

November 2021

The Aggregate Impact of Consumer Reviews on Market Outcome in Differentiated Products Market

Jun B. Kim

Seoul National University, Republic of Korea, junbkim@snu.ac.kr

Follow this and additional works at: <https://amj.kma.re.kr/journal>



Part of the [Advertising and Promotion Management Commons](#), [E-Commerce Commons](#), [Marketing Commons](#), and the [Other Business Commons](#)

Recommended Citation

Kim, Jun B. (2021) "The Aggregate Impact of Consumer Reviews on Market Outcome in Differentiated Products Market," *Asia Marketing Journal*: Vol. 23 : Iss. 3 , Article 1.

DOI: 10.53728/2765-6500.1578

Available at: <https://amj.kma.re.kr/journal/vol23/iss3/1>

This Article is brought to you for free and open access by Asia Marketing Journal. It has been accepted for inclusion in Asia Marketing Journal by an authorized editor of Asia Marketing Journal.

The Aggregate Impact of Consumer Reviews on Market Outcome in Differentiated Products Market

Jun B. Kim

Seoul National University, South Korea

Abstract

We investigate the aggregate impact of consumer reviews on market outcome in a differentiated product category. We model consumers as Bayesian learners who use online consumer reviews to learn and update their beliefs on product quality before their choice. For our empirical analysis, we use aggregate-level, longitudinal data from Amazon.com in the digital camcorder category and estimate the demand parameters.

Using model estimates, we conduct two simulation studies and quantify the impact of consumer reviews on the market outcome. We report that the products experience heterogeneous market share changes: the standard deviation of market share changes across products and time is 16.7%, ranging from -40% to 20% . In addition, consistent with the previous findings in experience goods, the marginal effect of low consumer ratings is greater than that of high consumer ratings. We discuss model limitations and offer directions for further research.

Keywords: Consumer reviews, Bayesian learning, Choice model, Aggregate demand model, Differentiated products

1. Introduction

A large body of empirical research reports significant effects of online consumer reviews on consumer demand in books (e.g., [Chevalier and Mayzlin 2006](#); [Pathak et al. 2010](#); [Zhao et al. 2013](#)), in movies (e.g., [Chakravarty et al. 2010](#); [Yu et al. 2012](#); [Chen et al. 2019](#); [Chintagunta et al. 2010](#)), and in video games (e.g., [Zhu and Zhang 2010](#); [Cui et al. 2012](#)). As the underlying mechanism, it is postulated that imperfectly informed consumers depend on past consumers' opinions in experience products (e.g., [Senecal and Nantel 2004](#)) and on their own experiences in consumer packaged goods (e.g., [Erdem and Keane 1996](#)) to resolve product uncertainty prior to their purchase decisions. On the other hand, other studies report a non-significant relationship between online consumer reviews and sales (e.g., [Liu 2006](#); [Duan et al. 2008](#)).

This paper follows the literature and posits that intangible and experience attributes are important to consumers in differentiated products market (e.g., [Chen and Xie 2008](#); [Huang et al. 2009](#)). We model consumers as Bayesian learners who use online

consumer reviews to learn about and update their beliefs on product quality before their choice. We further simulate and quantify the aggregate impact of consumer reviews on the market outcome. Doing so, we aim to make the following contributions to the empirical literature on consumer reviews.

First, we study the effects of online consumer reviews in differentiated products. Critically note that most of the past empirical research has focused on experience goods such as books and movies and there is a lack of empirical research in differentiated products. As a consequence, our understanding of consumer reviews on this topic may be limited.¹ Nonetheless, we expect consumer reviews to play a significant role in differentiated products because durable goods purchase usually involves high costs ([You et al. 2015](#)) and shoppers may not be familiar with all the product attributes ([Huang et al. 2009](#)).

Second, we explicitly incorporate the consumer learning model into a choice-based aggregate demand model in the context of consumer reviews. While most of the past research has adopted reduced-form models and studied the effect of consumer review, we develop and apply the

Received 23 August 2021; revised 3 October 2021; accepted 21 October 2021.

Available online 14 November 2021

E-mail address: junbkim@snu.ac.kr.

¹ The only exception is [Gu et al. \(2012\)](#), who studied the effect of online reviews among two manufacturers in the camera category.

<https://doi.org/10.53728/2765-6500.1578>

2765-6500/© 2021 Korean Marketing Association (KMA).

Bayesian learning model to aggregate-level data. An explicit model of consumer behaviors will allow us to estimate how informative consumer reviews are and how consumers use them to resolve product uncertainties during their choice.

Last, we conduct a set of simulation studies and quantify the aggregate impact of consumer reviews on the market outcome under product competition. Note that most past research in experience goods investigated the effect of consumer reviews in isolation without considering product competition (e.g., [Chevalier and Mayzlin 2006](#)). However, since competition is a core feature in differentiated products markets, it is critical to include the feature in the empirical analysis of differentiated goods.

We apply the proposed aggregate learning and choice model to Amazon.com's longitudinal, store-level data in the digital camcorder category for our empirical application. Once we estimate the model parameters, we compute and contrast market shares for products across time when consumer reviews are all turned off for all products. We report that the products with low average consumer ratings lose market shares in the presence of consumer reviews. In addition, consumer reviews have a significant aggregate impact on market shares: the standard deviation of percentage-wise share differences is 16.7%, ranging between -40% and 20% across products and time. Therefore, our simulation study indirectly supports the earlier proposition that limited consumer information has a profound impact on the market outcome ([Nelson 1970](#)).

This paper is organized as follows. In section 2, we review the relevant literature. In section 3, we propose a Bayesian learning model within the random coefficients discrete choice model. In section 4, we briefly describe Amazon.com's longitudinal data set and conduct exploratory results on the effects of consumer reviews on sales rank. In section 5, we provide our empirical results and present the simulation studies in section 6. We conclude the paper in section 7.

2. Related literature

We discuss empirical literature on online consumer reviews followed by literature on consumer reviews and the market. A large body of empirical research documents the link between online consumer reviews and sales in experience goods category. We review a few key papers in this section. In their seminal paper, [Chevalier and Mayzlin \(2006\)](#) use sales rank data from Amazon.com and investigate the incremental effects of consumer reviews on aggregate book sales. Among their key findings is

that consumers find one-star reviews more informative than five-star reviews. [Liu \(2006\)](#) and [Duan et al. \(2008\)](#) use box office data and report that it is not the consumer ratings (valence) but the volume that can be used as a predictor for box office performance. In contrast, [Chintagunta et al. \(2010\)](#) use consumer review data from Yahoo Movie website and box office data in the US market and report the opposite: they report that, after various control mechanisms, it is the valence of the consumer reviews and not the review volumes that predicts movie performance. [Zhu and Zhang \(2010\)](#) use data on the video game industry and report that online reviews impact sales and the effect is more substantial with consumers who have more internet experience.

