



The influence of online word-of-mouth on long tail formation

Bin Gu^b, Qian Tang^{a,*}, Andrew B. Whinston^a

^a Department of Information, Risk, and Operations Management, McCombs School of Business, the University of Texas at Austin, 1 University Station, B6500, Austin, TX 78712, United States

^b Department of Information Systems, W P Carey School of Business, Arizona State University, P O Box 874606, Tempe, AZ 85287, United States

ARTICLE INFO

Available online 8 November 2012

Keywords:

Long tail
Word-of-mouth (WOM)
Customer review
Information cascade

ABSTRACT

The long tail phenomenon has been attributed to both supply side and demand side economies. While the cause on the supply side is well-known, research on the demand side has largely focused on the awareness effect of online information that helps consumers discover new and often niche products. This study expands the demand side factors by showing that online information also influences the long tail phenomenon through the informative effect, which affects consumers' evaluation of product quality. We examine the informative effect in the context of online WOM. Two sets of theories suggest opposite directions for the implication of the informative effect. Information search and information cascade literatures indicate that WOM provides additional information to consumers, reduces the occurrence of information cascade, and encourages the formation of long tail. Studies on behavior heuristics, however, suggest that consumers tend to ignore online information inconsistent with their prior beliefs, which leads to a rich-gets-richer effect for popular products and curtail the formation of the long tail. We empirically examine the conflict by analyzing different impacts of online WOM across product popularity and WOM ratings. Using a panel data collected from Amazon.com, we show that positive reviews improve the sales of popular products more than the sales of niche products, while negative reviews hurt niche products more than popular products. The results are consistent with the prediction of the behavior heuristic and suggest that online WOM restrains the formation of long tail.

Published by Elsevier B.V.

1. Introduction

Online retailers have long noticed that the sales of non-hit or niche products account for a significant portion of overall product sales. This phenomenon is known as the long tail [1] and has attracted attention of both industry practitioners and academic researchers. Several factors are conducive to the emergence of the long tail phenomena in online markets. The supply side drivers include lower storage costs, faster distribution and the aggregation of geographically dispersed niche markets [5]. These enable online retailers to carry a far wider range of products than their physical counterparts. On the demand side, studies note that the use of information sharing mechanisms such as recommender systems, co-purchase networks, search tools, and blog sites, enables customers to discover new and niche products more easily and encourages the long tail formation [1,4,5,15,34,42].

Although prior demand side studies focus on how online information increases awareness of less known products, the influence of online information often goes beyond helping consumers discover new products. Consumers are known to use a two-stage process in making purchase decisions [28]. In the first stage, consumers identify a set of

potential products for further consideration. In this stage, consumers use IT-enabled tools both actively and passively to identify products of potential interest. Search tools and recommendation systems help narrow down possible choices according to consumer preferences. In the second stage, consumers obtain detailed information on each product in the consideration set to form a purchase decision. Product awareness created by searching tools and recommendation systems is important in the first stage. However, it is the detailed product quality information that ultimately shapes consumers' purchase decisions in the second stage. Such information increasingly comes from online sources such as online WOM. However, little is known on how these information sources influence the long tail phenomenon. The goal of this study is to take the first step to understand this influence.

Two sets of theories suggest opposite directions for the impact of online WOM on the long tail phenomenon. A key challenge for electronic commerce is that customers are often uncertain about product quality before purchase. The lack of information leads to herding where consumers follow the choices of others, as suggested by information search and informational cascade literatures [3]. In an informational cascade, it is optimal for a user to follow predecessors' behavior and ignore his private information. As such, a few "hit" products take up the most sales. Such behavior could lead to "bad" herding where popular products are not necessarily of high quality. One of the key conditions for informational cascade is that users do not have access to the private information of others. However, online review allows consumers to reveal private information, helps future consumers to learn about the true

* Corresponding author at: Department of Information Systems, W. P. Carey School of Business, Arizona State University, PO Box 874606, Tempe, AZ 87287-4606, United States. Tel.: +1 480 965 1870.

E-mail addresses: bin.gu@asu.edu (B. Gu), qian.tang@phd.mcombs.utexas.edu (Q. Tang), abw@uts.cc.utexas.edu (A.B. Whinston).

quality before purchase, and reduces the occurrence of “bad” herding [36]. In particular, the arrival of negative WOM could stop the positive informational cascade of a popular but low quality product, while the arrival of positive WOM could stop the negative cascade of an unpopular but high quality product [3]. From this respect, WOM information encourages the formation of the long tail.

However, the opposite is implied by behavior heuristics studies. Behavior heuristics theories believe that individuals are biased in making adjustments to accommodate new information. Information consistent with prior belief often leads to overreaction, while contradictory information leads to under-reaction [26,36]. Since product popularity commonly serves as a starting point for prior belief, consumers tend to overreact to positive WOM on popular products and negative WOM on unpopular products, and underreact to negative WOM on popular products and positive WOM on unpopular products, leading to a rich-get-richer effect. In this respect, WOM constrains the formation of the long tail.

To examine the conflicting hypotheses, we note that the hypotheses suggest that the influence of a customer review depends on the interaction between review valence and product popularity. As such, testing the hypotheses requires an analysis at the customer review level instead of the product level. This is in contrast with most prior WOM research that often uses aggregate WOM measures such as average review rating to study the effect of WOM at the product level. Compared to the product level analysis, analysis at the review level allows us to capture consumers' reaction to each specific review and how their reaction varies for different combinations of review valence and product popularity. To facilitate the analysis, we use a first difference approach. This approach allows us to analyze the influence of each customer review on changes in product sales and provides significant flexibility in modeling factors that moderate the influence of individual customer reviews. We apply the model to a panel data collected from Amazon.com on 3000 books over 200 days. The results indicate that positive WOM benefits popular products more than niche products, while negative WOM hurts niche products more. Our finding supports the prediction of behavior heuristics theory and suggests that consumers are not fully rational in reacting to incoming customer reviews. Instead, they are biased toward their initial expectation. As a result, online WOM constrains the formation of the long tail.

