



Received: 18 April 2017  
Accepted: 19 July 2017  
First Published: 18 August 2017

\*Corresponding author: Øystein Moen,  
Department of Industrial Economics  
and Technology Management,  
Norwegian University of Science and  
Technology, Alfred Getz vei 3, 7491  
Trondheim, Norway  
E-mail: [Oeystein.moen@iot.ntnu.no](mailto:Oeystein.moen@iot.ntnu.no)

Reviewing editor:  
Len Tiu Wright, University of  
Huddersfield, UK

Additional information is available at  
the end of the article

## MARKETING | RESEARCH ARTICLE

# Online consumers reviews: Examining the moderating effects of product type and product popularity on the review impact on sales

Øystein Moen<sup>1\*</sup>, Lars Jaako Havro<sup>1</sup> and Einar Bjering<sup>1</sup>

**Abstract:** This paper aims to study the role product category plays as a moderating factor in online reviews, by introducing a novel method for product category classification using natural language processing (NLP). The study includes a wide variety of categories, based on a high number of products and number of reviews. The data-set presented includes 1.1 million unique reviews from 4,600 products in 30 different product categories. 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. This paper also reveals some evidence of rating biases as 60% of the 1.1 million reviews in our data-set show signs of bimodality. Based on the results we present “the review impact continuum”, a model mapping degree of subjectivity and product popularity enabling managers to assess the expected impact of online consumer reviews for their products.

### ABOUT THE AUTHORS

Øystein Moen is a professor at the Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology. He has worked within international marketing, both focusing internationalization processes of newly established firms and the use of information and communication technology in marketing activities. He has been in charge of several large research projects related to the maritime and energy industry in Norway and have published in journals, such as *International Business Review*, *Journal of International Marketing*, *Journal of Small Business Management*, *European Journal of Marketing* and the *International Marketing Review*. This manuscript is co-authored with researchers Lars Jaako Havro and Einar Bjering and build on the current activity with focus on how information technology influences marketing efforts by firms.

### PUBLIC INTEREST STATEMENT

User-generated online product reviews are important for choices made by consumers and for sales and revenues for many retailers. Our study investigate how product category and product popularity mediates the effect of reviews on sales. We divide product categories from search products where objective evaluations are possible before purchases (in example, USB memory sticks is evaluated based on speed and memory capacity) to experience products difficult to evaluate prior to purchase (in example, subjective categories as fashion clothes or movies). Our results document systematic differences in review effects on sales based on product type and popularity, subjectively evaluated products as well as less popular products see the largest relative effect of online product reviews. Based on the results we present the review impact continuum, a model for assessing the expected influence of user online reviews on sales based on product popularity and the type of product.

**Subjects: Business, Management and Accounting; Marketing; Advertising; Internet/Digital Marketing/e-Marketing; Marketing Communications; Marketing Management**

**Keywords: online product reviews; product type; popularity; sales**

## 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, like Amazon and Netflix use it to enhance their core model. The wide-spread 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 (Chevalier & Mayzlin, 2006; Dellarocas, Zhang, & Awad, 2007). In addition, a relationship between review volume and sales (Duan, Gu, & Whinston, 2008; Liu, 2006) has been established.

However, much of the research focus on a single category of products. For instance, Chevalier and Mayzlin (2006), Forman, Ghose, and Wiesenfeld (2008) and 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 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. 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 differences between review systems.

This paper aims to study the role product category plays as a moderating factor in 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.

The study includes a wide variety of categories, based on a high number of products and number of reviews. The data-set presented includes 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, jewellery and hardware.

### 1.1. Word-of-mouth and electronic word-of-mouth

In the marketing literature, an important concept is that of word-of-mouth communication (WOM). WOM may be defined as informal communications directed at other consumers about the ownership, usage or characteristics of particular goods and services or their sellers (Hu, Pavlou, & Zhang, 2006). WOM is a phenomenon that has been discussed for a century, with research attention gaining ground in the 1950s (Whyte, 1954). The effect has since been thoroughly researched, and is believed to be a strong influence on consumers purchase intentions (Arndt, 1967; Henning, Al-Thurau, Gwinner, Walsh, & Gremler, 2004).

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 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 as described by Dellarocas and Narayan (2006).

This new form of WOM communication has been dubbed electronic word-of-mouth, or eWOM (Henning et al., 2004; Racherla, Mandviwalla, & Connolly, 2012; Zhang, Craciun, & Shin, 2010). Henning et al. (2004) offer the following definition of eWOM: “a 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” (Henning et al., 2004, p. 39).

