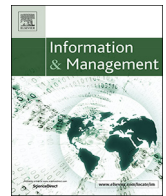




ELSEVIER

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

The effect of online reviews on product sales: A joint sentiment-topic analysis

Xiaolin Li^a, Chaojiang Wu^{b,*}, Feng Mai^c

^a Department of e-Business and Technology Management, College of Business and Economics, Towson University, Towson, MD, 21252, United States

^b Decision Sciences and MIS, LeBow College of Business, Drexel University, United States

^c School of Business, Stevens Institute of Technology, United States

ARTICLE INFO

Keywords:

eWOM
Electronic word-of-mouth
Star rating
Textual sentiment
Topical sentiment
Product reviews
Sales
Joint sentiment-topic analysis
JST
Mediation analysis

ABSTRACT

This research examines the business impact of online reviews. It empirically investigates the influence of numerical and textual reviews on product sales performance. We use a Joint Sentiment-Topic model to extract the topics and associated sentiments in review texts. We further propose that numerical rating mediates the effects of textual sentiments. Findings not only contribute to the knowledge of how eWOM impacts product sales, but also illustrate how numerical rating and textual reviews interplay while shaping product sales. In practice, the findings help online vendors strategize business analytics operations by focusing on more relevant aspects that ultimately drive sales.

1. Introduction

Online reviews play an essential role in shaping customers' awareness and perceptions about products [1–3]. As a major source of electronic word-of-mouth (eWOM), online reviews serve as a reliable source of information about the quality of goods, particularly goods that cannot be easily characterized before its use [1]. On e-commerce platforms, online product reviews enable customers to evaluate and compare alternatives before making purchase decisions [4]. Therefore, it is considered as a main driver for future product sales [5].

A considerable amount of research has studied the relationship between online reviews and product sales [5–9]. Although most evidence suggests that collectively eWOM has an impact on future sales, the findings are not always consistent. For example, Duan et al. [6] find that the volume of eWOM has a positive effect on future movie revenues, while Chintagunta et al. [7] show that only the valence (star ratings) of reviews matters. The key to resolving these conflicting findings is to understand how consumers process the information embedded in eWOM. As Hu et al. [5] point out, consumers pay attention to contextual information beyond the simple statistics such as ratings and volume. As a result, the influence of reviews on sales hinges on other factors such as the strength of the brand [10], reviewer reputation [5], reviewer location [11], and review text [12,13]. A better understanding

of how the information embedded in the reviews drives sales can help businesses exploit the value of eWOM through more accurate forecasting, promoting new products, and attracting and retaining shoppers.

We contribute to the literature by proposing a new mediation model, whereby numerical “star rating” partially mediates the relationship between review texts and product sales. A typical product review contains two types of information – the numerical rating and the review text. The numerical rating is a quantitative summary of the reviewer's experiences, attitudes, opinions, or sentiments toward a product or service, usually expressed as number of stars. The review text is an open-ended textual description of the reviewer's opinions toward the product or service [14,15]. Extant research on the economic impact of eWOM focuses on numerical ratings but rarely addresses textual reviews [16], in part due to the complexity of text analysis. Few studies that incorporate textual reviews use techniques such as sentiment polarity [12] or frequent noun phrases [13]. Yet, much of the value of product reviews lies in conveying “attributes and attribute dimensions using the ‘voice of the consumer’” [17]. Capturing the full economic impact of online reviews may require us to uncover the dimensions that consumers care about. New text analytics methods that go beyond sentiment analysis and counting phrase are needed.

We introduce a methodological tool to the online reviews literature

* Corresponding author.

E-mail addresses: xli@towson.edu (X. Li), cw578@drexel.edu (C. Wu), feng.mai@stevens.edu (F. Mai).

<https://doi.org/10.1016/j.im.2018.04.007>

Received 26 May 2017; Received in revised form 9 April 2018; Accepted 16 April 2018
0378-7206/ © 2018 Elsevier B.V. All rights reserved.

– the joint sentiment-topic model (JST) [18]. This unified machine learning model achieves two goals at the same time: it not only summarizes the sentiment in the review text, but also identifies the aspects of the product that the reviewer is happy with or critical of. JST provides a much richer representation of the qualitative review data. The outputs allow us to investigate, among other things, how positive or negative valence of specific product features lead to changes in future sales. We proceed to study the impact of textual reviews and numerical ratings on the actual sales of 312 tablet PC products using a panel dataset. Furthermore, we conduct a mediation analysis to study the interplay of textual review and numerical review ratings using Baron and Kenny's [19] approach.

The findings from JST and mediation model enhance the understanding of how online reviews provide information cues and shape product sales. We show that reviews with positive and negative valence focus on different sets of product aspects. More importantly, the positive and negative aspects have different impacts on sales performance. The numerical ratings mediate the effects of textual reviews that discuss negative aspects of a product. But the effects from textual reviews that carry positive valence persist in the mediation model. In a nutshell, reviews that highlight the positive aspects of the product provide an extra boost to sales that cannot be captured simply by a “5-star” rating. Our findings underscore the importance of analyzing social media data to e-commerce. Our research framework can also help online vendors strategize their business analytical initiatives by focusing on more relevant aspects of eWOM. Additionally, we demonstrate an innovative approach to analyzing textual data along with numerical data, which may be valuable for similar research in the future.

The rest of the paper is organized as follows. Section 2 discusses the theoretical and literature background of the research and formulates hypotheses. Section 3 presents the data and research method used for the study. Section 4 reports the analysis and results. Section 5 concludes the paper with a discussion of implications, limitations, and avenues for future research.

2. Literature review and hypothesis development

Many mechanisms can account for how eWOM affects future product sales. First, online reviews can serve as a signaling device in the context of imperfect information [9,20]. In online shopping, the prospective buyers usually lack the experience that product reviewers have. Through prior purchases and usage, reviewers possess valuable information about the product such as quality, value, and potential issues – the information that prospective buyer needs but lacks for comparing alternatives. Prior to making a purchase decision, therefore, the prospective buyer would seek various signals from product reviews and ratings. For example, high numerical rating or a large volume of reviews can be interpreted as signals of high quality [21]. Furthermore, review volume, valence, and sentiments in online reviews may directly affect customers' choice behavior. Higher valence, represented by higher average star rating, leads to higher choice probability [22]. Prior studies [6,8,17][e.g. 6,8,17] also found that online product reviews, either in numerical or textual format, may influence online shoppers' perception of the key factors in utility function such as brand, quality, price, values, and product attributes, which in turn, shapes sales performance of the product. An array of empirical studies have validated the relationship between online product reviews and sales performance, including the impact on the sales of books [5,8], movies [6], electronics [11,13], video games [9], to name a few. We refer readers to [23] and [24] for a more detailed literature review.

Our paper differs from prior work on the relationship between online reviews and sales in three respects. First, we propose a mediation model for the interplay of textual and numerical ratings. Instead of examining the direct relationship between eWOM variables and sales, we study the inter-relationship between qualitative review texts, quantitative numerical ratings, and sales. To this end, the work closest

to ours is that by Hu et al. [12]. They find that numerical ratings do not have any direct impact on sales, but instead have an indirect impact through sentiment in review texts. We provide empirical evidence contrary to their results and show that numerical ratings partially mediate the effects of textual reviews. In addition, Hu et al.'s model only includes review-level sentiments. Each review text is scored from strong negative to strong positive. We include aspect-level sentiments in the analysis. We find that changes in overall sentiments are not the only text feature that predicts sales – changes in sentiments relating to specific product dimensions also matter.

Second, we make methodological advancement by combining a novel machine learning model with the econometric analysis. Although using topic models on product reviews is not new [e.g. 17,25], most studies use it as a tool for exploratory analysis. To the best of our knowledge, no published study has examined the impact of the changes in positive or negative topics on product sales. Also, the JST model differs from the latent Dirichlet allocation (LDA) model [26] that is commonly employed in prior studies. Instead of assuming that all reviews share the same set of topics, JST extracts different topics under positive and negative reviews. The model reveals an important insight: consumers focus on different aspects when they express positive and negative sentiments.

Third, due to the difficulty in obtaining sales data, most extant research has used sales rankings of products as a proxy for actual sales. The often-cited rationale is that sales rank and actual sales follow a log-linear relationship [8], such that the marginal effect on sales rank can be interpreted as an effect on sales. This simplification can yield misleading findings for two reasons. First, the log-linear relationship is based on empirical observations that the ranks of books [27] and software [28] follow a Pareto distribution. Such inductive reasoning may not hold for all product categories. Second, even if a log-linear or other relationship between sales rank and sales holds empirically, the transformation is not exact. The interpretation of the effects of online product reviews is thus subject to measurement errors. We combine a reviews dataset with a proprietary dataset that records the exact sale quantities of products. This allows us to directly test the relationship instead of relying on proxies.

The research model we propose in this study is built upon information processing theory, which states that humans process the information they receive, rather than just respond to stimuli. Based on information processing theory articulated by Miller [29], qualitative information and quantitative information are processed differently: while qualitative information processing involves the use of language to represent concepts, quantitative information is processed to remember more items in working memory. The qualitative and quantitative components of information received often interact within the processing system. Our research framework in the present study builds on Miller's information processing theory to model the impacts of qualitative text reviews and quantitative star ratings on sales separately. In addition, we test their interplay on shoppers' purchase decisions (sales) through a mediation model.

2.1. Star ratings

Numerical rating, one of the most common formats in product reviews, is assigned by the reviewer to the product. It is commonly displayed in a star rating format ranging from one star or very negative, to five stars or very positive [30]. The numerical rating represents the reviewer's overall assessment towards the product. It is not only an indicator of product quality, but also may be a valuable reflector of product value [31]. Prior research has offered abundant empirical evidence of the impact of product reviews on the sales performance of a product. For example, with a dataset from Amazon.com, Jabr and Zheng [32] illustrated that product ratings influence product sales ranks within a competitive market. Similarly, Moe and Trusov [33] suggested that product ratings not only directly impact sales, but also

indirectly influence future ratings, which in turn, further affect sales. Therefore, we hypothesize:

H₁. Numerical ratings of a product's reviews will positively influence the sales performance of the product.

