to products higher up on the best sellers lists having better average ratings, supporting hypothesis 1A. Further, Table 6.4 shows that the fraction of 1-star reviews has a negative association with sales, the fraction of 5-star conversely showing a positive association with sales. However, the impact of 5-star reviews seems to be stronger than that of 1-star reviews, which contradicts earlier findings by Chevalier and Mayzlin (2006). Both regressions show support for hypothesis 1A.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	11.919	.072	-	164.429	.000
ln (price)	.089	.004	.063	20.216	.000
ln (no. of reviews)	-.668	.005	-.448	-143.200	.000
Average rating	-.286	.016	-.056	-17.984	.000
DEP: $\ln salesrank$, $R^2 = 0.206$, $N = 81{,}245$					

Table 6.3: Regression results: salesrank and average rating.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	11.133	.050		224.073	.000
ln (price)	.088	.004	.063	20.087	.000
ln (no. of reviews)	-.668	.005	-.448	-143.197	.000
Frac. 1-star	.305	.113	.011	2.695	.007
Frac. 5-star	-.704	.057	-.050	-12.287	.000
DEP: $\ln salesrank$, $R^2 = 0.206$, $N = 81{,}245$					

Table 6.4: Regression results: salesrank and star fractions.

When introducing the interaction terms of product category and rating as well as subjectivity and rating, we observe that product categories have significant differences in mean salesrank. As explained in Section 4.3.2, categories like books and movies that are top-level categories, will have their top 100 products occupy salesranks 1-100, whereas a lower level category such as screws or mailing envelopes will have its 100 best selling products placed further down the scale within some larger top-level category. Thus, it would make sense to control for category, in order to account for the bias in mean salesrank. However, introducing product type or subjectivity as a control variable results in severe multicollinearity issues, making the coefficients volatile and unreliable. To mitigate these issues, we center the average rating factor in the interaction terms around the mean average rating. This lessens some of the effects, but may still leave some of the coefficients in Table 6.5 and Table 6.6 unreliable.

Looking at Table 6.5, we see the hypothesized effects. Price and reviews maintain their expected signs, as does the average rating. The product type control shows how the mean salesranks differ, indicating in this case that the experience products in our sample hold a higher mean sales rank. The most interesting variable in this regression, however, is the interaction term between the average rating and product type. Hypothesis 1C states that experience products should see a greater effect from ratings than their search counterparts. The binary nature of the type variable means that the term only comes into play for experience products, meaning the experience products in our set with a given average rating will see a higher salesrank than a search product with the same rating. This supports hypothesis 1C, that product types moderate the effect of reviews on sales.

Performing similar regressions with our two subjectivity variables produce comparable results. The average rating terms see opposite signs, but for the rule-based subjectivity variable this effect is very small and not statistically significant. For the model-based subjectivity variable, the average rating has severe collinearity with the interaction term, which makes it hard to accurately say which sign is the correct. Removing the average rating produces a negative sign for the interaction term without any collinearity, but may be prone to omitted variable bias. The results in Table 6.6 are inconclusive, but give some support to hypothesis 1D, that subjectively evaluated products see larger effects from reviews than objectively evaluated products.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	12.792	0.113	-	113.257	.000
ln (price)	0.129	0.007	0.076	19.051	.000
ln (no. of reviews)	-0.532	0.006	-0.395	-95.35	.000
Type	-1.219	0.018	-0.278	-66.139	.000
Average rating	-0.44	0.025	-0.093	-17.416	.000
AR×type	-0.383	0.037	-0.055	-10.217	.000
DEP: $\ln salesrank$, $R^2 = 0.330$, N = 43,725					

Table 6.5: Regression results: salesrank and rating by product category. The AR variable in the interaction term has been centered around its mean.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	10.923	0.181	-	60.449	.000
ln (price)	0.092	0.004	0.066	21.131	.000
ln (no. of reviews)	-0.633	0.005	-0.426	-134.407	.000
Average rating	0.038	0.041	0.007	0.911	.362
Subj.(R)	-4.522	0.153	-0.095	-29.6	.000
AR×Subj.(R)	-2.586	0.345	-0.062	-7.497	.000
DEP: $\ln salesrank$, $R^2 = 0.214$, N = 81,149					
(Constant)	8.541	0.541	-	15.789	.000
ln (price)	0.105	0.004	0.075	24.222	.000
ln (no. of reviews)	-0.603	0.005	-0.406	-128.27	.000
Average rating*	1.095	0.124	0.217	8.812	.000
Subj.(M)	-7.961	0.156	-0.163	-51.11	.000
AR×Subj.(M)*	-3.762	0.351	-0.264	-10.727	.000
DEP: $\ln salesrank$, $R^2 = 0.231$, N = 81,149					

Table 6.6: Regression results: salesrank and rating by subjectivity. Top: Rule-based subjectivity. Bottom: Model-based subjectivity. AR has been centered for the interaction terms.
**Variable experiences strong collinearity*

So far, we see that the static model supports Hypotheses 1a and 1c. In testing 1d, the subjectivity variables introduce some multicollinearity issues, but give initial support to the notion that subjectively evaluated products see larger effects from ratings.

6.1.2 *Difference regression model*

Moving over to the difference regression model, we no longer measure the absolute values of sales ranks, but rather the change from the initial sales rank. This formulation allows us to cancel out the effect of biased mean in categories, as well as any other unobserved fixed effects. In addition, we will more accurately be able to ascribe the change in sales rank from a change in ratings, whereas the first model simply predicted a correlation between high sales and high ratings. Finally, we can also plug in the sales rank itself as a predictor variable, since our dependent variable now is $\Delta \ln salesrank$. This allows us to test whether any of the observed effects are stronger in certain segments of popularity.

Looking at the basic difference model, we test hypothesis 1b by regressing over our top 100 and random product sets. This produces the output in Table 6.7. We note that the change in price still retains its positive sign, indicating that *growth* in price leads to lower sales. This effect is statistically insignificant for the random set, however.

Contrary to the static model, we see that the growth in number of reviews now seem associated with lower sales. Although seemingly contradicting our previous findings, the change in sign can be reasonably explained with two things. First, the majority of products see a negative trend in sales, with over 60% of all the products recording a lower sales rank at the end of our data collection than at its commencement. Second, most of these products will, quite naturally, see an increase in reviews as time passes and more people review them. Thus, the dynamic model contributes little to the understanding of the causality between the volume of reviews and sales.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	.195	.003		57.199	.000
Δln (price)	.100	.008	.055	12.743	.000
Δln(no. of reviews)	.784	.018	.191	43.897	.000
Δaverage rating	.076	.044	.008	1.730	.084
DEP: $\Delta \ln salesrank$, $R^2 = 0.039$, N = 50,826					
(Constant)	.023	.004		5.907	.000
Δln (price)	.011	.015	.004	.694	.488
Δln (no. of reviews)	.567	.049	.068	11.588	.000
Δaverage rating	-.408	.076	-.031	-5.370	.000
DEP: $\Delta \ln salesrank$, $R^2 = 0.006$, N = 29,136					

Table 6.7: Regression results: change in salesrank on popularity. Top: Top 100 products. Bottom: Random products.

We also see a relatively low fit for the model, with R^2 values well below 0.1. These can be elevated by controlling for products' intial salesranks as well as with a binary variable indicating overall growth or decline in sales. This brings up the R^2 value to around .500. However, since the signs and magnitudes of our focus variables do not see any significant changes, we omit these variables for the sake of simplicity.

Looking at the change in average rating, we see differing signs for the two sets, with the top 100 showing a weak effect with the "wrong" sign, albeit with less statistical significance ($p < 0.1$). The change in rating for the random set, however, shows a relatively strong and statistically siginificant ($p < 0.01$) effect, with a negative sign. This suggests that the effect of ratings is stronger for less popular products, supporting hypothesis 1b, although further tests are necessary to conclusively determine the effect.

Exploiting the fact that the variable $\ln salesrank$ can now be used in the set of predictor variables, we use it in an interaction term with the average rating to see if the effect of the ratings increase with increased sales rank (less popular products). Table 6.8 summarizes the results, showing a negative sign with statistical significance for the interaction term. The negative sign means that an increase in the term leads to higher sales. This indicates that a change in rating at a given level of sales rank will have a smaller effect than the same change in rating at a higher level of sales rank (less popular).

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	.129	.003		49.197	.000
Δln (price)	.087	.007	.044	12.561	.000
Δln(no. of reviews)	.826	.016	.176	50.500	.000
Δavg. rating$\times$ln(SR)	-.028	.004	-.022	-6.416	.000
DEP: $\Delta \ln salesrank$, $R^2 = 0.033$, $N = 79{,}961$					

Table 6.8: Regression results: change in salesrank by rating and popularity. *SR denotes salesrank*

Extending the previous argument, we perform several groupwise regressions for different rating variables with interaction from product categories and subjectivity. Specifically, we do regressions to check the different magnitudes for the coefficients for all products, products with a sales rank greater than or equal to 100, 1000, 10,000 as well as 100,000. The results of these regressions are summarized in Table 6.9. Regressing over all products (sales rank > 1, since $\ln 1 = 0$, and will produce errors in the data), we see similarities to the results in Table 6.8. With the best selling products included in the set, the effects seem inconclusive. This suggests that there are other factors in play for these products, with reviews staking a smaller claim of the total purchase decision making process. This is in accordance to the theory presented when formulating hypothesis 1B, which contends that there is a relative abundance of available information about the most popular products. There could also be other phenomena impacting the purchase decisions for these products, such as fashion and hype, or external marketing campaigns. Combined, these other phenomena may contribute to diminishing the importance of consumer generated reviews.

As such, it makes sense to see an increase in the effect of reviews as we exclude more and more of the best selling products. Indeed, we see that the effect of ΔAR grows for every step as we exclude more of the top-selling products. This supports hypothesis 1b, the effect of ratings is larger for less popular products. Likewise, we see that the coefficients for the interaction terms between the change in average rating and category specific variables all increase in magnitude as we move lower in product popularity, behaving in accordance with the two-dimensional aspect of the review impact continuum. In addition, the standardized coefficients of the interaction terms are almost exclusively larger than for the change in rating alone, as shown graphically in Figure 6.1. These findings support hypothesis 1b, 1c as well as 1d, implying that the effect of reviews is both larger for less popular products, as well as for experience (or subjectively evaluated) products.

VARIABLE	>1	⩾100	⩾1000	⩾10000	⩾100000
ΔAR	-0.047	-0.180**	-0.194**	-0.356**	-1.058**
ΔAR×type	0.072	-0.436**	-0.764**	-0.840**	-6.021*
ΔAR×subj.(M)	-0.123	-0.612**	-0.734**	-1.199**	-3.258**
ΔAR×subj.(R)	0.103	-1.518**	-1.963**	-3.040**	-12.485**
ΔAR	-0.004	-0.017	-0.018	-0.033	-0.115
ΔAR×type	0.005	-0.026	-0.039	-0.036	-0.050
ΔAR×subj.(M)	-0.004	-0.02	-0.023	-0.037	-0.115
ΔAR×subj.(R)	0.001	-0.019	-0.022	-0.062	-0.124
DEP: $\Delta \ln \mathsf{salesrank}$					

Table 6.9: Regression results: coefficients and standardized coefficients for effect of rating, rating×category and rating×subjectivity for different segments of salesranks. Standardized coefficients in the lower table. *AR denotes average rating.*
* - $p < 0.05$, ** - $p < 0.01$

6.1.3 *Conclusion*

In summary, all hypotheses in this section finds support. Using the static regression model, we test for and find an association between average rating and sales. We also see that the category moderates this effect, as well as subjectivity, although collinearity issues cloud the conclusiveness

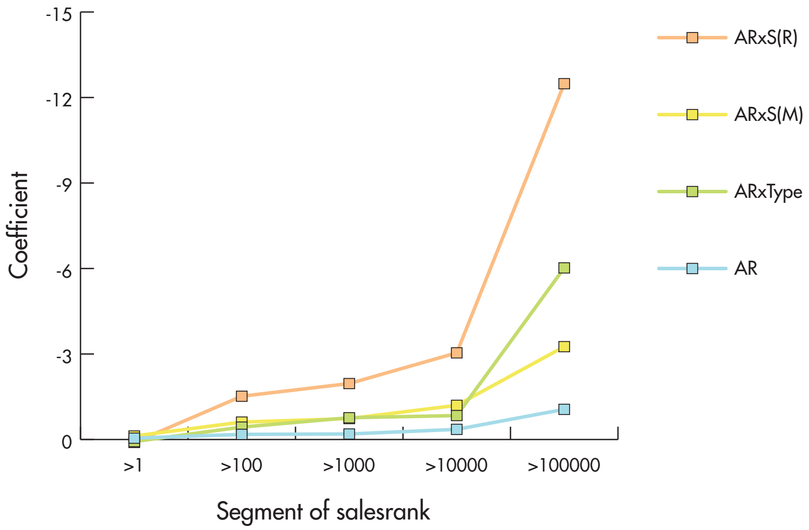

Figure 6.1: Coefficients for AR, AR×type and AR×subjectivity.

somewhat. The number of reviews is found to strongly correlate with sales. Using the difference regression model, we find further support for the effect of ratings on sales. We test the moderating factor of popularity and find that less popular products see a larger effect of reviews. Finally, we strengthen both the produuct type-specific, subjectivity-specific and popularity-specific hypotheses by a group-wise regression through different segments of sales ranks.

6.1.4 *Causality*

The regression analysis methods we have used to determine the association between changes in rating and sales used in this section do not provide any specific evidence of the causality of such a relation. Several other researchers however like Chevalier and Mayzlin (2006) and Luca (2011) have dealt with the issue of causality more in depth. Specifically Luca (2011), cleverly uses the thresholds for rounding to the nearest half star (as it is presented to the user on Yelp) to determine the causality issue. By looking at restaurants who have average ratings close to this threshold and the responses when they surpass it or gets under it, he concludes that the rating causes the revenue change. We hold it as probable that the same is the case for Amazon, which is also the conclusion of Chevalier and Mayzlin (2006), but do not present any evidence of our own.

6.2 RATING DISTRIBUTIONS AND REVIEW BIASES

In this section we look at hypotheses 3 and 4, as shown in Table 6.10. For hypothesis 3, we use our logic test as shown in Equation 4.6. Tests for hypothesis 4 employs the negative exponential model as formulated by Li and Hitt (2008), and in Equation 4.7.

#	HYPOTHESIS
3	The distribution of ratings for a product tends to be bimodal, with the low and high end of the scale as local modes.
4a	The average rating of a product tends to decrease over time before stabilizing at a long term value lower than the initial value.
b	Some products with a difference in initial average rating and long term average rating go through a ”undershooting” period after the initial period where the rating is lower than the long term average.
c	The undershooting effect is stronger for experience products than for search products.

Table 6.10: Hypotheses regarding distributions and biases.

6.2.1 *Rating distributions*

Hypothesis 3 states that the distribution of ratings tends to be bimodal, where 1 and 5 stars are minor and major modes, respectfully. To test for this, we constructed a simple set of requirements that needed to be fulfilled in order to show tendencies of bimodality. Remembering Equation 4.6:

$$\mathtt{Bimodality} = (\mathtt{f1} > \mathtt{f2}) \wedge ((\mathtt{f2} >= \mathtt{f3}) \vee (\mathtt{f3} <= \mathtt{f4})) \wedge (\mathtt{f4} < \mathtt{f5}) \quad (6.1)$$

In Equation 6.1, f1, f2, f3, f4 and f5 represents the frequency of the 1, 2, 3, 4, and 5 star rating, respectively. We have already briefly shown in Section 5.4 that the overall distribution of ratings exhibits signs of bimodality. Recall Figure 5.6, here with fitted U-shape.

However, to test our hypothesis, it needs to hold on a product level. We therefore test how many products in our dataset that show signs of bimodality. We limit our dataset to products with more than 20 reviews, which is similar to the limit set it Hu et al. (2006). Of the then 3,044 products that remain in our dataset, 1,814 show signs of being bimodal

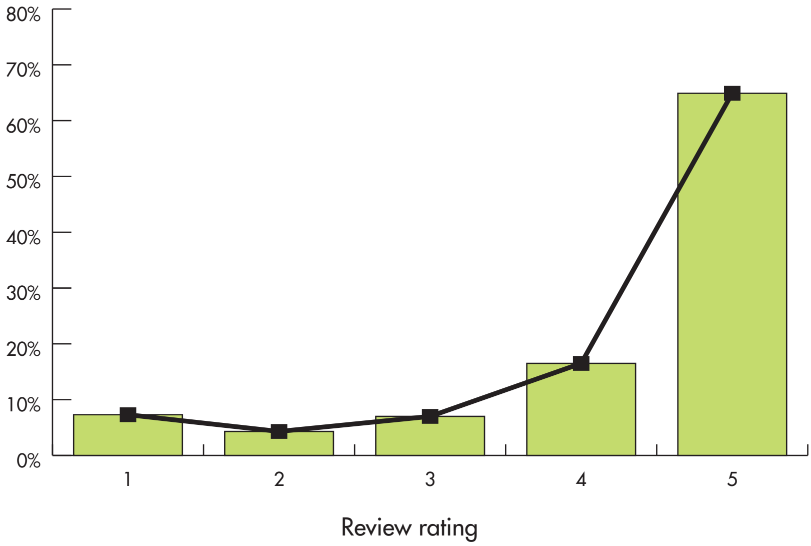

Figure 6.2: The distribution of ratings for all reveiws with U-curve.

in our simple test. This converts to a 59.6% share. This is slightly higher than the findings of Hu et al. (2006), who found that about 53% of products reviewed on Amazon have bimodal rating distributions. The most likely explanation for this is the more strict statistical approach of Hu. The review sets with statistical significant bimodality will be a subset of the ones identified by our test, thus it is to be expected that our results are slightly higher. Hu et al. (2006) use a DIP test (Hartigan and Hartigan, 1985), while we use a simpler logic test. An additional explanation could be that the difference stems from the difference in selection of categories. Further analysis shows for instance that books (1 of Hu's 3 categories) converges around a 54% bimodality, very similar to Hu's 53 %.

For robustness, we tested with different limits of number of ratings, to see of it affects the tendency of bimodality. The results in Figure 6.3 shows that the share increases as the limit increases, but seems to converge after with set with more than 110 reviews. The overall share is then at 70.0%.

This tells us that as the number of ratings increases, products rating distributions becomes increasingly bimodal. We also split the analysis on category. Figure 6.4 shows that there is a vast range between shoes at 27 % bimodality tendencies to ink and toner at 94 %.