### 1.2. Types of products

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. He further predicts that the recommendations of others will be used more and have greater impact for purchases of experience goods than search goods.

## 2. Hypotheses

The most obvious question when discussing online consumer reviews is of their efficacy, or rather, whether a positive review leads to more sales. 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, Liu and Zhang (2008), who analysed 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, Liu and Chang (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) used data from Yahoo Movies and Boxofficemojo.com and did not find that the ratings of online user reviews had 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 1: An increase in average rating on a site is associated with increased sales.

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.

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) did also find 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 hypothesis:

Hypothesis 2: 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 without 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) constructed an experiment allowing subjects to make purchasing decisions after receiving product recommendations, and did find that recommendations for experience products are significantly more influential than recommendations for search products. Park and Lee (2009) also showed 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 hypothesis becomes:

Hypothesis 3: 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. 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 & Schuff, 2010). 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 its 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.

Hypothesis 4: 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. (2008) 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 found 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 fifth hypothesis is:

Hypothesis 5: A comparatively high number of reviews on a site is associated with comparatively higher sales on that site.

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, 2008; Henning 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 6: The distribution of ratings for a product tends to be bimodal, with the low and high end of the scale as local modes.

### 3. Methodology

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 are 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, among others Schnapp and Allwine (2001) and Chevalier and Goolsbee (2003). Our statistical models build on methods presented in prior research as Chevalier and Mayzlin (2006), Li and Hitt (2008) and Mudambi and Schuff (2010). Amazon is selected as it is known as the world's largest online retailer and carries millions of products over hundreds of different categories. Combining in the fact that they have 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.

One of the services offered by Amazon Web Services is programmatic 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. In order to access the Amazon databases a set of methods were written in Java. With Java, Amazon allows developers access to the API using the SOAP request protocol.

We selected 30 different product categories, for each of these we included the 100 best selling products and 100 randomly picked products (not part of the top 100 products). 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. In order to choose the random 100 products a java script was used with random words (Wordnik/getRandomWords) and products was selected.

### **3.1. Collecting reviews**

A VBA script was created to download the required data. Our chosen method for measuring the subjectivity of reviews is through computerized sentiment analysis with subjectivity classification. For analysing the sentiments of reviews, we employed the OpinionFinder library, developed by researchers at the University of Pittsburgh, Cornell University and the University of Utah. The set of OpinionFinder classifiers have been widely used in previous research, reporting good results in classifying subjectivity (see for instance, He, MacDonald, & Ounis, 2008). The subjectivity classifiers included in the toolkit are based on the work by Riloff and Wiebe (2003) and Wiebe and Riloff (2005). The OpinionFinder toolkit includes two separate subjectivity classifiers. The first classifier is model-based, meaning it is based on a model that can be trained through machine learning. This classifier has a reported accuracy of 76%, subjective precision 79% and subjective recall of 76%. 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. The result is therefore, three classifications, objective, subjective or unknown. When calculating fraction of subjectivity with these result, we disregarded those classified as unknown, and employed the number of subjective sentences divided by those classified as either subjective or objective.

### **3.2. Brief description of the data-set**

A total of 1,147,488 primary reviews are included from the 30 product categories used as described in Table 1. Primary reviews is unique reviews while secondary reviews are duplicates of one in the primary set attached to a non-duplicate ASIN (Amazon Standard Identification Number) with distinct sales ranks. As an example, in the movies category one movie may be released on different platforms (DVD, Blu-ray, Amazon Instant), each of these versions may be assigned unique ASINs but reviews are shared.

They primary reviews are divided in 986,344 reviews of the top 100 products and 161,144 reviews from the random 100 products. The numbers of reviews vary from hobby fabrics (1,838) and screws (1,919) to movies (103,586) and books (216,361). The mean numbers of reviews per product also varies with movies as highest and envelopes as smallest. Reviews length was in average 160 characters, with 99% of reviews with less than 2,705 characters. Length mode is between 111 and 121 characters for all 30 product categories. Mean length varies from 650 (digital cameras) to 205 (jewellery). The product ratings are not evenly distributed, mean rating was 4.27 and 64.9% of all ratings are five-stars indicating many positive reviews. The subjective scale score (rule based) varies considerable as expected, with highest subjective scores on books and movies (0.22) and lowest scores on hard drives and copy paper (0.05).