2.2. Textual sentiments

Product reviews are multifaceted. In addition to numerical ratings, textual comments of product reviews are also likely to play an important or even determinant role in consumers' purchase choices [13]. While star ratings can indicate how much consumers like a product, textual review comments can reveal consumers' deep thoughts and detailed product experiences, which may lead to practically useful insights into why certain products succeed or fail [8]. Textual reviews are important not only because prospective customers do read them before making purchase decisions [8], but also because textual reviews are at least equally important in affecting the customers' purchase decisions. Numerical ratings are usually marked with bimodality – that is, a big portion of the ratings is either extremely high or extremely low [20]. The lack of reasonable variation in numerical ratings often makes them unable to reflect the true quality and value of the reviewed product, and thus undermine their role as a sole determinant of purchase decision-making [34]. As an important indicator of valence of online review [35], sentiments reflected in textual reviews not merely serve as unique cognitive appraisals from previous customers that offer useful information cues for the cognitive processing of prospective consumers [12,36], but also serve as an emotional contagion that passes the positive or negative emotions from previous customers to prospective customers [37]. For instance, Pavlou and Dimoka [38] revealed that the rich content of text comments plays an important role in building a buyer's trust in a seller's benevolence and credibility.

Some extant research has demonstrated, from different perspectives, the influence of textual sentiments on product sales. Jabr and Zheng [32] suggested that product reviewer opinions impact product sales rank within a competitive market. Yu et al. [39] found that the sentiments expressed in the product reviews have a significant impact on the future sales performance of the product. Ludwig et al. [40] suggested that affective contents of online reviews influence conversion rates. Similarly, Floh et al. [41] and Ketelaar et al. [42] suggested that textual reviews with stronger valence intensity lead to higher levels of purchase intentions. In other words, strong positive and negative comments have stronger impact on behavioral intentions than messages with mixed positive and negative opinions.

Based on the above analysis, we hypothesize:

H_{2a}. Reviewers' overall sentiments toward a product, expressed in textual reviews, will positively impact the sales performance of the product.

In addition to the effects of the overall sentiments, we posit that variations in sentimental topics have impacts on the sales performance. We define sentiment-topic aspects (aspect-level sentiments) [13,43] as the sentiments associated with product dimensions that are salient in product reviews. Information processing theory focuses on the cognitive process that occurs before a choice is made [35]. It can be used to explain consumer behavior in terms of cognitive operations [44]. Specifically, consumers are unlikely to consider the review text as a whole when making their choices. Instead, incoming information in the reviews will be processed and stored in active memory, and will later be retrieved when consumers make a purchasing decision. In consumer research, it is well known that a person's attitude towards an alternative is determined by the weighted sum of belief that the person has about the individual attributes of the alternative [45,46] – a framework known as the multi-attribute attitude model. Therefore, it is plausible that the information in reviews is processed and stored according to the multi-attribute model. It follows that the sentiments about the main

Table 1
Descriptive Statistics for Weekly Log(Sales).

Week	Mean	Std. Dev	Unique SKU's
2	1.059	0.782	231
3	1.056	0.807	240
4	0.950	0.809	264
5	0.966	0.794	276
6	0.951	0.809	279
7	0.913	0.787	282
8	0.895	0.781	287
9	0.848	0.787	288
10	0.850	0.767	288
11	0.795	0.757	291
12	0.787	0.765	292
13	0.757	0.756	297
14	0.798	0.742	297
15	0.798	0.768	301
16	0.765	0.752	303
17	0.715	0.748	304
18	0.693	0.742	304
19	0.723	0.755	305
20	0.731	0.768	309
21	0.720	0.757	309
22	0.682	0.739	309
23	0.638	0.707	312
24	0.633	0.709	312
Total	0.806	0.772	6680

Note: This table describes the statistics of Log(sales) for each week. Column 1 and 2 are respectively the average weekly Log(sales) and the standard deviations for all the products. Column 3 is the number of unique SKU's observed during each week.

product attributes (topics) in the reviews may determine the readers' attitude towards the product, and hence impact sales. A recent empirical study by Liang et al. [47] also supports this view. The authors examined how the sentiments of two major topics in online reviews, product quality and service quality, affect apps' sales rankings. They found that even though consumers' opinions on product quality occupies a larger portion of consumer reviews, their comments on service quality have a stronger effect on sales rankings. With the above theoretical and empirical evidence, we hypothesize:

H_{2b}. Different sentimental topics in textual reviews will have different influences on the sales performance of the product.

2.3. Interplay of textual sentiments and star ratings

Although eWOM—both numerical ratings and text reviews—can influence product sales, it is the interplay among them that shapes eventual consumer purchase decisions. According to information processing theory, numerical ratings and quantitative text reviews are processed in different manners: while text reviews use language to represent abstract concepts, numerical ratings use quantitative summaries to process information on the reviewed product [29]; further, based on information processing theory, text and numerical components of a product review would often interact within the processing system. As numerical and text reviews product reviews are presented on a website simultaneously, they may not be independent of each other when influencing review readers' perceptions towards the product, and ultimately, their purchase decisions. Investigating their interplay is thus particularly meaningful [35].

Prior research indicated that numerical ratings and textual comments might work separately or in combination [41]. On the one hand, star ratings may first decide whether a reader will read the reviews. The large body of texts of the reviews for each product practically creates difficulties for the reader to choose which textual reviews to read. One may choose to read a subset of reviews prior to making purchasing decisions. The star rating and its distributions could affect how the

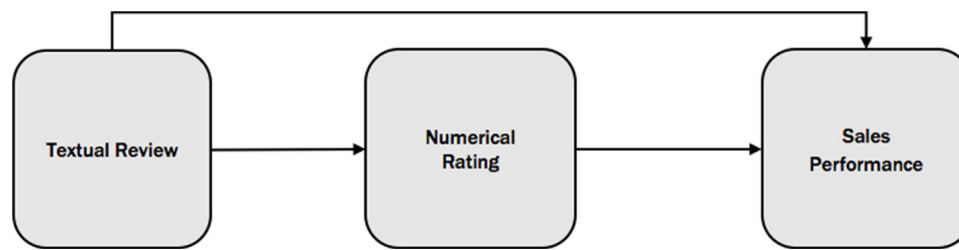


Fig. 1. Research Framework.

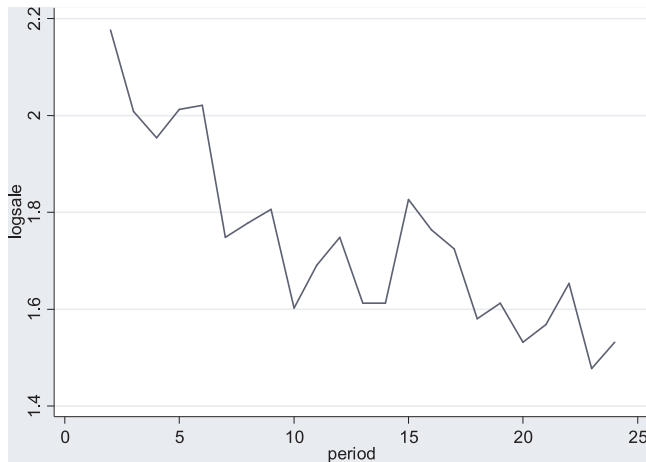


Fig. 2. Weekly Log(Sales) for a Product Introduced in a period (year 2010) before the observational period (item 116).

reader decides to read the textual reviews further. The reader may choose a few representative reviews for each star rating to read. Thus, star rating may bridge the effect of review sentiments on purchasing decisions and product sales. Some studies in the literature [32,40,41] revealed that review sentiments impact sales. Other studies [48] suggested that review sentiments affect star ratings. As another example, Gan et al. [49] examined the influence of review attributes and sentiments on restaurant star ratings and confirmed that five attributes and sentiments in text reviews—food, service, context, price, and ambience—significantly impact star ratings. Still, others found star ratings influence sales [32,33]. With these findings on the relationships among review sentiments, star ratings, and sales in prior research, one may logically wonder whether text review sentiments and star rating interplay in their influences on sales. One plausible way of such interplay lies in the possibility that the impact of review comments on sales is partially or even completely mediated by star ratings.

No research has been conducted to examine the meditational role of star ratings in the relationship between review sentiments and sales. But the findings of some prior research have signaled that such mediation might exist. For instance, Tang et al. [50] examined the effects of star ratings and emotions in text reviews on profitability (an indicator of sales) and they found that star ratings significantly, positively, affect profitability. Further, they suggested positive emotions statistically predict higher profitability and negative emotions predict lower profitability. With the findings of this study, we cannot help but wonder: is a part or all of the effects of emotions on profit channeled through star ratings?

Moon et al. [16] examined the interplay through proposing and comparing different models that integrate text reviews into star ratings-sales regression model. They found that the introduction of sentiments in text reviews improves the predictive power of a linear regression model with the numeric ratings as an independent variable to explain the product sales as the dependent variable, implying that sentiments

may have a direct and/or mediated (by star rating) effect on sales.

In the present research, our joint sentiment-topic model decomposes the review texts into parts with positive and negative sentiments, under which different topics are nested. The sentiment of the review texts is usually congruent with the overall star rating. Thus, the star rating can often reflect the level of satisfaction of the reviewer. For example, a reviewer's satisfaction expressed in a textual product review will be reflected in a high star rating.

Based on the analysis above, we believe that review sentiments—both overall and sentiment topic—may have an indirect effect on sales, bridged by star ratings. Thus, we formulate the following hypotheses:

H_{3a}. The effect of reviewers' overall sentiments on sales performance will be mediated by star ratings.