The rest of the paper is organized as follows. Section 2 is literature review. Section 3 introduces the theoretical background and develops key hypotheses. Section 4 develops a first difference model that allows WOM's influence to differ across product popularity and WOM valence. Section 5 describes the data collection process. Section 6 presents the empirical results. Section 7 discusses the implication of our findings and concludes the study.

2. Literature review

2.1. Online WOM

Online WOM plays an important role in electronic commerce. A survey of Bizrate.com found that 44% of users consulted opinion sites prior to making a purchase [7]. This survey also found that 59% of respondents considered consumer-generated reviews to be more valuable than expert reviews. A number of studies have been conducted to examine the impact of WOM on consumer purchase decisions [2,9,13,16,17,20,23,30,40,42]. These studies reveal that online WOM has both awareness effect and informative effect. The awareness effect is reflected in the influence of WOM volume on product sales. WOM volume captures the underlying dispersion of WOM within and across online communities and studies show that WOM volume predicts future product sales [13,17,23,27].

The informative effect for product quality, on the other hand, is mainly reflected in review ratings and their impact on product sales. Zhu and Zhang [41] find that online customer reviews are important resource for consumers to seek product quality information. Moon et al. [28] use new movie ratings as a measure of movie performance

and argue that positive ratings can enhance the effectiveness of advertising spending to raise movie revenues. Chevalier and Mayzlin [9] compare book sales across two online retailers and find that improvement in a book's average review rating at one site leads to an increase in relative sales at that site. Similarly, Zhang et al. [40] study the relationship between movie ratings and box office sales and find that online consumer ratings have significant impact on movie revenue. Moe and Trusov [31] decompose the rating effect on sales into a baseline component and a social dynamic component and show both the direct impact of baseline component and the indirect impact through social dynamics influence product sales. Clemons et al. [12] suggest that the informative effect of WOM plays a key role in redefining a firm's product strategy. We extend this stream of literature by noting that the influence of WOM could vary with the interaction between WOM and product popularity due to economic or behavioral motivations. We also show that such variations could have significant implications for the long tail phenomenon.

Our analysis complements the literature on the informative effect of WOM. For example, Sun [36] models high average rating as an indicator for high product quality and high rating variance as an indicator for niche products, which some consumers love and others hate. She finds that a higher standard deviation of ratings on Amazon improves a book's relative sales rank when the average rating is lower than 4.1 stars. In contrast, Chintagunta [10] find that review variance does not matter after controlling for movie specific and market specific effects.

Our analysis also complements recent studies that identify factors moderating the influence of WOM. Zhu and Zhang [41] claim that online reviews are more influential on less popular video games and games whose players have greater Internet experience. Chevalier and Mayzlin [9] reveal that negative WOM is more influential than is positive WOM. Chen et al. [8] and Duan et al. [18] find that WOM is more influential on less popular products. Hu et al. [29] and Chen et al. [8] show the importance of reviewer reputation in determining the influence of the review. Moon et al. [28] suggest that positive ratings can enhance the effectiveness of advertising spending to raise movie revenues. We complement these literatures by identifying new moderating factors based on information search and informational cascade theory and behavior heuristic theory. Our hypotheses suggest that consumers infer product quality information from both WOM and product popularity and that whether the two sets of information are consistent with or contradictory to each other could have significant influence on consumer purchase decisions.

2.2. The long tail

This study also contributes to the emerging literature on the long tail phenomenon. Brynjolfsson et al. [6] find that the sales distribution of a retailer's online channel is less concentrated than that of its traditional channels, indicating that a reduction in search costs contributes to the long tail phenomenon. Rahman and Hahn [35] find that the long tail phenomenon is more prevalent in search goods than in experience goods, and Wimble et al. [39] find the product dispersion of purchases made by households with broadband access is more flattened than that by households without broadband access. Both studies suggest the importance of search costs in the long tail formation. A number of long tail studies take a step further to identify the underlying information sources that facilitate the long tail phenomenon. Oestreicher-Singer and Sundararajan [33] show that the long tail phenomenon is influenced by the presence of co-purchase networks which allows online consumers to observe products commonly bought together. Dewan and Ramaprasad [15] find that blogging activities lead to more purchases of niche products and contribute to the long tail phenomenon. Goh and Bockstedt [24] show that product sampling plays an important role in the long tail phenomenon.

Prior studies of the long tail phenomenon also reveal the presence of both long tail and superstar phenomenon. Most studies consider

the impacts of WOM on popular and unpopular products respectively. Elberse and Oberholzer-Gee [19] analyze the distribution of US video sales from 2000 to 2005 and find both the long tail phenomenon where “the number of titles that sell only a few copies increased almost twofold” and superstar phenomenon where “an even smaller number of titles account for the bulk of sales”. Fleder and Hosanagar [21] study the influence of product recommendations provided by retailer using collaborate filtering algorithm and conclude that such recommendation may not promote the sales of less popular products as recommendations are based on historical product sales. Consequently, popular products are more likely to be recommended and purchased. Tucker and Zhang [37] and Duan et al. [18] indicate that the presence of product popularity information significantly increases the sales of popular products. Dellarocas and Narayan [14] suggest that popular products are more likely to be discussed by consumers, which contributes to the superstar phenomenon. Hervas-Drane [28] suggests that consumers with preference for mass market products are more likely to seek WOM because they are more likely to find other consumers with similar preference.

Complementary to these studies, we recognize that online information not only increases product awareness but also enables consumers to better evaluate product quality. The influence of this informative effect, however, could differ for popular products vs. niche products and affect the formation of the long tail.