**Table 1. Number of primary and secondary reviews for all categories**

| 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    |
| Jewellery         | 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 and 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 |

\*No random product list generated.

### 3.3. Model specification

For hypotheses 1 and 2 we defined two multiple regression models that aimed 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 the relationship between  $\ln$  SALES and  $\ln$  SALESRANK is approximately linear, making  $\ln$  SALESRANK an adequate substitute. A product's sales rank on Amazon is likely affected by several variables. Building on work by Chevalier and Mayzlin (2006), Luca (2011) and Forman et al. (2008), our regression model assumes that the sales rank 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 (Regression 1):

$$\ln \text{SALESRANK}_{pt} = \alpha + \beta_1 \ln \text{PRICE}_{pt} + \beta_2 \ln \text{AVGRATING}_{pt} + \beta_3 \ln \text{NUMREVIEWS}_{pt} + \beta_4 \ln \text{PRODUCTTYPE}_p + \beta_2 \text{AVGRATING}_{pt} \times \text{PRODUCTTYPE}_p + \varepsilon$$

With subjectivity variables (Regression 2):

$$\ln \text{SALESRANK}_{pt} = \alpha + \beta_1 \ln \text{PRICE}_{pt} + \beta_2 \ln \text{AVGRATING}_{pt} + \beta_3 \ln \text{NUMREVIEWS}_{pt} + \beta_4 \ln \text{SUBJECTIVITY}_p + \beta_2 \text{AVGRATING}_{pt} \times \text{SUBJECTIVITY}_p + \varepsilon$$

The subscript  $p$  denotes product and  $t$  denote time.  $\text{PRICE}_{pt}$  thus, denotes the price for product  $p$  at time  $t$ . The coefficient  $\beta_1$  may therefore, be seen as a measure of the effect of the product price on  $\ln \text{SALESRANK}$ , or in effect, a proxy for the price elasticity of the product.

$\text{AVGRATING}_{pt}$  represents the average star rating for a product  $p$  at time  $t$ . Since the sales rank data have been extracted at daily intervals, the average rating for any specific day includes all reviews submitted before or at that specific date. The coefficient  $\beta_2$  thus, represents the effect of the average star rating on the sales rank.

The variable  $\text{NUMREVIEWS}_{pt}$  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 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 percentage change in sales rank.

Further, we include the dummy variable  $\text{PRODUCTTYPE}_p$  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 this variable will only be used with those products.

To test for the interaction between product type and the average rating, we include the compound variable  $\text{AVGRATING}_{pt} \times \text{PRODUCTTYPE}_p$ . 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  $\text{AVGRATING}_{pt} \times \text{SUBJECTIVITY}_p$  variable is included to test if the degree of subjectivity really moderates the effect of ratings on sales.

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 sales rank. Substituting these two new variables into Regression 1 we get Regression 3:

$$\ln \text{SALESRANK}_{pt} = \alpha + \beta_1 \ln \text{PRICE}_{pt} + \beta_2 \ln \text{NUMREVIEWS}_{pt} + \beta_3 \text{PRODUCTTYPE}_p + \gamma_1 \ln \text{ONESTAR}_{pt} + \gamma_2 \text{FIVESTAR}_{pt} + \gamma_3 \text{ONESTAR}_{pt} \times \text{PRODUCTTYPE}_p + \gamma_4 \text{FIVESTAR}_{pt} \times \text{PRODUCTTYPE}_p + \varepsilon$$

Then,  $\text{ONESTAR}_{pt}$  now denotes the fraction of reviews with a rating of one star, and  $\text{FIVESTAR}_{pt}$  denotes the fraction of reviews with a rating of five stars. The coefficients  $\gamma_1$  and  $\gamma_2$  represent the effects of the respective fractional variables, or specifically, to what degree the one-star and five-star reviews affect sales.