H_{3b}. Star ratings will mediate the relationships between sentiment-topic aspects and sales.

Summarizing the above hypotheses, we propose our research model, which is illustrated in Fig. 1.

3. Research method

3.1. Data

The online reviews dataset is titled “Market Dynamics and User-Generated Content about Tablet Computers” and is provided by Wang, Mai and Chiang [51]. The dataset contains 88,901 consumer reviews on 794 tablet computer products or SKUs. The weekly market dynamics and reviews data were collected using a Java web crawler during a 24-week period from February 1 to July 11, 2012.¹ Each review contains the numerical rating ranging from 1 to 5, review text, and an indicator variable for whether the reviewers disclosed their real names. We also control for several product-related variables in the dataset, including the price and product attributes on the spec sheets (RAM, Processor, Screen Size). Note that the price of a product can change during the sample period due to promotions or pricing strategies. We aggregate all time-varying variables to product-week level to study the effect of eWOM on product sales.

We augment the data by obtaining actual sales.² After merging with sales data, we end up with 312 unique SKUs with sales records spanning 23 weeks, starting from the second week of the 24-week period. Table 1 summarizes the descriptive statistics of the weekly log-sales. There is on average a decreasing trend in the weekly sales for two reasons. On one hand, for hedonic products such as tablet computers, the life cycle of sales is relatively short, and the weekly sales can drop fast. On the other hand, new products were introduced in this period and their sales are low in the beginning. Fig. 2 shows the weekly sales for a product that was introduced to the market before the sample period. As expected,

¹ The 24-week panel dataset is comparable in length with prior studies. For example, [8] use 2 periods spanning two years, [6] include daily observations for two weeks, and [12] use a panel dataset of 26 periods.

² We thank an online vendor for providing us with the data. The sales data do not contain Amazon products such as Kindle Fire.

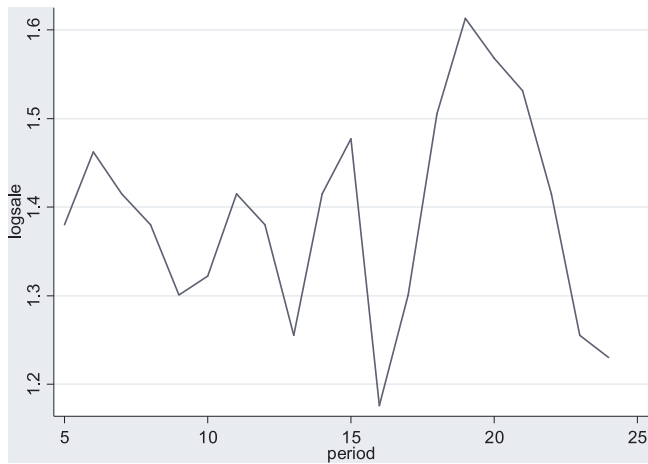


Fig. 3. Weekly Log(Sales) for a Product Introduced in Week 5 (item 119) of the observational period.

the product's sales quantity shows a decreasing trend over the 24-week period. Fig. 3 illustrates the weekly sales for a relatively new product, which is launched in the week 5 of the 24-week period. Its sales quantity exhibits significant weekly fluctuations. To account for such heterogeneity in our study, we control for the number of weeks since the product's introduction to the market (*WeekIntro*). Negative numbers indicate the introduction time is after the first week of observation.

Historically the sales data are very difficult to obtain so the sales rank data are often used as a proxy [27]. The relationship of sales and rank are usually estimated using a log-linear model

$$\log(\text{sales}) = \beta_0 + \beta_1 \log(\text{rank}) + \varepsilon.$$

The estimation is usually done with a standard ordinary least square method. While the coefficients may depend on the applications such as product type, the R^2 is usually as high as around 0.8. Brynjolfsson, Hu and Smith [27] and Brynjolfsson, Hu and Simester [52] found that there is a long tail on online markets, meaning sales distribution is less concentrated and niche products sales can be realized more so than offline markets, partially due to product availability and partially due to lower search costs on Internet.

Since we observe the actual sales and the sales rank for a class of products, we can directly examine whether the empirical evidence supports the log-linear model in our setting. Fig. 4 plots the scatter plot of log-sales and log-sales rank. The log-sales drop faster in $\log(\text{rank})$ than a log-linear model would suggest. This means that for lower ranked products, the sales are much lower than higher ranked products. Fig. 4 also shows that the variance of the residuals increases with the sales rank, suggesting that it is less accurate to impute sales from sales rank using the log-linear model. For the above reasons, using actual sales as the dependent variable captures the relationship between ratings, reviews, and sales more accurately.

3.2. Mining sentiments and topics from textual reviews

We decompose the sentiment expressed through review texts into topics with different sentiments, which correspond to major quality dimensions of the products that reviewers are happy or unhappy with. The measure we use in our empirical model combines the information from both sentiment mining and unsupervised text clustering approach. In conventional sentiment mining, the goal is to derive the information on whether a review contains positive or negative opinions – yet it does not reveal what the reviewer likes or dislikes about the product. On the other hand, in an unsupervised text clustering (e.g. LSA) or topic modeling (e.g. LDA) approach, the review texts are categorized into latent clusters or topics that correspond to quality aspects. The outputs

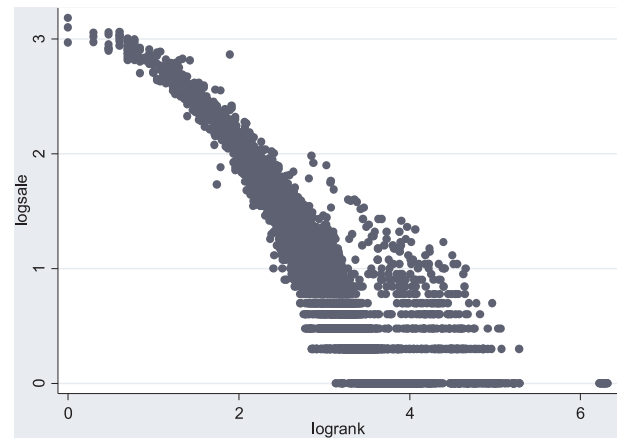


Fig. 4. Scatter Plot of Log(Sales) and Log(Sales Rank).

help us understand what aspects of the product the reviewers are commenting, but not how they feel (like/dislike) about those aspects.

In order to construct the measures, we need an automated method to identify the sentiments of reviews, the major product quality aspects (topics) nested within the sentiment labels, and the valence of the topics. We choose the Joint Sentiment-Topic (JST) model [18]. JST model is an extension of the popular latent Dirichlet allocation (LDA), which is used to discover the latent ideas contained in the documents and identify reviewers' sentiments/opinions on those ideas. JST assumes that the act of user generating reviews can be decomposed into a number of simple probabilistic steps:

1. Reviewers have, in general, two sentiment polarities. We label them as $l \in \{\text{pos}, \text{neg}\}$.
2. After experiencing a product, the reviewer decides how much she likes a product and writes a review accordingly. For example, if she likes the product overall, she might write 90% positive and 10% negative in the review. Using subscript d as the review index, we denote this distribution as π_d .
3. The reviewer then decides to be more specific and lays out what and how much she likes and dislikes about the product. Note that the aspects she likes may or may not correspond with what she does not like. In other words, JST allows the topics nested under sentiment labels to be different. We denote this valence towards different aspects as topic distribution $\theta_{d,l}$.
4. Based on step 2 and 3, the reviewer chooses her words such that some words are more likely to be used to describe a complaint towards an aspect, and other words be used for a compliment towards another aspect. More specifically, before writing each word w_i in the review d , the reviewer decides to

- a Choose a sentiment label $l_i \sim \text{Multinomial}(\pi_d)$,
- b Choose a topic label $z_i \sim \text{Multinomial}(\theta_{d,l_i})$,
- c Choose a word w_i from φ_{l_i, z_i} , which is a distribution that governs word usage based on the chosen sentiment label and topic label.

JST is built upon the foundation of Bayesian statistical inference and therefore has a principled model fitting and selection procedures. The model estimation can be completed using a Gibbs sampling procedure. The outputs of JST offer "soft" classification for the review dataset. For each review, it automatically extracts the valence of different aspects that the reviewers like and dislike about the product, in addition to the overall satisfaction towards the product.

Before training JST, one assumption is that there are positive and negative sentiment labels. We need prior information to start the model so that φ_{l_i, z_i} are reasonably initiated. We use the MPQA Subjectivity Lexicon [53] for this purpose. The MPQA lexicon provides 8222

Table 2
Topics from the Joint Sentiment-Topic Model.

	Mean	S.D.	Top Keywords
Positive Topics	0.52	0.25	
Hedonic Experience	0.63	0.36	game, movie, happy, enjoy, fast, music
Hardware	0.37	0.36	camera, battery life, USB, keyboard, SD card
Negative Topics	0.48	0.25	
Interface	0.58	0.30	access, browser, page, touch, content, button
Logistics&Service	0.42	0.30	charge, order, call, receive, week, item, ship
N	40741		

Note: This table describes the results for the Joint Sentiment-Topic model of individual reviews. We list the average prevalence (mean) of the two positive topics and two negative topics and the variations (S.D.) in reviews. The table also lists the top keywords for each sentiment-topic combination. We label the two positive topics as Hedonic Experience and Hardware and the two negative topics as Interface and Logistics & Service. These topics are nested under positive and negative sentiments. For example; conditional on a piece of review text with positive sentiment; 63% of its content is describing hedonic experience; and 37% of its content is describing hardware.

subjective words labeled according to polarity (positive/negative) and strength (strong/weak). We assign a prior probability of 0.9 to the strong subjective words in the direction of their sentiment polarity, and 0.7 for the weakly subjective words. For example, the word “complicated” is weakly negative according to MPQA, therefore, it has a prior probability of 0.7 being negative in a review. On the contrary, “fantastic” is a strong positive word according to MPQA, so we assign a prior probability of 0.9 for its positive tendency in a review.