3. Theoretical background and hypothesis

Prior e-commerce studies have noted that informational cascades often drive consumer purchase decisions. An informational cascade occurs when an individual, having observed the actions of those ahead of him, follows the behavior of predecessors without considering his own private information [3]. Informational cascade in e-commerce arises because each online consumer faces two sources of product quality information. One is product popularity information (e.g. sales rank) provided by online retailers based on historical consumer purchase decisions. Product popularity information is relatively objective. However, the quality that consumers infer from this information may not represent the true product quality since product sales are often influenced by other factors such as price or information cascade. The other information source is WOM, or customer reviews in most cases [18]. Customer reviews reveal consumers' perception of product quality directly and thus provide access to the private information of others. However, they are relatively subjective and can be affected by reviewers' characteristics [32]. Consumers combine these two information sources to arrive at purchase decisions. If the two information sources are contradictory to each other, consumer decision is determined by the relative strengths of the two sources. When customer reviews are scarce, consumers are more likely to follow the step of their predecessors, leading to an informational cascade. The increasing availability of customer reviews improves product information quality, leading to a reduction in informational cascades and facilitating the formation of the long tail.

Informational cascade theory also suggests that an information cascade is fragile if predecessors' private information is accessible to successors. In particular, the arrival of new information contradictory to the direction of information cascades can break the informational cascade process [3]. Informational cascade could be either positive or negative, resulting in either extremely popular or unpopular products. Without private information, consumers infer from the observed sales that popular products are likely to be high quality and unpopular products are likely to be low quality. However, as more customer reviews come in, revealing purchasers' private information, rational consumers can infer whether the popularity reflects the true quality or just “bad” cascades, which cause low quality products to sell well or high quality products to remain unknown. From this point, positive and negative reviews have different impacts on popular and unpopular products. Specifically, positive reviews on a popular product

confirm the high quality of the product and thus sustain the informational cascade, while negative reviews on a popular product suggest that the popular product is low quality and can break the informational cascade. On the other hand, positive reviews on an unpopular product indicate that the product is high quality and break the negative informational cascade, while negative reviews on an unpopular product sustain the status quo. As such, positive reviews on hit products and negative reviews on niche products sustain the “good” informational cascade and keep the status quo, while negative reviews on hit products and positive reviews on niche products could stop the “bad” informational cascades, reducing the sales increase on hit products and accelerating the sales increase on niche products. The above analysis suggests that positive reviews would benefit niche products more, while the negative review would hurt popular products more, leading to a more flattened sales distribution, i.e., the long tail. Accordingly, we propose the following hypothesis:

H1a. The impact of a positive (negative) customer review on niche (popular) products will be greater than its impact on popular (niche) products.

While informational cascade theory suggests the positive impact of WOM on the formation of the long tail, literatures on behavior economics suggests that the impact could be in the opposite direction. According to behavior economics, individuals use various heuristics in information processing and decision making [38]. While informational cascade theory assumes rationality, behavior economics argues that individuals are not perfectly rational. Instead, certain bias exists during the information update process. A commonly used behavior heuristic is anchoring and adjustment, i.e. people make adjustment incrementally from their prior expectation and often make insufficient adjustment when new evidence is inconsistent with their prior expectation [26]. As aforementioned, online consumers have two sources of information: popularity information provided by the websites and product information from customer reviews. Consumers typically form their initial expectations based on popularity information, and then adjust their expectations according to customer reviews. Because of the heuristic bias, consumers tend to ignore information that is inconsistent with their prior beliefs, while they pay more attention to information that is consistent with their beliefs. This bias results in over-reaction to evidence consistent with their expectations and under-reaction to information contradictory to their expectations [34]. Therefore, this stream of theories suggests different impacts of positive and negative reviews on hit and niche products but in a different direction. In particular, a positive review on a popular product confirms consumers' prior expectation and encourages sales, while a negative review on a popular product tends to be ignored or at least downplayed because it contradicts consumers' initial expectation. For an unpopular product, a negative review confirms consumers' expectation and discourages sales, while a positive review tends to be ignored or downplayed. As a result, we expect that a positive incremental review have more impact on popular products than on niche products while a negative incremental review hurts niche products more. Therefore, based on behavior heuristics, we derive the opposite hypothesis:

H1b. The impact of a positive (negative) review on popular (niche) products will be greater than its impact on niche (popular) products.

H1a and H1b are two hypotheses derived from two streams of theories (Table 1). By testing the two competing hypotheses, we can test whether online WOM facilitates or inhibits the formation of the long tail and whether consumers are rational or biased in making purchase decisions. If H1a is supported, we can conclude that online WOM contributes to the formation of the long tail and consumers are able to process information rationally. If H1b is supported, we

Table 1
Research hypotheses.

Hypothesis	Positive reviews	Negative reviews
H1a: online customer review facilitates long tail formation.	Impact <i>niche</i> products more	Impact <i>popular</i> products more
H1b: online customer review inhibits long tail formation.	Impact <i>popular</i> products more	Impact <i>niche</i> products more

would arrive at the opposite conclusion that online WOM shifts sales distribution against the long tail and consumers are biased toward the starting point in updating their expectations using incoming WOM.

4. Empirical model

A unique challenge in assessing the influence of online WOM is that its influence is cumulative and long lasting. Product sales today are not only affected by today's customer reviews but also by all previously posted customer reviews about the product. As a result, prior studies often use a variety of aggregation functions to summarize customer reviews to date on a given product as a way to study their influence on product sales. An analysis at the aggregate level, however, is not desirable here because the hypotheses indicate that the influence of each piece of online customer review could vary depending on the rating and product popularity. To address this empirical challenge, we take a new approach by adopting a first-difference model that focuses on the influence of the arrival of each customer review on changes in product sales. This approach allows greater flexibility in modeling the influence of WOM on product sales. In particular, it allows the influence of individual consumer reviews to vary with product popularity, product age, review rating, review volume, and the interactions between WOM and product popularity.

4.1. First difference model with heterogeneous influence

While this paper focuses on the informative aspect of WOM, it is necessary to incorporate both awareness and informative effects in the empirical model. In prior literature, the informative and awareness effects are captured by the average review rating and review volume respectively [9] using a log-linear demand function as follows:

$$LgSales_{it} = \beta AvgRating_{it} + \gamma LgNumOfReviews_{it} + \delta LgPrice_{it} + \eta X_{it} + \mu_i + v_t + \epsilon_{it}. \quad (1)$$

In the above equation, β measures the influence of review ratings on product sales, capturing the informative effect, while γ measures the influence of review volume on product sales, capturing the awareness effect. δ assesses the price elasticity. X_{it} is a vector containing all the control variables and η captures the influence of the control variables. The model also controls for product and time fixed effects.

