Estimating the Socio-Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics

Anindya Ghose
New York University
aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
New York University
panos@nyu.edu

September 1, 2008

Abstract

With the rapid growth of the Internet, the ability of users to create and publish content has created active electronic communities that provide a wealth of product information. However, the high volume of reviews that are typically published for a single product makes harder for individuals as well as manufacturers to locate the best reviews and understand the true underlying quality of a product. In this paper, we re-examine the impact of reviews on economic outcomes like product sales and see how different factors affect social outcomes like the extent of their perceived usefulness. Our approach explores multiple aspects of review text, such as *lexical*, *grammatical*, *semantic*, and *stylistic* levels to identify important text-based features. In addition, we also examine multiple reviewer-level features such as average usefulness of past reviews and the self-disclosed identity measures of reviewers that are displayed next to a review. Our econometric analysis reveals that the extent of subjectivity, informativeness, readability, and linguistic correctness in reviews matters in influencing sales and perceived usefulness. Reviews that have a mixture of objective, and highly subjective sentences have a negative effect on product sales, compared to reviews that tend to include only subjective or only objective information. However, such reviews are considered more informative (or helpful) by the users. By using Random Forest based classifiers, we show that we can accurately predict the impact of reviews on sales and their perceived usefulness. Reviews for products that have received widely fluctuating reviews, also have reviews of widely fluctuating helpfulness. In particular, we find that highly detailed and readable reviews can have low helpfulness votes in cases when users tend to vote negatively not because they disapprove of the review quality but rather to convey their disapproval of the review polarity. We examine the relative importance of the three broad feature categories: 'reviewer-related' features, 'review subjectivity' features, and 'review readability' features, and find that using any of the three feature sets results in a statistically equivalent performance as in the case of using all available features. This paper is the first study that integrates econometric, text mining, and predictive modeling techniques toward a more complete analysis of the information captured by user-generated online reviews in order to estimate their socio-economic impact. Our results can have implications for judicious design of opinion forums.

1 Introduction

With the rapid growth of the Internet, product related word-of-mouth conversations have migrated to online markets, creating active electronic communities that provide a wealth of information. Reviewers contribute time and energy to generate reviews, enabling a social structure that provides benefits both for the users and the firms that host electronic markets. Indeed, the provision of a forum facilitating information and social exchanges in the form of user-generated product reviews is an important part of many electronic markets, such as Amazon.com. In such a context, 'who' says 'what' and 'how' they say it, should matter in influencing socio-economic outcomes.

Unfortunately, a large number of reviews for a single product may also make it harder for individuals to track the gist of users' discussions and evaluate the true underlying quality of a product. Recent work has shown that the distribution of an overwhelming majority of reviews posted in online markets is bimodal [29].

Reviews are either allotted an extremely high rating or an extremely low rating. In such situations, the average numerical star rating assigned to a product may not convey a lot of information to a prospective buyer. Instead, the reader has to read the actual reviews to examine which of the positive and which of the negative attributes of a product are of interest. Similarly, product manufacturers are interested in reading user-generated reviews to identify which attributes of a product most influence consumer decisions.

So far, the best effort for ranking reviews for consumers comes in the form of "peer reviewing" in review forums, where customers give "helpful" votes to other reviews in order to signal their informativeness. Unfortunately, the helpful votes are not a useful feature for ranking *recent* reviews: the helpful votes are accumulated over a long period of time, and hence cannot be used for review placement in a short- or medium-term time frame. Similarly, merchants need to know what aspects of reviews are the most informative from consumers' perspective. Such reviews are likely to be the most helpful for merchants, as they contain valuable information about what aspects of the product are driving the sales up or down. Lack of this information often can lead to missed opportunities for feature-based targeting and product promotions, especially for new product launches. Thus, researchers and practitioners alike are interested in understanding the relationship between online consumer product reviews and user behavior, but a comprehensive explanation of which text-based features of reviews influence sales and usefulness remains elusive. This is primarily due to the fact that the extant literature has largely subsumed the information embedded in reviews with the valence and volume of reviews [9] and the disclosure of reviewer level information [18]. Compounding this issue of parsing the user-generated information is the fact that the language used by consumers on the Web does not often comply with standard grammatical and expositional rules. This necessitates the use of text-mining techniques to mine the content, and this is one of the goals of our paper.

In this paper, we propose techniques for predicting the helpfulness and importance of a review so that we can have:

- a consumer-oriented mechanism which can potentially rank the reviews according to their *expected helpfulness*, and

- a manufacturer-oriented ranking mechanism, which can potentially rank the reviews according to their *expected influence* on sales.

To understand better what are the factors that influence consumers perception of usefulness and what factors affect consumers most, we conduct a two-level study. First, we perform an explanatory econometric analysis, trying to identify what aspects of a review (and of a reviewer) are important determinants of its usefulness and impact. Then, at the second level we build a predictive model using Random Forests that offer significant predictive power and allow us to predict with high accuracy how peer consumers are going to rate a review and how sales will be affected by the posted review.

Our algorithms are based on the idea that the style of the review plays an important role in determining the perceived helpfulness by other fellow customers and the degree of influencing purchase decisions. In our work, we perform multiple levels of automatic text analysis to identify characteristics of the review that are important. We perform our analysis at the *lexical*, *grammatical*, *semantic*, and at the *stylistic* levels to identify text features that have high predictive power in identifying the perceived helpfulness and the economic impact of a review. Furthermore, we examine whether the past history and characteristics of a reviewer can be a useful predictor for the usefulness and impact of a review. We present an extensive experimental analysis using a real data set of 411 products, monitored over a 15-month period on Amazon.com. Our analysis indicates that we can predict accurately the helpfulness and influence of product reviews.

The rest of the paper is structured as follows. First, Section 2 discusses related work and provides the theoretical framework for generating the variables for our analysis. Then, in Section 3, we describe our data set and discuss how we extract the variables that we use to predict the usefulness and impact of a review. In Section 4, we present our explanatory econometric analysis for estimating the influence of the different variables and in Section 5 we describe the experimental results of our predictive modeling that uses Random Forest classifiers. Finally, Section 6 provides some additional discussion and concludes the paper.

2 Theoretical Framework and Related Literature

From a business perspective, consumer product reviews are most influential if they affect product sales and the online behavior of users of the word-of-mouth forum. Two streams of literatures discuss how online consumer product reviews influence sales. The first relevant stream of literature takes the customer's perspective to assess the effects of online product reviews on sales. Research in this vein has generally assumed that the primary reason that such reviews influence sales is because they provide information about the product or the vendor to potential consumers. For example, research has demonstrated an association between how positively a book is rated by consumers on a site (review valence) and subsequent sales of the book on that site [9, 13, 44] or between review volume and sales [16, 18, 37]. Indeed, to the extent that better products receive more positive reviews, there should be a positive relationship between review valence and sales. Sales may also be positively related to review valence due to the influence of the reviews themselves. In sum, this stream of work suggests that reviews and sales may be positively related even when underlying product quality is controlled [9, 18].

So far, prior work in information systems and marketing has not looked at how the sentiments and the textual characteristics of a review affect sales. Intuitively, reviews of reasonable length, that are easy to read, and lack spelling and grammar errors should be, ceteris paribus, more helpful and influential compared to other reviews that are difficult to read and have errors.

However, no prior work that we know of, has examined such factors systematically and no prior work has combined opinion mining techniques with econometrics and predictive modeling techniques to evaluate how the textual content of user-generated reviews impacts sales and online behavior of users in a community. Our hypothesis is that sentiments embedded in the text of product reviews will affect sales over and beyond the numerical information such as review valence and volume. Such sentiments can comprise of "subjective opinions" that portray reviewers' emotions about product features or more "objective statements" that portray factual data about product features, or a mix of both. Therefore, *style and content* of the review are also important features. Keeping these in mind, we formulate four potential constructs for text-based features that are likely to have an impact: (i) the average level of subjectivity in the reviews, (ii) the range and mix of subjective and objective comments, (iii) the extent to which the content is easy to read, and (iv) the proportion of spelling errors in the review.

A second stream of related research on word-of-mouth suggests that perceived attributes of the reviewer may shape consumer response to reviews [18]. In the social psychology literature, message source characteristics have been found to influence judgment and behavior [3, 7, 8, 24], and it has been often suggested that source characteristics might shape product attitudes and purchase propensity. Indeed, Forman et al. [18] draw on the information processing literature to suggest that product sales will be affected by reviewer disclosure of identity-related information. Prior research on computer mediated communication (CMC) suggests that online community members communicate information about product evaluations with an intent to influence others' purchase decisions as well as provide social information about contributing members themselves [30, 46]. Research concerning the motivations of content creators in online contexts highlights the role of identity motives in defining why users provide social information about themselves (e.g., [15, 26, 38, 39]).

Increasingly, we have seen that both identity-descriptive information about reviewers and product information is prominently displayed on the websites of online retailers. Prior research on self-verification in online contexts has pointed out the use of persistent labeling, defined as using a single, consistent way of identifying oneself such as 'real name' in the Amazon context, and self-presentation, defined as ways of presenting oneself online that may help others to identify one, such as posting geographic location or a personal profile in the Amazon context [38] as important phenomena in the online world. Indeed, information about product reviewers is often highly salient. Visitors to the site can see more professional aspects of reviewers such as their badges (e.g., "top-50 reviewer", "top-100 reviewer" badges) and ranks ("reviewer rank") as well as personal information about reviewers ranging from their real name to where they live, their nick names, hobbies, professional interests, pictures and other posted links. In addition, users have the opportunity to examine more "professional" aspects of a reviewer such as the proportion of helpful votes given by other users not only for a given review but across all the reviews of all other products posted by a reviewer. Further, interested users can also read the actual content of all reviews generated by a reviewer across all products.