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 Regression 1 (disregarding coefficients) the following relation appears:

$$\begin{aligned} \ln \text{SALESRANK}_{pt} - \ln \text{SALESRANK}_{p0} &= \ln \text{PRICE}_{pt} - \ln \text{PRICE}_{p0} + \ln \text{NUMREVIEWS}_{pt} \\ &\quad - \ln \text{NUMREVIEWS}_{p0} + \text{PRODUCTTYPE}_p - \text{PRODUCTTYPE}_{p0} \\ &\quad + \text{AVGRATING}_{pt} - \text{AVGRATING}_{p0} + \text{AVGRATING}_{pt} \\ &\quad \times \text{PRODUCTTYPE}_p - \text{AVGRATING}_{p0} \times \text{PRODUCTTYPE}_p + \varepsilon \end{aligned}$$

Performing all operations, this gives us Regression 4:

$$\begin{aligned} \Delta \ln \text{SALESRANK}_{pt} &= \Delta \ln \text{PRICE}_{pt} + \Delta \ln \text{NUMREVIEWS}_{pt} + \Delta \text{AVGRATING}_{pt} + \Delta \text{AVGRATING}_{pt} \\ &\quad \times \text{PRODUCTTYPE}_p + \varepsilon \end{aligned}$$

We see that the variable  $\text{PRODUCTTYPE}_p$  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 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.

### 3.4. Rating distributions

To test for the existence of an under-reporting bias and thus, a U-shaped rating distribution in hypothesis 6, we 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 hypothesized U-shape, showing a tendency towards bimodality, we acknowledge that the number of one-star ratings need to be larger than two-star rating, and the number of five stars need to be larger than four stars. Lastly, we want to exclude those with at spike of ratings in the middle of the distribution. Our test is as follows:

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

where  $f1$ ,  $f2$ ,  $f3$ ,  $f4$  and  $f5$  represents the frequency of the one-, two-, three-, four- and five-star ratings, 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. Results

This section details the testing of hypotheses, we begin by testing the suppositions with our static regression model, before delving deeper with the difference regression model. 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. 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 data-set was considered to contain a sufficient number of observations. For the same reason, we have not included products with missing reviews.

### 4.1. Static regression model

The results of the regression using our static model are shown in Tables 2 and 3.

**Table 2. Regression results, sales rank and average rating**

| Variable            | Coeff. | Std. Err. | Std. Coeff. | t        | Sig.  |
|---------------------|--------|-----------|-------------|----------|-------|
| (Constant)          | 11.919 | 0.072     | -           | 164.429  | 0.000 |
| ln (price)          | 0.089  | 0.004     | 0.063       | 20.216   | 0.000 |
| ln (No. of reviews) | -0.668 | 0.005     | -0.448      | -143.200 | 0.000 |
| Average Rating      | -0.286 | 0.016     | -0.056      | -17.984  | 0.000 |

DEP. ln SALESRANK,  $R^2 = 0.206$ ,  $N = 81,245$

**Table 3. Regression results, sales rank and star fraction**

| Variable            | Coeff. | Std. Err. | Std. Coeff. | t        | Sig.  |
|---------------------|--------|-----------|-------------|----------|-------|
| (Constant)          | 11.133 | 0.050     | -           | 224.073  | 0.000 |
| ln (price)          | 0.088  | 0.004     | 0.063       | 20.087   | 0.000 |
| ln (No. of reviews) | -0.668 | 0.005     | -0.448      | -143.197 | 0.000 |
| Frac. one star      | 0.305  | 0.113     | 0.011       | 2.695    | 0.007 |
| Frac. five star     | -0.704 | 0.057     | -0.050      | -12.287  | 0.000 |

DEP. ln SALESRANK,  $R^2 = 0.206$ ,  $N = 81,245$

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

When introducing the interaction terms of product category and rating as well as subjectivity and rating (Tables 4 and 5), we observe that product categories have significant differences in mean sales rank.

**Table 4. Regression results: sales rank and rating by product category**

| Variable            | Coeff. | Std. Err. | Std. Coeff. | t       | Sig.  |
|---------------------|--------|-----------|-------------|---------|-------|
| (Constant)          | 12.792 | 0.113     | -           | 113.257 | 0.000 |
| ln (price)          | 0.129  | 0.007     | 0.076       | 19.051  | 0.000 |
| ln (No. of reviews) | -0.532 | 0.006     | -0.395      | -95.350 | 0.000 |
| Type                | -1.219 | 0.018     | -0.278      | -66.139 | 0.000 |
| Average rating*     | -0.44  | 0.025     | -0.093      | -17.416 | 0.000 |
| AR × type           | -0.383 | 0.037     | -0.055      | -10.217 | 0.000 |