As we mentioned earlier, Eq. (1) imposes constraints on the relationship between WOM and product sales. In particular, the model implies that each customer review has the same magnitude of influence on a product's sales and the influence of customer reviews is the same across products with different popularity and age. These assumptions allow parsimony in assessing the overall influence of WOM but are overly restrictive to assess the influence of individual pieces of WOM. To address this limitation, we propose a new approach to allow more flexibility in estimating the influence of WOM. Our objective is to shift the consideration of the aggregate effect of all WOM on a product to focusing on how newly arrived WOM on a given day affects changes in product sales using a first difference approach. This approach enables us to cancel out the influence of previous WOM and focus on the

influence of newly arrived WOM. To transform the aggregate level model to the disaggregate level, we start with taking the first difference between t and $t + 1$ of Eq. (1).

$$\Delta LgSales_{it} = \beta \Delta AvgRating_{it} + \gamma \Delta LgNumOfReviews_{it} + \delta \Delta LgPrice_{it} + \eta \Delta X_{it} + \Delta v_t + \epsilon_{it}. \quad (2)$$

Eq. (2) indicates that, on days without new WOM, changes in average review rating and changes in number of reviews are both zero. In this case, sales changes are entirely driven by changes in product prices, time effect and other changes in the control variables. On the other hand, on days with newly arriving WOM, WOM has two effects on sales changes. First, the arrival of WOM may influence average review rating, thus affect future product sales. Second, the very presence of new WOM indicates underlying WOM dispersion that may increase product awareness and affect future product sales. In Eq. (2), β captures the incremental informative effect of WOM, while γ captures the incremental awareness effect of WOM. The first-difference approach shifts the focus from the influence of overall WOM in Eq. (1) to the influence of individual pieces of WOM that arrives on a given day in Eq. (2). This disaggregated approach provides a foundation that allows us to further extend the model to consider how the influence of these individual pieces of WOM could be moderated by factors proposed in the hypotheses.

4.2. Moderating effect on WOM influence

We first extend the model to allow the influence of positive reviews and negative reviews to differ. We divide reviews into two categories – negative reviews and positive reviews – based on their influence on product average rating. If the arrival of a customer review reduces a product's average review rating, we define it as a negative review. If its rating is greater than the previous average, we define it as a positive review. A positive review will bring an improvement in a product's review rating and a negative one will impose a negative impact on review rating. To allow the influences to differ, we split the review rating variable into two variables: one for positive changes in customer reviews rating and the other for negative changes in customer reviews rating.

$$[\Delta AvgRating_{it}]^+ = \begin{cases} AvgRating_{it}, & \text{if } AvgRating_{it} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$[\Delta AvgRating_{it}]^- = \begin{cases} AvgRating_{it}, & \text{if } AvgRating_{it} < 0 \\ 0, & \text{otherwise} \end{cases}$$

We use coefficients β^+ and β^- to identify their influences:

$$\Delta LgSales_{it} = \beta^+ [\Delta AvgRating_{it}]^+ + \beta^- [\Delta AvgRating_{it}]^- + \gamma \Delta LgNumOfReviews_{it} + \delta \Delta LgPrice_{it} + \eta \Delta X_{it} + \Delta v_t + \epsilon_{it}. \quad (3)$$

Hypotheses 1a and 1b suggest that the influence of positive reviews and negative reviews is moderated by the interactions between review rating and product popularity. **H1a** suggests that reviews with rating contradictory to product popularity are more influential while **H1b** suggests that reviews with rating consistent with product popularity are more influential. The hypotheses indicate that the influence of positive reviews and negative reviews is not constant. Rather, it is a function of product popularity. We therefore extend Eq. (3) by making β^+ and β^- functions of product popularity. That is,

$$\beta^+ = \beta_1^+ + \beta_2^+ \times LgSales_{i,t-1} \quad (4)$$

$$\beta^- = \beta_1^- + \beta_2^- \times LgSales_{i,t-1}. \quad (5)$$

Substitute Eqs. (4) and (5) into Eq. (3), we obtain:

$$\begin{aligned} \Delta LgSales_{it} = & \beta_1^+ [\Delta AvgRating_{it}]^+ + \beta_1^- [\Delta AvgRating_{it}]^- + \beta_2^+ [\Delta AvgRating_{it}]^+ \\ & \times LgSales_{i,t-1} + \beta_2^- [\Delta AvgRating_{it}]^- \times LgSales_{i,t-1} \\ & + \gamma_1 \Delta LgNumOfReviews_{it} + \delta_1 \Delta LgPrice_{it} + \gamma_2 \Delta LgPrice_{it} \\ & + \eta \Delta X_{it} + \Delta v_t + \epsilon_{it}. \end{aligned} \quad (6)$$

In Eq. (6), the impact of WOM is able to vary with product popularity and review rating. H1a suggests that β_2^+ be negative and β_2^- be positive because informational cascade theory suggests that positive reviews have more influence on niche products and negative reviews have more influence on popular products. H1b suggests the opposite: β_2^+ be positive and β_2^- be negative because of the influence of behavior heuristics.

4.3. Control variables

The influence of WOM is not only affected by customer review ratings and product popularity but also by other factors. We consider the following variables that may moderate the influence of WOM.