The second body of related literature theoretically or empirically examines the implications of online reviews on consumer welfare and seller's strategy. [Li and Hitt \(2008\)](#) theorize that online consumer reviews are subject to bias, leading to potential bias in future demand. That is, online product reviews can influence the evolution of subsequent online reviews, which can affect downstream sales and consumer surplus ([Park et al. 2016](#)). [Chen and Xie \(2008\)](#) empirically validate that online consumer reviews are informative of product quality and complementary to objective information provided by the manufacturers. [Jiang and Chen \(2007\)](#) develop a theoretical model and predict that consumer reviews and ratings improve consumer surplus, vendor profitability, and social welfare. They also conjecture that a seller may have incentives to induce higher product ratings by underpricing the products during the early product life cycle. [Wu et al. \(2015\)](#) use data set from a Chinese food review site, develop a consumer learning model, and estimate the impact of online reviews on consumers and sellers. In a series of simulation studies, they report that the monetary value for each consumer is about 7 CNY and about 8.6 CNY from each consumer for the reviewed restaurants. In our paper, we assess and quantify the aggregate impact of consumer reviews on the market outcome. Doing so, we aim to contribute to the research stream that empirically studies the market-level implications of online customer reviews.

Another related research stream is consumer learning. In their seminal paper, [Erdem and Keane \(1996\)](#) posit that consumers in the fast-moving consumer goods (FMCG) market are forward-looking and use their own past consumption experiences as a major information source to learn about and update the true product quality. For a comprehensive review on this topic, please refer to [Ching et al. \(2013\)](#). Most closely related to our paper

regarding the learning part of the model is [Narayanan et al. \(2005\)](#), who use aggregate-level data and model physicians' learning behaviors in the prescription market. Their paper models that physicians receive noisy signals on true drug quality from multiple sources such as detailing and feedback from the past patients. They report that the learning rate is heterogeneous across physicians. Another related paper is [Zhao et al. \(2013\)](#) who develop a learning model for consumer reviews for individual-level data and investigated their implications on retailers' profit. They estimate that consumers learn more from other consumers' reviews on the focal book than their own experience from the same book genre.

Last, our model is an extension of [Kim \(2019\)](#), who estimated consumer demand for durable goods using data from Amazon.com. The key difference is that while [Kim \(2019\)](#) adopted a reduced form approach and directly incorporated review valence and volume in the utility specification, we explicitly incorporate consumers' learning process in the choice model and estimate learning-related parameters. By doing so, we can understand how informative consumer reviews are to consumers and use the learning parameters to explain how valence and volume of consumer reviews affect consumer choice.

3. Model

3.1. Consumer choice

A product is broadly characterized by two sets of attributes ([Nelson 1970](#); [Li and Hitt 2008](#)). First, search attributes are product characteristics consumers can fully evaluate before purchase. Second, experience or intangible attributes are product characteristics consumers can fully evaluate only after purchase and consumption.² In our empirical context of digital camcorders, brand names and technical attributes such as zoom and screen size are examples of search attributes. Ease of operation and usage scenarios are examples of experience attributes. While shoppers can fully learn about the values of search attributes prior to choice, they must resort to alternative information sources to resolve uncertainty on experience attributes. In our empirical context, we postulate that consumers use online consumer reviews as an alternative information source. Our premise is aligned with past research. [Chen and Xie \(2008\)](#) report that online consumer reviews are informative of experience attributes and

that they are complementary to information provided by the manufacturers. In their analysis, consumer reviews indeed contain very little information on search attributes because they provide very little information on technical specification (i.e., of 0.032).

We construct consumer utility with two sets of attributes. In the next subsection, we discuss in detail our learning model on experience attributes from online consumer reviews. We express the utility of a risk-neutral shopper $i = 1, \dots, I$, for product $j = 1, \dots, J$, at time $t = 1, \dots, T$ as,

$$\begin{aligned} u_{ijt} &= V_{ijt} + \bar{Q}_{ijt} + e_{ijt} \\ &= \gamma_j + X_j \cdot \beta_{1i} + Z_{jt} \cdot \beta_{2i} - \alpha_i P_{jt} + \bar{Q}_{ijt} + e_{ijt}, \end{aligned} \quad (1)$$

where γ_j is the product-specific intercept, X_j and Z_{jt} are $[1 \times K1]$ and $[1 \times K2]$ row vectors representing time-invariant and time-varying search attributes, and P_{jt} price. Term e_{ijt} is an i.i.d. error term across i , j , and t , and represents idiosyncratic consumer tastes. \bar{Q}_{ijt} is i 's expectation of experience attribute j conditional on information available at t . It is formally defined as,

$$\bar{Q}_{ijt} = E_l(Q_{ijl} | I_{jt})$$

where l indexes different usage or experience scenarios of users, Q_{ijl} is i 's experience value for j on usage case l , I_{jt} is all the information available for j at t . The expectation operator is across l . Shoppers do not have uncertainty about V_{ijt} in the utility function. However, shoppers are uncertain about Q_{ijl} , and they form and use its expectation during their choice process. Coefficient vector of $\beta_i = [\beta_{1i}; \beta_{2i}]$ is consumer sensitivities to time-invariant and time-varying product search attributes and represented as,

$$\beta_i \sim N(\beta_0, \sum_\beta),$$

where β_0 is the mean vector, and \sum_β is a diagonal variance-covariance matrix. Consistent with theory, we assume that price coefficients are always negative and subject to a log-normal distribution,

$$\exp(\alpha_i) \sim N(\alpha_p, \sigma_p^2).$$

We discuss two important roles of product-specific intercepts of γ_j in our utility specification. First, the intercept term captures all unobserved product heterogeneity (e.g., design) shoppers experience prior to choice. In contrast, the quantity of

² In this article, we sometimes use "intangible" attribute to indicate "experience" attributes.