DEP. ln SALESRANK  $R^2 = 0.330$ ,  $N = 43,725$

**Table 5. Regression results, sales rank and rating by subjectivity**

| Variable*   | Coeff. | Std. Err. | Std. Coeff. | t        | Sig.  |
|---|--------|-----------|-------------|----------|-------|
| (Constant)  | 10.923 | 0.181     | –           | 60.449   | 0.000 |
| ln (price)  | 0.092  | 0.004     | 0.066       | 21.131   | 0.000 |
| ln (No. of reviews)                                   | –0.633 | 0.005     | –0.426      | –134.407 | 0.000 |
| Average rating  | 0.038  | 0.041     | 0.007       | 0.911    | 0.362 |
| Subj. (R)   | –4.522 | 0.153     | –0.095      | –29.6    | 0.000 |
| AR × Subj. (R)  | –2.586 | 0.345     | –0.062      | –7.497   | 0.000 |
| DEP. ln SALESRANK, R <sup>2</sup> = 0.214, N = 81,149 |        |           |             |          |       |
| (Constant)  | 8.541  | 0.541     | –           | 15.789   | 0.000 |
| ln (price)  | 0.105  | 0.004     | 0.075       | 24.222   | 0.000 |
| ln (No. of reviews)                                   | –0.603 | 0.005     | –0.406      | –128.270 | 0.000 |
| Average rating**                                      | 1.095  | 0.124     | 0.217       | 8.812    | 0.000 |
| Subj. (M)   | –7.961 | 0.156     | –0.163      | –51.11   | 0.000 |
| AR × Subj. (M)**                                      | –3.762 | 0.351     | –0.264      | –10.727  | 0.000 |
| DEP. ln SALESRANK, R <sup>2</sup> = 0.231, N = 81,149 |        |           |             |          |       |

Top: Rule-based subjectivity. Bottom: Model-based subjectivity. AR has been centered for the interactive terms.

\*\*Variables experiencing strong collinearity.

Categories like books and movies that are top-level categories, will have their top 100 products occupy sales ranks 1–100, whereas a lower level category such as screws or mailing envelopes will have its 100 bestselling 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 sales rank. 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 centre 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 unreliable.

Looking at Table 4 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 sales ranks 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 3 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 sales rank than a search product with the same rating. This supports hypothesis 3, that product types moderate the effect of reviews on sales.

Performing similar regressions with our two subjectivity variables produce comparable results as presented in Table 5.

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 are inconclusive, but give some support to hypothesis 4, that subjectively evaluated products see larger effects from reviews than objectively evaluated products.

So far, we see that the static model supports hypotheses 1 and 3. In testing hypothesis 4, the subjectivity variables introduce some multi-collinearity issues, but give initial support to the notion that subjectively evaluated products see larger effects from ratings.

#### 4.2. 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  $\Delta \ln \text{SALESRANK}$ . 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 2 by regressing over our top 100 and random product sets. This produces the output in Table 6. 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 with our previous findings, the change in sign can be reasonably explained. 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.

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' initial sales ranks as well as with a binary variable indicating overall growth or decline in sales. This brings up the  $R^2$  value to around 0.500. 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 significant ( $p < 0.01$ ) effect, with a negative sign. This suggests that the effect of ratings is stronger for less

**Table 6. Regression results: change in sales rank on popularity**

| Variable*   | Coeff. | Std. Err. | Std. Coeff. | t      | Sig.  |
|---|--------|-----------|-------------|--------|-------|
| (Constant)  | 0.195  | 0.003     | -           | 57.199 | 0.000 |
| $\Delta \ln$ (price)  | 0.100  | 0.008     | 0.055       | 12.743 | 0.000 |
| $\Delta \ln$ (No. of reviews)                                     | 0.784  | 0.018     | 0.191       | 43.897 | 0.000 |
| $\Delta$ Average rating   | 0.076  | 0.044     | 0.008       | 1.730  | 0.084 |
| DEP. $\Delta \ln \text{SALESRANK}$ , $R^2 = 0.039$ , $N = 50,826$ |        |           |             |        |       |
| (Constant)  | 0.023  | 0.004     | -           | 5.907  | 0.000 |
| $\Delta \ln$ (Price)  | 0.011  | 0.015     | 0.004       | 0.694  | 0.488 |
| $\Delta \ln$ (No. of reviews)                                     | 0.567  | 0.049     | 0.068       | 11.588 | 0.000 |
| $\Delta$ Average rating   | -0.408 | 0.076     | -0.031      | -5.370 | 0.000 |
| DEP. $\Delta \ln \text{SALESRANK}$ , $R^2 = 0.006$ , $N = 29,136$ |        |           |             |        |       |

\*Top: top 100 products, bottom: random products.

popular products, supporting hypothesis 2, although further tests are necessary to conclusively determine the effect.