4.3.1. Review volume

Eq. (6) controls for the awareness effect by including changes in review volume as a control variable. The variable captures the underlying WOM diffusion process. Higher number of reviews suggests more interests from consumers to spread WOM and make others aware of the product. However, increased awareness does not necessarily lead to higher sales. In particular, the influence may vary across products with different popularity. A consumer who becomes aware of a popular product is more likely to make a purchase than a consumer who becomes aware of an unpopular product [18]. We therefore add the interaction between changes in review volume and product popularity as a control variable.

$$\begin{aligned} \Delta LgSales_{it} = & \beta_1^+ [\Delta AvgRating_{it}]^+ + \beta_1^- [\Delta AvgRating_{it}]^- + \beta_2^+ [\Delta AvgRating_{it}]^+ \\ & \times LgSales_{i,t-1} + \beta_2^- [\Delta AvgRating_{it}]^- \times LgSales_{i,t-1} \\ & + \gamma_1 \Delta LgNumOfReviews_{it} + \gamma_2 \Delta NumOfReviews_{it} \\ & \times LgSales_{i,t-1} + \delta_1 \Delta LgPrice_{it} + \Delta v_t + \epsilon_{it}. \end{aligned} \quad (7)$$

4.3.2. Product age

Product age could also have a significant moderating effect on the influence of WOM. For newly released products, product awareness and knowledge are often generated by vendor marketing campaigns (e.g. newly released movies), news, and press reports (e.g. Apple iPhone). The presence of these alternative information sources reduces the influence of WOM. On the other hand, marketing spending for and press interest in products that have been on the market for a long time is often limited. As a result, their sales become more dependent on WOM, which suggest that the influence of WOM could increase with product age. To control for the moderating effect of product age, we add interactive terms between product age and WOM rating and WOM volume to Eq. (7).

$$\begin{aligned} \Delta LgSales_{it} = & \beta_1^+ [\Delta AvgRating_{it}]^+ + \beta_1^- [\Delta AvgRating_{it}]^- + \beta_2^+ [\Delta AvgRating_{it}]^+ \\ & \times LgSales_{i,t-1} + \beta_2^- [\Delta AvgRating_{it}]^- \times LgSales_{i,t-1} \\ & + \gamma_1 \Delta LgNumOfReviews_{it} + \gamma_2 \Delta LgNumberofReviews_{it} \\ & \times LgSales_{i,t-1} + \theta_1^+ [\Delta AvgRating_{it}]^+ \times LgProductAge_{it} \\ & + \theta_1^- [\Delta AvgRating_{it}]^- \times LgProductAge_{it} \\ & + \theta_2 \Delta LgNumOfReviews_{it} \times LgProductAge_{it} + \delta_1 \Delta LgPrice_{it} \\ & + \Delta v_t + \epsilon_{it}. \end{aligned} \quad (8)$$

4.3.3. Other control variables

Besides the control variables that moderate the influence of WOM, we also add fixed product effect to the first difference model to allow different diffusion rates for different products [18]. Our final model is as follows:

$$\begin{aligned} \Delta LgSales_{it} = & \beta_1^+ [\Delta AvgRating_{it}]^+ + \beta_1^- [\Delta AvgRating_{it}]^- + \beta_2^+ [\Delta AvgRating_{it}]^+ \\ & \times LgSales_{i,t-1} + \beta_2^- [\Delta AvgRating_{it}]^- \times LgSales_{i,t-1} + \gamma_1 \\ & + \gamma_2 \Delta LgNumOfReviews_{it} \times LgSales_{i,t-1} + \theta_1^+ [\Delta AvgRating_{it}]^+ \\ & \times LgProductAge_{it} + \theta_1^- [\Delta AvgRating_{it}]^- \times LgProductAge_{it} \\ & + \theta_2 \Delta LgNumOfReviews_{it} \times LgProductAge_{it} + \delta_1 \Delta LgPrice_{it} \\ & + \Delta v_t + \tau_i + \epsilon_{it}. \end{aligned} \quad (9)$$

5. Data

We collected books and user review data daily from Amazon.com using its Electronic Commerce Service (ECS) platform to form a panel data of 3000 books over 220 days. To ensure unbiased sampling, we first collect ISBN numbers for all the books sold at Amazon. Amazon's ECS platform imposes a limit on the number of products returned for each query. To circumvent the limit, we leveraged the hierarchy of book categories. For instance, a book on baking bread is classified as follows:

Books > Subject > Cooking, Food & Wine > Baking > Cookies.

Each node in the hierarchy tree is attached with a unique browsing node number. We collected all the children nodes from each parent node recursively to establish the tree structure. We then obtained the list of all leaf nodes of books, attached corresponding ISBN numbers to these books, and deleted any duplications. This approach resulted in a total of over 3,700,000 unique books in a hierarchy tree with over 20,000 nodes. Among all these books, we drew a random sample of 3000 books for which we collected sales and review information on a daily basis.

The static information collected only once for each book includes ISBN, author, and release date. The dynamic information was collected daily for each book from September 2005 to April 2006 for 8 months. The daily information includes sales rank, Amazon price, and customer reviews (rating of each review and average rating for all reviews). We use sales volume to represent the popularity of a product. Amazon, however, only provides sales rank. Recent studies show that Amazon.com sales rank can be mapped into sales using a Pareto function [4,22,25]:

$$Sales = a \times SalesRank^b. \quad (10)$$

Log transformation of Eq. (10) suggests:

$$LgSales = lg(a) + b \times lg(SalesRank) \quad (11)$$

and thus

$$\Delta LgSales = b \times \Delta lg(SalesRank). \quad (12)$$

We use $b = 0.871$ estimated by Brynjolfsson et al. [4] to calculate sales in this study. Tables 2, 3a, and 3b provide the description, summary statistics and correlations for the key variables used in our model.

6. Results

In order to compare our results with the results of previous studies, we first present the regression results from the baseline first-difference

Table 2
Description of key variables.