\bar{Q}_{ijt} captures experience attribute that shoppers cannot observe before their choice (e.g., ease of operation). They can only form an expectation. Second, when studying the effect of consumer ratings on sales, we must address potential endogeneity since it is likely that a “high” quality simultaneously leads to higher sales and a higher consumer rating. By including product-specific intercept of γ_j we aim to control as much as possible product heterogeneity across options and avoid any spurious correlations between consumer ratings and demand (e.g., Chevalier and Mayzlin 2006). Under the assumption of GEV type I distribution for e_{ijt} , consumer i chooses j at time t with the following choice probability,(2)

$$p_{ijt} = \frac{\exp(\gamma_j + X_j \cdot \beta_{1i} + Z_{jt} \cdot \beta_{2i} - \alpha_i P_{jt} + \bar{Q}_{ijt})}{\sum_{k=1}^J \exp(\gamma_k + X_k \cdot \beta_{1i} + Z_{kt} \cdot \beta_{2i} - \alpha_i P_{kt} + \bar{Q}_{ikt})} \quad (2)$$

In the next section, we model how shoppers use consumer reviews and form \bar{Q}_{ijt} during their choice.

3.2. Consuming and learning from consumer reviews

In this subsection, we first model how Bayesian shoppers consume *one* consumer review and learn about and update their belief on experience attribute. We then extend our model to the case of *multiple* consumer review consumption. Doing so, we show that our approach for learning leads to a parsimonious model involving average consumer rating (or numeric rating) and the number of reviews, which are common in past empirical research. A consumer review consists of two components: a numeric rating (e.g., the number of stars at Amazon.com ranging from 1 to 5) and review text. While the former summarizes consumers’ overall evaluation of a product, the latter typically describes the pros and cons of the product. Departing from past empirical research, we assume that consumers use both components when they consume a review and learn about products. Our key modeling premise is that, upon consuming *one* online review, shoppers receive a noisy signal around the true mean value of the experience attribute. That is, while the numeric rating summarizes the mean experience attribute value, the review text adds noise around the mean value.

We now formalize the above proposition. First, we assume that shoppers at $t = 0$ share an initial belief for the experience attribute value of \bar{Q}_{ij0} that is common across all products,

$$\bar{Q}_{ij0} = N(Q_0, \sigma_0^2), \quad (3)$$

where tilde (\sim) represents a random variable, and its randomness comes from the idiosyncratic shoppers’ needs such as different usage scenarios. We model that shopper i , consuming an online review of $cr_j = \{r_j, txt_j\}$ where r_j is a numeric rating and txt_j is a review text, receives a noisy signal on the value of experience attributes as,

$$\tilde{\lambda}_{ij} = N(\Phi_0 + \Phi_1 r_j, \sigma_\omega^2) \quad (4)$$

where Φ_0 is the base parameter, Φ_1 is the mean sensitivity with respect to numeric rating, and σ_ω^2 is the variance capturing the accuracy of a signal. Note that the mean value of the signal is proportional to r_j . The randomness of a signal comes from multiple sources in the review text. First, there can be an information gap between the contents of the review text (txt_j) and the information needs of shoppers. In addition, there can be idiosyncratic taste differences between review providers and review consumers. For instance, the preferences of early buyers may be different from those of later buyers (Li and Hitt 2008).

We now extend our model to the consumption of *multiple* consumer reviews at t . We model that the set of consumer reviews provides shopper i at t with a set of independent signals,

$$R_{ijt} = \{ \tilde{\lambda}_{ij1}, \tilde{\lambda}_{ij2}, \dots, \tilde{\lambda}_{ij\tau} \}, \quad (5)$$

where $\tilde{\lambda}_{ij\tau}$ is defined in Eq (4) and time subscript τ means that the review was generated at $\tau < t$. Following the literature, we assume that the signals are independent of one another (Erdem and Keane 1996; Narayanan et al. 2005). For this part of the learning model development, we closely follow the Bayesian learning model (e.g. Narayanan et al. 2005). Facing the review set of R_{ijt} , shopper i update their expectation on j ’s experience value by computing the posterior belief as,

$$Q_{ijt} | R_{ijt} \sim N(\bar{Q}_{ijt}, \sigma_{jt}^2) \quad (6)$$

where,

$$\bar{Q}_{ijt} = \frac{\left(\frac{1}{\sigma_0^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} \cdot Q_0 + \frac{\left(\frac{n_{jt}}{\sigma_\omega^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} \cdot \left(\sum_{\tau < t} \tilde{\lambda}_{ij\tau}\right), \quad (7)$$

where n_{jt} is the number of reviews at t or the number of elements in the set R_{ijt} . The variance of the posterior experience, σ_{jt}^2 is,

$$\sigma_{jt}^2 = \frac{1}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} \tag{8}$$

Using Eq (4) subject to the standard property for the mean of i.i.d. normal distribution, we can rewrite the summation term in Eq (7) as,

$$\sum_{\tau < t} \frac{\tilde{\lambda}_{ij\tau}}{n_{jt}} \equiv \lambda_{ijt} \sim N\left(\Phi_0 + \Phi_1 \cdot s_{jt}, \frac{\sigma_\omega^2}{n_{jt}}\right), \tag{9}$$

where s_{jt} is the average numeric rating for j at t and is defined as,

$$s_{jt} = \frac{1}{n_{jt}} \cdot \sum_{\tau < t} r_{j\tau},$$

in which $r_{j\tau}$ is the numeric rating of $cr_{j\tau}$ generated at $\tau < t$. Now, we can rewrite Eq (7),

$$\bar{Q}_{ijt} = \frac{\left(\frac{1}{\sigma_0^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} \cdot Q_0 + \frac{\left(\frac{n_{jt}}{\sigma_\omega^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} \lambda_{ijt}, \tag{10}$$

Therefore, \bar{Q}_{ijt} is subject to a distribution of

$$\bar{Q}_{ijt} \sim N\left(m_{jt}, \sigma_{m_{jt}}^2\right), \tag{11}$$

where,

$$m_{jt} = \frac{\left(\frac{1}{\sigma_0^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} Q_0 + \frac{\left(\frac{n_{jt}}{\sigma_\omega^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)} (\Phi_0 + \Phi_1 \cdot s_{jt}), \tag{12}$$

$$\sigma_{m_{jt}}^2 = \frac{\left(\frac{n_{jt}}{\sigma_\omega^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n_{jt}}{\sigma_\omega^2}\right)^2} \tag{13}$$

We discuss our learning model. For the case of multiple review consumption in Equations (12) and (13), our approach leads to a parsimonious learning model characterized by the average numeric rating (s_{jt}) and the number of reviews (n_{jt}). This is an attractive feature and the use of these two quantities is common in past research based on reduced form approaches (e.g., Chevalier and Mayzlin 2006). Our approach offers more rich interpretation and introduces uncertainty in the learning model. By doing so, we explicitly model how informative each review is to shoppers during their choice process.