With regard to the benefits reviewers derive, work on online user-generated content has primarily focused on

the consequences of peer recognition rather than on its antecedents [1, 21]. Its only recently that [18] evaluated the influence of reviewers' disclosure of information about themselves on the extent of peer recognition of reviewers and their interactions with the review valence by drawing on the social psychology literature. We hypothesize that after controlling for features examined in prior work such as reviewer disclosure of identity information and the valence of reviews, the actual text of the review matters in determining the extent to which users find the review useful. In particular, we focus on four constructs namely subjectiveness, informativeness, readability, and proportion of spelling errors. Our paper thus contributes to the existing stream of work by examining text-based antecedents of peer recognition in online word-of-mouth forums.

This paper builds on our previous work [19, 20]. In [19, 20] we examined just the effect of subjectivity, while in the current work, we expanding our data to include more product categories and examine a significantly increased number of features, such as different readability metrics, information about the reviewer history, different features of reviewer disclosure and so on. The present paper is unique in looking at how lexical, grammatical, semantic, and stylistic levels in the text of reviews affect product sales and the perceived helpfulness of these reviews.

In parallel with our work, researchers in the natural language processing field have examined the task of predicting review helpfulness [33, 36, 53], also using Amazon reviews as training and test data. Our work uses a superset of the features used in the past for helpfulness prediction (e.g., reviewer history and disclosure, deviation of subjectivity in the review, and so on). Also, none of these studies attempts to predict the influence of reviews on product sales. Also a differentiating factor of our approach is the two-pronged approach building on methodologies from economics and from data mining, building both explanatory and predictive models to understand better the impact of different factors. Interestingly, all prior use support vector machines (in a binary classification and in regression mode), which we observed to perform worse than Random Forests (as we discuss in Section 5). Predicting the helpfulness of a review is also related to the task of evaluating the quality of web posts or the quality of answers to posted questions [27, 31, 48, 49], although there there are more cues (e.g., clickstream data) that can be used to estimate the perceived quality of a posting.

Prior work in computer science has extensively analyzed and classified sentiments in online opinions [28, 32, 34, 35, 41, 41, 47], however they have not examined the economic impact of the textual content of reviews on sales or helpfulness of those reviews. Our research thus aims to make a contribution by bridging these two streams of work, and contributes to the emerging stream of work that examines the economic impact of user-generated content such as reviews, blogs and reputation systems [1, 14, 21].

3   Data Set and Variables

A major goal of this paper is to explore how the user-generated textual content of a review and the self-reported characteristics of the reviewer who generated the review can influence economic transactions (such as product sales) and online community behavior (such as peer recognition in the form of helpful votes). To examine this, we collected data about the economic transactions on Amazon.com and analyzed the associated review system. In this section, we describe the data that we collected from Amazon; furthermore, we discuss how we computed the variabes to perform our analysis, based on the discussion of Section 2.

3.1   Product and Sales Data

To conduct our study, we created a panel data set of products belonging to three product categories:

1. Audio and video players (144 products),

2. Digital cameras (109 products), and

3. DVDs (158 products).

We picked the products by selecting all the items that appeared in the "Top-100" list of Amazon over a period of 3 months from January 2005 to March 2005. We decided to use popular products, in order to have products in our study with a significant number of reviews. Then, using Amazon web services, from March 2005 until May 2006 we collected the information for these products described below.

Type	Variable	Explanation
Product and Sales Data	Retail Price	The retail price at Amazon.com
	Sales Rank	The sales rank within the product category
	Average Rating	Average rating of the posted reviews
	Number of Reviews	Number of reviews posted for the product
	Elapsed Date	Number of days since the release of the product
Individual Review	Moderate Review	Does the Review have a moderate rating (3 star rating) or not
	Helpful Votes	The number of helpful votes for the review
	Total Votes	The total number of votes for the review
	Helpfulness	$\frac{HelpfulVotes}{TotalVotes}$
Reviewer Characteristics	Reviewer Rank	The reviewer rank according to Amazon
	Top-10 Reviewer	Is the reviewer a Top-10 reviewer?
	Top-50 Reviewer	Is the reviewer a Top-50 reviewer?
	Top-100 Reviewer	Is the reviewer a Top-100 reviewer?
	Top-500 Reviewer	Is the reviewer a Top-500 reviewer?
	Real Name	Has the reviewer disclosed his/her real name?
	Nickname	Does the reviewer have a nickname listed in the profile?
	Hobbies	Does the reviewer have an "about me" section in the profile?
	Birthday	Does the reviewer list his/her birthday?
	Location	Does the reviewer disclose its location?
	Web Page	Does the reviewer have a home page listed?
	Interests	Does the reviewer list his/her interests?
	Snippet	Does the reviewer has a description in the reviewer profile?
	Any Disclosure	Does the reviewer list *any of the above* in the reviewer profile?
Reviewer History	Number of Past Reviews	Number of reviews posted by the reviewer
	Reviewer History (Macro)	Average past review helpfulness (macro-averaged)
	Reviewer History (Micro)	Average past review helpfulness (micro-averaged)
	Past Helpful Votes	Number of helpful votes accumulated in the past from the reviewer
	Past Total Votes	Number of total votes on the reviews posted in the past for the reviewer
Review Readability	Length (Chars)	The length of the review in characters
	Length (Words)	The length of the review in words
	Length (Sentences)	The length of the review in sentences
	Spelling Errors	The number of spelling errors in the review
	ARI	The Automated Readability Index (ARI) for the review
	Gunning Index	The Gunning–Fog index for the review
	Coleman–Liau index	The Coleman–Liau index for the review
	Flesch Reading Ease	The Flesch Reading Ease score for the review
	Flesch–Kincaid Grade Level	The Flesch–Kincaid Grade Level for the review
	SMOG	The Simple Measure of Gobbledygook score for the review
Review Subjectivity	AvgProb	The average probability of a sentence in the review being subjective
	DevProb	The standard deviation of the subjectivity probability

Table 1: The variables collected for our study. The panel data set contains data collected over a period of 15 months; we collected the variables daily and we capture the variability over time for the variables that change over time (e.g., sales rank, price, reviewer characteristics and so on).

We collected various product-specific characteristics over time. Specifically, we collected the manufacturer suggested *list price* of the product, its Amazon *retail price*, its Amazon *sales rank* (which serves as a proxy for units of demand [22], as we will describe later).

Together with sales and price data, we also collected other data that may influence the purchasing behavior of consumers. For example, we collected the date the product was released into the market, to compute the *elapsed time* from the date of product release, since products released long time ago tend to see a decrease in sales over time. We also collected the *number of reviews* and the *average review rating* of the product over time.

3.2   Individual Review Data

Beyond the product-specific data, we also collected all reviews of a product since the product was released into the market. For each review, we retrieve the actual textual content of the review and the *review rating* of the product given by the reviewer. The rating that a reviewer allocates to the reviewed product is denoted by a number of stars on a scale of 1 to 5. From the textual content, we generated a set of variables at the lexical, grammatical, and at the stylistic level. We describe these variables in detail in Section 3.4, when we describe the textual analysis that we conducted.

Review Helpfulness: Amazon has a voting system whereby community members provide helpful votes to rate the reviews of other community members. Previous peer ratings appear immediately above the posted review, in the form, "[*number of helpful votes*] out of [*number of members who voted*] found the following review helpful." These *helpful* and *total* votes enable us to compute the fraction of votes that evaluated the review as helpful. To have as much accurate representation of the percentage of customers that found the review helpful, we collected the votes in December 2007, ensuring that there is a significant time period after the time the review was posted and that there is a significant number of peer rating votes accumulated for the review.

3.3   Reviewer Characteristics

Reviewer Disclosure: While review valence is likely to influence consumers, there is reason to believe that social information about reviewers themselves (rather than the product or vendor) is likely to be an important predictor of consumers' buying decisions [18]. On many sites, social information about the reviewer is at least as prominent as product information. For example, on sites such as Amazon, information about product reviewers is graphically depicted, highly salient, and sometimes more detailed and voluminous than information on the products they review: the "Top-1000" reviewers have special tags displayed next to their names, the reviewers that disclose their real name[1] are also highlighted and so on. Given the extent and salience of available social information regarding product reviewers, it seems important to control for the impact of such information on online product sales and review helpfulness. Amazon has a procedure by which reviewers can disclose personal information about themselves. There are several types of information that users can disclose: we focus our analysis on the categories most commonly indicated by users: whether the user disclosed their *real name*, their *location, nickname*, and *hobbies*. With real name, we refer to a registration procedure that Amazon provides for users to indicate their actual name by providing verification with a credit card, as mentioned above. Reviewers may also post additional information in their profiles such as geographical location, disclose additional information (e.g., "Hobbies") or use a nickname (e.g., "Gadget King"). We use these data to control for the impact of self-descriptive identity claims. We encode this information as binary variables. We also constructed an additional dummy variable, labeled "*any disclosure*"; this variable captures each instance where the reviewer has engaged in any one of the four kinds of self-disclosure. We also collected the *reviewer rank* of the reviewer as published on Amazon.

Reviewer History: Since one of our goal is to predict the future usefulness of a review, we wanted to examine whether the past history of a reviewer can be used to predict the usefulness of the future reviews written by the same reviewer. For this, we collected the past reviews for each reviewer, and collected the helpful and total votes for each of the past reviews. Using this information, we constructed *for each reviewer and for each point in time* the past performance of a reviewer. Specifically, we created the variable *reviewer*

---

[1]Amazon compares the name of the reviewer with the name listed in the credit card on file before assigning the "Real Name" tag.

history macro, which was the ratio of all past helpful votes divided by the total number of votes. Similarly, we also created the variable *reviewer history micro*, in which we first computed the average helpfulness for each of the past reviews and then computed the average across all past reviews. The difference with the macro and micro versions is that the micro version gives equal weight to the helpfulness of all past reviews, while the macro version weights more heavily the importance of reviews that received a large number of votes.

3.4   Textual Analysis of Reviews

Our approach is based on the hypothesis that the actual text of the review matters. Previous text mining approaches focused on extracting automatically the polarity of the review [10, 12, 25, 28, 32, 40–43, 45, 47, 51]. In our setting, the numerical rating score already gives the (approximate) polarity of the review,[2] so we look in the text to extract features that are not possible to observe using simple numeric ratings.