Variable	Definition
$LgSales_{it}$	Log value of the sales volume of book i at time t
$LgPrice_{it}$	Log value of the price of book i at time t
$AvgRating_{it}$	Average review rating for book i at time t
$[\Delta AvgRating_{it}]^{+a}$	$= \begin{cases} \Delta AvgRating_{it}, & \text{if } \Delta AvgRating_{it} > 0 \\ 0, & \text{otherwise} \end{cases}$
$[\Delta AvgRating_{it}]^{-}$	$= \begin{cases} \Delta AvgRating_{it}, & \text{if } \Delta AvgRating_{it} < 0 \\ 0, & \text{otherwise} \end{cases}$
$LgNumOfReviews_{it}$	Log value of cumulative total number of reviews for book i at time t
$LgProductAge_{it}$	Log value of product age of book i at time t measured in days

^a $[\Delta AvgRating_{it}]^{+}$ measures the increase in average review rating for book i in period t , while $[\Delta AvgRating_{it}]^{-}$ measures the decrease in average review rating for book i in period t .

model without the moderating effect (Eq. (2)) in Column 1 of Table 4. To control for possible autocorrelation and heteroscedesity, we report Newey–West standard errors when applicable. The estimation of Eq. (2) suggests that the coefficients of the equation are the same as the traditional model used by prior studies, thus making it a good starting point for our analysis. The results show that an increase in review rating has a positive impact on incremental sales and the increase in the number of reviews also improves product sales. Both impacts are statistically significant. This result is consistent with the previous research, indicating that both average review rating and cumulative number of reviews positively affect the sales volume of the title and that price is negatively correlated with the sales.

Earlier WOM studies also show that negative WOM is more influential than is positive WOM [9], indicating that the influence WOM could vary across WOM valence. We validate this finding by splitting the variable for changes in average review rating ($\Delta AvgRating$) into two variables, one for positive changes in average rating and the other for negative changes in average rating. The coefficient on the positive rating change variable indicates the magnitude of the influence of positive WOM, while the coefficient on the negative rating change variable indicates the magnitude of the influence of negative WOM. We reanalyze Eq. (2) with the new variables in Column 2 of Table 4. The result confirms that negative WOM significantly reduces product sales while positive WOM has little influence, which is consistent with prior studies.

Our hypotheses are tested based on results in Table 5. column 1 of Table 5 reports the result from Eq. (6), which allows WOM influence to vary with the interactions between product popularity and WOM rating and volume. Given the presence of interactive terms with product popularity, the coefficients on the direct effect of positive and negative WOM represent their influences on the sales of the least popular products, while the coefficients on the interactive terms reveal how the influence changes with product popularity. The four coefficients show that positive WOM has insignificant influence on the sales of unpopular products but the influence of positive WOM increases significantly with product popularity. The result is exactly the opposite for

Table 3a
Summary statistics for key variables.

Variable	Mean	Std. dev.	Min	Max
$LgSales_{it}$	0.902	1.553	−2.630	10.526
$LgPrice_{it}$	2.704	0.625	0.693	5.700
$[\Delta AverageRating_{it}]^{+}$	0.001	0.031	0	3.000
$[\Delta AverageRating_{it}]^{-}$	−0.002	0.034	−2.500	0
$LgNumOfReviews_{it}$	0.921	2.103	−2.303	8.080
$LgProductAge_{it}$	4.034	3.454	−6.908	10.222

Table 3b
Correlations of key variables.

Variable	1	2	3	4	5	6
$LgSales_{it}$	1					
$LgPrice_{it}$	−0.076	1				
$[\Delta AverageRating_{it}]^{+}$	0.027	−0.001	1			
$[\Delta AverageRating_{it}]^{-}$	−0.035	0.000	0.002	1		
$LgNumOfReviews_{it}$	0.434	−0.231	0.002	−0.005	1	
$LgProductAge_{it}$	0.041	−0.018	−0.010	0.007	0.185	1

negative WOM. Negative WOM hurts the sales of unpopular products, but the influence decreases significantly with product popularity. In addition, the analysis reveals that the influence of WOM volume on sales also increases with product popularity. The results provide strong support for predictions from behavior heuristics theory (H1b).

To assess the robustness of the analysis, column 2 of Table 5 incorporates product age and the interactions between product age and WOM as control variables (Eq. (9)). These control variables allow the influence of WOM to vary with product age. The result in column 2 shows that most coefficients on these control variables are insignificant and the coefficients on the interactions between WOM and product popularity remain unchanged.

The result that popular products benefit more from positive online reviews while niche products suffer more from negative reviews indicates that WOM helps consumers to converge to the most popular books and leads to a rich-get-richer situation. Therefore, our result shows that the informative effect of WOM curtails the formation of long tail. This conclusion is consistent with the prediction of behavior economics that suggests individuals make insufficient adjustment when newly arrived information contradicts their prior beliefs. Our study also offers a possible explanation of the presence of superstar phenomenon identified in earlier long tail studies.

7. Discussion and conclusion

This paper provides a new perspective on how online information, specifically online WOM, influences the long tail phenomenon. Most prior studies focus on the awareness effect of online information and suggest that increasing awareness of niche products drives the long tail phenomenon. In this paper, we suggest that online information also influences consumers' evaluation of product quality and this informative effect could have a significant influence on consumer purchase decisions. Our analysis reveals that consumers are more receptive to positive WOM on popular products and negative WOM on unpopular products. This disparity leads to a rich-get-richer situation, constraining the formation of the long-tail.

Table 4
First difference model estimation of WOM influence.

	Model 1	Model 2
$\Delta LgPrice_{it}$	−0.426 (0.078)***	−0.426 (0.078)***
$\Delta AvgRating_{it}$	0.045 (0.020)**	
$[\Delta AvgRating_{it}]^{+}$		0.014 (0.031)
$[\Delta AvgRating_{it}]^{-}$		0.071 (0.028)***
$\Delta LnNumOfReviews_{it}$	0.058 (0.019)***	0.066 (0.020)***
Product fixed effect	Included	Included
Time fixed effect	Included	Included
Autocorrelation parameter	−0.354 (0.002)***	−0.354 (0.002)***
R-square	12.666%	12.666%
Number of observations	506,051	506,051

Newey–West standard errors are reported.

*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.10$.

Table 5
Decaying impact with moderating effect: how WOM affect long tail.