Another important parameter of JST is the number of topics. We trained JST models with the number of topics from 2 to 5 on 90% of the data and evaluated the perplexity on the 10% of the held-out set. We chose 2 topics for each sentiment label (4 total topics) because it offered the good perplexity performance on the held-out set and has high interpretability at the same time.

Table 2 summarizes the major dimensions that consumers cared about for Tablet PCs, manifested as four topics in the reviews. The total proportions of the positive and negative content are about equal (52%–48%). The results further illustrate that, when expressing positive sentiments toward a tablet, reviewers tend to comment on their hedonic experiences (63%) and hardware features (37%) of the products. When complaining about a product, however, reviewers tend to focus on the user interface of the tablet (58%) and the non-physical, or “augmented” features of the product—logistics and customer service (42%).³

The results suggest that sentiment alone is insufficient for capturing the full range of information in textual reviews. Besides expressing whether they are happy with the product in general, reviewers often discuss why, and to what extent, a specific aspect of the purchase prompts them to leave the review. Further, the positive and negative aspects of the textual review are asymmetric. For instance, reviewers rarely praise a product because the logistics and service are exceptional. They are, however, much likely to leave a negative review when the shipping is late. These observations lead us to our next inquiry: how do sentiments and aspects of the textual review influence future sales? And what is the role of numerical ratings? We use a set of econometric models to answer these questions and test our hypotheses. The variables of interests are the (weekly average) star ratings of reviews, the overall sentiment in the textual reviews (total proportion of the positive topics), and the proportion of four specific topics nested under either positive and negative sentiment. The four topic variables *Hedonic*,

³ The percentage here is conditional on the sentiment. The unconditional percentage of review texts that is complaining about logistics and customer service would be $48\% \times 0.42\% = 20.16\%$.

Hardware, *Interface* and *Logistics & Service* are respectively the positive and negative sentimental topics (Table 2) mined from the text reviews. Table 3 summarizes all the variables used in our econometric analysis. Table 4 presents the summary statistics and correlation matrix for the variables.

4. Empirical analysis and result

4.1. Empirical model and estimation

To establish the relationship of numerical ratings and textual reviews on sales, we follow [13], and use a dynamic panel data (DPD) model with the following estimation equation

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \gamma z_{it} + \varepsilon_{it}, \quad (1)$$

$$\varepsilon_{it} = \zeta_i + u_{it} \quad (2)$$

where y_{it} is $\log(\text{sale}_{it})$, or the log-sales for product i at week t . x_{it} is the review variables, which could be numerical ratings, overall sentiment, or positive and negative aspects of the products. z_{it} is a vector of control variables such as price, and measure of product newness such as time of the introduction. The error structure ε_{it} is given in (2) where ζ_i models the time-invariant product-specific unobserved effects and u_{it} is the usual idiosyncratic error term.⁴ The DPD model contains the lag of the dependent variable on its right-hand side, allowing for the modeling of partial adjustment mechanism. Such partial adjustment addresses endogeneity arising from the fact that word-of-mouth can be influenced by past sales. The parameter α reflects the persistence in the process of partial adjustment and $|\alpha| < 1$. Similar to [13], we made sure the numerical ratings and sentiments are at least one period before the sales.

The above DPD model cannot be consistently estimated using the least square dummy variable estimator. That is because the within transformation which is used to get rid of product-specific fixed effect ζ_i leads to correlation between the lagged dependent variable and the error term. The resulting correlation creates the ‘Nickell bias’ [54], which cannot be mitigated by large sample of product units. Anderson and Hsiao [55] suggest using first difference transformation to remove the fixed effect

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta x_{it} \beta + \Delta z_{it} \gamma + \Delta u_{it}$$

and then use instrument variables such as the lagged dependent variable $y_{i,t-2}$, uncorrelated with the error term Δu_{it} , to obtain a consistent estimator. However, the above approach does not employ all orthogonality conditions in the sample and thus is inefficient.

Within the generalized method of moments (GMM) framework, Arellano and Bond [56] developed a difference GMM estimator for the above model. The Arellano-Bond estimator allows for the inclusion of all the lagged values of $y_{i,t-1}$ and x_{it} to generate orthogonality conditions, leading to the improvement on the efficiency of the Anderson-Hsiao estimator. However, the original Arellano-Bond estimator is potentially weakened by the fact that lagged level variables are often rather poor instruments for first differenced variables. Arellano and Bover [57] and further expand the Arellano-Bond estimator by using the lagged differences and the lagged levels. The expanded estimator is called System GMM estimator, in contrast to the original difference GMM estimator. The implementation of System GMM is available in STATA package *xtabond2*, as documented with great details in Roodman [58]. We apply the finite sample correction [59] to correct the standard errors in two-step estimation, without which the standard errors are downward biased. For robust statistical inference, we use heteroskedasticity and autocorrelation robust standard errors.

To understand the relationship of the joint effect of star ratings and

⁴ For clarity of the mathematics, we present here only the product fixed effect. We have also added a time fixed effect η_t to remove time trends, which captures the common time trend of all products.

Table 3
Variable Definitions.

Variable	Definition
<i>Log (Sales)</i>	The natural Log of weekly sales quantity of a product
<i>StarRating</i>	The average star rating of the reviews for a product
<i>OverallSentiment</i>	The average sentiment of the reviews estimated using the joint sentiment-topic model.
<i>Hedonic(+)</i>	The average proportion of review texts with positive sentiment on hedonic experience
<i>Hardware(+)</i>	The average proportion of review texts with positive sentiment on hardware
<i>Interface(-)</i>	The average proportion of review texts with negative sentiment on user interface
<i>Logistics&Service(-)</i>	The average proportion of review texts with negative sentiment on logistics and service
<i>N_Reviews</i>	Total number of reviews for a product in a week
<i>Price (\$100)</i>	Price of the product in the week
<i>WeekIntro</i>	Number of weeks since the product was introduced
<i>RealNamePct</i>	Proportion of reviews that display a “Real Name” badge next to the review, revealing the identity of the reviewer
<i>RAM (MB), Processor (GHz), Screen Size (inches)</i>	Major physical product attributes of the tablet computer on the spec sheet

Table 4
Descriptive Statistics and Correlation.

Variable	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10
1 <i>Log (Sales)</i>	0.81	0.77	0.00	3.18										
2 <i>StarRating</i>	3.24	0.99	1.00	5.00	0.41*									
3 <i>OverallSentiment</i>	0.50	0.12	0.07	0.93	0.18*	0.57*								
4 <i>Hedonic(+)</i>	0.44	0.22	0.01	0.97	0.06*	-0.13*	-0.33*							
5 <i>Hardware(+)</i>	0.56	0.22	0.03	0.99	-0.06*	0.13*	0.33*	-1.00*						
6 <i>Interface(-)</i>	0.46	0.18	0.02	0.98	0.19*	0.36*	0.33*	-0.26*	0.26*					
7 <i>Logistics&Service(-)</i>	0.54	0.18	0.02	0.98	-0.19*	-0.36*	-0.33*	0.26*	-0.26*	-1.00*				
8 <i>N_Reviews</i>	45.78	94.10	1	690	0.51*	0.19*	0.05*	0.01	-0.01	0.16*	-0.16*			
9 <i>Price (\$100)</i>	3.67	3.58	0.50	31.10	0.07*	0.37*	0.27*	-0.40*	0.40*	0.28*	-0.28*	-0.03*		
10 <i>WeekIntro</i>	56.47	64.04	-21.00	456.00	0.19*	0.19*	-0.02	0.14*	-0.14*	-0.01	0.01	0.06*	0.11*	
11 <i>RealNamePct</i>	0.12	0.28	0.00	1.00	0.06*	0.27*	0.28*	-0.09*	0.09*	0.09*	-0.09*	0.03*	0.08*	0.02

Note: This table presents the descriptive statistics and the correlation matrix of the main variables. For each product, we observe the listed variables for 24 weeks. N = 6368.

* p < 0.05.

textual reviews on sales, this study adopts the mediation analysis framework of Baron and Kenny [19]. We follow the four steps of mediation analysis:

- Step 1: show independent variable directly affects the dependent variable. To show this, we regress the dependent variable (sales) on the independent variables (sentiments and topics). We call it Path A.
- Step 2: show that the independent variable directly affects the mediator variable. To show this, we regress the mediator (star) on independent variables. We call it Path B.
- Step 3: show that the mediator variable affects the dependent variables. To show this, we regress the dependent variable on both the independent variables and the mediator. We call it Path C.
- Step 4: establish a complete mediation of the mediator on the relationship of the independent variables on the dependent variable. The effect of independent variables should be zero, after controlling for the mediator effect. Step 3 and 4 are implemented in the same regression equation (Path C).

Note that in Path A and Path C, in order to establish the direct effect of independent variables and/or mediator on the dependent variable, we use the dynamic panel model to account for the product heterogeneity and time trend. To correct for the heterogeneity of products that may vary over time, we control for previous sales, product price, product newness, and time-specific fixed effect. In Path B, we use regression to establish the relationship between numerical rating and sentiment. Because of the bounded, J-shaped distribution of *StarRating*, we use a beta regression model, which is appropriate for dependent variable between 0 and 1. *StarRating*, the dependent variable, is defined as (star rating - 1)/4. Results are qualitatively the same without the transformations. To correct for heteroskedasticity and autocorrelations of the errors, we use robust standard errors clustered by products in all models.