Exploiting the fact that the variable  $\ln$  SALESRANK now may 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 7 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).

Extending the previous argument, we perform several group-wise 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, 1,000, 10,000 as well as 100,000. The results of these regressions are summarized in Table 8. With the bestselling 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 2, 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.

**Table 7. Regression results: change in sales rank by rating and popularity**

| Variable                               | Coeff. | Std. Err. | Std. Coeff. | t      | Sig.  |
|--|--------|-----------|-------------|--------|-------|
| (Constant)                             | 0.129  | 0.003     | -           | 49.197 | 0.000 |
| $\Delta \ln$ (price)                   | 0.087  | 0.007     | 0.044       | 12.561 | 0.000 |
| $\Delta \ln$ (No. of reviews)          | 0.826  | 0.016     | 0.176       | 50.500 | 0.000 |
| $\Delta$ avg. rating $\times \ln$ (SR) | -0.028 | 0.004     | -0.022      | -6.416 | 0.000 |

DEP.  $\Delta \ln$  SALESRANK,  $R^2 = 0.033$ ,  $N = 79,961$

Note: SR: salesrank.

**Table 8. Regression results; coefficients and standardized coefficients for effect of rating, rating  $\times$  category and rating  $\times$  subjectivity for different segments of sales rank**

| Variable*                      | >1     | >100     | >1,000   | >10,000  | >100,000  |
|--------------------------------|--------|----------|----------|----------|-----------|
| $\Delta$ AR                    | -0.047 | -0.180** | -0.194** | -0.356** | -1.058**  |
| $\Delta$ AR $\times$ type      | 0.072  | -0.436** | -0.764** | -0.840** | -6.021*   |
| $\Delta$ AR $\times$ Subj. (M) | -0.123 | -0.612** | -0.734** | -1.199** | -3.258**  |
| $\Delta$ AR $\times$ Subj. (R) | 0.103  | -1.518** | -1.963** | -3.040** | -12.485** |
| $\Delta$ AR                    | -0.004 | -0.017   | -0.018   | -0.033   | -0.115    |
| $\Delta$ AR $\times$ type      | 0.005  | -0.026   | -0.039   | -0.036   | -0.050    |
| $\Delta$ AR $\times$ Subj. (M) | -0.004 | -0.020   | -0.023   | -0.037   | -0.115    |
| $\Delta$ AR $\times$ Subj. (R) | 0.001  | -0.019   | -0.062   | -0.062   | -0.124    |

DEP.  $\Delta \ln$  SALESRANK

Notes: Standardized coefficients are at the bottom of the table. AR: average ratings.

\* $p < 0.05$ .

\*\* $p < 0.01$ .

As such, it makes sense to see an increase in the effect of reviews as we exclude more and more of the bestselling products. Indeed, we see that the effect of  $\Delta AR$  grows for every step as we exclude more of the top-selling products. This supports hypothesis 2, 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. 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 1. These findings support hypotheses 2, 3 as well as 4, implying that the effect of reviews is both larger for less popular products, as well as for experience (or subjectively evaluated) products.

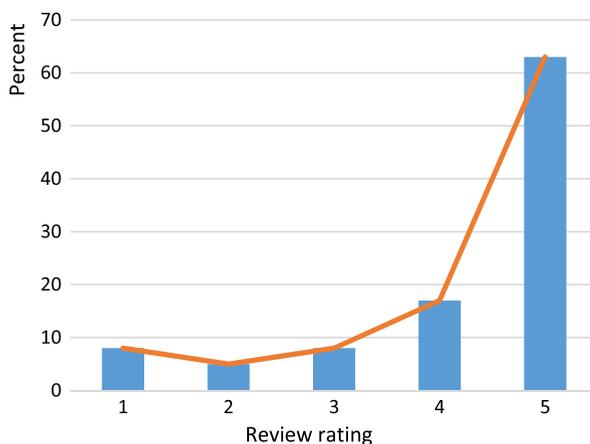
Hypothesis 6 states that the distribution of ratings tends to be bimodal, where 1 and 5 stars are minor and major modes, respectfully. To test for this, we constructed a simple set of requirements that needed to be fulfilled in order to show tendencies of bimodality. Figure 1 show the distribution of all ratings in the data-set. Visual inspection of this distribution indicate bimodality.