Variables	Model 3	Model 4
$\Delta LgPrice_{it}$	-0.424 (0.078)***	-0.424 (0.078)***
$[\Delta AvgRating_{it}]^+$	-0.006 (0.031)	-0.025 (0.045)
$[\Delta AvgRating_{it}]^-$	0.096 (0.040)***	0.106 (0.043)**
$[\Delta AvgRating_{it}]^+ * LgSales_{i,t-1}$	0.065 (0.014)***	0.067 (0.015)***
$[\Delta AvgRating_{it}]^- * LgSales_{i,t-1}$	-0.029 (0.013)***	-0.029 (0.014)**
$[\Delta AvgRating_{it}]^+ * LgProductAge_{it}$		0.005 (0.009)
$[\Delta AvgRating_{it}]^- * LgProductAge_{it}$		0.003 (0.008)
$\Delta LgNumOfReviews_{it}$	0.111 (0.014)***	0.072 (0.032)***
$\Delta NumOfReviews_{it} * LgSales_{i,t-1}$	0.040 (0.020)***	0.044 (0.006)***
$\Delta NumOfReviews_{it} * LgProductAge_{it}$		0.011 (0.006)*
Product fixed effect	Included	Included
Time fixed effect	Included	Included
Autocorrelation parameter	-0.354 (0.002)***	-0.354 (0.002)***
R-square	12.680%	12.682%
Number of Observations	506,051	506,051

Newey–West standard errors are reported.

*** p<0.01.

** p<0.05.

* p<0.10.

While businesses are growing in their ability to use long tail strategies to sell a wider range of goods in smaller quantities [11], our findings suggest that the increasing availability of online WOM is against the sales of niche products while supports the sales of popular products. This finding has important implications for product vendors and retailers. In particular, our finding suggests that vendors and retailers do not need to worry much about occasional negative reviews on popular products, but they need to take care of unsatisfied buyers of niche products as negative WOM is particularly damaging to niche products. At the same time, our finding suggests that it is important for product vendors and retailers to attract positive reviews on popular products as such reviews can further enhanced their sales. Our findings also indicate the importance of consumers' initial expectations and the difficulty of changing consumers' beliefs once expectations have been set. Business can leverage this behavior bias to develop better marketing strategies on advertising, promotion, and customer relationship management. One important implication, for instance, is that existing customers are more tolerant of negative reviews than new customers and thus making new customers satisfied is essential for business growth.

This study is the first to analyze the influence of online WOM using behavior heuristics. Most prior studies ignore consumer behavior that may affect the way expectations are formed and decisions are made. Different from the economic assumption that consumers are perfectly rational, behavior heuristics theories suggest that consumers are biased toward their initial expectation when making decisions heuristically. This bias serves as an underlying mechanism driving the influence of online WOM on product sales. Extending existing research on WOM, we use the first difference approach to provide more flexibility in modeling the impact of WOM on product sales. Our approach focuses on the influence of newly generated WOM and allows the influence to vary with the interaction between WOM and product popularity. The findings provide a more refined understanding of the WOM effect.

This study also has some limitations. First, while we develop our hypotheses based on behavior heuristics theory and show that our result is consistent with its prediction, we do not survey consumers for their underlying motivations and there may exist alternative explanations for consumers' biased responses to WOM. Future studies could benefit from a combination of secondary data from consumer survey to take a more in-depth look at the phenomenon. Second, our analysis focuses on the informative effect of WOM. In reality, WOM influences the long tail phenomenon in multiple ways. Notably, Dellarocas and Narayan [14] suggest that consumers are much more likely to discuss popular products, implying that the awareness effect of WOM also favors popular products. Future studies can benefit from combining multiple effects

of WOM and providing a more comprehensive modeling of the influence of WOM on the long tail formation. This paper also bears several limitations in the data collection. The data are solely collected from Amazon.com. In future research, using data sets from multiple online retailers can help generalize our research on WOM effects and the long tail phenomenon. In addition, information provision and website design could also have significant impacts on WOM effect and the formation of the long tail. We leave those concerns for future investigation.

Acknowledgements

The authors would like to thank Xin Zhao for his work on earlier versions of this paper and participants at the 2008 Conference on Information Systems and Technology (CIST 2008) and the 5th Symposium on Statistical Challenges in Electronic Commerce Research (SCECR 2009) for their valuable feedback.

References

- [1] C. Anderson, The long tail, *Wired* 12 (10) (2004) 170–177.
- [2] S. Basuroy, S.S. Chatterjee, A. Ravid, How critical are critical reviews? The box office effects of film critics, star power, and budgets, *The Journal of Marketing* 67 (4) (2003) 103–117.
- [3] S. Bikhchandani, D. Hirshleifer, I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy* 100 (5) (1992) 992–1026.
- [4] E. Brynjolfsson, M.D. Smith, Y.J. Hu, Consumption surplus in the digital economy: estimating the value of increased product variety at online booksellers, *Management Science* 49 (2003) 1563–1579.
- [5] E. Brynjolfsson, M.D. Smith, Y.J. Hu, From niches to riches: anatomy of the long tail, *Sloan Management Review* 47 (4) (2006) 67–71.
- [6] E. Brynjolfsson, M.D. Smith, Y.J. Hu, Goodbye pareto principle, hello long tail: the effect of search costs on the concentration of product sales, in: MIT working paper, 2007.
- [7] P.Y. Chen, S. Dhanasobhon, M.D. Smith, All reviews are not created equal: the disaggregate impact of reviews and reviewers at Amazon.com, in: Paper, 55, Heinz Research, 2001, (Available at <http://repository.cmu.edu/heinzworks/55>).
- [8] Y. Chen, S. Fay, Q. Wang, Marketing implications of online consumer product reviews, in: University of Florida Working Paper, 2003.
- [9] J. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, *Journal of Marketing Research* 43 (2006) 345–354.
- [10] 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, *Marketing Science* 29 (5) (2010) 944–957.
- [11] E.K. Clemons, P.F. Nunes, Carrying your long tail: delighting your consumers and managing your operations, *Decision Support Systems* 51 (4) (2011) 884–893.
- [12] E.K. Clemons, G. Gao, L. Hitt, When online reviews meet hyperdifferentiation: a study of the craft beer industry, *Journal of Management Information Systems* 23 (2) (2006) 149–171.
- [13] C. Dellarocas, N.F. Awad, X. Zhang, Exploring the value of online product reviews in forecasting sales: the case of motion pictures, *Journal of Interactive Marketing* 21 (4) (2007) 23–45.
- [14] C. Dellarocas, G. Gao, R. Naraya, Are consumers more likely to contribute online reviews for hit or niche products? *Journal of Management Information Systems* 27 (2) (2010) 127–158.
- [15] S. Dewan, J. Ramaprasad, Impact of blogging on music sales: the long tail effect, in: Working paper, 2007.
- [16] T. Dierkes, M. Bichler, R. Krishnan, Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks, *Decision Support Systems* 51 (3) (2011) 361–371.
- [17] W. Duan, B. Gu, A.B. Whinston, 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.
- [18] W. Duan, B. Gu, A.B. Whinston, Information cascades and software adoption on the internet: an empirical investigation, *MIS Quarterly* 33 (1) (2009) 23–48.
- [19] Elberse, F. Oberholzer-Gee, Superstars and underdogs: an examination of the long tail phenomenon in video sales, in: Harvard Business School Working Paper Series, 07–015, 2007.
- [20] J. Eliashberg, S.M. Shugan, Film critics: influencers or predictors? *The Journal of Marketing* 61 (2) (1997) 68–78.
- [21] D. Fleder, K. Hosanagar, Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity, in: Net Institute Working Paper, 2008.
- [22] A. Ghose, M.D. Smith, R. Telang, Internet exchanges for used books: an empirical analysis of product cannibalization and welfare, *Information Systems Research* 17 (1) (2006) 3–19.
- [23] D. Godes, D. Mayzlin, Using online conversations to study word-of-mouth communication, *Marketing Science* 23 (24) (2004) 545–560.
- [24] K.H. Goh, J. Bockstedt, Unbundling and the long tail: new evidence on the consumption of information goods, in: Working Paper, 2008.