We see that most digital products seem to have high shares of bimodal rating distributions, while simpler analog products seem to have lower. We can only speculate in these differences, but it could be that the share of bimodal distributions is correlated with the chance of misuse

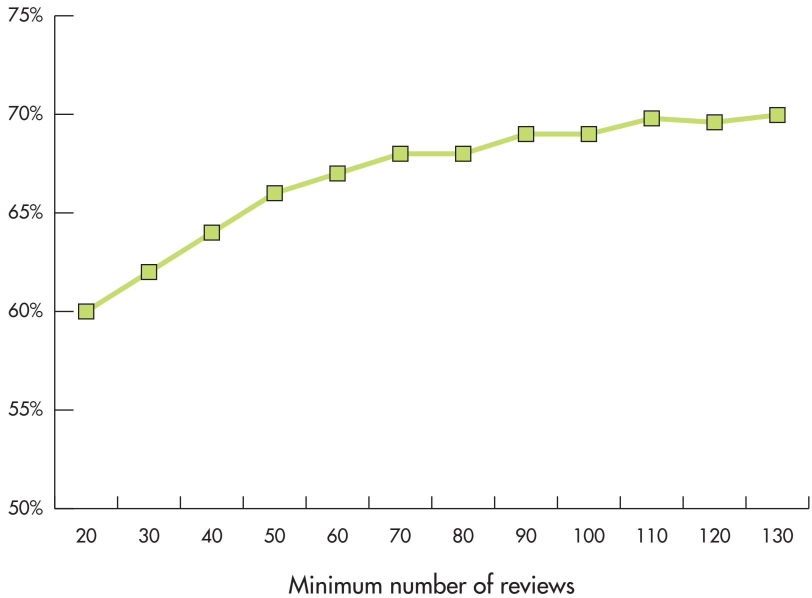

Figure 6.3: The convergence of share of bimodality.

of the product. Further analysis shows that the requirement that most often fails a distribution from being classified as bimodal is the $f1 > f2$ requirement. It is perhaps so that the ”spike” in 1 ratings comes from users that have somehow not been able to use the product properly, and is thus frustrated and rewards it with a 1 star rating. Since proportionally fewer people might experience this with shoes and envelopes than with software and hard drives, it could explain the differences.

In conclusion, we see support for hypothesis 3, that the distribution of ratings tends to be bimodal, and find evidence of this in 60% of our products, increasing to 70% as the number of reviews increases. The implication of this result is that the average rating displayed is not a reliable representation of the opinions posted, but rather an unstable balance point between extreme ratings in most cases, which is also argued by Hu et al. (2006). It is worth noting, however, that one could argue that this does not hold for all categories individually. 11 categories have less than 50 % bimodal distributions when the limit is 20 reviews. Further research is needed to determine predictors of which categories are exhibiting large degrees of bimodality and which are showing little.

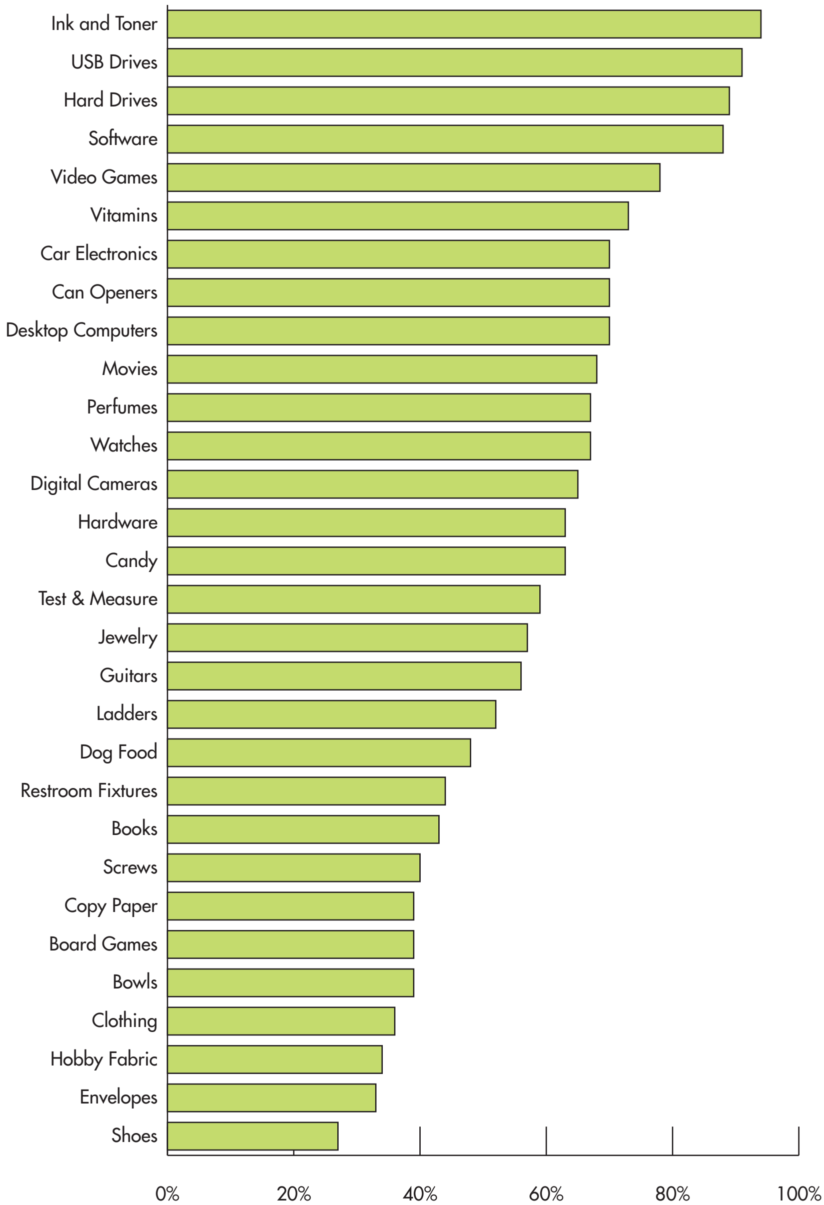

Figure 6.4: The share of bimodal distributions for all categories.

6.2.2 *Self-selection bias*

To test for occurrences of self-selection in our database of reviews, we first run the enhanced negative exponential model on the whole set, as per Li and Hitt (2008). Products with less than three reviews were removed, in order to require each product to have a meaningful average. This limits the total amount of products to 4,342, which renders the following optimal solution:

$$\begin{aligned}&\mathrm{AVGRATING}_{pt} = \\ &\quad 4.102 + 0.290\exp(-0.04 \cdot T_{pt})\cos(7.126\mathrm{E}-6 \cdot T_{pt}) + u_p + e_{pt}\end{aligned} \tag{6.2}$$

The equation shows a clear downward trend for the set as a whole ($f1f2 > 0$), as seen in Figure 6.5. This supports the notion that self-selection is a phenomenon that occurs with online consumer reviews. The cosine term is quite small, which means the characteristic self-selection "dip" is not immediately visible when regressing over all products at once. This mirrors the findings in Li and Hitt (2008). A more detailed analysis is required to assess the prevalence of undershooting, however.

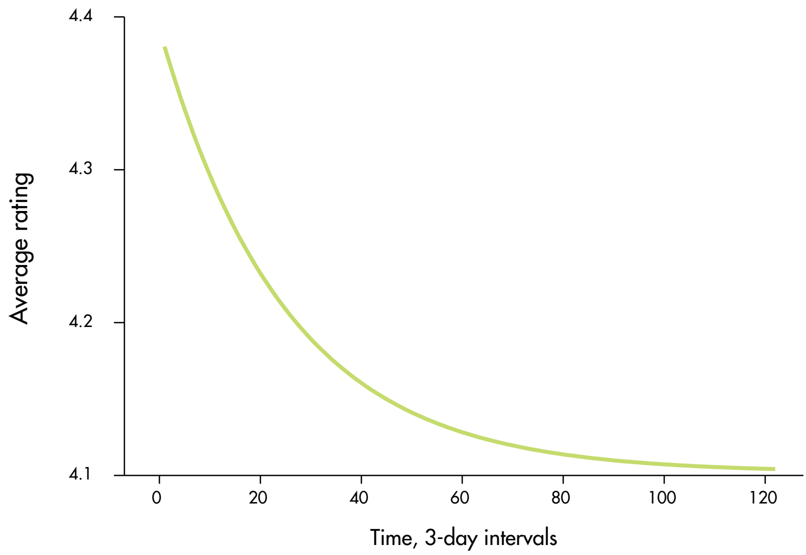

Figure 6.5: Average rating plotted against time.

When repeating the regression for each individual product, a significant amount did not have enough review activity in the first several time intervals, resulting in failed regressions. 866 products thus returned no results, leaving us 3,476 completed regressions. The results of the product-level regressions are listed in Table 6.11.

	DECLINING	INCREASING	UNDERSH.	NOT SIGN.
Number	293	104	240	3079
Pct.	8.43	2.99	6.90	88.57
	N=3476, significance measured at $p < 0{,}1$			

Table 6.11: Regression: self-selection and undershooting on product-level.

As we can determine from the table, the overwhelming majority, or nearly 90%, of regressions did not return any significant results. This is likely a consequence of fixed effects and other "noise" in the data, as well as inaccurate starting values for the regression parameters. Li and Hitt (2008) mention having performed several statistical techniques for nonlinear fixed effects models in order to mitigate the problem of noisy data, as well as estimating sound starting values for each product-specific regression. We have not been able to replicate these techniques for this thesis, which regrettably limits the confidence with which we can draw conclusions from the data.

We do find several individual instances of declining curves for average rating, which supports hypothesis 4, predicting the existence of self-selection. Out of the 397 cases with significant trends (either increasing or declining rating averages), 293 products, or roughly 74%, display evidence of the self-selection effect. 104 cases, or approximately 26% show a growing trend.

Out of the 293 products that show a declining trend, 240, or almost 82% also have an f3 parameter that differs significantly from 0, meaning they test positive for undershooting. How many of these that actually show visible undershooting is not entirely clear, however, since many hold relatively small values. In addition, comparing with the results reported by Li and Hitt (2008) is difficult, since they do not disclose any threshold value. Notwithstanding, the existence of undershooting finds some evidence, which lends partial support to hypothesis 4b.

In order to look at differences in undershooting across categories, we attempted two different strategies. First, we performed category-specific regressions to see if any clear differences existed on the category level. All but four categories returned insignificant results (all four with declining trends), which hardly gives any basis on which to speculate about differences across categories. The second strategy attempted to look closer at the products that did return significance for undershooting in the product-specific regressions. Magnitudes of the significant undershooting parameters (f3) were plotted against categoryspecific variables. As Figure 6.6 shows, the significant cases of undershooting, split on search and experience products, are virtually indistinguishable, showing a close to 50/50 split on cases and with very similar magnitudes.

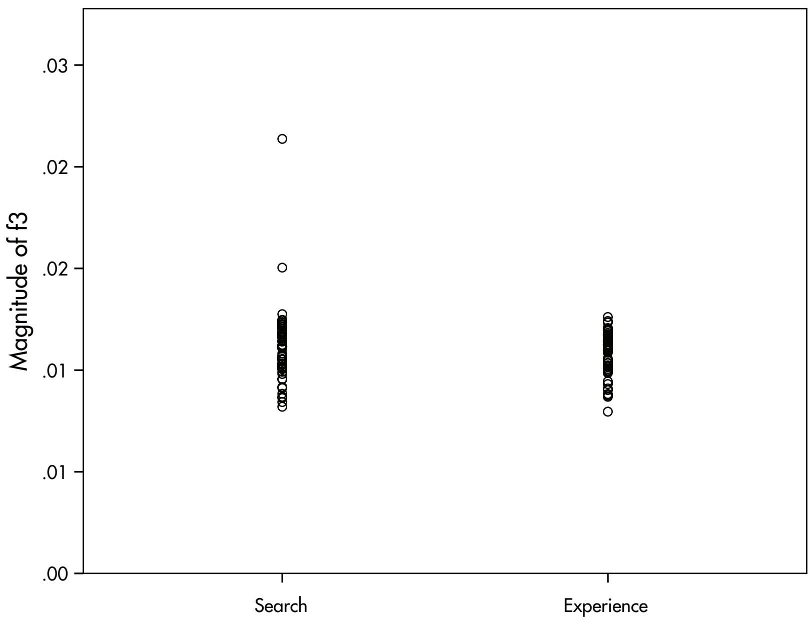

Figure 6.6: Magnitudes of undershooting parameter against product type.

Likewise, plotting the magnitudes across the model-based subjectivity values shows no clear trend either; the linear curve estimation shows a near horizontal relation, see Figure 6.7. We do not find any results with which to accept hypothesis 4c. If anything, the plots suggest that there exists no difference in undershooting across product categories. However, since there are only 240 cases of f3 having any significance from our initial set of some 5,000 products, there is little to draw any conclusions from.

6.2.3 *Conclusion*

In summary, we find mixed support for our hypotheses regarding review distributions and biases. Our logic test reveals strong signs of bimodality and under-reporting bias, with numbers similar to previous research. We also find large differences between categories. The tests for self-selection, however, likely suffer from noisy data, rendering few results from which to draw conclusions. We find evidence of self-selection bias in the dataset as a whole, but product-speficic regressions only return a few results. We find some evidence of the existence of undershooting, but are not able

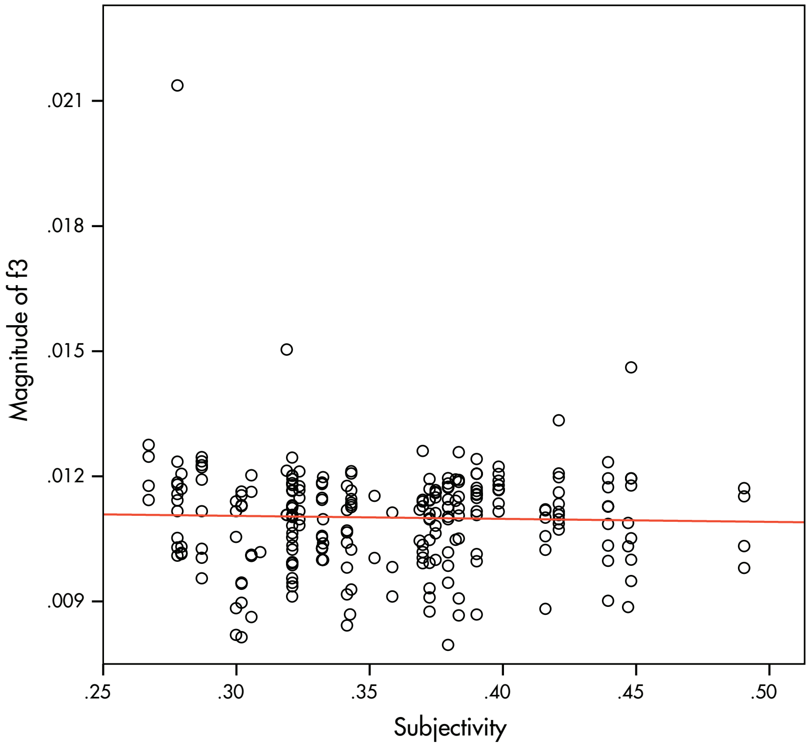

Figure 6.7: Magnitudes of undershooting parameter against degree of subjectivity.

to make estimations of its prevalence. We also have insufficient results to infer any differences across categories.

6.3 REVIEW HELPFULNESS

For the final section of this chapter we look at hypotheses 5 and 6, as shown in Table 6.12. For these hypotheses we rely on Equation 4.8 and Equation 4.9 that are based on work by Mudambi and Schuff (2010).

When structuring the data for these tests, it soon became obvious that there was a strong bias towards reviews with 100% or 0% helpful votes, as seen in Figure 6.8. These mostly consisted of cases with single votes of either helpful or not, severely skewing the results without having much data to back it. Therefore, we pruned these cases from our data, which resulted in a much more even distribution, resembling a gaussian curve, as seen in Figure 6.9. This still includes all recorded values between a perfect 0 and 100%, which should suffice to determine any effects for review helpfulness in general.

#	HYPOTHESIS
5a	Reviews that are perceived as helpful tend to be longer than other reviews.
b	The association between helpfulness and the length of the review is stronger for search goods than for experience goods.
c	The association between helpfulness and the length of the review is stronger for products which tend to be objectively evaluated than for those which tend to be subjectively evaluated.
6a	Reviews written by reviewers that use their real name are perceived as more helpful than other reviews.
b	Reviews written by reviewers with verified purchases are perceived as more helpful than other reviews.

Table 6.12: Hypotheses regarding review helpfulness.

For our first regression, we simply replicate the test as performed by Mudambi and Schuff (2010), albeit with a few omitted interaction terms describing the effects of rating with product category as well as the square of the rating. As these interaction terms did not affect neither the sign nor magnitude of our focus variables, they were omitted in order to reduce contextually redundant information. Results of the regressions are shown in Table 6.13, and tests the suppositions described for hypotheses 5 and 6.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	48.011	0.294	-	163.361	.000
Rating	2.792	0.035	0.22	79.649	.000
Total votes	0.024	0.001	0.065	23.295	.000
Ver. purchase	2.308	0.119	0.055	19.34	.000
Real name	1.304	0.137	0.026	9.52	.000
Chars	0.005	0	0.246	16.744	.000
Type	-8.694	0.271	-0.113	-32.085	.000
Type×chars	-0.002	0	-0.082	-5.481	.000
DEP: Helpfulness (%), $R^2 = 0.098$, N = 120,552					

Table 6.13: Regression results: helpfulness and product category.

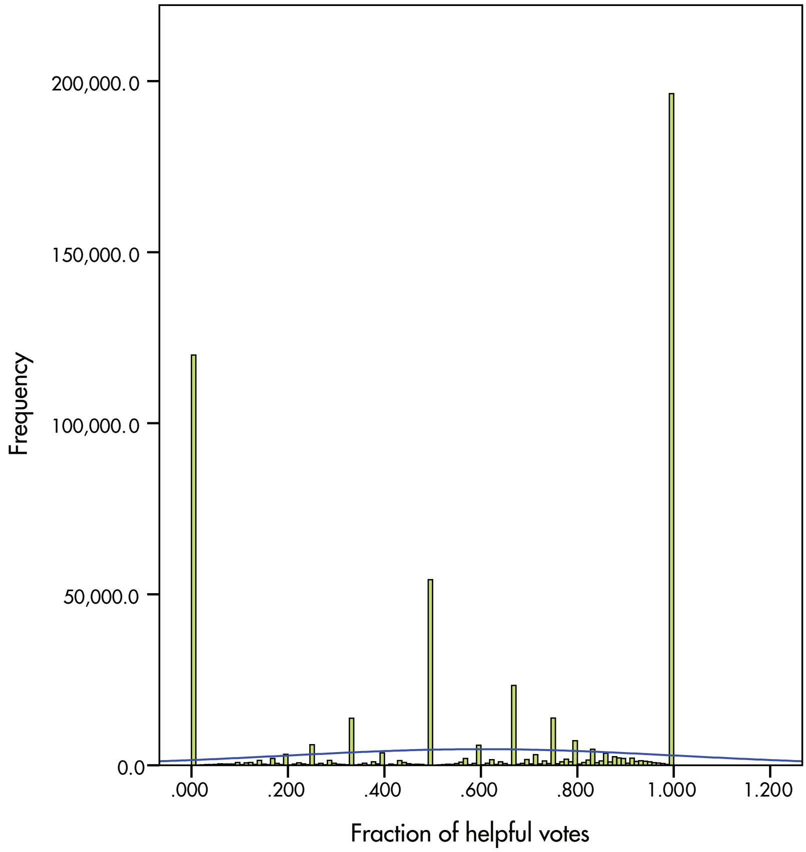

Figure 6.8: Distribution of the fraction of helpful votes.