In total, there are five parameters to estimate in our learning part of the model, $\{Q_0, \sigma_0^2, \Phi_0, \Phi_1, \sigma_\omega^2\}$. For identification purpose, we set $Q_0 = 0$ during

estimation. Note that Eqs (11)–(13) will be used in computing Eq (1). Using our individual-level model of Eq (2), we predict j 's market share at t as,

$$\hat{s}_{jt} = \int \left(\frac{\exp\left(\gamma_j + X_j \beta_i - \alpha_j P_{jt} + \bar{Q}_{ijt}\right)}{\sum_k^J \exp\left(\gamma_k + X_k \beta_i - \alpha_k P_{kt} + \bar{Q}_{ikt}\right)} \right) f\left(\beta_i, \alpha_i, \gamma_j, \Phi | \Omega\right) d(\Omega), \tag{14}$$

where $\Phi = \{Q_0, \sigma_0^2, \Phi_0, \Phi_1, \sigma_\omega^2\}$ are model parameters associated with consumer learning and $\Omega = \{\beta_0, \Sigma_\beta, \alpha_p, \sigma_p^2\}$ are hyper-parameters for consumer preference. In estimating our model using aggregate-level data, we follow the common approach in choice-based aggregate demand models (e.g., Berry et al. 1995; Nevo 2000) and numerically integrate Eq. (3) over consumer distributions using simulation.

4. Data

4.1. Description

We use aggregate level, longitudinal data set in the camcorder category from Amazon.com for our empirical application. Our empirical data in this paper overlap those in Kim (2019). For a detailed description of the data, please refer to Kim (2019). In this subsection, we provide a short description of the data. Data were collected, once every other day, for about ten months, starting from August 2006. The long data collection time window is attractive for our empirical analysis since we observe product entries and the complete trajectories of sales and consumer reviews. In the data, we observe daily sales rank, price, consumer reviews, and detailed characteristics of camcorders sold at Amazon.com. The average number of products in the choice set across time is about 80 with a minimum of 61 and a maximum of 103. We also have product availability information in the data. “Seller (Amazon.com)” and “Seller (3rd party)” indicate that a product is available for purchase from Amazon and other 3rd party vendors, respectively. “Seller (Request)” means that the product is unavailable but consumers can submit a request to participating vendors.

4.2. Exploratory analysis

Before estimating the proposed model, we briefly assess the effects of consumer reviews in an exploratory study. We first present the trajectories of two selected products in Figs. 1 and 2. We show the evolution of sales rank, average star rating, and number of reviews in each figure. In the figure, a

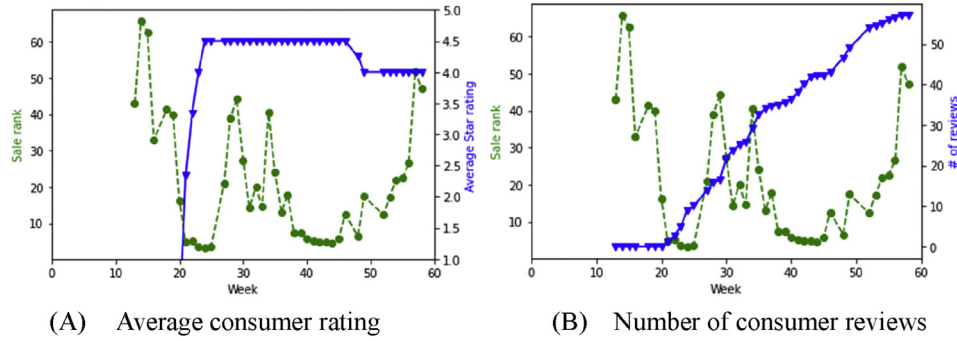


Fig. 1. Product with many reviews. Panel (A) shows the evolution of the sale rank (circle, left vertical axis) and the average consumer rating (inverted triangle, right vertical axis). Lower sales rank indicates higher popularity. Panel (B) shows the evolution of the sale rank (circle, left vertical axis) and the number of reviews (inverted triangle, right vertical axis).

smaller rank value on the y-axis means higher popularity. While Fig. 1 shows the trajectory of a product with many reviews, Fig. 2 shows that of a product with a small number of reviews. In Fig. 1, the correlation between the sales rank and average rating (number of reviews) is -0.55 (-0.23). In Fig. 2, the correlation between the sales rank and average rating (number of reviews) is -0.70 (0.69). In Fig. 2, the valence of reviews seems to matter but not the volume of reviews. Although these figures help visualize the evolutions of two products, other product characteristics may also affect the product sales. Therefore, we use a linear regression and offer model-free evidence for the effects of consumer reviews on sales.

To that end, we use the following specification for our linear regression,

$$\log\left(\frac{1}{rank_{jt}}\right) = \beta_t + \alpha_j + \beta_1 \cdot \log(price_{jt}) + \beta_2 \cdot \log(s_{jt} + 1) + \beta_3 \cdot \log(n_{jt} + 1) + \beta_4 \cdot Seller_{Amzn_{jt}} + \beta_5 \cdot Seller_{Partner_{jt}} + e_{jt},$$

$$j = 1, \dots, J, t = 1, \dots, T$$

where $rank_{jt}$ is j 's raw sales rank at t , β_t and α_j are time and product fixed effects. In addition, s_{jt} is the average numeric rating and n_{jt} is the number of consumer reviews at t . Note that α_j captures the effects of all time-invariant product characteristics such as brand name and number of pixels. The dependent variable in our model is the log of the inverse of sales rank. In the absence of sales data, the inverse of sales rank can be used as a proxy for sales (Chevalier and Mayzlin 2006). Taking the log of all time-varying independent variables, we can directly interpret the parameter estimates as elasticity.