4.2. Results

Table 5 presents the results of our analysis. We add the controls and main variables of interests hierarchically. Model (1) reports the results with control variables only. Model (2) presents the regression results, given the existence of the controls, while using *StarRating* as the only predictor of *LogSales*. Model (3) reports the regression results, given the existence of the controls, while using *OverallSentiment* as the only predictor of *LogSales*. Model (4a) and (4b) add the sentimental topic variables. Note that we do not include all four topics in the same regression. This is because, by construction, the proportions of two positive (negative) topics sum to 1. We omit one topic at a time to avoid multicollinearity problem in the regressions.

As expected, Model (1) shows that most of the controls, including previous sales (in log form), the number of reviews, and *WeekIntro* are found to influence sales. The significant and positive coefficients ($\beta = 0.071, p < 0.001$) of star rating in Model (2) offers support for our Hypothesis 1, suggesting that *StarRating* positively impact sales, after controlling for past period sales, number of reviews, price, introduction time, and two-way fixed effects. Results reported in Model (3) indicate that the overall sentiments significantly, positively, influence sales ($\beta = 0.174, p < 0.001$), after controlling for previous period sales, number of reviews, price, introduction time, and two-way fixed effects, supporting Hypothesis 2a.

We hypothesize that different sentimental topics will have different impacts on sales and such impacts will be mediated by star ratings. Table 6 reports the beta regression results concerning the direct effects of the independent variables, sentiment-topic aspects, on the mediator, *StarRating*. The standard errors of the coefficients are adjusted by clustering on products. Model (6) in Table 6 presents results with only controls. Model (7) adds *OverallSentiment*. Model (8a) and (8b) add the aspects from JST. In all models, the four aspects mined from the textual

Table 5
Mediation Analysis (Sentiments & Star Rating → Sales).

Variable	(1) LogSales	(2) LogSales	(3) LogSales	(4a) LogSales	(4b) LogSales	(5a) LogSales	(5b) LogSales
<i>Lag (LogSales)</i>	0.739*** (0.032)	0.708*** (0.034)	0.734*** (0.033)	0.401*** (0.051)	0.401*** (0.051)	0.378*** (0.051)	0.378*** (0.051)
<i>StarRating</i>		0.071*** (0.013)				0.094*** (0.023)	0.094*** (0.023)
<i>OverallSentiment</i>			0.174* (0.069)	0.492*** (0.149)	0.492*** (0.149)	0.228 (0.162)	0.228 (0.162)
<i>Hedonic(+)</i>				0.267** (0.099)		0.212* (0.096)	
<i>Hardware(+)</i>					-0.267** (0.099)		-0.212* (0.096)
<i>Interface(-)</i>				-0.024 (0.102)		-0.074 (0.103)	
<i>Logistics&Service(-)</i>					0.024 (0.102)		0.074 (0.103)
<i>N_Reviews</i>	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
<i>Price (\$100)</i>	0.002 (0.002)	-0.003 (0.003)	0.001 (0.003)	0.012 (0.007)	0.012 (0.007)	0.006 (0.006)	0.006 (0.006)
<i>WeekIntro</i>	0.001*** (0.000)	0.000** (0.000)	0.001*** (0.000)	0.001* (0.000)	0.001* (0.000)	0.001 (0.000)	0.001* (0.000)
<i>RealNamePct</i>	0.181** (0.065)	0.001 (0.073)	0.129 (0.070)	0.119 (0.120)	0.119 (0.120)	-0.045 (0.128)	-0.045 (0.128)
<i>Constant</i>	0.153*** (0.030)	0.005 (0.035)	0.078* (0.037)	-0.121 (0.091)	0.122 (0.114)	-0.090 (0.096)	0.049 (0.117)
<i>Observations</i>	6368	6368	6368	6368	6368	6368	6368
<i>Number of items</i>	312	312	312	312	312	312	312
<i>Product FE</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Week FE</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents part of the mediation analyses results. Dependent variable is Log(Sales). Independent variables are lagged log-sales, numerical star rating, overall sentiment, Sentiment for each sentiment-topic combination (Hedonic, Hardware, Interface, and Logistics & Service), number of reviews, price in \$100, and number of weeks of introduction for the product. Robust standard errors in parentheses.

*** $p < 0.001$

** $p < 0.01$.

* $p < 0.05$.

reviews are all highly correlated with *StarRating*.

With a given level of overall sentiment, a lower level of positive discussions on *hedonic* aspects implies a higher level of positive discussion on hardware aspects. A similar relationship holds for the two negative sentiment-topic aspects, *interface* and *logistics&service*. As indicated in Table 4 Model (4a) and (4b), given *OverallSentiment*, the two negative sentiment-topic aspects, *interface* and *logistics&service*, are found to be statistically insignificant, indicating that whichever of the two aspects is discussed more in the reviews makes no difference in product sales performance. However, the story is totally different with the two positive sentiment-topic aspects, *hedonic* and *hardware*. Given *OverallSentiment*, *hedonic* is found to be positively associated with sales ($\beta = 0.267$, $p < 0.01$), suggesting that, given overall sentiment, more positive discussion on hedonic experience and less on *hardware* result in higher sales. The finding is intuitive. From the perspective of shoppers, it is the information they do not already know or cannot easily find in the product description that is more likely to influence their shopping decisions. The hardware information of a tablet is readily available in the product description, and further, because the tablet industry is quite mature and most hardware components like hard drive and memory are highly standardized and their specifications and capacities are highly quantifiable, information from the product review pertaining to hardware brings the shoppers relatively less value and thus has little impact on their purchase decisions. On the other hand, information regarding hedonic experience, which is very crucial for an entertainment-oriented consumer product like a tablet, is not readily available in the product description; shoppers thus may rely on textual product reviews to gain such information to confirm a higher star rating and use it in purchase decisions. The disparities in the influences of sentiment-topic aspects on sales support our prediction articulated in Hypothesis 2b; that is,

different sentimental topics have different influences on sales performance.

Tables 5 and 6 jointly make up the Path A, B, and C in Baron and Kenny [19] for our mediation analysis. The results in Models (4a–4b), (8a–8b) and (5a–5b) suggest *StarRating* mediates the relationship between *OverallSentiment* and *LogSales*, supporting Hypothesis 3a. In particular, the positive coefficient of *OverallSentiment* ($\beta = 0.492$, $p < 0.001$) in Model (4a) for Path A establishes the relationship of overall sentiment with sales. The significant coefficient of *OverallSentiment* ($\beta = 2.995$, $p < 0.001$) in Model (8a) indicates overall sentiment directly correlates with the star ratings (Path B). In column (5a), when *StarRating* is present, *OverallSentiment* is no longer significant, indicating the effect of *OverallSentiment* on sales is completely mediated by *StarRating* (Path C).

With regard to the positive sentimental topics, the significant coefficient of *hedonic* ($\beta = 1.037$, $p < 0.001$) in Model 8a indicates hedonic experience sentiment directly correlates with the star ratings (Baron and Kenny Path B). In Model (5a), when *StarRating* is present, *hedonic* is still significant ($\beta = 0.212$, $p < 0.05$) but both the magnitude and significance level has decreased. These suggest that the influence of *hedonic* on *LogSales* is partially mediated by *StarRating*. Similarly, regression coefficients of *hardware* in Model (8b, 4b, and 5b) ($\beta = -1.037$, -0.267 , -0.212 respectively; and $p < 0.001$, 0.01 , 0.05 respectively), suggesting that *StarRating* partially mediates the linkage between *hardware* and *LogSales*. These provide statistical support for our Hypothesis 3b: *StarRating* mediates the relationships between sentimental topics and sales. We note that the coefficients of hedonic and hardware are symmetric in magnitude but opposite in the signs. This is expected in the Joint Sentiment-Topic model. As mentioned above, by construction, the proportions of two positive

Table 6
Mediation Analysis (Sentiments → Star Rating).

Variable	(6) StarRating	(7) StarRating	(8a) StarRating	(8b) StarRating
<i>OverallSentiment</i>		2.830*** (0.470)	2.995*** (0.476)	2.995*** (0.476)
<i>Hedonic(+)</i>			1.037*** (0.256)	
<i>Hardware(+)</i>				-1.037*** (0.256)
<i>Interface(-)</i>			0.963** (0.297)	
<i>Logistics&Service(-)</i>				-0.963** (0.297)
<i>N_Reviews</i>	0.001** (0.000)	0.001** (0.000)	0.001* (0.000)	0.001* (0.000)
<i>Price (\$100)</i>	0.045 (0.032)	0.040 (0.027)	0.038* (0.017)	0.038* (0.017)
<i>WeekIntro</i>	0.002* (0.001)	0.002* (0.001)	0.002* (0.001)	0.002* (0.001)
<i>RealNamePct</i>	3.367*** (0.418)	2.509*** (0.401)	2.273*** (0.339)	2.273*** (0.339)
<i>RAM (MB)</i>	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
<i>Processor (GHz)</i>	0.281 (0.211)	0.121 (0.185)	0.269 (0.167)	0.269 (0.167)
<i>Screen Size (inches)</i>	-0.009 (0.032)	0.024 (0.030)	0.024 (0.028)	0.024 (0.028)
<i>Constant</i>	-0.670* (0.260)	-2.042*** (0.324)	-3.146*** (0.392)	-1.146*** (0.343)
Observations	5944	5944	5944	5944
Week FE	Yes	Yes	Yes	Yes

Note: This table presents mediation analyses results when the dependent variable is Log(sales rank). Robust standard errors in parentheses.

*** p < 0.001.

** p < 0.01.

* p < 0.05.

(negative) topics sum to 1. In other words, the two pairs of hedonic/hardware and interface/logistic&service are always perfectly correlated but with opposite signs. We omit one topic at a time to avoid multicollinearity problem in the regressions.

As a robustness check, we conduct similar mediating analyses without adding the week fixed effect. The results are qualitatively similar. That is, *StarRating* completely mediates the relationship between the *OverallSentiment* and partially mediates the relationship between the positive topic sentiments (*hedonic* and *hardware*) and sales.