We see several expected effects. The two most prevalent of which are the effect of the review's star rating, as well as the effect of the length of the review, measured in the amount of characters. The signs are both positive, indicating that a more positive rating, as well as a longer review, tends to see higher percentages of helpfulness.

The effect of a positive review might be attributed to how the users approach reviews. A typical consumer will surf review systems in order to help them in a purchase decision. In many cases, one would assume that the consumer has a certain need and is trying to find a remedy to take care of that need. The goal is to find the right product. In this case, a positive review might reassure a consumer with doubt, whereas a negative review might dispel the consumer from making a choice, and in effect, prolonging the search process. Our interpretation of the positive regression coefficient is that consumers value reviews helping them make good decisions more than reviews helping them avoid bad ones.

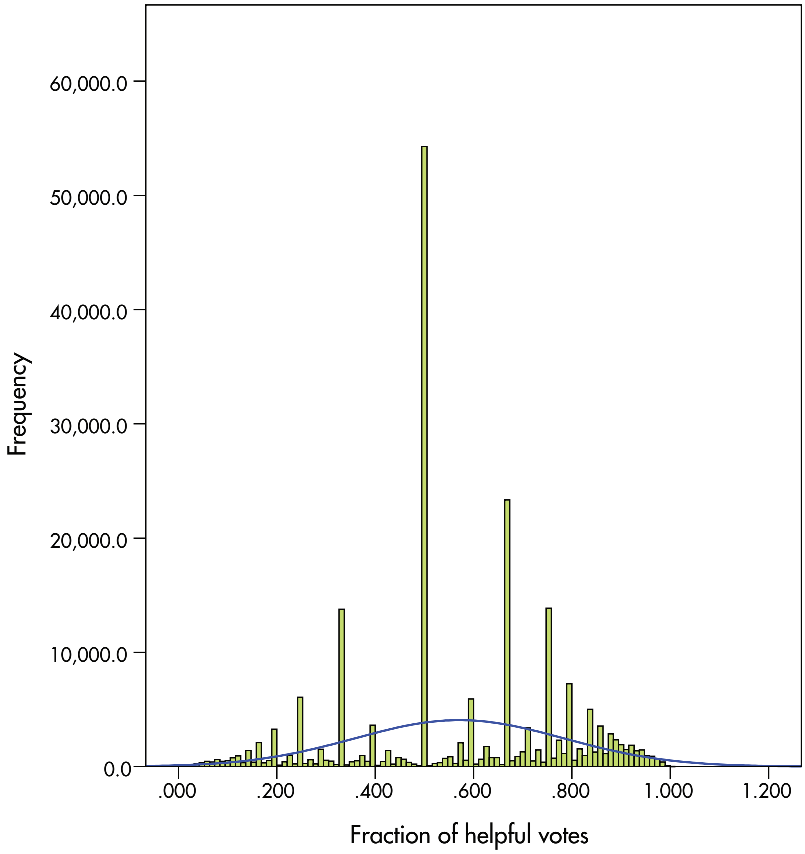

Figure 6.9: Distribution of the fraction of helpful votes after pruning.

The interpretation of the positive coefficient for the amount of characters is less convoluted. As outlined in Section 3.3, longer reviews tend to contain more information, and often include both negative and positive sides to a product. This implies that consumers value reviews that describe products and their use in detail, rather than short sentiments that do not provide much information. We therefore find support for hypothesis 5a.

Further, we note that the total votes are positively associated with higher helpfulness. This is somewhat in contrast to previous findings (Mudambi and Schuff, 2010). However, our result has a higher statistical significance ($p < 0.001$) but a smaller relative magnitude. We interpret the sign of total votes to be an effect of increased visibility. As a review garners several helpful votes, it may eventually reach the frontpage for that product, where typically the 8 most helpful reviews are presented. These reviews will likely see a massive amount of exposure compared to reviews with smaller amounts of helpfulness. As time passes, these reviews will then see ever larger amounts of total votes, and since they've already been voted up as the most helpful, they will in all likelihood continue to be voted relatively helpful. This process will, according to our understanding, give the most helpful reviews a larger amount of votes, sometimes by significant amounts. This may therefore skew the effect of the total votes in favour of reviews found most helpful, resulting in a positive coefficient for our total votes variable.

We also note that the coefficients for the verified purchase and real name badges are positive and significant. Consumers seem to appreciate information that lessens the risk of fraudulent or manipulated reviews. This is in accordance with the hypothesized effect, and supports hypotheses 6a and b.

The main focus of the test, however, is the interaction term between product type and the length of reviews, as measured in the number of characters. The sign is negative, supporting hypothesis 5b, which means consumers rate longer reviews more helpful for search products. This effect is attributed to the fact that search product reviews tend to be short and factual in nature, sometimes written in a bullet list format.

Substituting the binary product type variable for our continuous subjectivity variables, we would expect to see similar results as for the first test. Indeed, as shown in Table 6.14 both variants of the subjectivity classification seem to hold up, lending support to hypothesis 5C. Objectively evaluated products are associated with a stronger effect from the length of reviews on helpfulness.

6.3.1 *Conclusion*

Summing up, we find support for all hypotheses regarding review helfulness. The length of the review is in all cases associated with larger values of heplfulness. The verified purchase and real name badges also show positive signs in all regressions, all highly significant. Replicating the test for product type by Mudambi and Schuff (2010), we practically mirror their findings, with the exception of the total votes variable, which switches signs. The most important finding, however, is that our subjectivity variable matches the expected results, showing stronger effect of the length of review for objectively evaluated products.

VARIABLE	COEFF.	STD. ERR	STD. C	T	SIG.
(Constant)	50.094	0.207		241.805	.000
Rating	2.619	0.03	0.203	88.313	.000
Total votes	0.009	0	0.049	21.422	.000
Ver. purchase	1.607	0.103	0.038	15.57	.000
Real name	1.833	0.115	0.037	15.946	.000
Chars	0.004	0	0.204	30.563	.000
Subj. (R)	-44.849	1.004	-0.127	-44.678	.000
Chars×subj. (R)	-0.002	0.001	-0.022	-3.177	.001
DEP: Helpfulness (%), $R^2 = 0.100$, N = 172,663					
(Constant)	63.391	0.546		116.098	.000
Rating	2.646	0.03	0.205	88.878	.000
Total votes	0.009	0	0.05	21.637	.000
Ver. purchase	2.167	0.101	0.051	21.408	.000
Real name	1.759	0.115	0.035	15.268	.000
Chars	0.006	0	0.285	13.098	.000
Subj. (M)	-53.021	1.367	-0.11	-38.784	.000
Chars×subj. (M)	-0.005	0.001	-0.101	-4.644	.000
DEP: Helpfulness (%), $R^2 = 0.097$, N = 172,663					

Table 6.14: Regression results: helpfulness and subjectivity.

6.4 SUMMARY OF FINDINGS

Table 6.15 summarizes the findings from our analyses. We see that most hypotheses find support in our tests, except two of the hypotheses regarding self-selection. The data used for these tests were likely suffering from fixed effects and other noise, which renders the tests inconclusive. We also note that the subjectivity variables returned expected and significant results for all other hypotheses.

#	FINDINGS
1a	Supported
b	Supported
c	Supported
d	Supported
2	Supported
3	Supported for most categories
4a	Supported
b	Inconclusive
c	Inconclusive
5a	Supported
b	Supported
c	Supported
6a	Supported
b	Supported

Table 6.15: Overview of findings.

7

DISCUSSION

This section reviews how the results presented in Chapter 6 match the hypothesized review impact continuum, as outlined in Chapter 2. We also discuss the possible underlying driver for the model, and how this may predict similar interaction effects between reviews and other phenomena. We continue by addressing some potential concerns regarding the validity of our results due to the existence of self-selection bias. We then delve deeper into some assumptions made for the helpfulness tests, and argue how some of the results may need different interpretations. Finally, we present a new model with which to assess and allocate marketing resources, based on the potential exposure of eWOM for products and businesses.

7.1 REVIEW IMPACT CONTINUUM REVISITED

7.1.1 *Popularity*

When constructing the method for this thesis, one of our goals was to test the expected effect of ratings on different levels of popularity, and whether products aimed at the masses differ from those serving more niche markets in this regard. Our analyses performed to look at this aspect suggest that the lower we venture into the sales hierarchy, the larger the relative effect of reviews. Testing the relative change in $\ln salesrank$ against the top 100 and random sets, we saw a strong association between ratings and sales for the random sets. The top 100 products, on the other hand, showed a very weak and less significant effect in the “wrong” direction. Performing group-wise difference regressions for increasingly higher levels of sales rank, we saw strictly increasing magnitudes in the coefficients for the variables denoting change in average rating, as well as interaction terms with average rating and product type, or subjectivity. These results support the presupposition of the popularity axis in the review impact continuum, as shown in Figure 7.1

Consumers in the market for less popular products will seemingly take larger notice of the experiences of previous customers. As explained in Section 2.3, these products are often left more to their own devices, without massive marketing campaigns or hype surrounding them. They are more often from smaller, lesser known producers that also do not enjoy the positive associations that come from having a strong brand. The review impact continuum, backed up by our results, suggests businesses

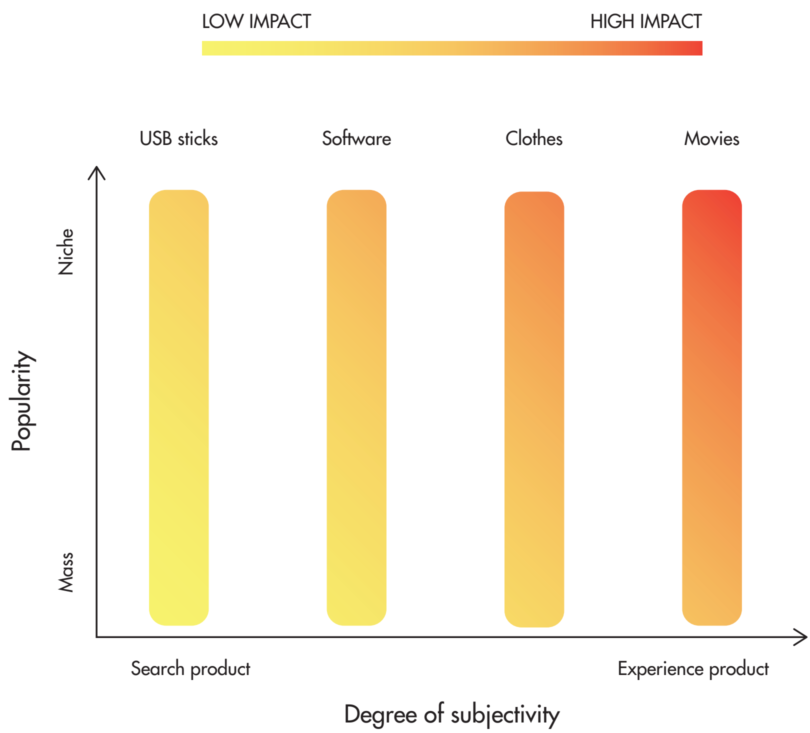

Figure 7.1: The Review Impact Continuum.

serving these markets should double down on efforts design to increase and stimulate WOM activity surrounding their products. With the current trend in business strategy being that presence on social media is required, our findings suggest that this may be a more nuanced issue. We do not suggest mass market businesses should not prioritize social media and WOM marketing, but rather emphasize this importance for niche products without mass market penetration.

7.1.2 *Subjectivity*

The statistical analyses performed in this thesis fit the expected effects for subjectivity. Using our novel subjectivity variables, we matched the findings produced by the established categorization of search and experience goods. Sales do seem to be affected more by reviews in categories with high levels of subjectivity, than in categories with low levels of subjectivity. The differences found for helpfulness in reviews for search and experience goods are also found for products with varying degrees of subjectivity. This strongly supports the idea of subjectivity as presented in the review impact continuum.

Nelson (1970) based his classification of product type on when consumers no longer would incur the cost of search to determine the product quality, opting instead for experience as evaluation. However, with the advent of eWOM, one could argue that fewer and fewer products in fact are experience products in the original sense; consumers are to a larger extent able to evaluate their quality by reading other consumers' experiences. Their search cost is lowered. Pure experience products such as fiction novels or movies will still not be evaluated equally by the entire population, but with a sufficiently large review mass, surprises in terms of experienced quality should be fewer. As such, several researchers nowadays focus on attributes or qualities (for instance Mudambi and Schuff, 2010) that describe products within the two groups instead of trying to mathematically measure the search/experience threshold. This usually means researchers stick to products with unequivocal classifications when conducting research across product categories. Since human interpretation of a vaguely defined set of attributes is required, the chance of differing labels for a product across research is not insignificant. As a result, products with an ambiguous set of search and experience qualities may be avoided altogether, possibly painting an oversimplified picture.

In this regard, the notion of a subjectivity variable is superior. Not only does it allow for classification of all products and categories, but by utilizing computerized language processing techniques, we can drastically reduce ambiguity. In Chapter 4 we outlined our method of assigning each product category a subjectivity score, which resulted in two different subjectivity variables. With these, we were able to place each product category along the horizontal axis in the review impact continuum as presented in Section 2.3. The subjectivity scores fit the ostensible conventions of search and experience products, placing most all categories in the expected positions.

All in all, the subjectivity variables seem to be a worthy addition to our understanding of how reviews, and by extension, how eWOM affects consumers' purchase decisions for different products. Nevertheless, the question still remains whether the measured subjectivity in the review content is the underlying driver of the increased effect of ratings, or if it is a proxy for some other, more fundamental phenomenon. It should be noted that we cannot separate the possible effect of subjective reviews themselves from the product categories. That is, because of the way the subjectivity variables are measured, we have to acknowledge the possibility that it is the subjective reviews that account for the increased effect of reviews, rather than aspects about the products. This would imply, however, that the majority of consumers write inefficient, i.e. objective, reviews about search products. We have no reason to believe this is the case, but future research should attempt to validate the subjectivity scale by measuring the relative effect of subjective and objective reviews.

7.1.3 *Magnitude of impacts*

We have in this thesis shown a relative difference in the magnitudes of review impacts for different categories of products, as well as for products with differing levels of popularity. The results show a stronger response in subjectively evaluated products and niche products. However, these impacts have not been quantified in terms of actual change in sales with increased ratings. This has not been the focus of this study, and actual sales numbers would likely be needed to properly assess the size of this effect. However, Luca (2011) finds that a one star increase on Yelp leads to a 5-9% increase in sales for independent restaurants. Since restaurants would be considered a classic experience good, we expect that this corresponds to the top right corner of our review impact continuum. Luca also finds that chain restaurants see no effects of reviews. This suggests that the impacts of reviews in our continuum may range from negligible up to somewhere around a 9% increase with a one star improvement in ratings.

7.2 SCARCITY AND AVAILABILITY OF INFORMATION

The common denominator for the effects predicted by the review impact continuum seems to be the need for information about products required to properly assess a product's quality. Presumably, products for which there is less objective, or rather, actionable information available, make consumers to a larger extent dependent on information from alternative channels, such as reviews from other consumers. Similarly, while there is an abundance of information available about the most popular products, useful information about niche products are harder to come by, making the information that exists proportionally more influential. In addition, because the niche consumer is possibly more specific in his preferences, more information is needed for him to make a purchasing decision than for the average consumer. Luca (2011, p. 2) denotes that "existing mechanisms [like chain affiliations] aimed at solving information problems are imperfect" and that consumers reviews therefore might complement or substitute for existing information sources. He goes on to use this as the primary explanation of why the market share of chain affiliations have dropped as the use of consumer reviews has grown.

In this paper, we have demonstrated two arguably important drivers or dimensions of scarcity of information, and a natural extension is to search for more. We take a look into our dataset and propose 3 more:

7.2.1 *Price level*

Another factor that affects the influence reviews has on sales could be the price, or rather the price level. Our regression in Section 6.1.2 showed that a change in price is associated with changes in sales (see for instance Table 6.6), which follows the conventional law of demand. However, we are interested in whether reviews have stronger impact on cheaper products, and weaker on more expensive. This speculation is based on two assumptions:

1. That there may in general be less available alternative (other than reviews) sources of information for cheap products, making the reviews relatively more influential. This is a similar argument used to explain the differences between hit and niche products, but with a different driver. The assumption is based on a general notion that there will be less eWOM and expert reviews for many cheap products, because investments in spreading such information will not be deemed profitable for most products, except for some big brands. However, Luca (2011) has already proposed that consumers seem to prefer review information over brand recognition.

2. Consumers are willing to invest more search time to validate impressions from reviews to reduce risk, by finding other sources of information for more expensive products. The bigger the purchasing investment, the higher the standards for the research required before a purchasing decision is made. Typically, if you are in the market for a box of screws, you want the best screws available given a price level, but may not be willing to do any more research outside the review system. Still, in the absolute absence of other information, consumers might simply pick the highest rated screws. On the other hand, if you are in the market for a drill, some consumers might be willing to invest time in doing some research before purchasing. Both products are arguably in the same ”category”, and assuming the goal is an average drill, none are distinct hit or niche products.

To investigate this further, we grouped the products in our sample in 3 groups, namely products priced at less than \$10, between \$10 and \$100 and between \$100 and \$1000. We then constructed a simple regression test. The results, presented in Table 7.1, gives initial support for the supposition.