The parameter estimates are shown in Table 1. The signs of key coefficients are all intuitive. Among them, the coefficients of price and average numeric rating are significant, while that of the number of consumer reviews is not. Our log–log model estimates the price elasticity at -1.66 . Concerning consumer reviews, the average star rating elasticity is significant at 0.46 , while volume elasticity is insignificant at 0.02 . Our results on consumer review variables are opposite from those in Gu et al. (2012): they report an insignificant effect of valence and a significant effect of review volume. We speculate

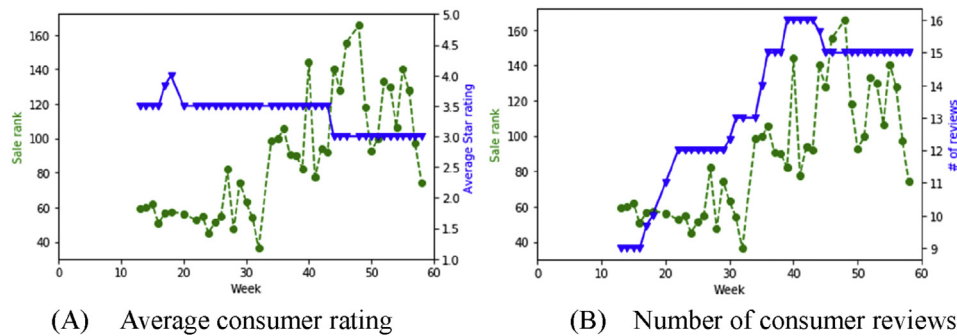


Fig. 2. Product with smaller number of reviews. Panel (A) shows the evolution of the sale rank (circle, left vertical axis) and the average consumer rating (inverted triangle, right vertical axis). Panel (B) shows the evolution of the sale rank (circle, left vertical axis) and the number of reviews (inverted triangle, right vertical axis).

Table 1. Regression estimation result.

Variables	Estimates (std. err.)
log (price)	-1.66 (0.06)
log (s _{jt} +1)	0.46 (0.03)
log (n _{jt} +1)	0.02 (0.02)
Seller_amazon _{jt}	0.42 (0.27)
Seller_partner _{jt}	0.13 (0.27)
Product FE	Yes
Time FE	Yes
R ²	0.81
Number of observations	5142

that the different results may be due to different empirical settings between the two papers. We include a larger number of brands in the analysis as we include the top five manufacturers over 10 months. They have camera data from the top two manufacturers for five months. Given the difference in the number of brands, consumer reviews may play a lesser role when the number of brands is small, primarily when the data are confined to well-known brands (Kostyra et al. 2016). In addition, there are differences in the selection of the independent variables in the analysis. While we condition our analysis to consumer review data from Amazon.com, they include a richer set of consumer review data from many online sources. Lastly, our estimated elasticity of average numeric rating falls within the range reported in the meta-analysis (You et al. 2015). Nonetheless, the difference between these two papers may support our view that we need more empirical research in differentiated products.

Although our regression analysis is informative, it has a key limitation. Our regression analysis ignores the competitive effects since the specification does not reflect product competition in the analysis. In addition, the regression analysis does not allow us to study the mechanism, such as how informative consumer reviews are and how shoppers use them to resolve product uncertainty (Zhao et al. 2013). Our learning model within the choice model addresses such limitations by explicitly modeling consumer learning behaviors.

5. Empirical analysis

5.1. Overview

The estimation of the proposed model overall follows the approach outlined in Kim (2019). The key innovation we introduce to the estimation in this paper is the additional estimation layer that corresponds to consumer learning. For computation of \bar{Q}_{ijt} in Eq (3) during estimation, we use Eqs

(11)–(13). Given the model parameters and (s_{jt}, n_{jt}) , we compute the mean and variance of \bar{Q}_{ijt} using Eq (12) and Eq (13). Then we repeatedly draw \bar{Q}_{ijt} using Eq (11) as i 's expected experience value in the simulation process. Once we simulate \bar{Q}_{ijt} , we can compute individual choice probability using Eq (2).

The dependent variables in our empirical analysis are sales rank. We convert daily sales rank data into a set of pairwise indicator variables. In our approach, the dependent variables of $\{I_{jkt}\}$ are defined as,

$$I_{jkt} = \begin{cases} 1 & \text{if } sr_{jt} > sr_{kt} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where sr_{jt} is j 's sales rank at t , $sr_{jt} > sr_{kt}$ means that j is more popular than k at t .

We link the predicted market share to the observed sales rank following the recipe in Kim (2019). The prediction of j 's market share at t , \hat{s}_{jt} , is the sum of the true, unobserved market share of s_{jt} and an error term,

$$\hat{s}_{jt} = s_{jt} - \varepsilon_{jt}, \quad (16)$$

where $\varepsilon_{jt} \sim N(0, \frac{\sigma_v^2}{2})$ is an i.i.d. random variable. Then, we express the probability of observing a pairwise rank inequality between j and k at t as,

$$\Pr(I_{jkt} = 1) = \Phi\left(\frac{\hat{s}_{jt}(\Theta, \Omega, \mathbf{X}) - \hat{s}_{kt}(\Theta, \Omega, \mathbf{X})}{\sigma_v}\right), \quad (17)$$

where $\Phi(\cdot)$ is CDF of standard normal distribution. Our likelihood function and the corresponding MLE is,

$$\{\Theta^*, \Omega^*, \sigma_v^*\} = \arg \max_{\{\Theta, \Omega, \sigma_v\}} \prod_t \prod_j \prod_{k \neq j} \Phi\left(\frac{\hat{s}_{jt}(\Theta, \Omega, \mathbf{X}) - \hat{s}_{kt}(\Theta, \Omega, \mathbf{X})}{\sigma_v}\right)$$

where $j, k = 1, \dots, J_t$ and $t = 1 \dots, T$.

We briefly discuss the identification of key model parameters. As the choice-based aggregate demand model serves as the backbone of the proposed model, our key consumer preference parameters are identified through the conventional mechanism documented in the literature (e.g., Berry et al. 1995). The mean consumer preference parameters are identified by the correlation between product attribute values and sales popularity across products and time. Consumer heterogeneity parameters are identified by the correlation between product attribute values and the difference between the

predicted and actual sales popularity (Berry et al. 1995). That is, given the candidate of mean consumer parameters, we can predict sales popularity. Consumer heterogeneity absorbs the difference between the predicted and actual popularity across products and time in the data. Among the consumer learning parameters, Φ_1 is identified by the correlation between consumer numeric ratings and sales popularity across products and time conditional on consumer preference parameters. Lastly, the correlation between the changes in sales popularity and the changes in the number of consumer reviews across products and time identifies the consumer review variance of σ_w^2 .