Given the same level of *OverallSentiment*, the two negative topical sentiments, *interface* and *logistics&sales* are not found to have any impacts, direct or mediated, on sales. One possible reason may be that the effects of negative sentimental topics are heavily factored in the *OverallSentiment*, so when *OverallSentiment* is controlled for, it is hard to detect any additional effect. Another plausible reason is that, although the two sentimental topics, *interface*, and *logistic&service*, are the most discussed negative sentiments, they may not be essential in tablet purchase decisions. For that reason, caution may be needed when generalizing this finding to other types of products.

Table 7 summarizes all hypothesis-testing results. For similar purposes, Fig. 5 visually illustrates the relationships we have hypothesized in our research model.

Fig. 5 further illustrates the mediating effect of star ratings on the relationship between sentiments and sales. The direct effect of overall sentiment on Log(Sales) is 0.492 and indirect effect of overall sentiment on log-sales is 0.228 (statistically insignificant). The direct effect of positive aspects on log-sales is 0.267 and indirect effect of 0.212, a reduced magnitude and significance (still significant at 5%). These results are intuitive. When reading online reviews, online shoppers are more likely to first look at the product's star rating to form an initial impression, and then move forward to digest the text comments and

Table 7
Summary of Hypothesis Testing Results.

Hypothesis	Var. Relationship	Relationship Significance	Hypothesis Support
H ₁	Star Rating → Sales	Y	Y
H _{2a}	Overall Sentiment → Sales	Y	Y
H _{2b}	Topical Sentiments → Sales		Y
	Hedonic → Sales	Y	
	Hardware → Sales	Y	
	Interface → Sales	N	
	Logistics & Service	N	
H _{3a}	Overall Sentiment → Star Rating → Sales	Y	Y
H _{3b}	Topical Sentiments → Star Rating → Sales		Y
	Hedonic → Star Rating → Sales	Y	
	Hardware → Star Rating → Sales	Y	

Note: This table summarizes the results of the hypotheses in our study, including the hypothesized relationship of variables, the significance of the test and whether or not it is supported.

seek confirmation/or disconfirmation for such impression. When the overall review rating is favorable, they turn to the text review for positive comments that can be used to support the ratings. The impact of overall sentiments in the text review on shopping decision (ultimately, sales), therefore, is completely bridged by star ratings.

Finally, we offer a comparison of results when sales rank (*LogRank*) is used as the proxy of actual sales. Models in Table 8 mirrors the models in Table 5, with the only change being the dependent variable. The coefficients are reversed as products with higher actual sales have lower sales ranks. We find higher star rating and more reviews have direct, significant, effects on future sale ranks. Also, the overall sentiment expressed in textual reviews spurs future sales, but the effect is mediated by star rating. Moreover, the impact of positive aspects in the review texts is not fully captured by either the sentiment or the star rating – the coefficients remains significant in Model (6) and (7). All these are consistent with our prior findings using log sales. Meanwhile, we also notice subtle, but important differences when using sales rank as the proxy. First, the coefficient on lagged sales ranks, which captures the correlation between historical sales and future sales, is exacerbated in all models in Table 8. Relatedly, the coefficients on eWOM variables, such as star rating, the number of reviews, and sentiment are all smaller. The net effect of these two differences may lead to miscalibrated models and managerial decisions that underestimate the effects of online reviews.

5. Discussion

This study examines the influence of online word-of-mouth on the sales performance of products. In particular, we investigate the impacts of numerical star ratings and sentiments expressed in textual reviews on sales. We use JST to mine the topics as well as the sentiments associated with each topic. The results reveal two prominent positive topics – “hedonic experience” and “hardware” and two dominant negative topics – “user interface” and “logistics and customer service”. The results suggest that, when expressing positive sentiments toward a tablet, online product reviewers often comment on their hedonic experiences and hardware features of the products. When complaining about a product, however, reviewers often focus on the user interface of the tablet as well as the non-physical, or “augmented” features of the product (logistics and customer service).

Further, we use JST to mine the sentiments associated with those topics and use them along with star ratings to predict the sales of

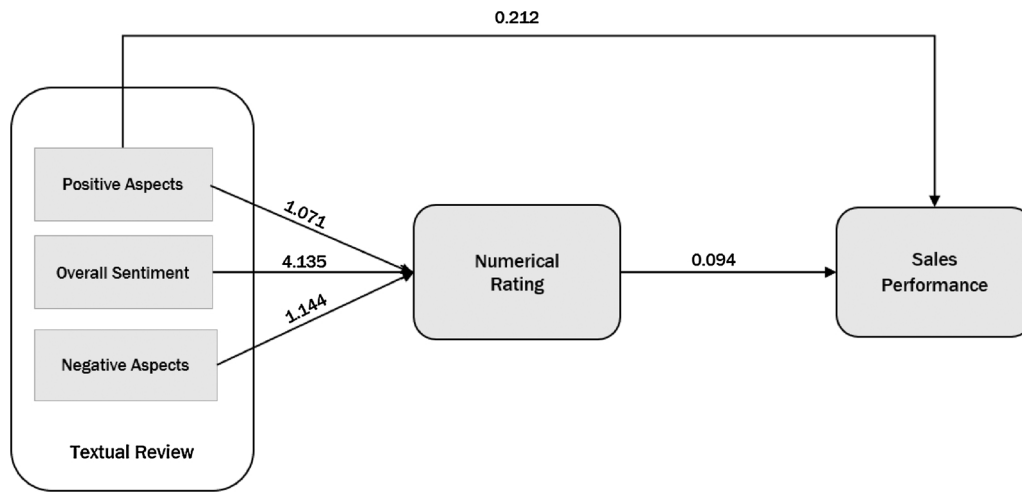


Fig. 5. Final Mediating Model.

products. Our regression analysis reveals that both numerical ratings and sentiment expressed in textual reviews have significant impacts on the sales performance of products. In addition, the analysis reveals that given the overall sentiment in the text, the two negative sentiment-topic aspects, *interface* and *logistics&service*, are statistically insignificant, indicating that whichever of the two aspects is discussed more in the reviews makes no difference in product sales performance. However, given overall sentiment, the positive sentiment-topic aspects, including *hedonic experience* and *hardware*, are found to influence sales. More discussion on *hedonic experience* and less on *hardware* result in higher

sales.

Our mediation analysis suggests that the influence of overall sentiment on sales is completely mediated by star rating, while the effects of the two positive sentiment-topic aspects (*hedonic experience* and *hardware*) on sales are partially mediated by star rating.

5.1. Implications for theory

Most extant research examines the impact of numerical ratings and, to a less extent, textual reviews on business performance. Little is done

Table 8
Mediation Analysis of Sales Rank (Sentiments & Star Rating → Sales Rank).

Variable	(1) LogRank	(2) LogRank	(3) LogRank	(4a) LogRank	(4b) LogRank	(5a) LogRank	(5b) LogRank
Lag (LogRank)	0.769*** (0.029)	0.740*** (0.032)	0.763*** (0.030)	0.445*** (0.051)	0.445*** (0.052)	0.547*** (0.048)	0.547*** (0.048)
StarRating		-0.066*** (0.015)				-0.076** (0.024)	-0.076** (0.024)
OverallSentiment			-0.182 [†] (0.083)	-0.568** (0.201)	-0.568** (0.201)	-0.251 (0.162)	-0.251 (0.162)
Hedonic(+)				-0.327*** (0.099)		-0.224** (0.075)	
Hardware(+)					0.327*** (0.099)		0.224** (0.075)
Interface(-)				0.109 (0.110)		0.115 (0.087)	
Logistics&Service(-)					-0.109 (0.110)		-0.115 (0.087)
N_reviews	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.004*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)
Price (\$100)	-0.002 (0.003)	0.003 (0.003)	-0.001 (0.003)	-0.015 (0.008)	-0.015 (0.008)	-0.007 (0.006)	-0.007 (0.006)
WeekIntro	-0.000** (0.000)	-0.000** (0.000)	-0.000*** (0.000)	-0.001 [†] (0.000)	-0.001 [†] (0.000)	-0.000 (0.000)	-0.000 (0.000)
RealNamePct	-0.121 [†] (0.055)	0.047 (0.066)	-0.066 (0.060)	-0.030 (0.117)	-0.030 (0.117)	0.102 (0.105)	0.102 (0.105)
Constant	0.000 (0.000)	1.026*** (0.135)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Observations	5937	5937	5937	5937	5937	5937	5937
Number of items	300	300	300	300	300	300	300
Product FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents part of the mediation analyses results. Dependent variable is Log(sales rank). Independent variables are lagged Log(sales rank), numerical star rating, overall sentiment, Sentiment for each sentiment-topic combination (Hedonic, Hardware, Interface, and Logistics & Service), number of reviews, price in \$100, and number of weeks of introduction for the product. Robust standard errors in parentheses.

*** p < 0.001.

** p < 0.01.

* p < 0.05.

to evaluate how numerical ratings interplay with text reviews while shaping consumer purchase decisions and product sales. The present study demonstrates that star rating mediates the influence of review sentiments expressed in text reviews on product sales performance.

This study contributes to academic research with a new data analytic approach. We first use joint sentiment/topic modeling technique to mine text reviews and extract reviewers' sentiments toward products, and then analyze how text sentiments interplay with numerical ratings in their influences on sales. This approach is novel in that it effectively combines text mining with numeric product ratings to assess the roles of different review information cues in shaping product sales performance. Prior research has found that J-shape distribution of online ratings poses problems for firms [60]. In other words, mean rating alone is insufficient at capturing product quality. We provide a way to overcome this limitation. We demonstrate JST as a method to summarize eWOM by revealing the aspects of products that consumers are happy or concerned about.