Table 7.1 shows that the correlation between sales and average rating gets smaller as the prices goes up. This could indicate that consumers are in fact using more sources of information when purchasing pricey products, and the impact of the online consumer reviews then gets propotionally smaller. We plotted the coefficients on a trend curve suggesting the

VARIABLE	>$10	>$100	>$1000
AR (coeff)	-0.356**	-0.240**	-0.135*
AR (std. coeff)	-0.072	-0.050	-0.028
DEP: $\ln salesrank$			

Table 7.1: Regression results: coefficients and standardized coefficients for effect of average rating on $\ln salesrank$ for different segments of price. *AR denotes average rating.*
* - $p < 0.05$, ** - $p < 0.01$

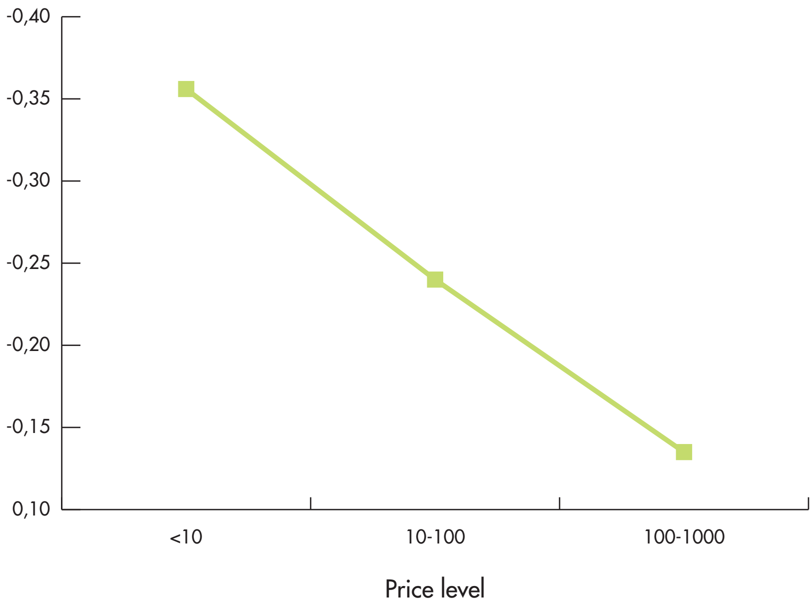

Figure 7.2: Regression coefficients for effect of average rating with increasing price levels.

relationship between reviews impact on sales and prices, demonstrated in Figure 7.2.

Although these results are promising, more research is needed to investigate the role price plays as a possible moderating factor on the impact of online consumer reviews. We recommend a study where other drivers like popularity and degree of subjectivity is properly addressed in the data selection, or controlled for in the analysis.

7.2.2 *Product complexity*

When looking at mean rating in Figure 5.7, we noted that there are some differences across product categories, and suggest this may stem from an unequal risk of misuse of certain products. Especially software stands out with a particularly low rating. Another possible explanation for this could be that it is in fact the product complexity that drives the risk of misuse, and that this is a separate driver for the need of quality information.

Although not their main focus, Chen and Xie (2008) also note that reviews will have the most effect for complex, or high-tech products, for which it is difficult to attain the requisite knowledge for correct use. Chen and Xie further argue that novice users will be unsure if the product matches their preferences or suits their needs, and will benefit from reading reviews written by more expert users.

Considering this, we briefly tested one proposed method for quantifying product complexity using the number of distinct words in a review set adjusted for the number of reviews. This notion was based on presumption that complexity requires a wider vocabulary, and that different reviews will explain the complexity differently - while reviews tend to be more alike in word selection for less complex products. However, this classification did not provide any sensible results, and was quickly abandoned. Future research should try different approaches in quantifying complexity, and assess whether this is a driver of its own for review impact.

7.2.3 *Credence goods*

Later research into the search and experience categorization by Nelson (1970) produced a third product type. Darby and Karni (1973) introduced the notion of credence goods; products for which the quality was difficult to assess even after consumption. It has been previously argued that credence goods may see larger benefits from brands, client relationships and word-of-mouth (Darby and Karni, 1973; Bjering and Havro, 2013). This stems from a difficulty for the consumer to both gather information, as well as to verify it. This could create a scarcity of information that gives larger review-related effects.

For this study, arguably a couple of examples of credence goods were also included, namely vitamin supplements and dog food. Both product categories were placed in the upper half of the subjectivity scale, but did not differentiate themselves in terms of effects of reviews. This is in our view a too small set to make any sort of inferences, so a deeper look is needed still to conclusively determine if credence goods see any additional effects.

7.3 SELF-SELECTION AND HYPE - COINCIDING EFFECTS?

In hypothesis 1a, we observe that the change in average rating is associated with a change in relative sales. Later, in hypothesis 4a we present evidence that the products ratings over time in many cases is lowered from the initial average rating due to the phenomenon of self-selection. Because of our selection method in choosing top 100 best selling products, products that possibly are at the top of their "hype" period and usually are fairly recently released, one could argue that these products have small possibilities for increased sales, and are likely to experience a decrease in sales in the forthcoming weeks independent of how the average rating develops. Considered together with the self-selection effect, which predicts that products who are recently released will drop in average rating, one could thus further argue that effects we find is perhaps simply a coincidence of two different mechanisms, and that they have little or nothing to do with each other.

We note that in fact about 63% of our top 100 observations have a drop in sales rank from the beginning our data collection period. This lends credibility to the first basis of this argument i.e. the claim that the top 100 segment is dropping in sales rank to large degree than it is increasing. However, only 32% of the top 100 segment observations exhibit a drop in average rating, while 39% have an increased average rating (29% are unchanged). The share of observations that trends upward is larger than the share that trends downwards. This deprives credibility from the second basis for this argument, that top 100 segment is generally falling in average rating. Note that these numbers are not products but observations, which are the input for our regression.

In addition, when looking at the random segment of products in Table 6.7 we find that the association between average rating and sales rank holds for this segment independently. In fact, it is the top 100 segment that seems to be the most troublesome, producing a positive coefficient (indicating that an increase in rating is associated with decreased sales). However, the results from the top 100 segment are not highly significant ($p = .084$), while the results from the random segment are ($p = .000$). While it is difficult to assess the meaning of the top 100 analysis, this at least shows that there are several drivers to the overall effect.

This, combined with the mismatch in share of observations that trends downwards in sales and downwards in ratings, makes the linkage between the self-selection effect and hype an insufficient explanation for our findings in our view. We cannot reject completely the possibility of some coincided aspects of two different phenomenons, but we regard the coincided explanation to be too limited to account for the effect as a whole. To conclusively show that the association between sales and rating exists outside of the self selection period, a regression on only products older than a year could be run. This was outside the scope for our study, and our data set is not ideal for such a test (large potion of relatively new products).

An alternative interpretation is that these two phenomenons are not in fact different mechanisms but rather two ways that the same phenomenon exhibits itself. It is possible that by looking on the development in average rating, that one could predict when a hype is coming to an end. For this study however, it is sufficient for us to argue that our results are not solely based on coinciding drops in rating and average rating for newly released products.

7.4 EXAMINING HELPFULNESS ACROSS CATEGORIES

In Section 6.3 we found the expected effect of review helpfulness paired with product categories. Search products are assumed to see larger effects of review length on review helpfulness, since search product reviews tend to be short and factual in nature (Mudambi and Schuff, 2010). We examine this last notion in Table 7.2, where we display the means for our search and experience categories, respectively. The numbers give some support to this supposition, as a t-test for the groups as wholes reveal that experience products see approximately 90 characters more per review than search products (374.86 vs. 283.53 characters).

However, recalling the distribution of lengths of reviews in Figure 5.2, the vast majority of reviews are quite short, with a minority of long reviews skewing the mean value upwards. The mode, however, is remarkably similar across categories, as shown in Table 7.2. Further, looking at the 5% trimmed means, where the top and bottom 5% are cut, the difference is only 38.39 (238.80 vs 277.19). This suggests that the effect measured in our tests concerns a relatively small fraction of the sample, specifically those that lie to the right of the mode. Also, it somewhat challenges the assumption that search product reviews are inherently shorter, since it's mostly a case of the longest reviews being longer for certain of the experience products, such as books or video games.

All our tests, as well as those by Mudambi and Schuff (2010) show the positive effect of review length to be larger for search (objectively evaluated) products, but we contend there may be different ways to interpret this. As the simple length argument, asserting that search product reviews are shorter, seems slightly weakened, our suggestion will be to more closely examine differences in the content of the reviews, to see if there may be other factors at play. Mudambi and Schuff (2010) argue that the incremental value of more depth to a search good review will be larger than for a experience good review, basing this on the fact that an increase in subjective information has less informational value. This may very well be the case, and indeed, we have not seen contradictory

CATEGORY	MODE	MEAN
Bowls	116	284.85
Can Openers	113	311.53
Copy Paper	111	263.79
Envelopes	125	239.38
Hard Drives	112	413.41
Ink and Toner	114	253.21
Screws	120	306.89
USB Drives	113	287.15
Board Games	115	394.30
Books	118	433.02
Candy	123	293.43
Movies	113	381.39
Perfumes	114	293.95
Video Games	112	452.99
Vitamins	113	380.09

Table 7.2: Mode and mean of characters per review for search and experience goods. Search goods in the top half, experience goods in the bottom half.

evidence. However, we propose that one should also consider that reviews for search goods may increase in subjectivity level as the length of the review increases, simply because the objective facts have all been stated. If this is the case, one would have to consider the fact that the increased effect of length on helpfulness may stem from subjective statements holding a higher informational value in an otherwise objective review.

7.5 PRACTICAL APPLICATIONS OF THE SUBJECTIVITY SCORE

In Section 6.1 we found that sales are more influenced by reviews in categories with more subjective review content. It is important to note that this effect goes both ways, good products might enjoy an increase in sales because of higher ratings, but lowered ratings will then consequently lead to a decline in sales. Thus, the subjectivity scale is really demonstrating the level of exposure subjectively evaluated product categories have to effects caused by what is written about them online. We propose a general model for estimating WOM exposure in Figure 7.3.

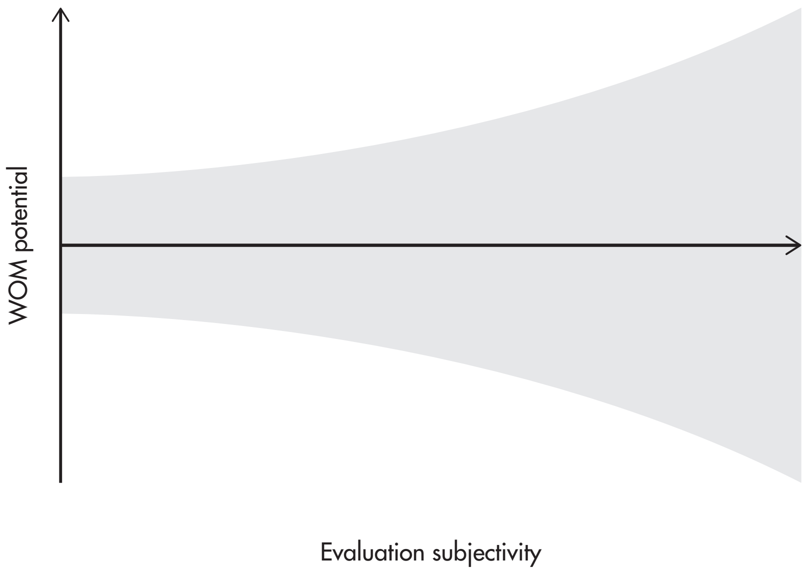

Figure 7.3: Subjectivity based assessment of WOM exposure. The gray area corresponds to the sample space.

Figure 7.3 illustrates that as the level of subjective evaluation increases, so does the potential influence from WOM. The model also captures that this influence could be both good and bad. Considering that online consumer reviews is just a particular form of eWOM, with this we propose that the correlation between subjective WOM (like reviews) and exposure to the effects of that WOM might go beyond online review systems. We contend that the model makes intuitive sense, because it seems reasonable that a primary driver for word of mouth effects is the degree of subjective experiences the consumers have had using a product. Looking at the number of reviews per product in Figure 5.1, it seems like consumers to some degree are more likely to tell about their experiences for highly subjectively evaluated products. Also, as we have previously discussed, other consumers are more likely to be influenced for these reviews.

The input for the subjectivity analysis need not be online reviews, but could be comments on social media, blogs, twitter or the buzz on the web in general. However, if the input is anything other than reviews, the result is not necessarily comparable to the scores we have presented in this paper. A reference frame of subjectivity values need to be established for other types of input.

Using our estimation of the level of subjectivity in reviews, we can place some important categories in the model.

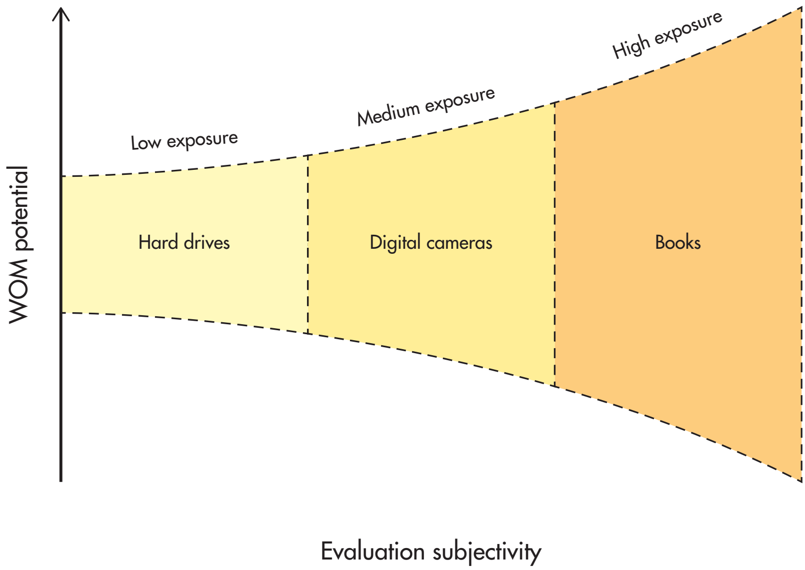

Figure 7.4: Subjectivity based WOM exposure with categories.

In Figure 7.4 we have placed sample categories or markets in the model, asserting the assumed WOM exposure. That books have a greater potential WOM influence than hard drives is hardly novel, but using this method a business could easily place its own products or the markets in the model and quickly estimate the WOM potential for a range of other more unconventional and less investigated categories or products.

This could provide benefits in smarter data driven resource allocation and prioritization. For instance, businesses producing both digital cameras and hard drives can allocate more resources on customer support for digital cameras since the possible negative (and positive) feedback mechanisms are greater. It could also help determine the level of online presence a business should have on social media like Facebook. For businesses in markets with very small WOM potential, it might make sense to choose not to be present in such media, while businesses in markets with lots of potential effects might need to be present in order to monitor and control the flow of WOM written about them, as well as to stimulate positive WOM. The model could be applied on individual products, or on businesses as a whole.

Extending the use of NLP even further, sentiment classifiers similar to the ones used in this paper can be utilized to assess the level of positive and negative content in WOM. This could offer a flash insight into the ”mood” of WOM for a product or brand at any given time. However, more research is needed to investigate this potential further.

8 CONCLUSION AND IMPLICATIONS

8.1 CONCLUDING REMARKS

Our study confirms several previous findings regarding online consumer reviews. We find evidence for reviews having an effect on sales, and that this effect interacts with other factors, most notably the product category as well as product popularity. We find that subjectively evaluated products, as well as less popular products see the largest relative effect of WOM. To the authors' knowledge, this is the first study that encompasses both of these effects simultaneously. Our findings give initial support to the hypothesized model to explain the relative impact of online reviews, dubbed the review impact continuum. We also speculate in the existence of at least two more dimensions affecting the impact, namely price and product complexity.

In this study, we also introduce a novel way of categorizing products, using natural language processing with subjectivity classification to measure the degree of subjective sentences used by consumers when evaluating the products. This subjectivity variable is used throughout our study, complementing and possibly replacing the heretofore standard categorization of search and experience products. Our subjectivity variable holds up remarkably well, matching previous findings whilst including significantly larger sets of products and reviews, as well as products that have previously been difficult to classify.

We mirror previous findings in Mudambi and Schuff (2010) for the perceived helpfulness of reviews and find that consumers seem to rate longer reviews as more helpful. This effect is found to be stronger for objectively evaluated products. In addition, information verifying the validity of the review, such as the use of real names by reviewer, or a badge confirming that the reviewer has purchased the product, are found to affect helpfulness positively.

This paper also reveals some evidence of rating biases. About 60% of the 1.1 million reviews in our dataset show signs of bimodality, meaning the average rating displayed is not a reliable representation of the opinions posted, but rather an unstable balance point between extreme ratings. We also find signs of self-selection bias, meaning average ratings may be inflated until a sufficient amount of reviewers have contributed.

Although our study has been performed with data from Amazon, we believe the results should hold for other online retailers and review systems as well.

8.2 IMPLICATIONS FOR RESEARCHERS

We assert that future research on the impact of online consumer reviews need to properly treat product category and popularity as a factor, and that this can be done using our proposed NLP based subjectivity score, and actual sales numbers or other proxies for it, like sales rank. This could make it possible to eventually compare the relative effects of review systems that sell different products, and to better identify best practices in this market. More research is needed to identify other possible drivers, we therefore propose development of quantifiable measurements for product complexity and further studies of the impact price has on the effect of online consumer reviews.

Although our results and previous research regarding product type implies that the differences in effects of reviews stem from attributes of the products, we cannot conclusively rule out the possibility that the increased effect is related to the subjectivity of the reviews themselves. If this were the case, objective reviews would be less effective. To rule this possibility out, future research should attempt to validate the subjectivity scale by measuring the relative effect of subjective and objective reviews on products with equal subjectivity scores. If there is no discernible differences in the effect of these reviews, one can assume that the larger effect of reviews observed for subjectively evaluated products is a result of the products - not the reviews. This would both validate previous findings with search and experience classifications, as well as strengthening the validity of the subjectivity scale.

We also encourage further NLP studies to develop our proposition to use subjectivity in assessing the WOM exposure for businesses. Such studies should among other things focus on systematising possible differences in subjectivity score when using online input from different sources to expand the reliability and utility of the model. In addition, exploration is needed to asses the potential NLP holds as a WOM monitor tool, and the implications this could have for a contemporary approach for businesses to control WOM. We believe this area holds a significant potential.

To properly address the causality questions that remain, especially for the causation for the association between the volume of reviews and sales, we propose a regression analysis with time lagged dependent or predictor variables. This could conceivably be able to isolate growth in either the dependent variable or the predictor, and identify a related response in the affected variable. In addition, we contend that such an analysis could show even stronger correlation between ratings and sales.

8.3 IMPLICATIONS FOR MANAGERS

The results of this thesis have several applications for managers. Using the review impact continuum, it is possible to quickly evaluate the expected impact of online consumer reviews on their business, and take appropriate actions. It could strengthen understandings of the basic mechanisms, and provide a framework for better customizing marketing approaches for different products, dependent on their expected influence from reviews. In particular, we propose that managers for businesses selling niche products and services utilize the greater potential influence of eWOM for their offerings. This may aid them in conducting smarter campaigns, gaining the most out of their budgets.

Many businesses might also experience considerable effects in addressing the unhappy consumers responsible for the minor mode in the bimodal distribution of ratings caused by the under-reporting bias. Assuming these are customers with particular challenges in the usage of the products, addressing them inside the review systems could help solve their difficulties and thus lower the share of 1 star reviews and increase the average rating for the product, positively impacting sales. Indeed, we see signs that this thinking is getting traction, particularly within the mobile app market, but we contend that gains could be achieved in other markets as well.