Readability Analysis: We are interested to examine what types of reviews affect most sales and what types of reviews are most helpful to the users. For example, everything else being equal, a review that is easy to read will be more helpful than another that has spelling mistakes and is difficult to read.

As a first, low-level variable, we measured the number of spelling mistakes within each review, and we normalized the number by dividing with the length of the review (in characters). Furthermore, to measure the cognitive effort that a user needs in order to read a review, we measured the length of a review in sentences, words, and characters.

Beyond these basic features, we also used the extensive results from research on *readability*. Past research has shown that easy-reading text improves comprehension, retention, and reading speed, and that the average reading level of the US adult population is at the eighth grade level [50]. Therefore, a review that can be read easily by a large number of users is also expected to be rated by more users. Today there are numerous metrics for measuring the readability of a text, and while none of them is perfect, the computed measures correlate well with the actual difficulty of reading a text. To avoid idiosyncratic errors peculiar to a specific readability metric, we computed a set of metrics for each review. Specifically, we computed the following:

- Automated Readability Index

- Coleman–Liau Index

- Flesch Reading Ease

- Flesch–Kincaid Grade Level

- Gunning fog index

- SMOG

(See [17] for detailed description on how to compute each of these metrics.) Based on research in readability, these metrics are useful metrics for measuring how easy is for a user to read a review.

Subjectivity Analysis: Beyond the low-level spelling and readability analysis, we also expect that there are stylistic choices that affect the perceived helpfulness of a review. We observed empirically that there are two types of listed information, from the stylistic point of view. There are reviews that list "objective" information, listing the characteristics of the product, and giving an alternate product description that confirms (or rejects) the description given by the merchant. The other types of reviews are the reviews with "subjective," sentimental information, in which the reviewers give a very personal description of the product, and give information that typically does not appear in the official description of the product.

As a first step towards understanding the impact of the style of the reviews on helpfulness and product sales, we rely on existing literature of subjectivity estimation from computational linguistics [42]. Specifically, Pang and Lee [42] described a technique that identifies which sentences in a text convey objective information, and which of them contain subjective elements. Pang and Lee applied their techniques in a data set with movie review data set, in which they considered as objective information the movie plot, and as subjective the information that appeared in the reviews. In our scenario, we follow the same paradigm. In particular,

---

[2]We should note, though, that the numeric rating does not capture all the polarity information that appears in the review [1].

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	5699	151.33	130.57	0	3299.99
Sales Rank	7352	7667.42	51039.42	0	2090308
Log (Elapsed Date)	7352	5.12	1.09	0	7.63
Average Rating	7352	3.86	1.41	1	5
Number of Reviews	7352	195.07	138.76	0	522
Moderate Review	7352	0.093	0.29	0	1
Any Disclosure	7352	0.52	0.49	0	1
Helpful Votes	7352	5.51	11.7	0	744
Total Votes	7352	8.38	14.05	0	893
Log(Spelling Errors)	7352	-3.85	0.74	-6.67	-1.34
Readability (Gunning Fog Index)	7352	12.46	13.31	1.36	277.95
AvgProb	7352	0.58	0.05	0.37	0.83
DevProb	7352	0.047	0.024	0	0.18
Reviewer History Macro	3076	0.69	0.23	0	1

Table 2: Descriptive statistics of audio and video players for Econometric Analysis

objective information is considered the information that also appears in the product description, and subjective is everything else.

Using this definition, we then generated a training set with two classes of documents:

- A set of "objective" documents that contains the product descriptions of each of the products in our data set.

- A set of "subjective" documents that contains randomly retrieved reviews.

Since we deal with a rather diverse data set, we constructed separate subjectivity classifiers for each of our product categories. We trained the classifier using a Dynamic Language Model classifier with n-grams ($n = 8$) from the LingPipe toolkit[3].

After constructing the classifiers for each product category, we used the resulting classification models in the remaining, *unseen* reviews. Instead of classifying each review as subjective or objective, we instead classified each *sentence* in each review as either "objective" or "subjective," keeping the probability being subjective $Pr_{subj}(s)$ for each sentence s. Hence, for each review, we have a "subjectivity" score for each of the sentences.

Based on the classification scores for the sentences in each review, we derived the average probability $AvgProb(r)$ of the review r being subjective defined as:

$$AvgProb(r) = \frac{1}{n} \sum_{i=1}^{n} Pr_{subj}(s_i) \tag{1}$$

where n is the number of sentences in review r and $s_1, \ldots, s_n$ are the sentences that appear in review r. Since the same review may be a mixture of objective and subjective sentences, we also kept of standard deviation $DevProb(r)$ of the subjectivity scores for each review, defined as:

$$DevProb(r) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \Big(Pr_{subj}(s_i) - AvgProb(r) \Big)^2} \tag{2}$$

The summary statistics of the data for audio-video players, digital cameras and DVDs are given in Table 2, Table 3 and Table 4, respectively.

---
[3] http://www.alias-i.com/lingpipe/

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	1690	159.94	351.62	0	7999.99
Sales Rank	2730	395.27	1418.04	0	38353
Log(Elapsed Date)	2730	5.46	0.55	1.38	7.02
Average Rating	2730	4.06	1.31	1	5
Number of Reviews	2730	84.69	492.62	0	3795
Moderate Review	2730	0.36	0.48	0	1
Any Disclosure	2730	0.58	0.49	0	1
Helpful Votes	2730	17.68	58.58	0	1669
Total Votes	2730	21.67	59.68	0	1689
Log(Spelling Errors)	2730	-4.23	0.86	-7.03	-0.84
Readability (Gunning Fog Index)	2730	13.07	8.25	1.2	117.44
AvgProb	2730	0.55	0.02	0.46	0.67
DevProb	2730	0.03	0.01	0.001	0.17
Reviewer History Macro	1085	0.78	0.22	0	1

Table 3: Descriptive statistics of digital cameras for Econometric Analysis

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	2005	21.30	7.81	0	71.96
Sales Rank	2002	3824.28	6168.97	6	49868
Log (Elapsed Date)	1698	4.38	1.29	0	7.55
Average Rating	2018	4.05	0.62	1	5
Number of Reviews	2018	84.54	88.36	0	431
Moderate Review	2018	0.45	.49	0	1
Any Disclosure	2018	0.74	0.43	0	1
Helpful Votes	2018	6.35	13.626	0	246
Total Votes	2018	10.498	19.31	0	320
Log(Spelling Errors)	2018	-4.34	0.91	-7.44	-1.003
Readability (Gunning Fog Index)	2018	13.09	6.47	0.8	136.83
AvgProb	2018	0.52	0.012	0.46	0.56
DevProb	1946	0.016	.01	0.00003	0.07
Reviewer History Macro	1450	0.68	0.21	0	1

Table 4: Descriptive statistics of DVD for Econometric Analysis

4 Explanatory Econometric Analysis

So far, we have explain the different types of data that we collected, that have the potential, according the various hypotheses, to affect the impact and usefulness of the reviews. In this section, we present the results of our *explanatory* econometric analysis, which will examine the importance of each factor. Through our analysis, we aim to provide a better understanding of how customers are affected by the reviews. (In the next section, we will describe our *predictive* model, based on machine learning techniques.) In Section 4.1 we analyze the effect of the different review characteristics and their impact on product sales. Our results show what factors are important for a merchant to observe. Then, in Section 4.2 we presents our analysis on how different factors affect the helpfulness of a review.

4.1 Effect on Product Sales

We first estimate the relationship between sales and stylistic elements of a review. Previous work has examined how price, number of reviews, and review valence influence product sales on Amazon and Barnes and Noble [9]. Recent work by Forman et al. [18] also describes how reviewer disclosure of identity descriptive information (e.g., *Real Name* or *Location*) affects product sales. Hence, to be consistent with prior work, we control for all these factors but focus mainly on the textual aspects of the review to see how they affect sales. In particular, we test the following hypotheses:

Hypothesis 1a: *All else equal, an increase in the proportion of subjective opinions in reviews will be associated with an increase in sales.*

Hypothesis 1b: *All else equal, a decreased prevalence of a mix of subjective and objective sentences in reviews will be positively related to sales.*

Hypothesis 1c: *All else equal, a decrease in the readability of reviews will be positively related to sales.*

Hypothesis 1d: *All else equal, a decrease in the proportion of spelling errors in reviews will be positively related to sales.*

4.1.1 Model Specification

We adopt a model similar to that used in [9] and [18], while incorporating measures for the quality and the content of the reviews. Chevalier and Mayzlin [9] and Forman, Ghose and Wiesenfeld [18] define the book's sales rank as a function of a book fixed effect and other factors that may impact the sales of a book. The unit of observation in our analysis is a product-date, and the dependent variable is $\ln(SalesRank)$, the log of sales rank of product k in time t. Specifically, to study the impact of reviews and the quality of reviews on sales, we estimate the following model:

$$
\begin{aligned}
\ln(SalesRank)_{kt} = {} & \alpha + \beta_1 \cdot \ln(AmazonPrice_{kt}) + \\
& \beta_2 \cdot AvgProb_{k(t-1)} + \\
& \beta_3 \cdot DevProb_{k(t-1)} + \\
& \beta_4 \cdot AverageReviewRating_{k(t-1)} + \\
& \beta_5 \cdot \ln(NumberofReviews_{k(t-1)}) + \\
& \beta_6 \cdot (Readability_{k(t-1)}) + \\
& \beta_6 \cdot \ln(SpellingErrors_{k(t-1)}) + \\
& \beta_7 \cdot (AnyDisclosure_{k(t-1)}) + \\
& \beta_8 \cdot \ln(ElapsedDate_{kt}) + \\
& \mu_k + \varepsilon_{kt}
\end{aligned}
\tag{3}
$$