5.2. Results

We present the estimated model parameters in Table 2. The estimated brand coefficients are intuitive. Among the mean brand coefficients, Sony has the greatest value (0.78), followed by Panasonic. Among the media formats, Mini-DV has a higher mean consumer preference of -0.73 than other formats. In addition, an average shopper prefers higher pixel numbers and higher zoom capability. We report a large heterogeneity for brands and media formats, consistent with the industry report that these are important attributes for shoppers. Next, we find that shoppers would prefer to buy directly from Amazon.com than from third-party vendors. Next, we discuss consumer learning parameters associated with consumer reviews. The

Table 2. Estimated model parameters.

Variables	mean effect (s.e.)	Heterogeneity (s.e.)
Sony	0.78 (0.18)	1.77 (0.43)
Panasonic	-0.16 (0.08)	1.77 (0.43)
Canon	-0.31 (0.09)	1.77 (0.43)
JVC	-0.84 (0.21)	1.77 (0.43)
MiniDV	-0.73 (0.23)	2.12 (0.49)
DVD	-1.02 (0.23)	2.12 (0.49)
FM	-1.12 (0.27)	2.12 (0.49)
Compact	-7.25 (1.88)	5.29 (1.33)
Hi-Def	0.01 (0.25)	2.27 (0.51)
Zoom	0.02 (0.02)	0.03 (0.03)
Pixel	0.42 (0.13)	0.17 (0.11)
Log (Price)	-6.25 (1.47)	1.16 (0.33)
Seller (Amazon)	0.98 (0.26)	0.03 (0.01)
Seller (Partner)	0.78 (0.20)	0.03 (0.01)
σ_0	0.94 (0.22)	NA
Φ_0	-0.87 (0.21)	NA
Φ_1	0.36 (0.03)	NA
σ_w	1.04 (0.24)	NA
σ_v	0.01 (0.12)	NA
Product FE		Yes
Number of inequalities		147,965
Loglikelihood		-47,132

coefficient of Φ_1 is positive and significant, implying that a higher average numeric rating (i.e., more stars at Amazon.com) makes a positive contribution to consumer utility. This finding is consistent with our regression result in section 4.2. From the estimated heterogeneity parameters, we find that the magnitudes of σ_0 (for prior mean experiential value) and σ_w (for each review) are similar. This implies that each consumer review text contains a large degree of uncertainty on the products' experience values and hence one review text is not as informative as the initial uncertainty level. However, the uncertainty level decreases rapidly with the increasing n_{jt} because the variance term is inversely proportional to n_{jt} in Eq (13).

In the next section, we provide a set of simulation studies and quantify the effects of customer reviews on market share changes subject to product competition in the market. Our approach contrasts with past literature that mainly focused on the effects of customer reviews without considering product competition (e.g., Chevalier and Mayzlin 2006). Although such an approach may be tenable in experience goods categories such as books and movies, we believe product competition must be fully considered for a comprehensive analysis of demand in differentiated products category.

6. Consumer reviews and market outcome

6.1. The effects of provision of reviews for a focal product

Our first simulation study investigates how the market share for focal product j would change when the consumer reviews are turned off for all products except j . That is, we compute and compare j 's market shares for the following two cases.

- (a) Case A: consumer reviews are turned on for product j only
- (b) Case B: consumer reviews are turned off for all products

We then repeat the above comparison for all products $j = 1, \dots, J_v$ and compare the percentage-wise share change for each product. This simulation study allows us to study the effect of consumer reviews for product j in isolation while suppressing the review effect for the rest of the product. We can therefore assess the value of consumer reviews for each product. For the market share simulation, we draw 10,000 consumers from the joint distribution and predict their choice probabilities. We then aggregate them to compute market shares at each period. We repeat this process under the above two scenarios.

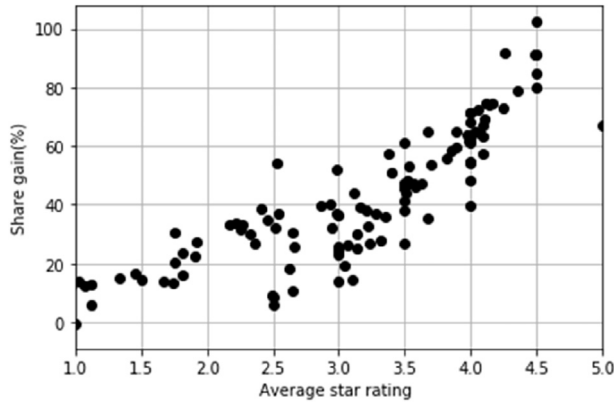


Fig. 3. The effect of consumer reviews in the first simulation study in which we keep the consumer reviews for the focal product while turning them off for the rest of the products. The graph shows that products with higher consumer ratings would experience additional market share gains when the platform shows the consumer ratings for the focal products while turning the consumer reviews off for the rest of the products.

Fig. 3 shows the simulation results. In the figure, each dot represents a product, the horizontal axis represents the average consumer rating (s_{jt}), and the vertical axis represents the simulated percentage-wise market share difference. The mean and standard deviation of share difference between the two cases are 39.7% and 24%, respectively. In this Figure, we note that the products with higher numeric ratings experience higher share improvement when reviews are exclusively available for them. In addition, the curve in the Figure shows a nonlinear pattern since the curve is relatively flat between 1 and 3 average star ratings but it becomes steeper after 3.5 average star ratings. This result implies that the provision of reviews further differentiates the focal product from the rest of the products, leading to a substantial impact on market share changes.

6.2. The effects of provision of consumer reviews on market shares

In this section, we quantify the aggregate effect of consumer reviews on the market outcomes. To that end, we compute the market shares for all products under the following two cases and compare their market share changes,

- (1) Case C: consumer reviews are turned on for all products across time
- (2) Case D: consumer reviews are turned off for all products across time

Case C is the current practice at Amazon.com, and case D is a hypothetical case in which consumer

reviews are unavailable for all products across time. After computing the market shares for all products under the two cases, we compare the percentage-wise share differences for all products across time. This simulation study is equivalent to quantifying the effects of the provision of customer reviews on market shares while we fully take product competition into account.

Fig. 4 visualizes the percentage-wise market share differences between two scenarios for all products across time. The positive share difference means that products gain market shares under consumer reviews and the negative difference indicates market share loss. This figure clearly shows that products with higher numeric ratings gain substantial, incremental market shares. In contrast, the aggregate effects from the number of reviews are quite muted compared to those from consumer ratings. For better visualization, Fig. 5 shows the two-dimensional view of Fig. 4.