However, to test our hypothesis, it needs to hold on a product level. We therefore, test how many products in our data-set that show signs of bimodality. We limit our data-set to products with more than 20 reviews, which is similar to the limit set in Hu et al. (2006). Of the 3,044 products that remain in our data-set, 1,814 show signs of being bimodal 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 & 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 (one of Hu's three 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 shows that the share increases as the limit increases, but seems to converge 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. Results shows that there is a vast range between shoes at a 27% bimodality tendency to ink and toner at a 94% bimodality share.

Most digital products seem to have high shares of bimodal rating distributions, while simpler analogue 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 of the product. Further

**Figure 1. Review rating distribution.**



analysis shows that the requirement that most often fails a distribution from being classified as bimodal is the  $f_1 > f_2$  requirement. It is perhaps so that the “spike” in one-star ratings comes from users that have somehow not been able to use the product properly, and is thus, frustrated and rewards it with a one-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 6, 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. A total of 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.

## 5. Discussion

When constructing the research design, 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. The statistical analyses performed in this study 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.

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 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. We outlined our method of assigning each product category a subjectivity score, which resulted in two different subjectivity variables. 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.

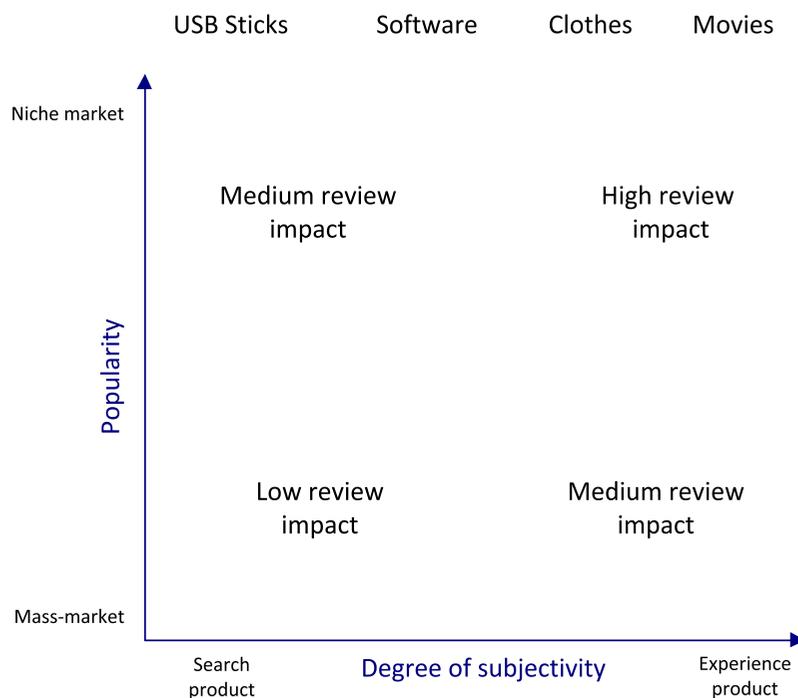
### 5.1. Introducing the review impact continuum

Based on our results, we would like to introduce a new concept: the review impact continuum (Figure 2). 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, Gao, & Hitt, 2006), video games (Zhu & Zhang, 2010), books (Chevalier & Mayzlin, 2006; Hu, Liu, & Chang, 2008; Li & Hitt, 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).

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. The popularity of the products marks the second axis in the figure.

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

Figure 2. The review impact continuum.



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 and 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 & 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.

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.

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 (Luca, 2011). 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. The movies reviewed in Duan et al. (2008) are a selection of the absolute highest grossing movies in the market. This is the extreme end of the popularity dimension, considering that many 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.

Similarly, 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.

## 6. Conclusions and implications

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.

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 standard categorization of search and experience products. Our subjectivity variable holds up remarkably well, matching previous findings, while including significantly larger sets of products and reviews, as well as products that have previously been difficult to classify.

This paper also reveals some evidence of rating biases. About 60% of the 1.1 million reviews in our data-set 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. 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.

### **6.1. 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 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 systematizing 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 assess 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.

### **6.2. Implications for managers**

The results of this study 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. 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 one-star reviews and increase the average rating for the product, positively impacting sales. Indeed, some companies have started with this type of customer responses, 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 negative WOM or exploiting positive WOM.

#### Funding

The authors received no direct funding for this research.