Variable	Audio Video	Digital Camera	DVD
Retail Price	0.095 (0.007)***	0.099 (0.01)***	0.212 (0.051)**
AvgProb	-1.49 (0.77)*	-0.163 (1.5)	-2.66 (1.68)
DevProb	2.31 (1.03)**	-0.689 (2.08)	5.24 (2.59)**
Average Rating	-0.32 (0.078)***	0.15 (0.11)	-0.04 (0.11)
Log(Num of Reviews)	-0.618 (0.042)	-0.298 (0.066)***	-0.33 (0.08)***
Readability	0.002 (0.004)	-0.014 (0.006)**	0.004 (0.005)
Log(Spelling Errors)	-0.012 (0.024)	-0.015 (0.034)	0.041 (0.024)*
Any Disclosure	-0.047 (0.030)	-0.023 (0.056)	-0.036 (0.049)
Log(Elapsed Date)	0.120 (0.070)	0.255 (0.12)**	0.831 (0.032)***
Number of Observations	7352	2730	2018
R-square (with fixed effects)	0.91	0.77	0.87

Table 5: These are OLS regressions with product-level fixed effects. The dependent variable is Log (Salesrank). Robust standard errors are listed in parenthesis; ***, ** and * denote significance at 1%, 5% and 10%, respectively. The R-square includes fixed effects in R-square computation.

where μ_k is a product *fixed effect* that controls for unobserved heterogeneity across products. Note that increases in sales rank mean lower sales, so a negative coefficient on a variable implies that an increase in that variable increases sales. The control variables used in our model include the Amazon retail price, the difference between the date of data collection and the release date of the product (*Elapsed Date*), the average numeric rating of the product (*Rating*), and the log of the number of reviews posted for that product (*Number of Reviews*). This is consistent with prior work such as Chevalier and Mayzlin [9] and Forman et al. [18] To account for potential non-linearities and to smooth large values, we take the log of the dependent variable and some of the control variables such as Amazon Price, volume of reviews and days elapsed consistent with the literature [18, 22]. For these regressions in which we examine the relationship between review sentiments and product sales, we aggregate data to the weekly level. By aggregating data in this way, we smooth potential day-to-day volatility in sales rank.[4]

We estimate product-level fixed effects to control for differences in average sales rank across products. These fixed effects are algebraically equivalent to including a dummy for every product in our sample, and so this enables us to control for differences in the average quality of products. Thus, any relationship between sales rank and review valence will not reflect differences in average quality across products, but rather will be identified off changes over time in sales rank and review valence within products, diminishing the possibility that our results reflect differences in average unobserved book quality rather than aspects of the reviews themselves [18].

Our primary interest is in examining the association between textual variables in user-generated reviews and sales. To maintain consistency with prior work, we also examine the association between average review valence and sales. However, prior work has shown that review valence may be correlated with product-level unobservables that may be correlated with sales. In our setting, though we control for differences in the average quality of products through our fixed effects, it is possible that changes in the popularity of the product over time may be correlated with changes in review valence. Thus, this parameter reflects not only the information content of reviews but also may reflect exogenous shocks that may influence product popularity [18]. Similarly, the variable *Number of Reviews* will also capture changes in product popularity or perceived product quality over time; thus, β_5 may reflect the combined effects of a causal relationship between number of reviews and sales [16] and changes in unobserved book popularity over time.[5]

---

[4]As a robustness check, we also ran regressions at the daily and fortnightly level, and find that the qualitative nature of most of our results remain the same.

[5]Note that prior work in this domain has often transformed the dependent variable (sales rank) into quantities using the specification similar to Ghose et al. [22]. That was usually done because those papers were interested in demand estimation and imputing price elasticities. However, in this case we are not interested in estimating demand, and hence we do not need to make the actual transformation. In this regard, our paper is more closely related to [18].

4.1.2 Empirical Results

The sign on the coefficient of *AvgProb* suggests that an increase in the average subjectivity of reviews leads to an increase in sales for products, although the estimate is statistically significant only for audio-video players and digital cameras (see Table 5). It is statistically insignificant for DVDs. Our conjecture is that customers prefer to read reviews that describe the individual experiences of other consumers and buy products with significant such (subjective) information available only for *search goods* (such as camreas and audio-video players) but not for *experience goods*.[6]

The coefficient of *DevProb* has a positive and statistically significant relationship with sales rank in audio-video players and DVDs, but is statistically insignificant for digital cameras. In general this suggests that a decrease in the deviation of the probability of subjective comments leads to a decrease in sales rank, i.e., an increase in product sales. This means that reviews that have a mixture of objective, and highly subjective sentences have a negative effect on product sales, compared to reviews that tend to include only subjective or only objective information.

The coefficient of *Spelling Errors* is positive and statistically significant for DVDs suggesting that an increase in the proportion of spelling mistakes in the content of the reviews decreases product sales for some products whose quality can be assessed only after purchase. However, for hedonic like products such as audio-video players and digital cameras whose quality can be assessed prior to purchase, the proportion of spelling errors in reviews does not have a statistically significant impact on sales. For all three categories, we find that this result is robust to different specifications of normalizing the number of spelling errors such as normalizing by the number of characters, words or sentences in a given review.

The coefficient of the *Readability* is negative and statistically significant for digital cameras suggesting that reviews that have higher Readability scores are associated with higher sales. This is likely to happen if such reviews are written in more authoritative and sophisticated language which enhances the credibility and informativeness of such reviews. Our results are robust to the use of other Readability metrics described in Table 1 such as ARI, Coleman–Liau index, Flesch–Reading Ease, Flesch–Kincaid Grade Level and the SMOG index. In sum, our results provide support for Hypotheses 1a to 1d.

As expected, our control variables suggest that sales decrease as Amazon's price increases. Further, even though the coefficient of *Any Disclosure* is statistically insignificant, the negative sign implies that that the prevalence of reviewer disclosure of identity-descriptive information would be associated with higher subsequent sales. This consistent with prior research in the information processing literature supporting a direct effect for source characteristics on product evaluations and purchase intentions when information is processed heuristically [18]. Our results are robust to the use of other disclosure variables in the above regression. For example, instead of "Any Disclosure", if we were to use disclosures of the two most salient reviewer self-descriptive features (Real Name and Location), results are generally consistent with the existing ones.

We also find that an increase in the volume of reviews is positively associated with sales of DVDs and digital cameras. In contrast, average review valence has a statistically significant effect on sales for only audio-video players. These mixed findings are consistent with prior research which have found a statistically significant effect of review valence but not review volume on sales [9], and with others who have found a statistically significant effect of review volume but not valence on sales [18], [16], [37]. [7]

Using these results, it is now possible to generate a ranking scheme for presenting reviews to manufacturers of a product. The reviews that affect sales the most (either positively or negatively) are the reviews that should be presented first to the manufacturer. Such reviews tend to contain information that affects the perception of the customers for the product. Hence, the manufacturer can utilize such reviews, either by modifying future versions of the product or by modifying the existing marketing strategy (e.g., by emphasizing the good characteristics of the product). We should note that the reviews that affect sales most are not necessarily the same as the ones that customers find useful and are typically getting "spot-lighted" in review forums, like the one of Amazon. We present related evidence next.

---

[6]Search goods are goods with quality that can be observed before buying the product (e.g., electronics) while for experience goods, the consumers has to consume/experience the product in order to determine its quality (e.g., books, movies).

[7]Note that we do not have other variables such as "Reviewer Rank" or " Helpfulness" in this regression because of a concern that these variables will lead to biased and inconsistent estimates. Said simply, it is entirely possible that 'Reviewer Rank" or " Helpfulness" is correlated with other unobserved review-level attributes. Such correlations between regressors and error terms will lead to the well known endogeneity bias in OLS regressions [52].

4.2 Effect on Helpfulness

Next, we want to analyze the impact of review variables on the extent to which community members would rate reviews helpful after controlling for the presence of self-descriptive information. Recent work [18] describes how reviewer disclosure of identity descriptive information and the extent of equivocality of reviews (based on the review valence) affects perceived usefulness of reviews. Hence, to be consistent with prior work, we control for these factors but focus mainly on the textual aspects of the review and the reviewer history to see how they affect the usefulness of reviews. In particular, we test the following hypotheses:

Hypothesis 2a: *All else equal, an increase in the proportion of subjective opinions in a review will be associated with a decrease in perceived helpfulness of that review.*

Hypothesis 2b: *All else equal, an increased prevalence of a mix of subjective and objective sentences in a review will be positively related to perceived helpfulness of that review.*

Hypothesis 2c: *All else equal, an increase in the readability of a review will be positively related to perceived helpfulness of that review.*

Hypothesis 2d: *All else equal, a decrease in the proportion of spelling errors in a review will be positively related to perceived helpfulness of that review.*

Hypothesis 2e: *All else equal, an increase in the average helpfulness of a reviewer's historical reviews will be positively related to perceived helpfulness of a review posted by that reviewer.*

4.2.1 Model Specification

The dependent variable, $Helpfulness_{kr}$, is operationalized as the ratio of helpful votes to total votes received for a review r issued for product k. We use a well-known linear specification for our helpfulness estimation [18]:

$$\ln(Helpfulness)_{kr} = \alpha + \beta_1 \cdot (AvgProb)_{kr} +$$
$$\beta_2 \cdot (DevProb)_{kr} +$$
$$\beta_3 \cdot (AnyDisclosure)_{kr} +$$
$$\beta_4 \cdot (Readability)_{kr} +$$
$$\beta_5 \cdot (ReviewerHistoryMacro)_{kr} +$$
$$\beta_6 \cdot \ln(SpellingErrors)_{kr} +$$
$$\beta_7 \cdot (Moderate)_{kr} +$$
$$\beta_8 \cdot \ln(NumberofReviews)_{kr} +$$
$$\mu_k + \varepsilon_{kr} \tag{4}$$

The unit of observation in our analysis is a product-review and μ_k is a product fixed effect that controls for differences in the average helpfulness of reviews across products. We also constructed a dummy variable to differentiate between extreme reviews, which are unequivocal and therefore provide a great deal of information to inform purchase decisions, and moderate reviews which provide less information. Specifically, ratings of 3 were classified as *Moderate* reviews while ratings nearer the endpoints of the scale (1, 2, 4, 5) were classified as unequivocal [18].