Finally, this paper proposes a novel and cost effective method of assessing the WOM exposure for businesses, using NLP to measure the subjectivity level of the existing WOM. This could assist managers in allocating and prioritizing appropriate amounts of resources on either controlling or stimulating WOM, according to the expected ROI. Using NLP sentiment classification, we propose that it might be possible to get insight into the actual mood of WOM at any moment, and as such be able to act quickly on the current WOM at any time, i.e. limiting bad WOM or exploiting good WOM.

Part III

APPENDIX

This part contains detailed tables from our dataset of which most of the graphs is based in previous parts. It also includes the source code developed to obtain and process the dataset.

A MISCELLANEOUS

A.1 DETAILED INFORMATION OF PRODUCT CATEGORIES

For purposes of readability, the names of product categories have been shortened or otherwise modified throughout this thesis. Table A.1 displays the full list of categories with their edited and original name, as well as the browsenode ID used in Amazon's systems.

EDITED NAME	ORIGINAL NAME	ID
Board Games	Board Games	166225011
Books	Books	1000
Bowls	Bowls	367107011
Can Openers	Can Openers	289755
Candy	Candy & Chocolate	16322461
Car Electronics	Car Electronics	1077068
Clothing	Clothing	1036592
Restroom Fixtures	Commercial Restroom Fixtures	3310316011
Ink and Toner	Computer Printer Ink and Toner	172638
Copy Paper	Copy & Multipurpose Paper	1069712
Hobby Fabric	Craft & Hobby Fabric	12899121
Desktop Computers	Desktop Computers	565098
Digital Cameras	Digital Cameras	281052
Dog Food	Dog Food	2975359011
Guitars	Guitars	11971241
Hardware	Hardware	511228
Hard Drives	Internal Hard Drives	1254762011
Jewelry	Jewelry	3880591
Ladders	Ladders	553470
Envelopes	Mailing Envelopes	1069694
Movies	Movies	2625373011
Perfumes	Perfumes & Fragrances	11056591
Screws	Screws	16403521
Shoes	Shoes	672124011
Software	Software	491286
Test & Measure	Test, Measure & Inspect	256409011
USB Drives	USB Flash Drives	3151491
Video Games	Video Games	11846801
Vitamins	Vitamins & Dietary Supplements	3764441
Watches	Watches	378516011

Table A.1: Edited and Amazon category names with browsenode IDs.

A.2 DETAILED INFORMATION OF SELECTION OF PRODUCTS

This section provides detailed information on the number of products included in the study and information about those that were excluded. We have divided our products into two main groups: the primary and the secondary group. The primary group consist of products with unique ASINs and unique review sets. The secondary group have unique ASINs and therefor sales rank, but share review set with products in the primary group.

The third column ”Dup.” gives the number of products that were excluded because the did not have a unique ASIN in our set. For the random selection, this is mainly because of an error in duplicate checks in our random generator. For the top 100, it is because some products were best sellers in more than one category. For instance, one of the best sellers in hard drives were also best sellers in USB drives. In those cases, the product was included in what was considered the most precise and well-defined category.

The fourth column ”No. rev.” shows the number of products that satisfied all other criteria, but did not have any reviews at the end of our data collection period.

The fifth and last column in the top 100 table shows the products that for unknown reasons disappeared from Amazon during the collection period. No products from the random selection disappeared.

CATEGORY	PRIME.	SEC.	DUP.	NO REV.
Board Games	88	0	5	7
Books	99	0	0	1
Bowls	57	1	30	12
Can Openers	50	0	48	2
Candy	85	0	2	13
Car Electronics	79	1	10	10
Clothing	92	0	1	7
Copy Paper	33	1	41	25
Desktop Computers	57	0	20	23
Digital Cameras	83	0	13	4
Dog Food	85	2	8	5
Envelopes	59	0	20	21
Guitars	64	3	12	21
Hard Drives*	0	0	0	0
Hardware	91	0	1	8
Hobby Fabric	56	0	5	39
Ink and Toner*	0	0	0	0
Jewelry	79	0	2	19
Ladders	52	1	27	20
Movies	97	0	1	2
Perfumes	91	0	4	5
Restroom Fixtures	32	2	48	18
Screws	40	2	35	23
Shoes	90	0	1	9
Software	68	0	1	31
Test & Measure	77	0	3	20
USB Drives*	0	0	0	0
Video Games	90	0	1	9
Vitamins	93	0	4	3
Watches	71	0	6	23
Total	1,958	13	349	380

Table A.2: Number of products in the random selection for all categories.
**No random product list generated*

CATEGORY	PRIME.	SEC.	DUP.	NO REV.	NON-EXIST.
Board Games	98	1	0	0	1
Books	100	0	0	0	0
Bowls	79	19	0	2	0
Can Openers	89	11	0	0	0
Candy	95	5	0	0	0
Car Electronics	95	4	0	0	1
Clothing	98	0	0	0	2
Copy Paper	88	5	0	7	0
Desktop Computers	91	4	0	6	0
Digital Cameras	80	20	0	0	0
Dog Food	96	4	0	0	0
Envelopes	98	2	0	0	0
Guitars	82	18	0	0	0
Hard Drives	69	30	1	0	0
Hardware	97	1	2	0	0
Hobby Fabric	89	5	0	6	0
Ink and Toner	91	9	0	0	0
Jewelry	99	0	0	1	0
Ladders	90	10	0	0	0
Movies	51	49	0	0	0
Perfumes	99	1	0	0	0
Restroom Fixtures	79	16	0	5	0
Screws	86	7	0	7	0
Shoes	99	0	0	0	1
Software	89	11	0	0	0
Test & Measure	99	1	0	0	0
USB Drives	52	48	0	0	0
Video Games	81	15	0	4	0
Vitamins	96	4	0	0	0
Watches	98	0	0	0	2
Total	2,653	299	3	38	7

Table A.3: Number of products in the top 100 selection for all categories.

CATEGORY	PRIMARY	SECONDARY	ALL
Board Games	6,290	0	6,290
Books	15,500	0	15,500
Bowls	754	43	797
Can Openers	2,592	0	2,592
Candy	4,318	52	4,370
Candy	2,202	0	2,202
Clothing	10,125	0	10,125
Copy Paper	221	1	222
Desktop Computers	1,362	0	1,362
Digital Cameras	14,489	0	14,489
Dog Food	3,605	472	4,077
Envelopes	268	0	268
Guitars	1,242	289	1,531
Hard Drives*	0	0	0
Hardware	6,685	0	6,685
Hobby Fabric	353	0	353
Ink and Toner*	0	0	0
Jewelry	2,260	0	2,260
Ladders	1,969	105	2,074
Movies	34,088	0	34,088
Perfumes	5,543	0	5,543
Restroom Fixtures	399	74	473
Screws	246	15	261
Shoes	7,824	0	7,824
Software	3,023	0	3,023
Test & Measure	1,949	0	1,949
USB Drives*	0	0	0
Video Games	20,005	0	20,005
Vitamins	9,593	0	9,593
Watches	4,239	0	4,239
Total	161,144	1,051	162,195

Table A.4: Number of primary and secondary reviews in the random selection for all categories.
**No random product list generated*

CATEGORY	PRIMARY	SECONDARY	ALL
Board Games	28,117	395	28,512
Books	200,861	0	200,861
Bowls	3,362	2,789	6,151
Can Openers	19,476	4,984	24,460
Candy	11,591	1,973	13,564
Car Electronics	88,709	19,770	108,479
Clothing	58,341	0	58,341
Copy Paper	3,131	742	3,873
Desktop Computers	4,232	390	4,622
Digital Cameras	26,643	8,601	35,244
Dog Food	12,092	2,599	14,691
Envelopes	2,921	97	3,018
Guitars	7,997	5,004	13,001
Hard Drives	19,375	24,304	43,679
Hardware	33,409	293	33,702
Hobby Fabric	1,485	265	1,750
Ink and Toner	27,581	4,671	32,252
Jewelry	19,270	0	19,270
Ladders	10,991	3,285	14,276
Movies	69,498	127,393	196,891
Perfumes	16,181	358	16,539
Restroom Fixtures	2,336	894	3,230
Screws	1,673	105	1,778
Shoes	51,828	0	51,828
Software	38,146	10,400	48,546
Test & Measure	16,237	110	16,347
USB Drives	48,272	97,432	145,704
Video Games	70,953	18,784	89,737
Vitamins	64,223	5,017	69,240
Watches	27,413	0	27,413
Total	986,344	340,655	1,326,999

Table A.5: Number of primary and secondary reviews in the top 100 selection for all categories.

CATEGORY	N	MEAN	STD. DEV.
Board Games	88	71.48	69.08
Books	99	156.57	469.45
Bowls	57	13.23	19.57
Can Openers	50	51.84	61.93
Candy	85	25.91	39.68
Car Electronics	79	54.66	83.50
Clothing	92	110.05	330.74
Copy Paper	33	6.70	12.72
Desktop Computers	57	23.89	41.79
Digital Cameras	83	174.57	233.85
Dog Food	85	42.41	58.25
Envelopes	59	4.54	4.22
Guitars	64	19.41	38.50
Hard Drives	-	-	-
Hardware	91	73.46	197.55
Ink and toner	-	-	-
Hobby Fabric	56	6.30	15.14
Jewelry	79	28.61	39.81
Ladders	52	37.87	53.24
Movies	97	351.42	696.01
Perfumes	91	60.91	151.69
Restroom Fixtures	32	12.47	19.57
Screws	40	6.15	13.50
Shoes	90	86.93	169.85
Software	68	44.46	76.83
Test & Measure	77	25.31	37.10
USB Drives	-	-	-
Video Games	90	222.28	647.03
Vitamins	93	103.15	178.60
Watches	71	59.70	142.27
All	1,958	82.30	275.28

Table A.6: Mean number of reviews per product for random selection.

CATEGORY	N	MEAN	STD. DEV.
Board Games	98	286.91	278.91
Books	100	2008.61	3239.59
Bowls	79	42.56	50.16
Can Openers	89	218.83	315.06
Candy	95	122.01	130.92
Car Electronics	95	648.20	1394.13
Clothing	98	595.32	515.71
Copy Paper	88	35.58	76.82
Desktop Computers	90	47.02	72.06
Digital Cameras	80	333.04	300.33
Dog Food	96	125.96	236.41
Envelopes	98	29.81	31.15
Guitars	82	97.52	108.01
Hard Drives	69	280.80	449.81
Hardware	97	344.42	540.26
Hobby Fabric	89	16.69	19.38
Ink and Toner	91	303.09	263.31
Jewelry	99	194.65	327.36
Ladders	90	122.12	195.89
Movies	51	1362.71	1315.86
Perfumes	99	163.44	118.77
Restroom Fixtures	79	29.57	32.18
Screws	86	19.45	41.39
Shoes	99	523.52	430.66
Software	89	428.61	615.54
Test & Measure	99	164.01	159.91
USB Drives	52	928.31	1438.87
Video Games	82	865.28	1305.97
Vitamins	96	668.99	862.29
Watches	98	279.72	253.23
All	2,653	360.80	928.08

Table A.7: Mean number of reviews per product for top 100 selection.

CATEGORY	N	MODE	MEDIAN	MEAN
Board Games	34,407	115	240	394.30
Books	216,361	118	209	433.02
Bowls	4,116	116	197	284.85
Can Openers	22,068	113	216	311.53
Candy	13,793	123	196	293.43
Car Electronics	93,027	115	194	308.28
Clothing	68,466	121	185	297.84
Copy Paper	3,352	111	177	263.79
Desktop Computers	5,594	115	324	606.32
Digital Cameras	41,132	121	323	672.85
Dog Food	15,697	120	244	372.53
Envelopes	3,189	125	171	239.38
Guitars	9,239	117	244	425.87
Hard Drives	19,375	112	235	413.41
Hardware	40,094	117	220	350.62
Hobby Fabric	1,838	131	174.5	251.08
Ink and Toner	27,581	114	170	253.21
Jewelry	21,530	109	157	213.28
Ladders	12,960	120	233	358.80
Movies	103,586	113	175	381.39
Perfumes	21,724	114	172	293.95
Restroom Fixtures	2,735	129	244	414.06
Screws	1,919	120	194	306.89
Shoes	59,652	115	185	268.13
Software	41,169	118	224	432.88
Test & Measure	18,186	121	210.5	337.76
USB Drives	48,272	113	184	287.15
Video Games	90,958	112	183	452.99
Vitamins	73,816	113	243	380.09
Watches	31,652	120	194	323.47
All	1,147,488	113	201	375.26

Table A.8: Mode, median and mean of characters per review for all categories.

CATEGORY	1 STAR	2 STARS	3 STARS	4 STARS	5 STARS
Board Games	3.3	2.6	5.1	16.2	72.8
Books	4.5	3.6	6.7	15.8	69.5
Bowls	4.8	3.6	7.9	14.9	68.8
Can Openers	10.8	4.2	4.9	11.4	68.7
Candy	7.0	4.0	7.0	13.5	68.5
Car Electronics	10.2	5.3	7.2	16.3	61.0
Clothing	7.0	6.1	9.3	19.4	58.2
Copy Paper	5.4	3.5	6.2	15.5	69.4
Desktop Computers	10.4	4.8	7.3	18.7	58.8
Digital Cameras	6.7	4.4	6.5	18.5	63.9
Dog Food	5.3	3.0	4.9	11.2	75.6
Envelopes	4.7	3.9	7.0	17.3	67.0
Guitars	9.0	5.4	8.5	18.7	58.4
Hard Drives	8.0	3.1	4.6	15.3	68.9
Hardware	6.9	3.8	6.5	17.3	65.6
Hobby Fabric	4.7	3.4	7.0	14.8	70.1
Ink and Toner	12.6	4.1	6.8	16.5	59.9
Jewelry	6.6	4.4	8.7	17.1	63.2
Ladders	4.8	3.4	6.8	20.3	64.6
Movies	7.9	5.1	8.6	16.6	61.8
Perfumes	8.2	3.7	6.0	12.8	69.3
Restroom Fixtures	7.1	4.3	6.7	16.5	65.4
Screws	4.6	3.6	5.8	16.1	69.9
Shoes	3.7	4.1	6.9	17.9	67.4
Software	15.0	5.8	8.2	19.0	52.0
Test & Measure	8.7	4.6	7.6	20.1	58.9
USB Drives	6.2	3.3	5.6	16.4	68.4
Video Games	8.6	3.5	5.7	12.8	69.3
Vitamins	8.2	4.7	7.0	17.9	62.2
Watches	9.3	5.8	8.6	21.2	55.1
Total	7.3	4.3	7.0	16.5	64.9

Table A.9: Rating distribution for all categories in percentages.

CATEGORY	N	MEAN	STD. DEV
Board Games	34,407	4.53	0.95
Books	216,361	4.42	1.06
Bowls	4,092	4.39	1.09
Can Openers	22,068	4.23	1.35
Candy	13,785	4.33	1.20
Car Electronics	93,027	4.12	1.34
Clothing	68,455	4.16	1.23
Copy Paper	3,353	4.40	1.11
Desktop Computers	5,594	4.11	1.33
Digital Cameras	40,668	4.29	1.18
Dog Food	15,671	4.49	1.08
Envelopes	3,182	4.38	1.08
Guitars	9,239	4.12	1.29
Hard Drives	23,906	4.34	1.21
Hardware	40,611	4.31	1.18
Hobby Fabric	1,835	4.42	1.07
Ink and Toner	27,581	4.07	1.40
Jewelry	21,510	4.26	1.19
Ladders	12,960	4.37	1.07
Movies	103,586	4.19	1.26
Perfumes	21,714	4.31	1.24
Restroom Fixtures	2,870	4.29	1.20
Screws	1,873	4.43	1.07
Shoes	59,652	4.41	1.03
Software	41,169	3.87	1.47
Test & Measure	18,182	4.16	1.27
USB Drives	145,704	4.38	1.14
Video Games	85,561	4.31	1.26
Vitamins	73,816	4.21	1.26
Watches	31,652	4.07	1.30
Total	1,147,488	4.27	1.22

Table A.10: Mean rating and standard deviation for all categories.

CATEGORY	OBJ	SUBJ	TOTAL	FRAC
Board Games	3,050	1,951	5,001	0.390
Books	2,675	2,097	4,772	0.439
Bowls	2,692	1,272	3,964	0.321
Can Openers	2,918	1,513	4,431	0.341
Candy	2,477	1,470	3,947	0.372
Car Electronics	2,581	1,116	3,697	0.302
Clothing	2,335	1,452	3,787	0.383
Copy Paper	2,572	937	3,509	0.267
Desktop Computers	4,455	2,219	6,674	0.332
Digital Cameras	4,671	2,610	7,281	0.358
Dog Food	2,644	1,922	4,566	0.421
Envelopes	2,363	909	3,272	0.278
Guitars	3,042	1,777	4,819	0.369
Hard Drives	3,658	1,418	5,076	0.279
Hardware	2,905	1,517	4,422	0.343
Hobby Fabric	2,343	1,121	3,464	0.324
Ink and Toner	2,368	953	3,321	0.287
Jewelry	1,741	1,407	3,148	0.447
Ladders	2,803	1,856	4,659	0.398
Movies	2,129	1,516	3,645	0.416
Perfumes	2,246	1,373	3,619	0.379
Restroom Fixtures	3,423	1,702	5,125	0.332
Screws	2,587	1,211	3,798	0.319
Shoes	2,137	1,735	3,872	0.448
Software	3,499	1,564	5,063	0.309
Test & Measure	2,871	1,357	4,228	0.321
USB Drives	2,452	1,278	3,730	0.343
Video Games	2,901	1,795	4,696	0.382
Vitamins	2,828	1,660	4,488	0.370
Watches	2,600	1,558	4,158	0.375

Table A.11: Subjectivity scores from model-based classifier for all categories. The numbers are in number of identified sentences, and the frac is (subj/(subj+obj)).