In this figure, the horizontal axis represents the average consumer rating (s_{jt}) while the vertical axis represents the simulated percentage market share difference. In this graph, the dotted line represents the average percentage share difference at each level of numeric rating values. We note a few key points. First, the average percentage share difference across products and time is very substantial: the standard deviation of the percentage-wise share difference is 16.7% and the share changes range between -40% and 20% . This result implies that consumers resolve product uncertainty through consumer reviews and redirect their choices when consumer reviews are available. It also has a strong

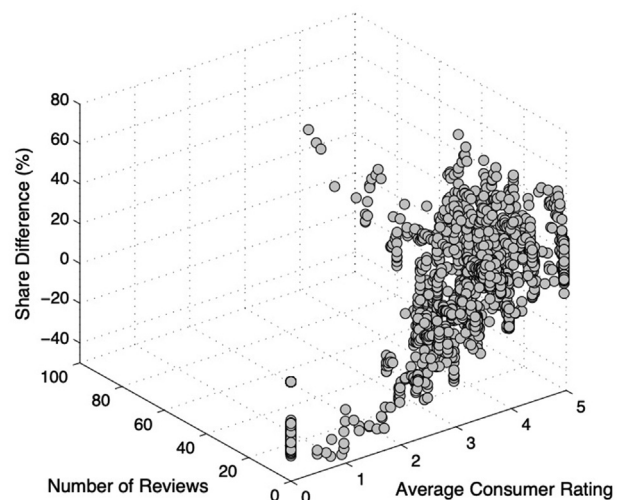


Fig. 4. 3D scatter plot of market share difference (%) with and without consumer reviews across products and time.

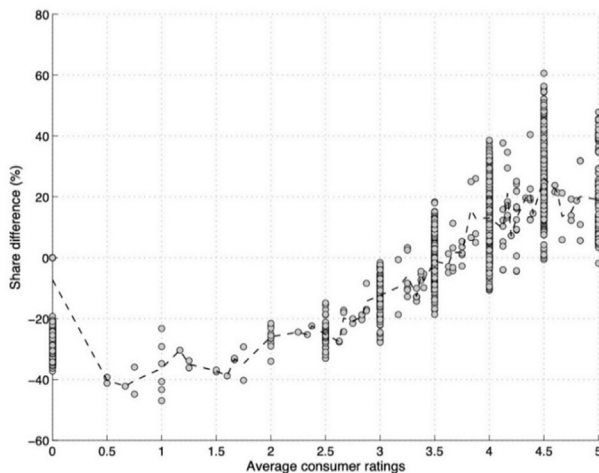


Fig. 5. 2D scatter plot of percentage market share differences with and without consumer reviews across products and time.

implication on consumer welfare, as predicted in Jiang and Chen (2007): consumer welfare may increase under the provision of consumer reviews. Next, products with average numeric ratings below 3.5 stars experience market share loss in the presence of consumer reviews. In contrast, products with numeric ratings above 3.5 stars are predicted to gain market shares. Therefore, shoppers may consider average ratings below the threshold of 3.5 as a negative signal. Our finding is consistent with a recent research that reports the importance of the negative word of mouth (WOM) on firm value (Jeon et al. 2020). In this figure, the negative signals have larger marginal effects since products with average ratings around 1 and 2 experience up to 40% share

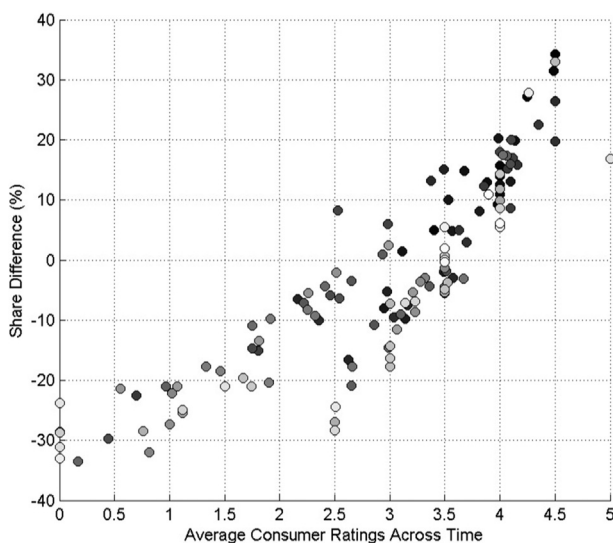


Fig. 6. Share difference for products with and without consumer reviews across time.

loss while products with four and above star ratings experience up to 20% share gain. Our simulation study provides much more detailed insights.

Last, Fig. 6 shows the percentage share difference (y-axis) with average numeric ratings (x-axis) across time for all products. Darker colors represent products with higher sales ranks (i.e., more popular ones). The correlation between sales rank and average consumer ratings is very high at 0.89 in the data. Combined with the results from the simulation study, we find that higher consumer ratings lead to higher market share gains. We came to this conclusion after fully controlling the product heterogeneity by introducing product fixed effects in the utility specification. Therefore, we infer that popular products generate positive reviews over time, which contributes to even higher sales in the future. In summary, consumer reviews serve as an amplification mechanism for popular products, reinforcing their popularities and polarizing the market outcome. Overall, both simulation studies inform us that the effects of consumer reviews on durable goods are substantial.

7. Conclusion

In this paper, we investigate and quantify the aggregate impacts of consumer reviews on the market outcome in differentiated products market. Our key premise in this paper is that differentiated products have both search and experience attributes and shoppers depend on consumer reviews to learn about and resolve product uncertainty during their choice. To that end, we develop and incorporate the Bayesian learning model into a choice-based aggregate demand model. By doing so, we aim to broaden our understanding of the role of consumer reviews in a differentiated product category.

From the simulation studies, we find that consumer numeric ratings of three stars or below negatively affect consumer utility and that these products would be better off without consumer reviews. After accounting for market competition, we also find that products with higher average star ratings gain market shares when online consumer reviews are available for all products. The standard deviation of percentage-wise share differences is 16.7%, and this considerable value implies a large impact of consumer reviews on consumer welfare. Lastly, we report that more popular products are likely to experience greater market share gains from favorable consumer reviews, leading to a polarized market structure.