The above equation can be estimated using a simple panel data fixed effects model. However, one concern with this strategy is that the posting of personal identity information such as Real Name or location may be correlated with some unobservable reviewer-specific characteristics that may influence review quality [18]. If some explanatory variables are correlated with errors, then ordinary least squares regression gives biased and inconsistent estimates. To control for this potential problem, we use a Two Stage Least Squares (2SLS) regression with instrumental variables [52]. Under the 2SLS approach, in the first stage, each endogenous variable is regressed on all valid instruments, including the full set of exogenous variables in the main regression. Since the instruments are exogenous, these approximations of the endogenous covariates will not

Variable	Audio Video	Digital Camera	DVD
AvgProb	-1.184 (0.34)***	-1.284 (0.29)***	-1.440 (0.840)*
DevProb	0.77 (0.41)*	0.33 (0.3)	1.320 (0.950)
Disclosure	0.210 (0.12)*	0.374 (0.119)***	-0.360 (0.240)
Readability	0.003 (0.001)**	-0.001 (0.001)	0.016 (0.004)***
Past History	0.031 (0.035)	-0.063 (0.038)*	0.230 (0.060)***
Log (Spelling Errors)	-0.037 (0.006)***	0.010 (0.016)	-0.040 (0.010)***
Moderate	-0.01 (0.01)	-0.119 (0.018)***	-0.03 (0.018)***
Log (Number of reviews)	0.001 (0.003)	0.024 (0.008)***	0.01 (0.004)***
Number of Observations	3076	1085	1450
R-square	0.08	0.02	0.03

Table 6: These are 2SLS regressions with Instrument Variable. Fixed effects are at the product-level. The dependent variable is *Helpful*. Robust standard errors are listed in parenthesis; ***, ** and * denote significance at 1%, 5% and 10%, respectively. The p-values from the Sargan test of overidentifying restrictions confirm the validity of instruments.

be correlated with the error term. So, intuitively they provide a way to analyze the relationship between the dependant variable and the endogenous covariates. In the second stage, each endogenous covariate is replaced with its approximation estimated in the first stage and the regression is estimated as usual. The slope estimator thus obtained is consistent [52].

Specifically, we instrument for in the above equation using lagged values of disclosure, subjectivity and readability. We experimented with different combinations of these instruments and the find that the qualitative nature of our results are generally robust. The intuition behind the use of these instrument variables is that they are likely to be correlated with the relevant independent variables but uncorrelated with unobservable characteristics that may influence the dependent variable. For example, the use of a Real Name in prior reviews is likely to be correlated with the use of Real Name in the subsequent reviews but uncorrelated with unobservables that determine perceived helpfulness for a given review. Similarly, the presence of subjectivity in prior reviews is likely to be correlated with the presence of subjectivity in subsequent reviews but unlikely to be correlated with the error term that determines perceived helpfulness for the current review. Hence, these are valid instruments in our 2SLS estimation. This is consistent with prior work such as [18].

To ensure the validity of our instruments, we conducted the Sargan Test of overidentifying restrictions([52]). The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. For the 2SLS estimator, the test statistic is typically calculated as N*R-squared from a regression of the IV residuals on the full set of instruments. A rejection casts doubt on the validity of the instruments. Based on the p-values from these tests, we are unable to reject the null hypothesis for each of the three categories, thereby confirming the validity of our instruments.

4.2.2 Empirical Results

With regard to the usefulness of reviews, our analysis reveals that for product categories such as audio and video equipments, digital cameras, and DVDs the extent of subjectivity in a review has a statistically significant effect on the extent to which users perceive the review to be helpful. The coefficient of *AvgProb* is negative suggesting that highly subjective reviews are rated as being less helpful. Although *DevProb* is statistically significant for audio video products only, it always has a positive relationship with helpfulness votes. This results suggests that consumers find reviews that have a wide range of subjectivity/objectivity scores across sentences to be more helpful. In other words, reviews that have a mixture of sentences with objective and of sentences with extreme, subjective content are rated highly by users. It is worthwhile to mention that we observed the opposite effect for product sales, indicating that helpful reviews are not

necessarily the ones that lead to increases in sales.

The negative and statistically significant sign on the coefficient of the *Moderate* variable for two of the three product categories implies that as the content of the review becomes more moderate or equivocal, the review is considered less helpful by users. This result is consistent with the findings of [18] who analyze a panel of book reviews and find a similar negative relationship between equivocal reviews and perceived helpfulness. Increased disclosure of self-descriptive information *Disclosure* typically leads to more helpful votes as can be seen for audio-video players and digital cameras.

We also find that for audio-video players and DVDs, a higher readability score *Readability* is associated with a higher percentage of helpful votes. As with sales, these results are robust to the use of other Readability metrics described in Table 1 such as ARI, Coleman–Liau index, Flesch–Reading Ease, Flesch–Kincaid Grade Level and the SMOG index. In contrast, an increase in the proportion of spelling errors *Spelling Errors* is associated with a lower percentage of helpful votes for both audio-video players and DVDs. For all three categories, we find that this result is robust to different specifications of normalizing the number of spelling errors such as normalizing by the number of characters, words or sentences in a given review. Finally, the past historical information about reviewers *Past History* has a statistically significant effect on the perceived helpfulness of reviews of digital cameras and DVDs, but interestingly, the directional impact is quite mixed across these two categories. In sum, our results provide support for Hypotheses 2a to 2e.

Note that the within R-squared values of our models range between 0.02 and 0.08 across the four product categories. This is because these R-squared values are for the "within" (differenced) fixed effect estimator that estimates this regression by differencing out the average values across product-sellers. The R-squared reported is obtained by only fitting a mean deviated model where the effects of the groups (all of the dummy variables for the products) are assumed to be fixed quantities. So, all of the effects for the groups are simply subtracted out of the model and no attempt is made to quantify their overall effect on the fit of the model. This means that the calculated "within" R-squared values do not take into account the explanatory power of the fixed effects. If we estimate the fixed effects instead of differencing them out, the measured R-squared would be much higher. However this becomes computationally unattractive. This is consistent with prior work([18]).

Our econometric analyses imply that we can quickly estimate the helpfulness of a review by performing an automatic stylistic analysis in terms of subjectivity, readability and linguistic correctness. Hence, we can immediately identify reviews that are likely to have a significant impact on sales and are expected to be helpful to the customers. Therefore, we can immediately rank these reviews higher and display them first to the customers. This is similar to the "spotlight review" feature of Amazon which relies on the number of helpful votes posted for a review. However, a key limitation of this existing feature is that it because it relies on a sufficient number of people to vote on reviews, it requires a long time to elapse before identifying a helpful review.

5   Predictive Modeling

The explanatory study that we described above revealed what factors influence the helpfulness and impact of a review. In this section, we switch from explanatory modeling to *predictive* modeling. In other words, the main goal now is not to explain which factors affect helpfulness and impact, but to examine whether, given an existing review, how well can we predict the helpfulness and economic impact of an *unseen* review, i.e., of a review that was not included in the data used to train the predictive model.

5.1   Predicting Helpfulness

The *Helpfulness* of each review in our data set is defined by the votes of the peer customers, who decide whether a review is helpful or not. In our predictive framework, we could use a regression model, as in Section 4, or use a classification approach and build a binary prediction model that classifies a review as helpful or not. We attempted both approaches and the results were similar. Since we have already described a regression framework in Section 4 , we now focus instead on a binary prediction model for brevity. In the rest of the section, we first describe our methodology for converting the continuous helpfulness variable into binary. Then we describe the results of our experiments, using various machine learning approaches.
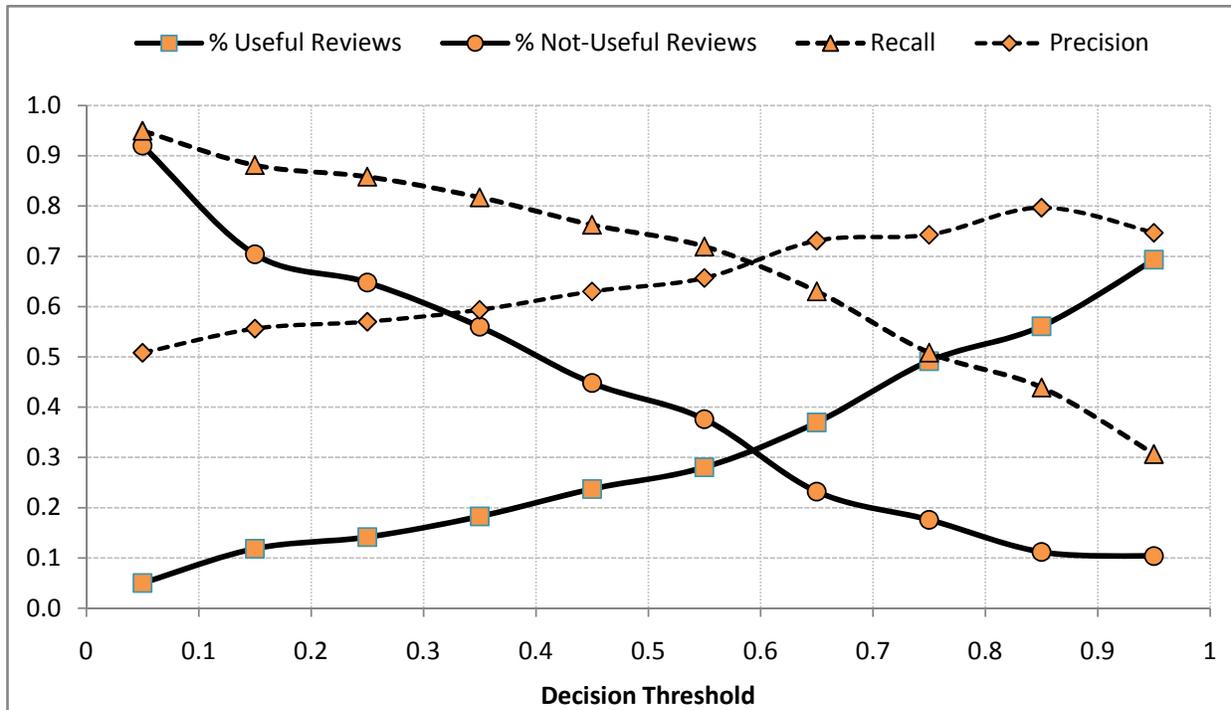
Figure 1: Picking a decision threshold that minimizes error rates for converting the continuous helpfulness variable into a binary one.