The present study also adds to prior research in the investigation of how firms should manage in the age of eWOM. The findings first provide strong empirical evidence for reviews-sales linkage. Because an increasing number of consumers post and read product reviews on a variety of online platforms, such as product discussion forums, online retailers' websites, and social networking sites, businesses must make it a strategic priority to monitor, capture, and analyze consumers' discussions. What is more, our study shows that monitoring the number and valence of reviews could be insufficient. The content of customers' discussion is equally important. More advanced analytical procedures and techniques must be developed to capture and process the vast volume of user-generated textual data. These procedures should seek to extract information pertaining to consumers' perceptions and decision-making. The information provided by such algorithms is crucial for targeted marketing campaigns, user-centered product designs, and effective customer services in the future. Overall, our findings point to the importance of business analytical initiatives for today's businesses.

In addition to what we have discussed above, the following implications are worth mentioning. First, the study offers an empirical comparison of using actual sales and its frequently used proxy, sales ranks as the dependent variable. The results reveal that sales ranks as a proxy could discount the influence of eWOM. The findings provide empirical evidence and offer an opportunity for academicians to start to think about the feasibility to use sales rank as a proxy in similar academic research. Second, this study contributes to research in business analytics by showcasing JST as an innovative approach for mining topics and sentiments from text documents. Finally, current marketing models (recommendation systems, conjoint analysis, etc.) assume that products share a set of features that consumers rate on. We find that the features are asymmetric on the positive and negative side. In addition, the positive and negative aspects have distinct impacts on sales. Thus, modifications of these models that take these disparities into account may be needed.

5.2. Implication for practice

The study suggests that overall sentiment expressed in textual reviews impacts product sales, and overall sentiment is mostly attributed to four most discussed review topics: two negative topics including *interface* and *customer service and logistics* and two positive topics including *hedonic experiences* and *hardware*. The findings on the role of negative sentiments are especially valuable because they reveal what aspects consumers tend to complain about. Negative sentiments in reviews fuel negative perceptions toward the product and pose a greater threat to product sales. It is thus vital for businesses to use effective web data monitoring and analytical techniques to detect negative online review sentiments. Businesses should proactively manage and mitigate such negative sentiments through timely responses or product recalls and redesign.

The two top negative topics discussed in reviews, revealed by the present study—*interface* and *customer service and logistics*—offer important insights, especially for manufacturers and vendors of tablets and similar entertainment oriented electronic consumer products. Given prevalent complaints on the user interface, product design and marketing campaigns for such products should focus on user-centered design and user accessibility. Also, non-core, peripheral features of products, such as logistics and customer service are found to be a key area of complaints in online reviews. Deliberate managerial attention is warranted to ensure logistics and customer service quality and to minimize the impact of negative online word-of-mouth.

Reviews that stress particular aspects of the product may have a stronger effect on sales, even though the rating and the overall sentiment of the review are the same. This study suggests that, given overall sentiment, the two top positive sentiment-topic aspects, *hedonic experience* and *hardware*, have different influences on sales: relatively more discussion on hedonic experience than on hardware leads to better sales performance. To ensure that business analytics contributes to the bottom line, online vendors and marketing firms need to strategize their business analytics initiatives by focusing on more relevant eWOM. For example, to best leverage the influence of eWOM, it is advisable for the vendor to make efforts to steer the reviewers' attention to their actual 'happy' experience rather than hardware features, which are readily available in product descriptions and specifications. Business analytics initiatives intended to boost sales should focus on whether and how reviewers are sharing their positive hedonic experience. Product design and marketing campaigns should center on the experiential content that product promises to deliver.

5.3. Limitations and avenues for future research

While highlighting the contributions of the research, we must point out a few major limitations, which leave opportunities for future research. First, the present study has only examined tablet computers, which are marked with unique features such as high-tech and hedonic orientation. Prospective shoppers of hedonic products are more likely than those of utilitarian products to be influenced by the subjective and emotional sentiments expressed in text reviews while making purchase decisions [16]. Although the overall research framework may apply to other products, caution is warranted when generalizing the specific findings to other products. Future research may be conducted to test the theories using other products. Second, numerical rating and sentiment are not sufficient for capturing the full effects of eWOM. Consumer's decision-making process is subtler. To predict long-term performance of a product through eWOM, more contextual factors need to be accounted for in future research. Such factors may include the reputation of the reviewers, the helpfulness of the reviews, and reviews from multiple platforms.

Acknowledgement

The first author would like to acknowledge the financial support of The National Natural Science Foundation of China (Project No. 71471125) for the completion of this paper.

References

- [1] X. Li, L. Hitt, Self-selection and information role of online product reviews, *Inf. Syst. Res.* 19 (2008) 456–474, <http://dx.doi.org/10.1287/isre.1070.0154>.
- [2] M.L. Jensen, J.M. Averbach, Z. Zhang, K.B. Wright, Credibility of anonymous online product reviews: a language expectancy perspective, *J. Manag. Inf. Syst.* 30 (2013) 293–324, <http://dx.doi.org/10.2753/MIS0742-1222300109>.
- [3] S. Xiao, C.P. Wei, M. Dong, Crowd intelligence: analyzing online product reviews for preference measurement, *Inf. Manag. Manag.* 53 (2016) 169–182, <http://dx.doi.org/10.1016/j.im.2015.09.010>.
- [4] Kexin Zhao, Antonis C. Stylianou, Yiming Zheng, Sources and impacts of social influence from online anonymous user reviews, *Inf. Manag.* 55.1 (2018) 16–30.
- [5] N. Hu, L. Liu, J.J. Zhang, Do online reviews affect product sales? The role of