#### Author details

Øystein Moen<sup>1</sup>

E-mail: [Oeystein.moen@iot.ntnu.no](mailto:Oeystein.moen@iot.ntnu.no)

ORCID ID: <http://orcid.org/0000-0002-3578-9000>

Lars Jaako Havro<sup>1</sup>

E-mail: [larsjh@gmail.com](mailto:larsjh@gmail.com)

Einar Bjering<sup>1</sup>

E-mail: [ebjering@gmail.com](mailto:ebjering@gmail.com)

<sup>1</sup> Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Alfred Getz vei 3, 7491 Trondheim, Norway.

#### Citation information

Cite this article as: Online consumers reviews: Examining the moderating effects of product type and product popularity on the review impact on sales, Øystein Moen, Lars Jaako Havro & Einar Bjering, *Cogent Business & Management* (2017), 4: 1368114.

#### References

- Anderson, C. (2008). *The long tail: Why the future of business is selling less of more*. New York, NY: Hyperion Books.
- Arndt, J. (1967). The role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research*, 4, 291–295. <https://doi.org/10.2307/3149462>
- Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26, 243–268. <https://doi.org/10.2307/4132332>
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54, 477–491. <https://doi.org/10.1287/mnsc.1070.0810>
- Chevalier, J. A., & Goolsbee, A. (2003). Measuring prices and price competition online: Amazon.com and BarnesandNoble.Com. *Quantitative Marketing and Economics*, 1, 203–222. <https://doi.org/10.1023/A:1024634613982>
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43, 345–354. <https://doi.org/10.1509/jmkr.43.3.345>
- Clemons, E., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23, 149–171. <https://doi.org/10.2753/MIS0742-1222230207>
- Dellarocas, C., & Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, 21, 277–285. <https://doi.org/10.1214/08834230600000169>
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21, 23–45. <https://doi.org/10.1002/dir.20087>
- Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, 45, 1007–1016. <https://doi.org/10.1016/j.dss.2008.04.001>
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 19, 291–313. <https://doi.org/10.1287/isre.1080.0193>
- Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13, 70–84. <https://doi.org/10.1214/aos/1176346577>
- He, B., MacDonald, C., & Ounis, I. (2008). Ranking opinionated blog post using OpinonFinder. In *Proceedings of the 31st annual international ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 727–728), ACM.
- Henning, E. T., Al-Thuray, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18, 38–52. <https://doi.org/10.1002/dir.10073>
- Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9, 201–214.
- Hu, N., Pavlou, P. A., & Zhang, J. (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* (pp. 324–330), ACM.
- Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19, 456–474. <https://doi.org/10.1287/isre.1070.0154>
- Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70, 74–89.
- Luca, M. (2011). *Reviews, reputation, and revenue* (Working Paper 12-016). Boston, MA: Harvard Business School.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34, 185–200.

- Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78, 311–329. <https://doi.org/10.1086/259630>
- Park, C., & Lee, T. M. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research*, 62, 61–67. <https://doi.org/10.1016/j.jbusres.2007.11.017>
- Racherla, P., Mandviwalla, M., & Connolly, D. J. (2012). Factors affecting consumers' trust in online product reviews. *Journal of Consumer Behaviour*, 11, 94–104. <https://doi.org/10.1002/cb.v11.2>
- Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on empirical methods in natural language processing* (pp. 105–112), Association for Computational Linguistics. <https://doi.org/10.3115/1119355>
- Schnapp, M., & Allwine, T. (2001). Mining of book data from Amazon.com. In *Presentation at the UCB/SIMS web mining conference*. Retrieved from <https://www.sims.berkeley.edu/resources/affiliates/workshops/webmining/slides/ORA>
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80, 159–169. <https://doi.org/10.1016/j.jretai.2004.04.001>
- Whyte, W. H. (1954). The web of word of mouth. *Fortune*, 50, 140–143.
- Wiebe, J., & Riloff, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. In *Computational Linguistics and Intelligent Text Processing* (pp. 486–497), Springer. <https://doi.org/10.1007/b105772>
- Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, 63, 1336–1341. <https://doi.org/10.1016/j.jbusres.2009.12.011>
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74, 133–148. <https://doi.org/10.1509/jmkg.74.2.133>



© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format  
Adapt — remix, transform, and build upon the material for any purpose, even commercially.  
The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.  
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.  
No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



**Cogent Business & Management (ISSN: 2331-1975) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