5.2 Converting Continuous Helpfulness to Binary

Converting the continuous variable *Helpfulness* into a binary one is, in principle, a straightforward process. Since *Helpfulness* goes from 0 to 1, we can simply select a threshold τ, and mark all reviews that have *helpfulness* $\geq \tau$ as *helpful* and the others as *not helpful*. However, selecting the proper value for the threshold τ is slightly trickier. What is the "best" value of τ for separating the "helpful" from the "not helpful" reviews? Setting τ too high would mean that helpful reviews would be classified as not helpful, and setting τ too low would have the opposite effect.

In order to select a good value for τ, we used two human coders do a content analysis on a sample of 1,000 reviews. The reviews were randomly chosen for each category. The main aim was to analyze whether the review was informative. For this, we asked the the coders to read each review and answer the question "*Is the review informative or not?*" The coders did not have access to the helpful and total votes that were casted for the review, but could see the star rating and the product that the review was referring to. We measured the inter-rater agreement across the two coders, using the kappa statistic. The analysis showed a substantial agreement, with $\kappa = 0.739$.

Our next step was to identify the optimal threshold (in terms of percentage of helpful votes) that separates the reviews that humans consider helpful from the non-helpful ones. We performed an ROC analysis, trying to balance the false positive rate and the false negative rate. Our analysis indicated that if we set the separation threshold at 0.6, then the error rates are minimized. In other words, if more than 60% of the votes indicate that the review is helpful, then we classify a review as "helpful". Otherwise, the review is classified as "not helpful" and this decision achieves a good balance between false positive errors and false negative errors.

Our analysis is presented in Figure 1. On the x-axis, we have the decision threshold τ, which is the percentage of useful votes out of all the votes received by a given review. Each review is marked as "useful" or "not-useful" by our coders, independently of the peer votes actually posted on Amazon.com. Based on the coder's classification, we compute the: (a) percentage of useful reviews that have an Amazon helpfulness rating below τ, (b) percentage of not-useful reviews that have an Amazon helpfulness rating above τ. (These

Data Set	Accuracy	AUC
DVD	78.79%	0.73
Audio & Video	87.57%	0.94
Digital Cameras	87.68%	0.94

Table 7: Accuracy and Area under the ROC curve for the Helpfulness Classifiers

values are essentially the error rates for the two classes if we set the decision threshold at τ.) Furthermore, by considering the "useful" class as the positive class, we compute the precision and recall metrics. We can see that if we set the separation threshold at 0.6, then the error rate in the classification is minimized. For this reason, we pick 0.6 as the threshold of separating the reviews as "useful" or not. In other words, if more than 60% of the votes indicate that the review is helpful, then we classify a review as "useful." Otherwise, the review is classified as "non-useful."

5.3   Building the Predictive Model

Once we are able to separate the reviews into two classes, we can then use any supervised learning technique to learn a model that classifies an unseen review as helpful or not.

We experimented with Support Vector Machines [4] and Random Forests [2]. Support Vector Machines have been reported to work well in the past for the problem of predicting review helpfulness. However, in all our experiments, SVM's consistently performed worse than Random Forests, for both our techniques and for the existing baselines. Furthermore, training time was significantly higher for SVM's compared to Random Forests. This empirical finding is consistent with recent comparative experiments [5, 6] that indicate that Random Forests are robust and perform better than SVM's for a variety of learning tasks. Therefore, in this experimental section we report only the results that we obtained using Random Forests.

In our experiments with Random Forests, we use 20 trees and we generate a different classifier for each product category. Our evaluation results are based on stratified 10-fold cross validation and we use as evaluation metrics the classification accuracy and the area under the ROC curve (AUC).

Using All Available Features: In our first experiment, we used all the features that we had available to build the classifiers. The resulting performance of the classifiers was quite high, as seen in Table 7.

One interesting result is the relatively lower predictive performance of the classifier that we constructed for the *DVD* data set. This can be explained by the nature of the goods: DVDs are *experience goods* whose quality is difficult to estimate in advance but can be ascertained after consumption. In contrast, digital cameras and audio & video are *search goods*, i.e., products with features and characteristics easily evaluated before purchase. Therefore the notion of *helpfulness* is more subjective for experience goods, as what constitutes a helpful review for one customer is not necessarily helpful for another. This contrasts with the search goods, in which a good review is one that allows customers to evaluate better, before the purchase, the quality of the underlying good.

Going beyond the aggregate results, we examined what kinds of reviews are the most difficult in terms of predicting their helpfulness. Interestingly, we observed a high correlation of classification error with the distribution of the underlying review ratings. Reviews for products that have received widely fluctuating reviews, also have reviews of widely fluctuating helpfulness. However, the different helpfulness scores do not necessarily correspond well with reviews of different inherent quality. Rather, in such cases, customers tend to vote not on the merits of the review per se, but rather to convey their approval or disapproval of the rating of the review. In such cases, an otherwise detailed and helpful review, may receive a bad helpfulness score. This effect was more pronounced in the DVD category, but also appeared in the digital camera and in the audio & video category.

Examining the Predictive Power of Features: The next step was to examine what is the power of the different features that we have generated. We have three broad feature categories: *reviewer-related features*, *review subjectivity features*, and *review readability features*. To examine their importance, we built classifiers using only subsets of the features.

We evaluated each classifier in the same way as above, using stratified 10-fold cross validation, and reporting the accuracy and the area under the ROC curve. One interesting result is that the classification

Data Set	Features	Accuracy	AUC
DVD	Reviewer	78.19%	0.71
	Subjectivity	77.95%	0.72
	Readability	77.23%	0.69
	Reviewer + Subjectivity	78.72%	0.73
	Reviewer + Readability	78.09%	0.72
	Subjectivity + Readability	78.14%	0.74
Audio & Video	Reviewer	83.07%	0.89
	Subjectivity	85.42%	0.91
	Readability	86.68%	0.94
	Reviewer + Subjectivity	86.82%	0.94
	Reviewer + Readability	87.11%	0.94
	Subjectivity + Readability	85.64%	0.93
Digital Cameras	Reviewer	82.99%	0.84
	Subjectivity	84.40%	0.86
	Readability	87.68%	0.93
	Reviewer + Subjectivity	86.47%	0.92
	Reviewer + Readability	89.04%	0.94
	Subjectivity + Readability	84.80%	0.90

Table 8: Accuracy and Area under the ROC curve for the Helpfulness Classifiers

performance did not decrease in a statistically significant manner. Using any of the feature sets resulted in a statistically equivalent performance as in the case of using all available features.

To explore further this puzzling result, we conducted an additional experiment: We examined whether we could predict the value of the features in one set, using the features from the other two feature sets (e.g., predict *Subjectivity* using the *Reviewer* and *Readability* features). We conducted the tests for all combinations. Surprisingly, the results indicated that the three feature sets are interchangeable. In other words, the information in the *Readability* and *Subjectivity* set is enough to predict the value of variables in the *Reviewer* set, and vice versa. Reviewers who have historical generated helpful reviews, tend to post reviews of specific readability levels, and with specific subjectivity mixtures in the reviews. Even though this may seem counterintuitive at first, it simply indicates that there is correlation between these variables, not causation. Identifying causality is rather difficult and is beyond the scope of this paper. What is of interest in our case is that the three feature sets are roughly equivalent in terms of learning a predictive model.

5.4 Predicting Impact on Sales

Our analysis so far, indicated that we can successfully predict whether a review is going to be rated as helpful by the peer customers or not. The next task that we wanted to examine was whether we can predict the impact of a review on sales.

Specifically, we examine whether the review characteristics can be used to predict whether the (comparative) sales of a product will go up or down after a review is published. So, we examine whether the difference

$$SalesRank_{t(r)+T} - SalesRank_{t(r)}$$

where $t(r)$ is the time the review is posted, is positive or negative. Since the effect of a review is not immediate, we examine variants of the problem for $T = 1$, $T = 3$, $T = 7$, $T = 14$ (in days). By having different time intervals, we wanted to examine how far in the future we can extend our prediction and still get reasonable results.

As we can see from the results in Table 9, the prediction accuracy is high, demonstrating that we can predict the direction of sales given the review information. While it is hard, at this point, to claim causality (it is unclear whether the reviews influence sales, or whether the reviews are just a manifestation of the underlying sales trend), it is definitely possible to show a strong correlation between the two. We also

Data Set	Features	Accuracy	AUC
DVD	$T = 1$	82.28%	0.91
	$T = 3$	82.28%	0.91
	$T = 7$	83.06%	0.92
	$T = 14$	84.35%	0.92
Audio & Video	$T = 1$	80.60%	0.91
	$T = 3$	81.28%	0.91
	$T = 7$	81.75%	0.92
	$T = 14$	83.24%	0.93
Digital Cameras	$T = 1$	73.98%	0.82
	$T = 3$	74.43%	0.86
	$T = 7$	77.93%	0.87
	$T = 14$	79.47%	0.89

Table 9: Accuracy and Area under the ROC curve for the Sales Impact Classifiers

observed that the predictive power increases slightly as T increases, indicating that the influence of a review is not immediate.

We also performed experiments with subsets of features as in the case of helpfulness. The results were very similar to the case of helpfulness: training models with subsets of the features results in similar predictive power. Given the helpfulness results, which we discussed above, this should not be a surprise.

6   Conclusions

In this paper, we build on our previous work [19, 20] by expanding our data to include multiple product categories and multiple textual features such as different readability metrics, information about the reviewer history, different features of reviewer disclosure and so on. The present paper is unique in looking at how lexical, grammatical, semantic, and stylistic levels in the text of reviews affect product sales and the perceived helpfulness of these reviews.