CATEGORY	OBJ	SUBJ	UNKN	TOTAL	FRAC
Board Games	1,557	265	3,179	5,001	0.145
Books	1,306	377	3,089	4,772	0.224
Bowls	1,289	142	2,533	3,964	0.099
Can Openers	1,465	105	2,861	4,431	0.067
Candy	1,093	160	2,694	3,947	0.128
Car Electronics	1,307	107	2,283	3,697	0.076
Clothing	1,101	197	2,489	3,787	0.152
Copy Paper	1,302	73	2,134	3,509	0.053
Desktop Computers	2,213	229	4,232	6,674	0.094
Digital Cameras	2,147	254	4,880	7,281	0.106
Dog Food	1,299	217	3,050	4,566	0.143
Envelopes	1,212	96	1,964	3,272	0.073
Guitars	1,515	164	3,140	4,819	0.098
Hard Drives	1,845	104	3,127	5,076	0.053
Hardware	1,343	150	2,929	4,422	0.100
Hobby Fabric	1,105	103	2,256	3,464	0.085
Ink and Toner	1,257	79	1,985	3,321	0.059
Jewelry	810	190	2,148	3,148	0.190
Ladders	1,092	274	3,293	4,659	0.201
Movies	1,008	288	2,349	3,645	0.222
Perfumes	1,198	195	2,226	3,619	0.140
Restroom Fixtures	1,570	173	3,382	5,125	0.099
Screws	1,222	116	2,460	3,798	0.087
Shoes	1,030	285	2,557	3,872	0.217
Software	1,896	155	3,012	5,063	0.076
Test & Measure	1,410	117	2,701	4,228	0.077
USB Drives	1,083	79	2,568	3,730	0.068
Video Games	1,486	243	2,967	4,696	0.141
Vitamins	1,474	139	2,875	4,488	0.086
Watches	1,313	224	2,621	4,158	0.146

Table A.12: Subjectivity scores from rule-based classifier for all categoreis. The numbers are in number of identified sentences, and the frac is (subj/(subj+obj))

B SOURCE CODE

B.1 REVIEW SCRAPING AND PROCESSING CODE

The process of obtaining the reviews is split into two parts; scraping the raw data by sending a GET-request to Amazon.coms servers, and structuring that unstructured data into different columns in Excel. The code is written in VBA, because of its integration with excel, and its lack of need of installations to run on other Windows computers. This enabled us to effortless run the code on different computers.

B.1.1 *Review Scraping*

The scraping code takes in a list of ASINs and iterates through all available pages of reviews using the web address

http://www.amazon.com/product-reviews/$asin$
/?ie=UTF8&pageNumber=$pagenumber$

where $asin$ and $pagenumber$ correspond to the specific ASIN and page number. The return object is stored in a HTML-document-object, before the DOM-element "productReviews" is extracted and stored in a string array list. This is the element that contains all the reviews. This, together with the ASIN is then written to Excel.

Listing B.1: Sub for scraping reviews from Amazon.com

```
Public Sub ScrapeReviews()
'This sub iterates through a list of ASINs in a ASIN-
    sheet and stores the unstructured data in a RawReview-
    sheet
'Needs stable web access and the Microsift Scripting
    Runtime and Microsoft HTML Object Library references
    enabled
'Running time depends on internet lag, but is in the area
     of 1.0-1.2 second per review page

    '***Buglist***
    'No known bugs

    'Declare objects
    Dim inSheet As Worksheet
```

```
Dim outsheet As Worksheet
Dim errLog As Worksheet
Dim webPage As HTMLDocument
Dim HTTPRequest As Object
Dim outputRange As Variant

'Declare help-variables
Dim startTime As Double
Dim endTime As Double
Dim asin As String
Dim isLastPageSet As Boolean
Dim lastPageString As String
Dim lastPageStringPart1 As String
Dim lastPageStringPart2 As String
Dim lastPageStringPos As Integer
Dim isMorePages As Boolean
Dim numOfScrapeErrors As Integer
Dim url As String
Dim reviews() As String
Dim reviewText As String

'Declare iteration indexes
Dim productPos As Integer
Dim lastProductPos As Integer
Dim pagePos As Integer
Dim lastPagePos As Integer
Dim reviewPos As Integer
Dim lastReviewPos As Integer
Dim outputPos As Long
Dim errLogPos As Integer

'Initialize objects
Set inSheet = ThisWorkbook.Sheets("ASIN")
Set outsheet = ThisWorkbook.Sheets("RawReviews")
Set errLog = ThisWorkbook.Sheets("ErrorLog")
Set webPage = New HTMLDocument

'Clear the error log
errLog.Cells.Clear
errLog.Cells(1, 1) = "ASIN"
errLog.Cells(1, 2) = "PagePos"
errLog.Cells(1, 3) = "ReviewPos"
errLog.Cells(1, 4) = "Errorcode"
errLog.Cells(1, 5) = "ErrorText"

'Initialize indexes
outputPos = outsheet.Cells(Rows.Count, 1).End(xlUp).
    Row + 1
```

```
lastProductPos = inSheet.Cells(Rows.Count, 2).End(
    xlUp).Row
errLogPos = 2

'Disable stuff to make it go faster
Application.EnableEvents = False
Application.ScreenUpdating = False
Application.Calculation = xlCalculationManual

'If an error occurs, go to the next review withour
    writing anything for the errorreview
On Error GoTo ErrorHandler

'**********************
'***Product iteration***
'**********************
For productPos = 2 To lastProductPos

    If inSheet.Cells(productPos, 5) Then GoTo
        NextProduct

    'Stores start time in variable "StartTime"
    startTime = Timer

    'Fetch ASIN and create product specific url
    asin = inSheet.Cells(productPos, 2)
    url = "http://www.amazon.com/product-reviews/" &
        asin & "/?ie=UTF8&pageNumber="

    'Reset product specific parameters
    isLastPageSet = False
    isMorePages = True

    If inSheet.Cells(productPos, 4) = "" Then
        pagePos = 1
    Else
        pagePos = inSheet.Cells(productPos, 4)
    End If

    If Not URLExists(url) Then GoTo NextProduct

    '*******************
    '***Page iteration***
    '*******************
    While isMorePages

        inSheet.Cells(productPos, 4) = pagePos
        numOfScrapeErrors = 0
```

```
ScrapePage:
            'Hands control over to the OS to compute high
                 priority tasks. Prevents Excel from
                freezing and craching.
            DoEvents

            'Fetch the web page and store it in web page
                object
            Set HTTPRequest = CreateObject("WinHttp.
                WinHttpRequest.5.1")
            HTTPRequest.Open "GET", url & pagePos, False
            HTTPRequest.send
            webPage.body.innerHTML = HTTPRequest.
                responseText

            'Finds the last page number and stores it,
                but only once for each product
            If Not isLastPageSet Then
                lastPageString = webPage.
                    getElementsByClassName("crAvgStars")
                    (0).innerText 'Fails
                lastPageStringPos = InStr(lastPageString,
                     "(")
                lastPageString = Mid(lastPageString,
                    lastPageStringPos + 1)
                lastPageStringPos = InStr(lastPageString,
                     "customer review")
                lastPageString = Mid(lastPageString, 1,
                    lastPageStringPos - 2)
                'Convert to int and avoid running this
                    before new product
                If InStr(lastPageString, ",") <> 0 Then
                    lastPageStringPos = InStr(
                        lastPageString, ",")
                    lastPageStringPart1 = Mid(
                        lastPageString, 1,
                        lastPageStringPos - 1)
                    lastPageStringPart2 = Mid(
                        lastPageString, lastPageStringPos
                        + 1)
                    lastPageString = lastPageStringPart1
                        & lastPageStringPart2
                End If
                lastPagePos = CInt(lastPageString)
                lastPagePos = Application.
                    WorksheetFunction.Ceiling((lastPagePos
                     / 10), 1)
```

```
            isLastPageSet = True

            inSheet.Cells(productPos, 6) =
                lastPagePos
        End If

        'Create string array with reviews
        reviews() = Split(webPage.getElementById("
            productReviews").innerText, "Comment
            Comment")
        lastReviewPos = UBound(reviews) - 1

        '**********************
        '***Review iteration***
        '**********************
        For reviewPos = 0 To lastReviewPos

            'inSheet.Cells(productPos, 5) = reviewPos

            reviewText = webPage.
                getElementsByClassName("reviewText")(
                reviewPos).innerText
            'Fix if text starts with =
            If Mid(reviewText, 1, 1) = "=" Then
                reviewText = "'" & reviewText

            outputRange = outsheet.Range("A" &
                outputPos, "C" & outputPos).Value2

            outputRange(1, 1) = "'" & asin
            outputRange(1, 2) = reviews(reviewPos)
            outputRange(1, 3) = reviewText

            outsheet.Range("A" & outputPos, "C" &
                outputPos).Value2 = outputRange

            outputPos = outputPos + 1

NextReview:
        Next reviewPos

        Application.StatusBar = "Progress: Page " &
            pagePos & " of " & lastPagePos & " |
            Product " & productPos - 1 & " of " &
            lastProductPos - 1
```

```
            'Check if this is the last page (this is
                neccessary cause in a from..to statement,
                the to index cannot be changed, and we do
                not know this index at the start of
                iteration)
            If pagePos = lastPagePos Then
                isMorePages = False
            Else
                pagePos = pagePos + 1
            End If



        'Next pagePos
        Wend

        'Stores end time in variable "EndTime"
        endTime = Timer
        inSheet.Cells(productPos, 8) = (endTime -
            startTime)
        inSheet.Cells(productPos, 7) = ((lastPagePos - 1)
             * 10) + (lastReviewPos + 1)
        'Mark product as scraped
        inSheet.Cells(productPos, 5) = True

NextProduct:
    Next productPos

    'Fix formating
    outsheet.Cells.ClearFormats

    'Enable stuff again
    Application.EnableEvents = True
    Application.ScreenUpdating = True
    Application.Calculation = xlCalculationAutomatic

    'Remove status bar
    Application.StatusBar = False

    MsgBox ("Reviews scraped with " & (errLogPos - 2) & "
         error(s). Check the log for more info")

    Exit Sub

'********************
'***Error handling***
'********************
ErrorHandler:
```

```
    errLog.Cells(errLogPos, 1) = asin
    errLog.Cells(errLogPos, 2) = pagePos
    errLog.Cells(errLogPos, 3) = reviewPos
    errLog.Cells(errLogPos, 4) = Err.Number
    errLog.Cells(errLogPos, 5) = Err.Description

    errLogPos = errLogPos + 1

    'If it is an "Object variable not set (Error 91)" the
         server probably did not respond or internet was
        down for a second or so
    'If so, wait 3 second, try scraping the page again,
        but only up to 3 times (indicating that something
        else is causing the 91 error)
    If Err.Number = 91 And numOfScrapeErrors < 3 Then
        numOfScrapeErrors = numOfScrapeErrors + 1
        Application.Wait (Now + TimeValue("0:00:03"))
        Resume ScrapePage
    Else
        Resume NextProduct
    End If

    'Resume processing on next review
    Resume NextReview

End Sub
```

B.1.2 *Review Processing*

The processing code uses a variety of pattern-recognition and textual markers approaches to identify the separators for the different data in a unprocessed review. It takes in the raw reviews scraped by the review scraping sub, which is basically just a long string with all the information. To account for all the possible differences in this data, the code becomes somewhat messy and complex, but it still works beautifully. For instance, the code recognizes that a date in the first part of the raw string on the format (case sensitive) ", Month" probably is the date of the review. But problems arise if a user for instance writes "Lorem Ipsum, May dolor sit amet" as a title of the review. The script therefore performes numerous checks to verify that that it is the actual date of the review that is has found. When this is confirmed, this position is used to split up the long string of data, and then new pieces of information are located. Using this approach, we were able to "restore" all original data available in the review and store it in a structured manner.

Listing B.2: Main sub and help methods for processing the unstructured scraped data from ScrapeReviews

```
Public Sub ProcessReviews()
    'This sub assumes RawReviews on the format provided
        by ScrapeReviews, and structures this data into
        different Excel-cells.
    'It stores the structured data in a Review-sheet

    '***Buglist***
    'No known bugs

    '************************************
    '***Section declare and initialize***
    '************************************
    'Declare objects
    Dim inSheet As Worksheet
    Dim outsheet As Worksheet
    Dim errLog As Worksheet
    Dim inputRange As Range
    Dim outputRange As Variant

    'Declare help-variables
    Dim startTime As Double
    Dim endTime As Double
    Dim manufacturerPos As Integer
    Dim manufacturerPos2 As Integer
    Dim amazonOffPos As Integer
    Dim amazonOffPos2 As Integer
    Dim permaPos As Integer
    Dim outputPos As Long
    Dim reviewPos As Long
    Dim lastReviewPos As Long
    Dim titleAndDate As String
    Dim wasHelpfulTemp As String
    Dim wasHelpful As String
    Dim totalHelpful As String
    Dim badgeString As String
    Dim starsPos As Integer
    Dim byPos As Integer
    Dim datePos As Integer
    Dim helpfulPos As Integer
    Dim helpfulStartPos As Integer
    Dim totalHelpfulVotesPos As Integer
    Dim ofPos As Integer
    Dim seeAllMyReviewsPos As Integer
    Dim reviewSrcStartPos As Integer
    Dim reviewTextStartPos As Integer
    Dim errLogPos As Integer
```

```
Dim reviewTextOrg As String

'Declare output fields
Dim RawReview As String
Dim asin As String
Dim reviewDate As String
Dim rating As Integer
Dim reviewerName As String
Dim reviewTitle As String
Dim reviewText As String
Dim numberOfChars As Integer
Dim numberOfWasHelpful As Integer
Dim numberOfTotalHelpful As Integer
Dim reviewSrc As String
Dim isVerifiedPurchase As Boolean
Dim isRealName As Boolean
Dim topReviewer As String
Dim isHallOfFame As Boolean
Dim isVineVoice As Boolean
Dim commentedByManufacturer As Boolean
Dim commentedByAmazon As Boolean
Dim isByPosSet As Boolean

'Stores start time in variable "StartTime"
startTime = Timer

'Initialize objects
Set inSheet = ThisWorkbook.Sheets("RawReviews")
Set outsheet = ThisWorkbook.Sheets("Reviews")
Set errLog = ThisWorkbook.Sheets("ErrorLog")

'Clear errorlog
errLog.Cells.Clear
errLog.Cells(1, 1) = "LineNr"
errLog.Cells(1, 2) = "Errorcode"
errLog.Cells(1, 3) = "ErrorText"

'Clear old review
Call ClearProcessedReviews(False)

'Initialize indexes
outputPos = 2
errLogPos = 2
lastReviewPos = inSheet.Cells(Rows.Count, 1).End(xlUp
    ).Row

'Disable stuff to make it go faster
Application.EnableEvents = False
```

```
    Application.ScreenUpdating = False
    Application.Calculation = xlCalculationManual

    'If an error occurs, go to the next review withour
        writing anything for the errorreview
    On Error GoTo ErrorHandler

    '************************
    '***Process raw reviews***
    '************************
    For reviewPos = 2 To lastReviewPos

        DoEvents ' Prevents it from freezing
        Application.StatusBar = "Progress: " & reviewPos
            & " of " & lastReviewPos & ": " & Format(
            reviewPos / lastReviewPos, "0%")

        'Fetch indata and store in range
        Set inputRange = inSheet.Range("A" & reviewPos, "
            C" & reviewPos)

        'Fetch the raw review
        RawReview = inputRange(1, 2).Value2
        'Check if the review is commented by the
            manufacurer, if so, trim the raw review
        commentedByManufacturer = False
        commentedByAmazon = False
RawReview:
        manufacturerPos = InStr(RawReview, "The
            manufacturer commented on the review belowSee
            comment")
        manufacturerPos2 = InStr(RawReview, "The
            manufacturer commented on this review")
        amazonOffPos = InStr(RawReview, "An Amazon.com
            official commented on the review belowSee
            comment")
        amazonOffPos2 = InStr(RawReview, "An Amazon.com
            official commented on this review")
        If manufacturerPos <> 0 Then
            commentedByManufacturer = True
            RawReview = Mid(RawReview, manufacturerPos +
                56)
            GoTo RawReview
        ElseIf manufacturerPos2 <> 0 Then
            commentedByManufacturer = True
            permaPos = InStr(RawReview, "Permalink")
            RawReview = Mid(RawReview, permaPos + 10)
            GoTo RawReview
```

```
ElseIf amazonOffPos <> 0 Then
    commentedByAmazon = True
    RawReview = Mid(RawReview, amazonOffPos + 62)
    GoTo RawReview
ElseIf amazonOffPos2 <> 0 Then
    commentedByAmazon = True
    permaPos = InStr(RawReview, "Permalink")
    RawReview = Mid(RawReview, permaPos + 10)
    GoTo RawReview
End If

'Asin
asin = inputRange(1, 1).Value2
'Fix if ASIN starts with 0
If Mid(asin, 1, 1) = "0" Then asin = "'" & asin

'ReviewDate
starsPos = InStr(RawReview, "stars")
isByPosSet = False
If (InStr(RawReview, "A Kid's Review") <> 0) Then
    byPos = InStr(RawReview, "A Kid's Review")
    isByPosSet = True
ElseIf (InStr(RawReview, "A Customer") <> 0) Then
    byPos = InStr(RawReview, "A Customer")
    isByPosSet = True
Else
    byPos = InStr(RawReview, "By ")
End If

seeAllMyReviewsPos = 0
seeAllMyReviewsPos = InStr(RawReview, "See all my
     reviews")
If seeAllMyReviewsPos = 0 Then seeAllMyReviewsPos
     = InStr(RawReview, "This review is from:")
If seeAllMyReviewsPos = 0 Then seeAllMyReviewsPos
     = byPos

titleAndDate = Mid(RawReview, 1, (
    seeAllMyReviewsPos - 2))
datePos = GetMonthPos(titleAndDate)

If Not isByPosSet Then
    While byPos < datePos
        byPos = InStr(byPos + 1, RawReview, "By "
            )
    Wend
End If
```

```
reviewDate = Mid(titleAndDate, datePos, (byPos -
    datePos - 3))

'Rating
rating = Mid(RawReview, starsPos - 13, 1)

'ReviewerName
If InStr(RawReview, "A Kid's Review") <> 0 Then
    reviewerName = "A Kid's Review"
ElseIf InStr(RawReview, "A Customer") <> 0 Then
    reviewerName = "A Customer"
Else
    reviewerName = Mid(RawReview, byPos + 3, (
        seeAllMyReviewsPos - (byPos + 3)) - 2)
End If
'Fix if starts with =
'If Mid(reviewerName, 1, 1) = "=" Then
    reviewerName = "'" & reviewerName
reviewerName = "'" & reviewerName

'ReviewTitle
reviewTitle = Mid(titleAndDate, starsPos + 6,
    datePos - (starsPos + 8))
'Fix if starts with =
'If Mid(reviewTitle, 1, 1) = "=" Then reviewTitle
    = "'" & reviewTitle
reviewTitle = "'" & reviewTitle

'ReeviewText
'Fix if text starts with =
reviewText = inputRange(1, 3).Value2
reviewText = "'" & reviewText

'Chars in reviewtext (without title)
numberOfChars = Len(reviewText)

'Number of people who found the review helpful
helpfulPos = 0
helpfulPos = InStr(RawReview, "people found the
    following review")
If helpfulPos <> 0 Then
    ofPos = InStr(RawReview, "of")
    wasHelpfulTemp = Mid(RawReview, 1, ofPos - 1)
    helpfulStartPos = InStr(wasHelpfulTemp, ")")
    wasHelpful = Mid(wasHelpfulTemp,
        helpfulStartPos + 9)
    If wasHelpful = "" Then wasHelpful = Mid(
        wasHelpfulTemp, 5)
```

```
Else
    wasHelpful = "0"
End If
numberOfWasHelpful = CInt(wasHelpful)

'Total number of helpfull votes
If helpfulPos <> 0 Then
    totalHelpful = Mid(RawReview, ofPos + 3, 7)
    totalHelpfulVotesPos = InStr(totalHelpful, "
        ")
    'While totalHelpfulVotesPos = 0
    '    totalHelpfulVotesPos = InStr(
        totalHelpfulVotesPos + 1, totalHelpful, "
        ")
    'Wend
    totalHelpful = Mid(RawReview, ofPos + 3,
        totalHelpfulVotesPos - 1)
Else
    totalHelpful = "0"
End If
numberOfTotalHelpful = CInt(totalHelpful)

'Source of review (i.e. DVD, Blu-ray, Amazon
    Instant Video)
reviewSrcStartPos = 0
reviewSrcStartPos = InStr(RawReview, "This review
     is from:")
reviewTextStartPos = InStr(reviewSrcStartPos + 1,
     RawReview, Mid(reviewText, 2, 40))
If reviewSrcStartPos <> 0 Then
    reviewSrc = Mid(RawReview, reviewSrcStartPos
        + 21, reviewTextStartPos - (
        reviewSrcStartPos + 21))
Else
    reviewSrc = "UNKNOWN"
End If

'Badges section
'Visit the link for a description of the
    available badges and what they mean
'http://www.amazon.com/gp/help/customer/display.
    html/ref=cm_rn_bdg_help?ie=UTF8&nodeId
    =14279681&pop-up=1#VN
'Extract the string with the badges
If reviewTextStartPos < byPos Then
    reviewTextStartPos = 1000
badgeString = Mid(RawReview, byPos,
    reviewTextStartPos - byPos)
```

```
'Verified purchase
If InStr(badgeString, "Amazon Verified Purchase")
     <> 0 Then
    isVerifiedPurchase = True
Else
    isVerifiedPurchase = False
End If

'Real name
If InStr(badgeString, "REAL NAME") <> 0 Then
    isRealName = True
Else
    isRealName = False
End If

'Top reviewer
'#1 REVIEWER, TOP 10 REVIEWER, TOP 50 REVIEWER,
    TOP 500 REVIEWER, TOP 1000 REVIEWER, HALL OF
    FAME REVIEWER
If InStr(badgeString, "TOP 1000 REVIEWER") <> 0
    Then
    topReviewer = "TOP 1000"
ElseIf InStr(badgeString, "TOP 500 REVIEWER") <>
    0 Then
    topReviewer = "TOP 500"
ElseIf InStr(badgeString, "TOP 50 REVIEWER") <> 0
     Then
    topReviewer = "TOP 50"
ElseIf InStr(badgeString, "TOP 10 REVIEWER") <> 0
     Then
    topReviewer = "TOP 10"
ElseIf InStr(badgeString, "#1 REVIEWER") <> 0
    Then
    topReviewer = "#1 REVIEWER"
Else
    topReviewer = "FALSE"
End If

'Hall of fame
If InStr(badgeString, "HALL OF FAME REVIEWER") <>
     0 Then
    isHallOfFame = True
Else
    isHallOfFame = False
End If

'Vine voice
```

```
        If InStr(badgeString, "VINE VOICE") <> 0 Then
            isVineVoice = True
        Else
            isVineVoice = False
        End If


        '***********************
        '***Section write data***
        '***********************
        outputRange = outsheet.Range("A" & outputPos, "U"
             & outputPos).Value2

        outputRange(1, 1) = asin
        outputRange(1, 2) = reviewDate
        outputRange(1, 3) = rating
        outputRange(1, 4) = reviewerName
        outputRange(1, 5) = reviewTitle
        outputRange(1, 6) = reviewText
        outputRange(1, 7) = numberOfChars
        outputRange(1, 8) = wasHelpful
        outputRange(1, 9) = totalHelpful
        outputRange(1, 10) = reviewSrc
        outputRange(1, 11) = isVerifiedPurchase
        outputRange(1, 12) = isRealName
        outputRange(1, 13) = topReviewer
        outputRange(1, 14) = isHallOfFame
        outputRange(1, 15) = isVineVoice
        outputRange(1, 26) = commentedByManufacturer
        outputRange(1, 27) = commentedByAmazon

        outsheet.Range("A" & outputPos, "Q" & outputPos).
            Value2 = outputRange


NextReview:
        outputPos = outputPos + 1

    Next reviewPos

    'Fix formating
    outsheet.Cells.ClearFormats
    outsheet.Columns("B").NumberFormat = "dd.mm.yyyy"
    outsheet.Columns("M").HorizontalAlignment = xlCenter
    outsheet.Columns("A:V").VerticalAlignment = xlTop
    Application.StatusBar = False

    'Enable stuff again
```

```
    Application.EnableEvents = True
    Application.ScreenUpdating = True
    Application.Calculation = xlCalculationAutomatic

    'Stores end time in variable "EndTime"
    endTime = Timer
    ThisWorkbook.Sheets("Main").Cells(6, 12) = (endTime -
        startTime)

    MsgBox ("Reviews processed with " & (errLogPos - 2) &
        " error(s). Check the log for more info")

'Avoid running error handler code
Exit Sub


'*******************
'***Error handling***
'*******************
ErrorHandler:
    errLog.Cells(errLogPos, 1) = reviewPos
    errLog.Cells(errLogPos, 2) = Err.Number
    errLog.Cells(errLogPos, 3) = Err.Description
    errLogPos = errLogPos + 1

    'Resume processing on next review
    Resume NextReview

End Sub

Function GetMonthPos(inputString As String) As Integer
    'Help-method for ProcessReviews
    'Returns the index of the first letter of a month in
       a string

    If GetMonthPos = 0 Then GetMonthPos = InStr(
       inputString, ", January")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
       inputString, ", February")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
       inputString, ", March")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
       inputString, ", April")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
       inputString, ", May")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
       inputString, ", June")
```

```
    If GetMonthPos = 0 Then GetMonthPos = InStr(
        inputString, ", July")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
        inputString, ", August")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
        inputString, ", September")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
        inputString, ", October")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
        inputString, ", November")
    If GetMonthPos = 0 Then GetMonthPos = InStr(
        inputString, ", December")

    'Correct from the comma to actual monthpos
    GetMonthPos = GetMonthPos + 2
End Function


Sub ClearRawReviews()
    'Clears previously scraped reviews and writes headers

    Dim outsheet As Worksheet
    Set outsheet = ThisWorkbook.Sheets("RawReviews")

    outsheet.Cells.Clear
    outsheet.Cells(1, 1) = "ASIN"
    outsheet.Cells(1, 2) = "RawReview"
    outsheet.Cells(1, 3) = "ReviewText"
    outsheet.Cells.ClearFormats

    Application.StatusBar = False
    MsgBox ("Raw reviews are cleared!")

End Sub


Sub ClearProcessedReviews(Optional GiveFeedback As
    Boolean = True)
    'Clears previously processed reviews and writes the
        headers

    Dim outsheet As Worksheet
    Set outsheet = ThisWorkbook.Sheets("Reviews")

    'Clear worksheet and write headers
    outsheet.Cells.Clear
    outsheet.Cells(1, 1) = "ASIN"
    outsheet.Cells(1, 2) = "Date"
```

```
    outsheet.Cells(1, 3) = "Rating"
    outsheet.Cells(1, 4) = "Reviewer"
    outsheet.Cells(1, 5) = "Title"
    outsheet.Cells(1, 6) = "Text"
    outsheet.Cells(1, 7) = "#Chars"
    outsheet.Cells(1, 8) = "#wasHelpful"
    outsheet.Cells(1, 9) = "#totalHelpful"
    outsheet.Cells(1, 10) = "Src"
    outsheet.Cells(1, 11) = "VerifiedPurchase"
    outsheet.Cells(1, 12) = "RealName"
    outsheet.Cells(1, 13) = "TopReviewer"
    outsheet.Cells(1, 14) = "HallOfFame"
    outsheet.Cells(1, 15) = "VineVoice"
    outsheet.Cells(1, 16) = "CommentedByManufacturer"
    outsheet.Cells(1, 17) = "CommentedByAmazon"
    outsheet.Cells.ClearFormats

    Application.StatusBar = False
    If GiveFeedback Then MsgBox ("Processed reviews are
        cleared!")

End Sub


Sub ClearErrorLog()

    Dim errLog As Worksheet
    Set errLog = ThisWorkbook.Sheets("ErrorLog")

    Application.StatusBar = False
    MsgBox ("Error log is cleared!")

End Sub
```