- reviewer characteristics and temporal effects, *Inf. Technol. Manag.* 9 (2008) 201–214.
- [6] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter?—An empirical investigation of panel data, *Decis. Support Syst.* 45 (2008) 1007–1016, <http://dx.doi.org/10.1016/j.dss.2008.04.001>.
- [7] P.K. Chintagunta, S. Gopinath, S. Venkataraman, The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets, *Mark. Sci.* 29 (2010) 944–957, <http://dx.doi.org/10.1287/mksc.1100.0572>.
- [8] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, *J. Mark. Res.* 43 (2006) 345–354, <http://dx.doi.org/10.1509/jmkr.43.3.345>.
- [9] F. Zhu, X. (Michael) Zhang, Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics, *J. Mark.* 74 (2010) 133–148, <http://dx.doi.org/10.1509/jmkg.74.2.133>.
- [10] N.N. Ho-Dac, S.J. Carson, W.L. Moore, The effects of positive and negative online customer reviews: do brand strength and category maturity matter? *J. Mark.* 77 (2013) 37–53, <http://dx.doi.org/10.1509/jm.11.0011>.
- [11] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, *Inf. Syst. Res.* 19 (2008) 291–313, <http://dx.doi.org/10.1287/isre.1080.0193>.
- [12] N. Hu, N.S. Koh, S.K. Reddy, Ratings lead you to the product, reviews help you clinch it: the dynamics and impact of online review sentiments on products sales the mediating role of online review sentiments on product sales, *Decis. Support Syst.* 57 (2013) 42–53.
- [13] N. Archak, A. Ghose, P.G. Ipeirotis, Deriving the pricing power of product features by mining consumer reviews, *Manage. Sci.* 57 (2011) 1485–1509, <http://dx.doi.org/10.1287/mnsc.1110.1370>.
- [14] S. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com, *MIS Q.* 34 (2010) 185–200. Article.
- [15] D.-H. Park, S. Kim, The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews, *Electron. Commer. Res. Appl.* 7 (2008) 399–410.
- [16] S. Moon, Y. Park, Y. Seog Kim, The impact of text product reviews on sales, *Eur. J. Mark.* 48 (2014) 2176–2197, <http://dx.doi.org/10.1108/EJM-06-2013-0291>.
- [17] T. Lee, E. Bradlow, Automated marketing research using online customer reviews, *J. Mark. Res.* (2011) <http://journals.ama.org/doi/abs/10.1509/jmkr.48.5.881>.
- [18] C. Lin, Y. He, Joint sentiment/topic model for sentiment analysis, *Proceeding 18th ACM Conf. Inf. Knowl. Manag. – CIKM '09*, ACM (2009), <http://dx.doi.org/10.1145/1645953.1646003> p. 375.
- [19] R.M. Baron, D.A. Kenny, The moderator-mediator variable distinction in social the moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations, *J. Pers. Soc. Psychol.* 51 (1986) 1173–1182, <http://dx.doi.org/10.1037/0022-3514.51.6.1173>.
- [20] N. Hu, P.A. Pavlou, J. Zhang, Can online word-of-mouth communication reveal true product quality? *Proc. 19th Int. Conf. Inf. Qual.* (2014) 175–203, <http://dx.doi.org/10.1145/1134707.1134743>.
- [21] L.L. Hellefs, R. Jacobson, Market share and customers' perceptions of quality: when can firms grow their way to higher versus lower quality? *J. Mark.* 63 (1999) 16–25, <http://dx.doi.org/10.2307/1251998>.
- [22] D.S. Kostyra, J. Reiner, M. Natter, D. Klapper, Decomposing the effects of online customer reviews on brand, price, and product attributes, *Int. J. Res. Mark.* 33 (2016) 11–26, <http://dx.doi.org/10.1016/j.ijresmar.2014.12.004>.
- [23] K. Floyd, R. Freling, S. Alhoqail, H.Y. Cho, T. Freling, How online product reviews affect retail sales, *J. Retail.* 90 (2014) 217–232 https://ac-els-cdn-com.proxy1.cl.msu.edu/S0022435914000293/1-s2.0-S0022435914000293-main.pdf?tid=4e574d8a-ab0c-11e7-b192-00000aac36&acdnat=1507345680_f0b31dcb9267743117061b7ba7991e0.
- [24] C.M.K. Cheung, D.R. Thadani, The impact of electronic word-of-mouth communication: a literature analysis and integrative model, *Decis. Support Syst.* 54 (2012) 461–470, <http://dx.doi.org/10.1016/j.dss.2012.06.008>.
- [25] O. Müller, I. Junglas, S. Debortoli, J. vom Brocke, Using text analytics to derive customer service management benefits from unstructured data, *MIS Q. Exec.* 15 (2016) 64–73.
- [26] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, *J. Mach. Learn. Res.* 3 (2003) 993–1022.
- [27] E. Brynjolfsson, Y. (Jeffrey) Hu, M.D. Smith, Consumer surplus in the digital economy: estimating the value of increased product variety at online booksellers, *Manage. Sci.* 49 (2003) 1580–1596, <http://dx.doi.org/10.1287/mnsc.49.11.1580.20580>.
- [28] A. Ghose, A. Sundararajan, Evaluating pricing strategy using e-commerce data: evidence and estimation challenges, *Stat. Sci.* 13 (2006) 131–142.
- [29] P. Miller, *Theories of Developmental Psychology*, 3rd ed., Macmillan, 1994, <http://dx.doi.org/10.1037/033859>.
- [30] D. Yin, S. Mitra, H. Zhang, When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth, *Inf. Syst. Res.* 27 (2016) 131–144, <http://dx.doi.org/10.1287/isre.2015.0617>.
- [31] X. Li, L.M. Hitt, Price effects in online product reviews: an analytical model and empirical analysis, *MIS Q.* 34 (2010), <http://dx.doi.org/10.2139/ssrn.1282303> 809–A5.
- [32] W. Jabr, Z. Zheng, Know yourself and know your enemy: an analysis of firm recommendations and consumer reviews in a competitive environment, *MIS Q.* 38 (2014) 635–654.
- [33] W.W. Moe, M. Trusov, The value of social dynamics in online product ratings forums, *J. Mark. Res.* 48 (2011) 444–456, <http://dx.doi.org/10.1509/jmkr.48.3.444>.
- [34] A. Ghose, P.G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics, *IEEE Trans. Knowl. Data Eng.* 23 (2011) 1498–1512.
- [35] A. Chong, B. Li, E. Ngai, E. Ch'ng, F. Lee, Predicting online product sales via online reviews, sentiments, and promotion strategies: a big data architecture and neural network approach, *Int. J. Oper. Prod. Manag.* 36 (2016), <http://dx.doi.org/10.1108/01443571111165839>.
- [36] D. Yin, S.D. Bond, H. Zhang, Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews, *MIS Q.* 38 (2014) 539–560 <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=95756002&lang=zh-cn&site=ehost-live>.
- [37] S.D. Pugh, Service with a smile: emotional contagion in the service encounter, *Acad. Manag. J.* 44 (2014) 1018–1027.
- [38] P.A. Pavlou, A. Dimoka, The nature and role of feedback text comments in online marketplaces: implica ..., *Inf. Syst. Res.* 17 (2006) 392–414, <http://dx.doi.org/10.1287/isre.1060.0106>.
- [39] X. Yu, Y. Liu, X. Huang, A. An, Mining online reviews for predicting sales performance: a case study in the movie domain, *IEEE Trans. Knowl. Data Eng.* 24 (2012) 720–734, <http://dx.doi.org/10.1109/TKDE.2010.269>.
- [40] S. Ludwig, K. de Ruyter, M. Friedman, E.C. Brügger, M. Wetzels, G. Pfann, More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates, *J. Mark.* 77 (2013) 87–103, <http://dx.doi.org/10.1509/jm.11.0560>.
- [41] A. Floh, M. Koller, A. Zauner, Journal of marketing management taking a deeper look at online reviews: the asymmetric effect of valence intensity on shopping behaviour taking a deeper look at online reviews: the asymmetric effect of valence intensity, *J. Mark. Manag.* 29 (2013) 37–41.
- [42] P.E. Ketelaar, L.M. Willemsen, L. Sleven, P. Kerkhof, The good, the bad, and the expert: how consumer expertise affects review valence effects on purchase intentions in online product reviews, *J. Comput. Commun.* 20 (2015) 649–666, <http://dx.doi.org/10.1111/jcc4.12139>.
- [43] R.Y.K. Lau, S.S.Y. Liao, K.-F. Wong, D.K.W. Chiu, Web 2.0 environmental scanning and adaptive decision support for business mergers and acquisitions, *MIS Q.* 36 (2012).
- [44] A.M. Tybout, B.J. Calder, B. Sternthal, Using information processing theory to design marketing strategies, *J. Mark. Res.* 18 (1981) 73–79, <http://dx.doi.org/10.2307/3151315>.
- [45] M. Fishbein, An Investigation of the Relationships between Beliefs about an Object and the Attitude toward that Object, *Hum. Relations* 16 (1963) 233–239, <http://dx.doi.org/10.1177/001872676301600302>.
- [46] A. Shocker, V. Srinivasan, Multiattribute approaches for product concept evaluation and generation: a critical review, *J. Mark. Res.* 16 (1979) 159–180 <http://www.jstor.org/stable/10.2307/3150681> (Accessed 11 July 2012).
- [47] T.P. Liang, X. Li, C.T. Yang, M. Wang, What in consumer reviews affects the sales of mobile apps: a multifacet sentiment analysis approach, *Int. J. Electron. Commer.* 20 (2015) 236–260, <http://dx.doi.org/10.1080/10864415.2016.1087823>.
- [48] F.V. Ordenes, S. Ludwig, K. De Ruyter, D. Grewal, M. Wetzels, Unveiling what is written in the stars: analyzing explicit, implicit, and discourse patterns of sentiment in social media, *J. Consum. Res.* 43 (2017) 875–894, <http://dx.doi.org/10.1093/jcr/ucw070>.
- [49] Q. Gan, B.H. Ferns, Y. Yu, L. Jin, A text mining and multidimensional sentiment analysis of online restaurant reviews, *J. Qual. Assur. Hosp. Tour.* 18 (2017) 465–492, <http://dx.doi.org/10.1080/1528008X.2016.1250243>.
- [50] C. Tang, M.R. Mehl, M.A. Eastlick, W. He, N.A. Card, A longitudinal exploration of the relations between electronic word-of-mouth indicators and firms' profitability: findings from the banking industry, *Int. J. Inf. Manage.* 36 (2014) 1124–1132, <http://dx.doi.org/10.1016/j.ijinfomgt.2016.03.015>.
- [51] X. Wang, F. Mai, R.H.L. Chiang, Database submission—market dynamics and user-generated content about tablet computers, *Mark. Sci.* 33 (2014) 449–458, <http://dx.doi.org/10.1287/mksc.2013.0821>.
- [52] E. Brynjolfsson, Y. (Jeffrey) Hu, D. Simester, Goodbye pareto principle, hello long tail: the effect of search costs on the concentration of product sales, *Manage. Sci.* 57 (2011) 1373–1386, <http://dx.doi.org/10.1287/mnsc.1110.1371>.
- [53] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, *Proc. Conf. Hum. Lang. Technol. Empir. Methods Nat. Lang. Process. – HLT '05*, Association for Computational Linguistics (2005) 347–354, <http://dx.doi.org/10.3115/1220575.1220619>.
- [54] S. Nickell, Biases in dynamic panel models with fixed effects, *Econometrica* 49 (1981) 1417–1426, <http://dx.doi.org/10.2307/1911408>.
- [55] T.W. Anderson, C. Hsiao, Estimation of dynamic models with error components, *J. Am. Stat. Assoc.* 76 (1981) 598–606.
- [56] M. Arellano, S. Bond, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Rev. Econ. Stud.* 58 (1991) 277, <http://dx.doi.org/10.2307/2297968>.
- [57] M. Arellano, O. Bover, Another look at the instrumental variable estimation of error components models, *J. Econom.* 68 (1995) 29–51.
- [58] D. Roodman, How to do xtabond2: an introduction to "difference" and "system" GMM in stata, *Cent. Glob. Dev. Stata J.* (2006) 1–48.
- [59] F. Windmeijer, A finite sample correction for the variance of linear two-step estimators, *J. Econom.* 126 (2000) 25–51.
- [60] N. Hu, J. Zhang, P.A. Pavlou, Overcoming the J-shaped distribution of product reviews, *Commun. ACM* 52 (2009) 144, <http://dx.doi.org/10.1145/1562764.1562800>.

Xiaolin Li is an Associate Professor of e-Business and Technology Management with the College of Business and Economics of Towson University. He received his Ph.D. in Management Systems from Kent State University. Dr. Li's current research emphases are

electronic commerce and big data. His research has been published in *Information Systems Research*, *Journal of the Association for Information Systems*, *Decision Sciences*, *International Journal of Production Economics*, *Electronic Commerce Research*, *Journal of Organizational Computing and Electronic Commerce*, *Journal of Computer Information Systems*, among others. He received the “Distinguished Research Paper Award” (in innovative education track) from the DSI Annual Conference in 2009 and Towson University College of Business and Economics Outstanding Scholarship Award in 2012.

Chaojiang Wu is Assistant Professor in LeBow College of Business at Drexel University. He obtained his PhD from the University of Cincinnati. His research focuses on data analytics and its business applications such as Finance and Information Systems. His

research develops quantitative methods in analyzing complex data for better decisions and applies such methods to applications in finance, online word of mouth, and scholarly communications. He has published in journals such as *Journal of Financial and Quantitative Analysis*, *IIE Transactions*, *Statistics and Computing*, and *Computational Statistics and Data Analysis*, among others.

Feng Mai is an Assistant Professor of Information Systems at the School of Business, Stevens Institute of Technology. He holds a Ph.D. in Business Administration from the University of Cincinnati. Professor Mai’s research interests are social media, marketing analytics, and FinTech. His work has been published in journals such as *Marketing Science*, *Journal of Consumer Research*, and *European Journal of Operational Research*.