Our key results can be summarized as follows:

- We find that an increase in the average subjectivity of reviews leads to an increase in sales for products. Further, a decrease in the deviation of the probability of subjective comments leads to an increase in product sales. This means that reviews that have a mixture of objective, and highly subjective sentences have a negative effect on product sales, compared to reviews that tend to include only subjective or only objective information.

- We find that in general, reviews which tend to include a mixture of subjective and objective elements are considered more informative (or helpful) by the users. In terms of subjectivity and effect on helpfulness, we observe that for feature-based goods, such as electronics, users prefer reviews to contain mainly objective information with a few subjective sentences. In other words, the users prefer reviews that mainly confirm the validity of the product description, giving a small number of comments (not giving comments decreases the usefulness of the review). For experience goods, such as movies, the marginally significant coefficient on subjectivity suggests that while users do prefer to see a brief description of the "objective" elements of the good (e.g., the plot), they do expect to see a personalized, highly sentimental positioning, describing aspects of the movie that are not captured by the product description provided by the producers.

- An increase in the readability of reviews has a positive impact on perceived helpfulness while an increase in the proportion of spelling errors has a negative impact on usefulness. In the same vein, we find that for some product categories an increase in the readability of reviews has a positive impact on sales while an increase in the proportion of spelling errors has a negative impact on sales.

- Using Random Forest classifiers, we find that for experience goods like DVDs, the classifiers have a lower performance while predicting the helpfulness of reviews, compared to that for search goods like electronics products. Furthermore, we observe a high correlation of classification error with the distribution of the underlying review ratings. Reviews for products that have received widely fluctuating reviews, also have reviews of widely fluctuating helpfulness. In particular, we found evidence that highly detailed and readable reviews can have low helpfulness votes in cases when users tend to vote negatively not because they disapprove of the review quality but rather to convey their disapproval of the review polarity.

- Finally, we examined the relative importance of the three broad feature categories: 'reviewer-related' features, 'review subjectivity' features, and 'review readability' features. We found that using any of the three feature sets resulted in a statistically equivalent performance as in the case of using all available features. Further, we find that the three feature sets are interchangeable. In other words, the information in the 'readability' and 'subjectivity' set is sufficient to predict the value of variables in the 'reviewer' set, and vice versa. Experiments with classifiers for predicting sales yield similar results in terms of the inter-changeability of the three broad feature sets.

Based on our findings, we can identify quickly reviews that are expected to be helpful to the users, and display them first, improving significantly the usefulness of the reviewing mechanism to the users of the electronic marketplace.

To summarize, the extant literature has explored the informational influence of consumer reviews on economic behavior such as sales and usefulness by focusing on review valence, volume and disclosure of self-descriptive information by reviewers. We contribute to this research by demonstrating that patterns of text may influence purchasing decisions and peer recognition, over and above the information contained in the numeric ratings and identity disclosure of reviewers. More broadly, we also add to an emerging stream of literature that combines economic methods with text mining [11, 21, 23]. For example, Das and Chen [11] examined bulletin boards on Yahoo! Finance to extract the sentiment of individual investors about tech companies and about the tech sector in general. They show that the aggregate tech sector sentiment predicts well the stock index movement, even though the sentiment cannot predict well the individual stock movements. Gruhl et al. [23] analyzed the correlation between online mentions of a product and sales of that product. Using sales rank information for books from Amazon.com, Gruhl et al. [23] demonstrate that, even though sales rank motion might be difficult to predict in general, online chatter can be used to successfully predict *spikes* in the sales rank. Ghose et al. [21] analyze the textual feedback of reputation feedback posted by buyers for sellers on Amazon's used good market in order to impute the different dimensions of sellers' reputations and quantify the average pricing power associated with each dimension.

While we have taken a first step examining the economic value of textual content in word-of-mouth forums, we acknowledge that our approach has several limitations, many of which are borne by the nature of the data itself. Some of the variables in our data are proxies for the actual measure that one would need for more advanced empirical modeling. For example, we use sales rank as a proxy for demand in accordance with prior work. Future work can look at real demand data. Our sample is also restricted in that our analysis focuses on the sales at one e-commerce retailer. The actual magnitude of the impact of textual information on sales may be different for a different retailer. Additional work in other on-line contexts will be needed to evaluate whether review text information has similar explanatory power that are similar to those we have obtained. Despite these limitations, we hope our paper motivates future research in this area.

Acknowledgments

References

[1] Archak, N., Ghose, A., and Ipeirotis, P. G. Show me the money! Deriving the pricing power of product features by mining consumer reviews. In *Proceedings of the Twelveth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2007)* (2007), pp. 56–65.

[2] Breiman, L. Random forests. *Machine Learning 45*, 1 (Oct. 2001), 5–32.

[3] Brown, J. J., and Reingen, P. H. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research 14*, 3 (Dec. 1987), 350–362.

[4] Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2*, 2 (June 1998), 121–167.

[5] Caruana, R., Karampatziakis, N., and Yessenalina, A. An empirical evaluation of supervised learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning (ICML 2008)* (2008).

[6] Caruana, R., and Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006)* (2006), pp. 161–168.

[7] Chaiken, S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology 39*, 5 (1980), 752–766.

[8] Chaiken, S. The heuristic model of persuasion. In *Social Influence: The Ontario Symposium, Volume 5*, M. P. Zanna, J. M. Olson, and C. P. Herman, Eds. Lawrence Erlbaum Associates, 1987, pp. 3–39.

[9] Chevalier, J. A., and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research 43*, 3 (Aug. 2006), 345–354.

[10] Cui, H., Mittal, V., and Datar, M. Comparative experiments on sentiment classification for online product reviews. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-2006)* (2006).

[11] Das, S. R., and Chen, M. Y. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science 53*, 9 (Sept. 2007), 1375–1388.

[12] Dave, K., Lawrence, S., and Pennock, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International World Wide Web Conference (WWW12)* (2003), pp. 519–528.

[13] Dellarocas, C., Awady, N. F., and Zhangz, X. M. Exploring the value of online product ratings in revenue forecasting: The case of motion pictures. Working Paper, Robert H. Smith School Research Paper, 2007.

[14] Dhar, V., and Chang, E. The impact of user-generated content on music sales. Tech. Rep. CeDER-07-06, New York University University, Stern School of Business, CeDER Working Papers, 2007.

[15] Dholakia, U. M., Bagozzi, R. P., and Pearo, L. K. A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing 21*, 3 (Sept. 2004), 241–263.

[16] Duan, W., Gu, B., , and Whinston, A. B. The dynamics of online word-of-mouth and product sales. An empirical investigation of the movie industry. *Journal of Retailing 84*, 2 (2008), 233–242.

[17] DuBay, W. H. *The Principles of Readability*. Impact Information, 2004. Available at `http://www.nald.ca/library/research/readab/readab.pdf`.

[18] Forman, C., Ghose, A., and Wiesenfeld, B. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research 19*, 3 (Sept. 2008).

[19] Ghose, A., and Ipeirotis, P. G. Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality. In *Workshop on Information Technology and Systems* (2006).

[20] Ghose, A., and Ipeirotis, P. G. Designing novel review ranking systems: Predicting the usefulness and impact of reviews. In *Proceedings of the 9th International Conference on Electronic Commerce (ICEC 2007)* (2007), pp. 303–310.

[21] Ghose, A., Ipeirotis, P. G., and Sundararajan, A. Opinion mining using econometrics: A case study on reputation systems. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL 2007)* (2007), pp. 416–423.

[22] Ghose, A., and Sundararajan, A. Evaluating pricing strategy using e-commerce data: Evidence and estimation challenges. *Statistical Science 21*, 2 (May 2006), 131–142.

[23] Gruhl, D., Guha, R., Kumar, R., Novak, J., and Tomkins, A. The predictive power of online chatter. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2005)* (2005), pp. 78–87.

[24] Hass, R. G. Effects of source characteristics on cognitive responses and persuasion. In *Cognitive responses in persuasion*, R. E. Petty, T. M. Ostrom, and T. C. Brock, Eds. Lawrence Erlbaum Associates, 1981, pp. 1–18.

[25] Hatzivassiloglou, V., and McKeown, K. R. Predicting the semantic orientation of adjectives. In *Proceedings*

of the 38th Annual Meeting of the Association for Computational Linguistics (ACL'97) (1997), pp. 174–181.

[26] HENNIG-THURAU, T., GWINNER, K. P., WALSH, G., AND GREMLER, D. D. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing 18*, 1 (2004), 38–52.

[27] HOANG, L., LEE, J.-T., SONG, Y.-I., AND RIM, H.-C. A model for evaluating the quality of user-created documents. In *Proceedings of the 4th Asia Infomation Retrieval Symposium (AIRS 2008)* (2008), pp. 496–501.

[28] HU, M., AND LIU, B. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)* (2004), pp. 168–177.

[29] HU, N., PAVLOU, P. A., AND ZHANG, J. 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 (EC'06)* (2006), pp. 324–330.

[30] JARVENPAA, S. L., AND LEIDNER, D. E. Communication and trust in global virtual teams. *Journal of Interactive Marketing 10*, 6 (Nov.-Dec. 1999), 791–815.

[31] JEON, J., CROFT, W. B., LEE, J. H., AND PARK, S. A framework to predict the quality of answers with non-textual features. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2006* (2006), pp. 228–235.

[32] KIM, S.-M., AND HOVY, E. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)* (2004), pp. 1367–1373.

[33] KIM, S.-M., PANTEL, P., CHKLOVSKI, T., AND PENNACCHIOTTI, M. Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)* (2006), pp. 423–430.

[34] LEE, T. Y., AND BRADLOW, E. T. Automatic construction of conjoint attributes and levels from online customer reviews. University of Pennsylvania, The Wharton School Working Paper OPIM WP 06-08-01, 2006.

[35] LIU, B., HU, M., AND CHENG, J. Opinion observer: Analyzing and comparing opinions on the Web. In *Proceedings of the 14th International World Wide Web Conference (WWW 2005)* (2005), pp. 342–351.