We discuss a few limitations of our study. First, we did not consider the outside options in our model

and estimation due to the data limitation and technical challenges. We have sales rank data as our dependent variable and cannot observe or reliably estimate the size of outside goods. Second, we assumed that consumers choose from the universal set of products. However, recent advances in empirical models of consumer search (e.g., Moe 2006; Ursu 2018) may allow us to impose limited consideration sets on consumer choice. We also note that consumers' choice set may be affected by seller-initiated actions such as product recommendations. Third, we do not model forward-looking consumers. This may be important since the products in consumer electronics categories often exhibit price declines over time, leading to dynamics in consumer demand. Last, we do not explicitly address price endogeneity as in Berry et al. (1995) due to our empirical setting and data. Instead, we expect the product-specific intercepts in the utility specification to mitigate, if not altogether remove, the price endogeneity.

There are a few natural extensions for this research. First, developing a full learning model may be desirable by incorporating other information sources such as past sales rank or price. Given the lack of empirical research on the relationship between market share and quality, such a model may enhance our understanding of this topic. Second, our implicit assumption that shoppers consume most of the consumer reviews may not be tenable in practice. With more detailed individual-level data, it may be possible to model selective consumption of online consumer reviews among shoppers, leading to a more realistic model. Last, the consumer review effect may be moderated by product characteristics such as price. We leave these for future research.

Funding

This work was supported by the New Faculty Startup Fund from Seoul National University.

References

- Berry, Steven, Levinsohn James, and Ariel Pakes (1995), "Automobile prices in market equilibrium," *Econometrica*, 63 (4), 841–90.
- Chakravarty, Anindita, Yong Liu, and Tridib Mazumdar (2010), "The differential effects of online word-of-mouth and critics' reviews on pre-release movie evaluation," *Journal of Interactive Marketing*, 24 (3), 185–97.
- Chen, Lian, Kang Jun Choi, and Jae Young Lee (2019), "The effect of rating dispersion on purchase of experience goods based on the Korean movie box office data," *Asia Marketing Journal*, 21 (1), 1–27.
- Chen, Yubo and Jinhong Xie (2008), "Online consumer review: A new element of marketing communications mix," *Management Science*, 54 (3), 477–91.
- Chevalier, Judith A. and Dina Mayzlin (2006), "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, 43 (3), 345–54.
- Ching, Andrew T., Tülin Erdem, and Michael P. Keane (2013), "Learning models: An assessment of progress, challenges, and new developments," *Marketing Science*, 32 (6), 913–38.
- Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets," *Marketing Science*, 29 (5), 944–57.
- Cui, Geng, Hon-Kwong Lui, and Xiaoning Guo (2012), "The effect of online consumer reviews on new product sales," *International Journal of Electronic Commerce*, 17 (1), 39–58.
- Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do online reviews matter? - an empirical investigation of panel data," *Decision Support Systems*, 45, 1007–16.
- Erdem, Tulin and Michael P. Keane (1996), "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets," *Marketing Science*, 15 (1), 1–20.
- Gu, Bin, Jaehong Park, and Prabhudev Konana (2012), "Research note—the impact of external word-of-mouth sources on retailer sales of high-involvement products," *Information Systems Research*, 23 (1), 182–96.
- Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra (2009), "Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods," *Journal of Marketing*, 73 (2), 55–69.
- Jeon, Jaihyun, Byung-Do Kim, and Junhee Seok (2020), "Impact of negative word of mouth on firm value," *Asia Marketing Journal*, 22 (3), 1–28.
- Jiang, Bao-Jun and Pei-Yu Chen (2007), "An economic analysis of online product reviews and ratings," Available at: SSRN: <http://ssrn.com/abstract=1023302>.
- Kim, Jun B. (2019), "Competitive analysis among multi-product firms," *Asia Marketing Journal*, 21 (3), 47–64.
- Kostyra, Daniel S., et al. (2016), "Decomposing the effects of online customer reviews on brand, price, and product attributes," *International Journal of Research in Marketing*, 33 (1), 11–26.
- Li, Xinxin and Lorin M. Hitt (2008), "Self-selection and information role of online product reviews," *Information Systems Research*, 19 (4), 456–74.
- Liu, Yong (2006), "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of Marketing*, 70 (3), 74–89.
- Moe, Wendy W. (2006), "An empirical two-stage choice model with varying decision rules applied to internet clickstream data," *Journal of Marketing Research*, 43 (4), 680–92.
- Narayanan, Sridhar, Puneet Manchanda, and Pradeep K. Chintagunta (2005), "Temporal differences in the role of marketing communication in new product categories," *Journal of Marketing Research*, 42 (3), 278–90.
- Nelson, Phillip (1970), "Information and consumer behavior," *Journal of Political Economy*, 78 (2), 311–29.
- Nevo, Aviv (2000), "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of Economics & Management Strategy*, 9 (4), 513–48.
- Park, Sungsik, Woochoel Shin, and Jinhong Xie (2016), "The first-review effect: Interdependence between volume and valence of online consumer reviews," Available at: SSRN 2824846.
- Pathak, Bhavik, Robert Garfinkel, Ram D. Gopal, Rajkumar Venkatesan, and Fang Yin (2010), "Empirical analysis of the impact of recommender systems on sales," *Journal of Management Information Systems*, 27 (2), 159–88.
- Senecal, Sylvain and Jacques Nantel (2004), "The influence of online product recommendations on consumers' online choices," *Journal of Retailing*, 80 (2), 159–69.
- Ursu, Raluca M. (2018), "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 37 (4), 530–52.

- Wu, Chunhua, Hai Che, Tat Y. Chan, and Xianghua Lu (2015), "The economic value of online reviews," *Marketing Science*, 34 (5), 739–54.
- You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015), "A meta-analysis of electronic word-of-mouth elasticity," *Journal of Marketing*, 79 (2), 19–39.
- Yu, Xiaohui, Yang Liu, Xiangji Huang, and Aijun An (2012), "Mining online reviews for predicting sales performance: A case study in the movie domain," *IEEE Transactions on Knowledge and Data Engineering*, 24 (4), 720–34.
- Zhao, Yi, Sha Yang, Vishal Narayan, and Ying Zhao (2013), "Modeling consumer learning from online product reviews," *Marketing Science*, 32 (1), 153–69.
- Zhu, Feng and Xiaoquan Zhang (2010), "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," *Journal of Marketing*, 74 (2), 133–48.