- [25] Goolsbee, J. Chevalier, Measuring prices and price competition online: amazon.com and barnesandnoble.com, *Quantitative Marketing and Economics* 1 (2) (2003) 203–222.
- [26] D. Griffin, A. Tversky, The weighing of evidence and the determinants of confidence, *Cognitive Psychology* 24 (1992) 411–435.
- [27] B. Gu, J. Park, P. Konana, The impact of external word-of-mouth sources on retailer sales of high-involvement products, *Information Systems Research* 23 (1) (2012) 182–196.
- [28] G. Haubl, V. Trifts, Consumer decision making in online shopping environments: the effects of interactive decision aids, *Marketing Science* 19 (1) (2000) 4–21.
- [29] A. Hervas-Drane, Word of mouth and recommender systems: a theory of the long tail, in: Working Paper, 2008.
- [30] N. Hu, L. Liu, J.J. Zhang, Do online reviews affect product sales? The role of reviewer characteristics and temporal effects, *Information Technology and Management* 9 (3) (2008) 201–214.
- [31] Y. Liu, Word of mouth for movies: its dynamics and impact on box office revenue, *The Journal of Marketing* 70 (2006) 74–89.
- [32] W.W. Moe, M. Trusov, The value of social dynamics in online product ratings forums, *Journal of Marketing Research* 48 (3) (2011) 444–456.
- [33] S. Moon, P.K. Bergey, D. Iacobucci, Dynamic effects among movie ratings, movie revenues, and viewer satisfaction, *The Journal of Marketing* 74 (1) (2010) 108–121.
- [34] G. Oestreicher-Singer, A. Sundararajan, Network structure and the long tail of electronic commerce, in: Working Paper, Stern School of Business, New York University, 2006.
- [35] M. Rabin, J.L. Schrag, First impressions matter: a model of confirmatory bias, *Quarterly Journal of Economics* 114 (1) (1999) 37–82.
- [36] M. Sun, The informational role of consumer disagreement, in: Working paper, 2008.
- [37] M. Sun, How does variance of product ratings matter? *Management Science* 58 (4) (2012) 696–707.
- [38] C. Tucker, J. Zhang, How does popularity information affect choices? A field experiment, *Management Science* 57 (5) (2011) 828–842.
- [39] Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, *Science* 185 (1974) 1124–1131.
- [40] M. Wible, J. Tripp, B. Phillips, B. Sambamurthy, The moderating role of search costs on long tail, in: Fifth Symposium on Statistical Challenges in Electronic Commerce Research, Pittsburgh, PA, 2009.
- [41] X. Zhang, C. Dellarocas, N.F. Awad, The impact of online movie reviews on box office performance, in: Workshop on Information Systems and Economics (WISE), 2004.
- [42] F. Zhu, X. Zhang, Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics, *The Journal of Marketing* 74 (2) (2010) 133–148.



Dr. Bin Gu is an associate professor at the W. P. Carey School of Business, Arizona State University. He received his Ph.D. from the Wharton School, University of Pennsylvania. His research focuses on electronic commerce, online social networks, information economics and IT strategy. His work has appeared in *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *Journal of Retailing*, *Decision Support Systems* and others. Bin's research won the 2012 Emerald Citations of Excellence Award, the 2008 ISR Best Published Paper Award and the Best Paper-in-Track Award of the 2007 International Conference on Information Systems (ICIS).



Ms. Qian Tang is a Ph.D. candidate in the Information, Risk, and Operations Management Department of McCombs School of Business at the University of Texas at Austin. She expects to receive her Ph.D. in 2013. Her research interests include electronic commerce, economics of information systems, social media, online Word of Mouth, reputation, and internet security. Her current research focuses on the issues in social media and internet security.



Dr. Andrew B. Whinston received his Ph.D. at Carnegie Mellon University and is currently a professor at The University of Texas at Austin where he holds the Hugh Roy Cullen Centennial Chair in Business Administration. He is also the director of the Center for Research in Electronic Commerce. He has published extensively on resource allocation issues and is currently working on Internet security. He has completed numerous research projects that investigate economics, internet technology, and operations research in the study of information systems issues. In 2011, he was rated as the most influential scholar in the Information Systems field by the h-index which measures scholarly influence.