[36] LIU, J., CAO, Y., LIN, C.-Y., HUANG, Y., AND ZHOU, M. Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (2007), pp. 334–342.

[37] LIU, Y. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing 70*, 3 (July 2006), 74–89.

[38] MA, M., AND AGARWAL, R. Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research 18*, 1 (Mar. 2007), 42–67.

[39] MCKENNA, K. Y. A., AND BARGH, J. A. Causes and consequences of social interaction on the internet: A conceptual framework. *Media Psychology 1*, 3 (Sept. 1999), 249–269.

[40] NIGAM, K., AND HURST, M. Towards a robust metric of opinion. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text* (2004), pp. 598–603.

[41] PANG, B., AND LEE, L. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)* (2002).

[42] PANG, B., AND LEE, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)* (2004), pp. 271–278.

[43] PANG, B., AND LEE, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)* (2005).

[44] REINSTEIN, D., AND SNYDER, C. M. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *Journal of Industrial Economics 53*, 1 (Mar. 2005), 27–51.

[45] SNYDER, B., AND BARZILAY, R. Multiple aspect ranking using the good grief algorithm. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL 2007)* (2007).

[46] SPEARS, R., AND LEA, M. Social influence and the influence of the 'social' in computer-mediated communication. In *Contexts of Computer-mediated Communication*, M. Lea, Ed. Harvester Wheatsheaf, June 1992, pp. 30–65.

[47] TURNEY, P. D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)* (2002), pp. 417–424.

[48] WEIMER, M., AND GUREVYCH, I. Predicting the perceived quality of web forum posts. In *Proceedings of the 2007 Conference on Recent Advances in Natural Language Processing, RANLP 2007* (2007).

[49] Weimer, M., Gurevych, I., and Mühlhäuser, M. Automatically assessing the post quality in online discussions on software. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL 2007)* (2007), pp. 125–128.

[50] White, S. The 2003 national assessment of adult literacy (naal). Tech. rep., National Center for Education Statistics (NCES), Institute of Education Sciences, U.S. Department of Education, Mar. 2003. Publication NCES 2003495rev. Also available at `http://nces.ed.gov/naal/`.

[51] Wilson, T., Wiebe, J., and Hwa, R. Recognizing strong and weak opinion clauses. *Computational Intelligence 22*, 2 (May 2006), 73–99.

[52] Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, 2001.

[53] Zhang, Z., and Varadarajan, B. Utility scoring of product reviews. In *Proceedings of the 2006 ACM Conference on Information and Knowledge Management (CIKM 2006)* (2006), pp. 51–57.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	7352	151.33	130.57	1	3299.99
Sales Rank	7352	7667.42	51039.42	1	2090308
Average Rating	7352	3.86	1.41	1	5
Number of Reviews	7352	195.07	138.76	1	522
Log (Elapsed Date)	7352	5.12	1.09	0.693	7.63
Moderate Review	7352	0.093	0.29	0	1
Helpful Votes	7352	5.51	11.7	0	744
Total Votes	7352	8.38	14.05	0	893
Reviewer Rating	7352	3.67	1.51	1	5
Reviewer Rank	7352	744617.2	710945.1	0	2931867
Top-10 Reviewer	7352	0	0	0	0
Top-50 Reviewer	7352	0.0005	0.02	0	1
Top-100 Reviewer	7352	0.001	0.034	0	1
Top-500 Reviewer	7352	0.005	0.067	0	1
Realname	7352	0.544	0.498	0	1
Nickname	7352	0.573	0.494	0	1
Hobbies	7352	0.081	0.272	0	1
Birthday	7352	0.203	0.402	0	1
Location	7352	0.787	0.409	0	1
Webpage	7352	0.011	0.106	0	1
Interests	7352	0.026	0.16	0	1
Snippet	7352	0.389	0.487	0	1
Any Disclosure	7352	0.52	0.49	0	1
Number of Past Reviews	3076	440.09	1065.63	1	3812
Reviewer History Macro	3076	0.69	0.23	0	1
Reviewer History Micro	3076	0.69	0.21	0	1
Past Helpful Votes	3076	2653.578	6315.109	0	20582
Past Total Votes	3076	4878.462	11744.19	0	39674
Length (Chars)	7352	790.62	786.62	33	9747
Length (Words)	7352	193.99	190.46	10	2336
Length (Sentences)	7352	10.76	10.64	1	143
Log(Spelling Errors)	7352	-3.85	0.74	-6.67	-1.34
Automated Readability Index	7352	9.206	16.616	-3.776	340.011
Gunning Index	7352	12.46	13.31	1.36	277.95
Coleman–Liau	7352	6.21	1.924	-3.156	17.147
Flesch–Reading Ease	7352	67.21	34.432	-606.119	113.1
Flesch–Kincaid Grade Level	7352	9.61	12.954	-0.347	268.22
SMOG	7352	10.117	3.31	3.13	47.747
AvgProb	7352	0.58	0.05	0.37	0.83
DevProb	7352	0.047	0.024	0	0.18

Table 10: Descriptive statistics of audio and video players for Prediction Analysis

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	2730	159.94	351.62	1	7999.99
Sales Rank	2730	395.27	1418.04	1	38353
Average Rating	2730	4.06	1.31	1	5
Number of Reviews	2730	84.69	492.62	1	3795
Log(Elapsed Date)	2730	5.46	0.55	1.38	7.02
Moderate Review	2730	0.36	0.48	0	1
Helpful Votes	2730	17.68	58.58	0	1669
Total Votes	2730	21.67	59.68	0	1689
Reviewer Rating	2730	4.062	1.306	1	5
Reviewer Rank	2730	444041.5	526740.6	0	2928764
Top-10 Reviewer	2730	0	0	0	0
Top-50 Reviewer	2730	0.001	0.033	0	1
Top-100 Reviewer	2730	0	0	0	0
Top-500 Reviewer	2730	0.004	0.063	0	1
Realname	2730	0.61	0.487	0	1
Nickname	2730	0.577	0.494	0	1
Hobbies	2730	0.075	0.264	0	1
Birthday	2730	0.237	0.425	0	1
Location	2730	0.817	0.386	0	1
Webpage	2730	0.022	0.149	0	1
Interests	2730	0.038	0.191	0	1
Snippet	2730	0.368	0.482	0	1
Any Disclosure	2730	0.58	0.49	0	1
Number of Past Reviews	1085	84.697	492.624	1	3795
Reviewer History Macro	1085	0.78	0.22	0	1
Reviewer History Micro	1085	0.77	0.21	0	1
Past Helpful Votes	1085	574.61	2836.533	0	20552
Past Total Votes	1085	982.993	5321.077	0	39512
Length (Chars)	2730	813.7	949.31	22	18545
Length (Words)	2730	194.32	221.62	4	4290
Length (Sentences)	2730	10.8	11.22	1	160
Log(Spelling Errors)	2730	-4.23	0.86	-7.03	-0.84
Automated Readability Index	2730	8.391	10.214	-4.158	139.592
Gunning Index	2730	13.07	8.25	1.2	117.44
Coleman–Liau Index	2730	6.883	1.954	-3.049	15.602
Flesch–Reading Ease	2730	63.063	22.294	-201.779	114.115
Flesch–Kincaid Grade Level	2730	9.498	7.984	-0.67	111.708
SMOG	2730	11.587	2.984	3.129	36.439
AvgProb	2730	0.55	0.02	0.46	0.67
DevProb	2730	0.03	0.01	0.001	0.17

Table 11: Descriptive statistics of digital cameras for Prediction Analysis

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	2018	21.30	7.81	1	71.96
Sales Rank	2018	3824.28	6168.97	6	49868
Average Rating	2018	4.05	0.62	1	5
Number of Reviews	2018	84.54	88.36	1	431
Log (Elapsed Date)	2018	4.38	1.29	0.693	7.55
Moderate Review	2018	0.45	.49	0	1
Helpful Votes	2018	6.35	13.63	0	246
Total Votes	2018	10.49	19.31	0	320
Reviewer Rating	2018	3.93	1.28	1	5
Reviewer Rank	2018	352120.6	595227.5	0	2932165
Top-10 Reviewer	2018	0.009	0.095	0	1
Top-50 Reviewer	2018	0.01	0.116	0	1
Top-100 Reviewer	2018	0.011	0.103	0	1
Top-500 Reviewer	2018	0.058	0.235	0	1
Realname	2018	0.513	0.499	0	1
Nickname	2018	0.748	0.433	0	1
Hobbies	2018	0.295	0.456	0	1
Birthday	2018	0.28	0.449	0	1
Location	2018	0.863	0.34	0	1
Webpage	2018	0.058	0.234	0	1
Interests	2018	0.11	0.315	0	1
Snippet	2018	0.42	0.49	0	1
Any Disclosure	2018	0.74	0.43	0	1
Number of Past Reviews	1450	110.637	466.792	1	3777
Reviewer History Macro	1450	0.68	0.21	0	1
Reviewer History Micro	1450	0.70	0.19	0	1
Past Helpful Votes	1450	743.29	2789.311	0	20441
Past Total Votes	1450	1224.388	5153.51	1	39369
Length (Chars)	2018	981.10	936.68	16	7969
Length (Words)	2018	221.49	207.29	4	1826
Length (Sentences)	2018	11.01	9.76	1	98
Log(Spelling Errors)	2018	-4.34	0.91	-7.44	-1.003
Automated Readability Index	2018	9.834	7.892	-1.876	164.429
Gunning Index	2018	13.09	6.47	0.8	136.83
Coleman–Liau Index	2018	8.239	2.372	-7.24	18.435
Flesch–Reading Ease	2018	60.717	19.272	-253.357	103.693
Flesch–Kincaid Grade Level	2018	10.085	6.193	-0.06	130.33
SMOG	2018	11.628	3.013	3.13	40.591
AvgProb	2018	0.52	0.012	0.46	0.56
DevProb	2018	0.016	.01	0.00003	0.07

Table 12: Descriptive statistics of DVD for Prediction Analysis