B.1.3 *Counting Distinct Words*

This section also includes the code used for counting distinct words in a set of reviews. This sub takes in a structured list of reviews, splits the review text into a list of words based on white-spaces, and then tries to adds these to a VBA-dictionary object for each review. Since the dictionary only holds one instance of each word, the built in count method gives us the number of distinct words. The dictionary is then reset each for each set of ASINs.

Listing B.3: Sub for counting distinct words for a review set

```
Sub CountDistinctWords()
```

```
'Counts unique (as in distinct) words in a set of
    reviews based on ASIN
'Need reference to Microsoft Scriting Runtime

Dim dicDistinct As Scripting.Dictionary
Dim reviewSheet As Worksheet
Dim asinSheet As Worksheet

Dim asin As String
Dim reviewPos As Long
Dim asinPos As Long
Dim lastReviewPos As Long
Dim words() As String
Dim numberOfWords As String
Dim wordPos As Integer
Dim review As String

Set reviewSheet = ThisWorkbook.Sheets("Reviews")
Set asinSheet = ThisWorkbook.Sheets("ASIN")
Set dicDistinct = CreateObject("Scripting.Dictionary"
    )

'Should not have errors
'On Error GoTo ErrorHandler

lastReviewPos = reviewSheet.Cells(Rows.Count, 2).End(
    xlUp).Row
asinPos = 2

asin = asinSheet.Cells(2, 2)

'Disable stuff to make it go faster
Application.EnableEvents = False
Application.ScreenUpdating = False
Application.Calculation = xlCalculationManual

'*********************
'***Review iteraton***
'*********************
For reviewPos = 2 To lastReviewPos

    DoEvents
    Application.StatusBar = "Progress: " & reviewPos
        & " of " & lastReviewPos & ": " & Format(
        reviewPos / lastReviewPos, "0%")

    If reviewSheet.Cells(reviewPos, 1) = "" Then
        GoTo NextReview
```

```
        ElseIf asin <> reviewSheet.Cells(reviewPos, 1)
            Then
            asin = reviewSheet.Cells(reviewPos, 1)
            dicDistinct.RemoveAll
NextASIN:
            asinPos = asinPos + 1
            If asinSheet.Cells(asinPos, 7) = 0 Then GoTo
                NextASIN

        End If

        review = UCase(reviewSheet.Cells(reviewPos, 6))
        words() = Split(review)
        numberOfWords = UBound(words())

        '*******************
        '***Word iteraton***
        '*******************
        For wordPos = 0 To numberOfWords

            If Not dicDistinct.Exists(words(wordPos))
                Then dicDistinct.Add words(wordPos),
                vbNullString

        Next wordPos

        asinSheet.Cells(asinPos, 10) = dicDistinct.Count

NextReview:
    Next reviewPos

    Application.StatusBar = False

    'Enable stuff again
    Application.EnableEvents = True
    Application.ScreenUpdating = True
    Application.Calculation = xlCalculationAutomatic

    'MsgBox ("Reviews processed with " & (errLogPos - 2)
        & " error(s). Check the log for more info")

    Exit Sub

ErrorHandler:
    Resume NextReview

End Sub
```

B.2 PRODUCT ADVERTISING API CODE

This section includes source code used to access Amazon's Product Advertising API. The code was written using Java in Eclipse (Kepler Service Release 2). Amazon allows developers to access their systems using the SOAP protocol (Wikipedia, 2014d). All API libraries were automatically generated using a WSDL file as supplied by Amazon.

Included code are for the two main classes; the class doing searches and product lookups for extraction of salesrank data, as well as the class written to generate lists of random products.

Also written, but not included, were classes for input (parsing) and output of CSV files using the open CSV-library, OpenCSV (Various developers, 2014b)

B.2.1 *Main class*

The main class in the application was named DataGenerator, which was used to fetch all sales rank data for this thesis. The DataGenerator class at one point references the handler class *AWSHandlerResolver* which is a class needed for correct signing of the requests to Amazon's servers. Oddly, this class (or an equivalent) was not supplied by Amazon in the API libraries, yet remains essential for the API access. AWSHandlerResolver was instead provided by the community of Product Advertising API developers (Various developers, 2014a).

Listing B.4: Main class for fetching sales data from the API

```java
package requests;

import java.math.BigInteger;

import com.ECS.client.jax.AWSECommerceService;
import com.ECS.client.jax.AWSECommerceServicePortType;
import com.ECS.client.jax.ItemLookupResponse;
import com.ECS.client.jax.ItemSearchResponse;

import java.util.ArrayList;
import java.util.List;
import java.util.ListIterator;

import java.text.DateFormat;
import java.text.SimpleDateFormat;
import java.util.Calendar;
import java.util.Date;
import java.util.concurrent.TimeUnit;
```

```
public class DataGenerator3 {

        //      Authentification strings
public static String SECRET_KEY = "REMOVED FOR SECURITY
    PURPOSES";
public static String AWS_KEY = "REMOVED FOR SECURITY
    PURPOSES";
public static String ASSOCIATE_TAG = "REMOVED FOR
    SECURITY PURPOSES";

public static void main(String[] args) {

        System.out.println("————————");
System.out.println("API request initiated.");

String inputFile = "lookup/generate.csv";
Input inputList = new Input(inputFile, 2);

List<String[]> inputFileList = inputList.getInputList();

for (ListIterator<String[]> iter = inputFileList.
    listIterator(); iter.hasNext(); ) {
        String[] element = iter.next();

        String lookupFile = element[0];
        int lookupType = Integer.parseInt(element[1]);

        System.out.println("————————");
System.out.println("Generating data for ASIN set: " +
    element[2]);
System.out.println("————————");

        generate(lookupFile, lookupType);

}
System.out.println("————————");
System.out.println("API request completed.");
System.out.println("————————");
}


public static void generate(String inputFile, int type ){
        //              Define file describing lookups
            and lookuptype. Type is 1 for top 100 search,
            2 for single ASIN lookup
//              3 is for generation of ASIN lists (
    RndAsinGenerator), 4 is for lookup of RND ASINs.
//              Define outputTarget for lookup searches
```

```
String lookupFile = inputFile;
int lookupType = type;

Input lookupList = new Input(lookupFile, lookupType);

if (lookupType == 1)
{
        List<String[]> itemSearchList = lookupList.
            getInputList();

        for (ListIterator<String[]> iter = itemSearchList
            .listIterator(); iter.hasNext(); ) {
                String[] element = iter.next();

                itemSearch(1, element[0]);
        }

}else if ( lookupType == 2 ||lookupType == 4){

        List<String[]> itemLookupList = lookupList.
            getInputList();

        for (ListIterator<String[]> iter = itemLookupList
            .listIterator(); iter.hasNext(); ) {
                String[] element = iter.next();

                System.out.println("Writing " + element
                    [1] + " ....");

        itemLookup(lookupType, element);

        System.out.println(element[1] + " written!");
        System.out.println("———————");

        // Amazon's servers return errors with too many
            rapid requests,
        // try sleeping for a few seconds in between if
            the API returns an error
        try {
                System.out.println("Sleeping ...");
                System.out.println("———————");

                TimeUnit.SECONDS.sleep(10);

        } catch (InterruptedException e) {
                //Handle exception
                System.out.println("DELAY ERROR!");
                        }
```

```
                }
        }
}

public static String removeLastChar(String input) {
        if (input == null || input.length() == 0) {
                return input;
        }
        return input.substring(0, input.length()-1);
}

public static String removeFirstChar(String input)
{
        if (input == null || input.length() == 0) {
                return input;
        }
        return input.substring(1);
}

public static String getDate(){

        // Create an instance of SimpleDateFormat used
            for formatting
// the string representation of date (month/day/year)
DateFormat df = new SimpleDateFormat("ddMM");

// Get the date today using Calendar object.
Date today = Calendar.getInstance().getTime();

// Using DateFormat format method we can create a string
// representation of a date with the defined format.
        String reportDate = df.format(today);

        return reportDate;
}

public static void itemSearch(int lookupType, String
    filename){

        //      Set serviceporttype and secret key to
            sign requests
//  Instantiate ItemElement with keys and associate tag
AWSECommerceService service = new AWSECommerceService();
service.setHandlerResolver(new AwsHandlerResolver(
    SECRET_KEY)); // important
AWSECommerceServicePortType port = service.
    getAWSECommerceServicePort();
```

```
// Instantiate Itemsearch element, set it with the
    correct keys
com.ECS.client.jax.ItemSearch ItemElement = new com.ECS.
    client.jax.ItemSearch();
ItemElement.setAWSAccessKeyId(AWS_KEY);
ItemElement.setAssociateTag(ASSOCIATE_TAG);

// Get the operation object:
// Set standard attributes, such as response group and
    sorting method
com.ECS.client.jax.ItemSearchRequest itemRequest = new
    com.ECS.client.jax.ItemSearchRequest();
itemRequest.getResponseGroup().add("SalesRank");
itemRequest.getResponseGroup().add("ItemAttributes");
itemRequest.getResponseGroup().add("OfferSummary");
itemRequest.setSort("salesrank");

// Create the input object
Input input = new Input(filename, lookupType);

// Load the input list
List<String[]> inputList = input.getInputList();

System.out.println("————");
System.out.println("API request initiated.");
System.out.println("————");

// Loop through the input list
// Search is performed for every line in input list and
    stored to file
// Every line has 10 pages of results, search is repeated
     for each page
for (ListIterator<String[]> iter = inputList.listIterator
    (); iter.hasNext(); ) {
        String[] element = iter.next();

        // Read SearchIndex, Browsenode and
            BrowsenodeName
itemRequest.setSearchIndex(element[0]);
itemRequest.setBrowseNode(element[1]);
String browseNodeName = element[2];

String outputTarget = browseNodeName;

// Clear the request to make sure you only perform one
    search at a time. IMPORTANT!
// Add the request to the ItemElement object
ItemElement.getRequest().clear();
```

```
ItemElement.getRequest().add(itemRequest);

for (BigInteger itemPage = BigInteger.valueOf(1);
                itemPage.compareTo(BigInteger.valueOf(10)
                    ) <= 0;
                itemPage = itemPage.add(BigInteger.ONE))
                    {

        // Set the page to lookup
        itemRequest.setItemPage(itemPage);

        // Call the Web service operation and store the
            response
        // in the response object:
        ItemSearchResponse response = port.itemSearch(
            ItemElement);

        // Send response object, browsenodename,
            lookuptype, and output target to output object
        // Write response to file
        Output output = new Output(response, outputTarget
            , browseNodeName, lookupType, itemPage);
        output.writeToFile();

        // Nullify output before next iteration
        output = null;
}
System.out.println("———————");
System.out.println(outputTarget + " completed.");
System.out.println("———————");
}
System.out.println("API request completed.");
System.out.println("———————");
}

public static void itemLookup(int lookupType, String[]
    lookupElement){

        //              Set serviceporttype and secret
            key to sign requests
//      Instantiate ItemElement with keys and associate
    tag
AWSECommerceService service = new AWSECommerceService();
service.setHandlerResolver(new AwsHandlerResolver(
    SECRET_KEY)); // important
AWSECommerceServicePortType port = service.
    getAWSECommerceServicePort();
```

```
// Instantiate Itemlookup element, set it with the
    correct keys
com.ECS.client.jax.ItemLookup ItemElement = new com.ECS.
    client.jax.ItemLookup();
ItemElement.setAWSAccessKeyId(AWS_KEY);
ItemElement.setAssociateTag(ASSOCIATE_TAG);

// Get the operation object:
// Set standard attributes, such as response group and
    sorting method
com.ECS.client.jax.ItemLookupRequest itemRequest = new
    com.ECS.client.jax.ItemLookupRequest();
itemRequest.getResponseGroup().add("Large");
itemRequest.getResponseGroup().add("OfferSummary");
itemRequest.setIdType("ASIN");

// Load the response list
List<ItemLookupResponse> responseList = new ArrayList<
    ItemLookupResponse>();

// Create the input object
Input input = new Input(lookupElement[0], lookupType);

// Load the input list
List<String[]> inputList = input.getInputList();


// Loop through the input list
// Search is performed for every 10 lines of input list
    and stored to file
for (int i = 0; i < inputList.size(); i = i+10) {
        String lookupString = null;
        int iter = 0;

        // Create string with the 10 ASINs to lookup
while ((i + iter) <= (inputList.size() - 1) && iter < 10)
     {

        String[] element = inputList.get(i + iter);

        if ( iter == 0){
                lookupString = element[0];
        }else{
                lookupString = lookupString + "," +
                    element[0];
        }
        iter++;
}
```

```
// Clear the list of ASINs from the previous lookup.
    IMPORTANT!
// The script will not work very long otherwise..
// Add the lookup string to the itemRequest
itemRequest.getItemId().clear();
itemRequest.getItemId().add(lookupString);

// Clear the request to make sure you only perform one
    search at a time. IMPORTANT!
// Add the request to the ItemElement object
ItemElement.getRequest().clear();
ItemElement.getRequest().add(itemRequest);


// Call the Web service operation and store the response
// in the response object:
// Store the response to the response list
                        com.ECS.client.jax.
                            ItemLookupResponse response =
                            port.itemLookup(ItemElement);
                        responseList.add(response);
                }
                Output output = new Output(responseList,
                    lookupElement[1], lookupType);
                output.writeToFile();
        }
}
```

B.2.2 *Random product generator class*

When generating lists of random products, the open API provided by Wordnik (2014) was used to fetch random search terms. The script fetched a total of 1000 words at a time, to limit the amount of requests to Wordnik's servers. An unfortunate oversight can be seen in the code, where the check for duplicates is only performed for the existing list of top 100 products, not for the fetched products during that session. As a result we experienced some duplicates from the script.

Listing B.5: Class for generating lists of random products

```
package requests;

import com.ECS.client.jax.AWSECommerceService;
import com.ECS.client.jax.AWSECommerceServicePortType;
import com.ECS.client.jax.Item;
import com.ECS.client.jax.Items;
```

```
import com.wordnik.client.api.*;
import com.wordnik.client.model.*;
import com.wordnik.client.common.*;

import java.util.ArrayList;
import java.util.Collections;
import java.util.Iterator;
import java.util.List;
import java.util.ListIterator;
import java.util.concurrent.TimeUnit;

public class RndASINGenerator {

public static String WORDNIK_API_KEY = "REMOVED FOR
    SECURITY PURPOSES";
public static String SECRET_KEY = "REMOVED FOR SECURITY
    PURPOSES";
public static String AWS_KEY = "REMOVED FOR SECURITY
    PURPOSES";
public static String ASSOCIATE_TAG = "REMOVED FOR
    SECURITY PURPOSES";

//      If Amazon's servers start returning errors, use a
     delay.
//      Delay length is determined in the search method
public static boolean DELAY = false;

public static void main(String[] args) {
//      Define file describing lookups and lookuptype.
    Type is 1 for top 100, 2 for single ASINs
//      Define outputTarget for lookup searches
String lookupFile = "lookup/BrowseNodes-Dupes2.csv";
int lookupType = 3;
Input lookupList = new Input(lookupFile, lookupType);

List<String[]> browseNodeList = lookupList.getInputList()
    ;

for (ListIterator<String[]> iter = browseNodeList.
    listIterator(); iter.hasNext(); ) {
        String[] element = iter.next();

        String browseNode = element[0];
        String outputTarget = element[1];
        String searchIndex = element[2];
        String top100 = element[4];
```

```
        List<String> asins = makeAsins(searchIndex,
            browseNode, outputTarget, top100);

        Output output = new Output(asins, outputTarget,
            browseNode, lookupType);
        output.writeToFile();

        System.out.println("————");
        System.out.println("Random ASINs successfully
            generated for category: " + outputTarget);
        System.out.println("————");
}
System.out.println("————");
System.out.println("Random ASIN generation has been
    completed.");
System.out.println("————");
}

public static List<String> getRndWords(){

        String key = WORDNIK_API_KEY;
        List<String> response = new ArrayList<String>();

        String include = "noun,adjective,verb,adverb";
try {
        WordsApi api = new WordsApi();
        api.getInvoker().addDefaultHeader("api_key", key)
            ;
List<WordObject> random = api.getRandomWords(
                include,
                null,
                null,
                null,
                null,
                null,
                null,
                null,
                null,
                null,
                null,
                1000); //Every search collects 1000 words
                    .

        for ( WordObject words : random){
                response.add(words.getWord());
        }
}
catch (ApiException e) {
```

```
        e.printStackTrace();
}
//Randomize the list of random words, originally returned
     in alphabetical order
Collections.shuffle(response);
        return response;
}

public static List<String> makeAsins(String searchIndex,
    String browseNode, String outputTarget, String top100)
    {

        boolean done = false;
        int limit = 100;

        List<String> asins = new ArrayList<String>();

        Input top100Input = new Input(top100, 3);
        List<String[]> top100ListArray = top100Input.
            getInputList();
        List<String> top100List = new ArrayList<String>()
            ;

        for (ListIterator<String[]> iter =
            top100ListArray.listIterator(); iter.hasNext()
            ; ) {
                String[] element = iter.next();

                top100List.add(element[0]);
        }

// Collect new random words for as long as (done = false)
while (done == false){

        List<String> rndWords = getRndWords();

// Iterate over current list of random words for as long
    as the list remains shorter than 100
Iterator<String> iter = rndWords.iterator();
while (asins.size() < limit && iter.hasNext()){
        String current = iter.next();

        String ASIN = search(current, browseNode,
            searchIndex, outputTarget, asins.size(),
            top100List);

        if( ASIN != null ){
                asins.add(ASIN);
```

```
                System.out.println("[Added ASIN (" +
                    asins.size() + "/" + limit + ") : " +
                    ASIN + "]");

                                if (asins.size() == limit
                                    ){
                                        done = true;
                                        break;
                                }
                        }
                }
        }
        return asins;
}

public static String search(String randomWord,
                String browseNode,
                String searchIndex,
                String outputTarget,
                int found,
                List<String> top100){

        System.out.println("—— " + outputTarget + " (" +
             found + "/100)" +" ——");

if (DELAY == true){
        try {

                System.out.println("Loading search ...");

        TimeUnit.SECONDS.sleep(2);

} catch (InterruptedException e) {
        //Handle exception
        System.out.println("DELAY ERROR!");
        }
}

System.out.println(randomWord);

String ASIN;

//      Set serviceporttype and secret key to sign
    requests
//    Instantiate ItemElement with keys and associate tag
AWSECommerceService service = new AWSECommerceService();
```

```
service.setHandlerResolver(new AwsHandlerResolver(
    SECRET_KEY)); // important
AWSECommerceServicePortType port = service.
    getAWSECommerceServicePort();

// Instantiate Itemsearch element, set it with the
    correct keys
com.ECS.client.jax.ItemSearch ItemElement = new com.ECS.
    client.jax.ItemSearch();
ItemElement.setAWSAccessKeyId(AWS_KEY);
ItemElement.setAssociateTag(ASSOCIATE_TAG);

// Get the operation object:
// Set response group, searchindex and browsenode. Add
    request element to itemElement
com.ECS.client.jax.ItemSearchRequest itemRequest = new
    com.ECS.client.jax.ItemSearchRequest();

itemRequest.getResponseGroup().add("Small");
itemRequest.getResponseGroup().add("SalesRank");
itemRequest.setSearchIndex(searchIndex);
itemRequest.setBrowseNode(browseNode);
itemRequest.setKeywords(randomWord);
itemRequest.setAvailability("Available");

ItemElement.getRequest().add(itemRequest);

// Call the Web service operation and store the response
// in the response object:
try{
        com.ECS.client.jax.ItemSearchResponse response =
            port.itemSearch(ItemElement);

        for (Items itemList : response.getItems()) {

                for (Item itemObj : itemList.getItem()) {

                        int salesRank = 0;

                        try{
                                salesRank = Integer.
                                    parseInt(itemObj.
                                    getSalesRank());
                        }
                        catch(Exception e){
                                salesRank = 0;
                        }
```

```
                        if ( itemObj.getSalesRank() !=
                            null && salesRank > 100 &&
                            salesRank <= 40000){
                                ASIN = itemObj.getASIN();
                                if( dupeCheck(top100,
                                    ASIN) == false){
                                        return ASIN;
                                }
                        }
                }
        }
}catch(Exception e){
        System.out.println("[Search failed]");
}
System.out.println("[No results]");
                return null;
        }

        public static Boolean dupeCheck(List<String>
            dupes, String ASIN) {
                Boolean result = false;

                for (String product : dupes) {
                        if ( product.contains(ASIN) ){
                                result = true;
                        }
                }
                return result;
        }
}
```

Amazon.com (2014a). *Amazon.com: About Badges*. URL: http://www.amazon.com/gp/help/customer/display.html/ref=hp_left_v4_sib?ie=UTF8&nodeId=14279681 (visited on May 10, 2014).

— (2014b). *Amazon.com Best Sellers: The most popular items on Amazon*. URL: http://www.amazon.com/Best-Sellers/zgbs/ref=zg_bs_unv_m_0_m_1 (visited on May 12, 2014).

— (2014c). *Product Advertising API*. URL: https://affiliate-program.amazon.com/gp/advertising/api/detail/main.html (visited on May 12, 2014).

Anderson, Eugene W (1998). "Customer satisfaction and word of mouth." In: *Journal of service research* 1.1, pp. 5–17.

Arndt, Johan (1967). "Role of Product-Related Conversations in the Diffusion of a New Product." In: *Journal of Marketing Research (JMR)* 4.3.

Ba, Sulin and Paul A Pavlou (2002). "Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior." In: *MIS quarterly*, pp. 243–268.

Bambauer-Sachse, Silke and Sabrina Mangold (2013). "Do consumers still believe what is said in online product reviews? A persuasion knowledge approach." In: *Journal of Retailing and Consumer Services* 20.4, pp. 373–381.

Bjering, E and L.J. Havro (2013). "Word of mouth in a digital age: The Case of Online Consumer Reviews." Pre-diploma Thesis. Norwegian University of Science and Technology.

BusinessWire (2010). *2010 Social Shopping Study Reveals Changes in Consumers' Online Shopping Habits and Usage of Customer Reviews*. URL: http://www.businesswire.com/news/home/20100503005110/en/2010-Social-Shopping-Study-Reveals-Consumers.

Chen, Yubo and Jinhong Xie (2008). "Online consumer review: Word-of-mouth as a new element of marketing communication mix." In: *Management Science* 54.3, pp. 477–491.

Chevalier, Judith A and Dina Mayzlin (2006). "The effect of word of mouth on sales: Online book reviews." In: *Journal of marketing research* 43.3, pp. 345–354.

Chevalier, Judith and Austan Goolsbee (2003). "Measuring prices and price competition online: Amazon. com and BarnesandNoble. com." In: *Quantitative Marketing and Economics* 1.2, pp. 203–222.

Clemons, Eric K, Guodong Gordon Gao, and Lorin M Hitt (2006). "When online reviews meet hyperdifferentiation: A study of the craft beer industry." In: *Journal of Management Information Systems* 23.2, pp. 149–171.

Darby, Michael R and Edi Karni (1973). "Free competition and the optimal amount of fraud." In: *Journal of Law and Economics* 16, p. 67.

Dellarocas, Chrysanthos (2006). "Strategic manipulation of internet opinion forums: Implications for consumers and firms." In: *Management Science* 52.10, pp. 1577–1593.

Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F Awad (2007). "Exploring the value of online product reviews in forecasting sales: The case of motion pictures." In: *Journal of Interactive marketing* 21.4, pp. 23–45.

Duan, Wenjing, Bin Gu, and Andrew B Whinston (2008). "Do online reviews matter?—An empirical investigation of panel data." In: *Decision Support Systems* 45.4, pp. 1007–1016.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008). "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets." In: *Information Systems Research* 19.3, pp. 291–313.

Hartigan, John A and PM Hartigan (1985). "The dip test of unimodality." In: *The Annals of Statistics*, pp. 70–84.

He, Ben, Craig Macdonald, and Iadh Ounis (2008). "Ranking opinionated blog posts using OpinionFinder." In: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, pp. 727–728.

Hennig-Thurau, Thorsten et al. (2004). "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?" In: *Journal of interactive marketing* 18.1, pp. 38–52.

Hu, Nan, Paul A Pavlou, and Jennifer Zhang (2006). "Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication." In: *Proceedings of the 7th ACM conference on Electronic commerce*. ACM, pp. 324–330.

Hu, Nan, Ling Liu, and Jie Jennifer Zhang (2008). "Do online reviews affect product sales? The role of reviewer characteristics and temporal effects." In: *Information Technology and Management* 9.3, pp. 201–214.

Hu, Nan, Ling Liu, and Vallabh Sambamurthy (2011). "Fraud detection in online consumer reviews." In: *Decision Support Systems* 50.3, pp. 614–626.

Kohn, Bob (2014). *How book publishers can beat Amazon*. URL: http://timesofindia.indiatimes.com/tech/tech-news/How-book-publishers-can-beat-Amazon/articleshow/35848068.cms (visited on June 3, 2014).

Li, Xinxin and Lorin M Hitt (2008). "Self-selection and information role of online product reviews." In: *Information Systems Research* 19.4, pp. 456–474.

Liddy, Elizabeth D (2001). *Natural language processing*. URL: http://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub (visited on May 22, 2014).

Liu, Yong (2006). "Word of mouth for movies: Its dynamics and impact on box office revenue." In: *Journal of marketing* 70.3, pp. 74–89.

Luca, Michael (2011). *Reviews, reputation, and revenue: The case of Yelp. com*. Tech. rep. Harvard Business School.

MPQA (2011). *OpinionFinder 2.x Release Page*. URL: http://mpqa.cs.pitt.edu/opinionfinder/opinionfinder_2/ (visited on May 22, 2014).

Mudambi, Susan M and David Schuff (2010). "What makes a helpful online review? A study of customer reviews on Amazon.com." In: *MIS quarterly* 34.1, pp. 185–200.

Nelson, Phillip (1970). "Information and consumer behavior." In: *The Journal of Political Economy*, pp. 311–329.

Pang, Bo and Lillian Lee (2008). "Opinion mining and sentiment analysis." In: *Foundations and trends in information retrieval* 2.1-2, pp. 1–135.

Park, Cheol and Thae Min Lee (2009). "Information direction, website reputation and eWOM effect: A moderating role of product type." In: *Journal of Business Research* 62.1, pp. 61–67.

Racherla, Pradeep, Munir Mandviwalla, and Daniel J Connolly (2012). "Factors affecting consumers' trust in online product reviews." In: *Journal of Consumer Behaviour* 11.2, pp. 94–104.

Riloff, Ellen and Janyce Wiebe (2003). "Learning extraction patterns for subjective expressions." In: *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, pp. 105–112.

Schnapp, Madeline and Tim Allwine (2001). "Mining of book data from Amazon.com." In: *Presentation at the UCB/SIMS web mining conference, http://www. sims.berkeley.edu/resources/affiliates/workshops/webmining/slides/ORA.ppt*.

Senecal, Sylvain and Jacques Nantel (2004). "The influence of online product recommendations on consumers' online choices." In: *Journal of retailing* 80.2, pp. 159–169.

Various developers (2014a). *AWSHandlerResolver.java*. URL: https://forums.aws.amazon.com/servlet/JiveServlet/download/9-35999-143146-2680/awshandlerresolver.java (visited on Mar. 18, 2014).

— (2014b). *OpenCSV - CSV (comma-separated values) parser library for Java*. URL: https://code.google.com/p/opencsv/ (visited on May 18, 2014).

Wang, Zhongmin (2010). "Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews." In: *The BE Journal of Economic Analysis & Policy* 10.1.

Whyte Jr, William H (1954). "The web of word of mouth." In: *Fortune* 50.1954, pp. 140–143.

Wiebe, Janyce and Ellen Riloff (2005). "Creating subjective and objective sentence classifiers from unannotated texts." In: *Computational Linguistics and Intelligent Text Processing*. Springer, pp. 486–497.

Wikipedia (2014a). *Amazon.com — Wikipedia, The Free Encyclopedia*. URL: http://en.wikipedia.org/w/index.php?title=Amazon.com&oldid=608242886 (visited on May 12, 2014).

— (2014b). *Amazon.com — Wikipedia, The Free Encyclopedia*. URL: http://en.wikipedia.org/w/index.php?title=Amazon.com&oldid=608242886 (visited on May 12, 2014).

— (2014c). *Machine learning — Wikipedia, The Free Encyclopedia*. URL: http://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=609547471 (visited on May 12, 2014).

— (2014d). *SOAP — Wikipedia, The Free Encyclopedia*. URL: http://en.wikipedia.org/w/index.php?title=SOAP&oldid=607539818" (visited on Apr. 7, 2012).

Wordnik (2014). *Wordnik - All the words*. URL: https://www.wordnik.com/ (visited on May 12, 2014).

Yin, Robert K (2009). *Case study research: Design and methods*. Vol. 5. sage.

Zhang, Jason Q, Georgiana Craciun, and Dongwoo Shin (2010). "When does electronic word-of-mouth matter? A study of consumer product reviews." In: *Journal of Business Research* 63.12, pp. 1336–1341.

Zhu, Feng and Xiaoquan Zhang (2010). "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics." In: *Journal of Marketing* 74.2, pp. 133–148.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

The style has been modified and updated by Lars Jaakko Havro to accomodate Futura as the typeface for headings and labels.

Final Version as of June 11, 2014 (`classicthesis` Norwegian University of Science).