NET Institute*

www.NETinst.org

Working Paper #14-03

September 2014

**Measuring Product Type with Dynamics of Online Product Review Variances:
A Theoretical Model and the Empirical Applications**

Yili Hong and Pei-yu Chen

W. P. Carey School of Business, Arizona State University

Lorin M. Hitt

Wharton School, University of Pennsylvania

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

Measuring Product Type with Dynamics of Online Product Review Variances:
A Theoretical Model and the Empirical Applications

Yili Hong^{*}, Pei-yu Chen^{} and Lorin M. Hitt^{***}**

^{*} Yili Hong (hong@asu.edu) is an assistant professor at W. P. Carey School of Business, Arizona State University

^{**}Pei-yu Chen (peiyu.chen@asu.edu) is an associate professor at W. P. Carey School of Business, Arizona State University

^{***}Lorin M. Hitt (lhitt@wharton.upenn.edu) is a Professor at the Wharton School, University of Pennsylvania

Measuring Product Type with Dynamics of Online Product Review Variances:

A Theoretical Model and the Empirical Applications¹

Abstract

A significant body of literature in information systems, marketing, and economics has shown the important implication of the distinction between experience products and search products (“product type”) on consumer information search, marketplace design, and firm strategy. However, how to empirically measure product types remains a challenge, and this challenge is further complicated by the growth of online commerce and the increasing availability of online reviews that have transformed the nature of many products and altered the traditional perception of these products.

The objective of this research is to propose an online product review-based measure that could accurately reflect consumers’ perception of a product, as search or experience dominated product. Based on the definitions of search and experience products — whether information can be easily transferred or not — we propose a data-driven method that can be used to infer product type from statistical analyses of online product reviews. Our theoretical analyses indicate that the variance of the ratings should decrease as more consumers rate a pure search product; for experience products however, the variance of the ratings may remain constant or increase depending on the importance of the experience attributes in determining consumer utility. We demonstrate the empirical applications of this approach at the category, product, and attribute levels using product reviews data from Amazon.com, Yelp.com, and Ctrip.com, respectively. In addition, a user study conducted on Amazon Mechanical Turk shows our review-based measure to outperform Nelson’s (1970) product classification, which historically has been the standard in determining product type. Overall, this new measure provides an easy to implement, less subjective and more accurate measure of product type. Therefore, researchers and practitioners can use this measure to better understand how consumers perceive products and to design strategies accordingly.

Keywords: product type, online product reviews, user-generated content, data-driven approach

¹ We gratefully acknowledge financial support from the NET Institute, www.NETinst.org.

Measuring Product Type with Dynamics of Online Product Review Variances: A Theoretical Model and the Empirical Applications

1. Introduction

The Internet has provided a convenient way for consumers to share their product and service consumption experiences through various online platforms. Over time, this reservoir of consumer evaluations has built a sizable base of information (known as online product reviews) and has become a dominant source of information for consumers in relation to purchasing and choosing from a consideration set. According to *Dimensional Research*, 90% of purchasing decisions are influenced by online reviews². These large-scale user-generated data in the form of online product reviews allow researchers and practitioners to obtain a deeper understanding of consumer perceptions and purchasing behaviors. Moreover, these data provide unprecedented opportunities to revisit classical theories and methods, which were developed under traditional constraints and assumptions that may no longer hold in the information age.

Scholars have utilized the wealth of online product review data to study the effects of online word of mouth (WOM) on firms' strategies (Chen and Xie 2008; Kwark et al. 2014) and consumers' purchasing decisions (Archak et al. 2011; Dellarocas 2003; Godes and Silva 2012; Li and Hitt 2010; Sun 2012; Wang et al. 2012). Online product review data also help researchers understand consumer perceptions and the way that these perceptions are communicated. In the offline world, consumer behavior data are difficult or impossible to observe on a large scale. This characteristic has led to the development of a substantial literature that connects measurable attributes of online reviews such as valence and volume (Chevalier and Mayzlin 2006; Dellarocas et al. 2007; Duan et al. 2008; Zhu and Zhang 2010), textual content (Archak et al. 2011; Ghose and Ipeirotis 2010; Pavlou and Dimoka 2006), and variance (Clemons et al. 2006; Kwark et al. 2014; Sun 2012) to economic outcomes such as consumer decisions, product market

²<http://www.reviewpro.com/survey-zendesk-mashable-dimensional-research-90-say-positive-reviews-impact-purchase-decisions-26016>

share, and sales growth. More recently, researchers have used this information to better understand the characteristics of products purchased online (Ghose and Ipeirotis 2010).

We deviate from the orientation of these prior studies because our interest is in the implication of this wealth of data on classical theories or methods and the use of publicly available data to refine and advance these theories and methods. In this study, we develop a *review-based, data-driven method that can be used to identify product type*. The development is premised on the idea that the dynamics of online reviews reveal information on the perception of consumers of a product: as a “search” product or as an “experience” product.

Nelson (1970) was the first to propose the notion of search versus experience products based on a search model (Stigler 1961). Prior to Nelson (1970), search models focused on the “unit” of search cost. Nelson brought a new lens, *product type*, to understand consumers’ search behavior based on the extent with which “search” helps in informing consumers about the utility of the product. Nelson (1981) generalized the dichotomous distinction and suggested that products can be described by both search and experience attributes. Search attributes are characteristics that can be unambiguously described and are therefore amenable to identification through consumer search. Experience attributes are characteristics that cannot be effectively ascertained without the actual use of a product by a consumer. Moreover, different consumers may perceive the same experience attributes differently because of idiosyncratic consumer preferences. An example offered by Nelson was “canned tuna fish”. Even with all the detailed descriptions provided on the label, consumers would not know how much they would like the product without tasting it. A product can be classified as a search product according to the extent that search attributes dominate consumer choice, whereas it can be categorized as an experience product when experience attributes drive choice. We will refer to this distinction throughout the paper as “*Product Type*”.

ESTIMATES OF THE RATIO OF REPAIR EXPENDITURES TO SALES BY GOOD, 1963

Goods	Nonmerchandise Receipts*	Total Repair†
Experience goods:		
Jewelry	12.6	17.3
Typewriters	8.8	...
Radio, television	8.1	24.3‡
Tire, battery	7.5	...
Aircraft, boats, motorcycles	7.0	...
Heating and plumbing	6.4	...
Bicycles	6.0	...
Automobiles	5.7	30.0§
Music instruments	4.9	10.7
Appliances	4.4	14.0
Search goods:		
Floor covering	3.8	...
Garden implements	3.5	...
Sporting goods	2.8	...
Household trailers	2.5	...
Cameras	2.4	...
Furniture	2.1	7.6
Paint and mirrors	1.9	...
China, glassware	1.9	...
Hardware	1.6	...
Hobbies, games	0.8	...

* The percentage of nonmerchandise receipts to sales of stores specializing in the sale of the specified good.
† To the percentage of nonmerchandise receipts, we add the ratio of the receipts of service specialists to the sales of the good for those goods whose service specialists are specified in the U.S. Bureau of the Census (1966b).
‡ The U.S. Commerce Department estimates this percentage as roughly 29% (U.S. Department of Commerce 1964).
§ Commerce estimates for this percentage is roughly 23%.
|| Furniture stores have a percentage of nonmerchandise receipts of 2.7. But they also sell television and appliances. The tabulated figure adjusts for the proportion of sales in the three categories.

Figure 1. Original Classification by Nelson (1970, page 319, JPE)

This distinction between search and experience products makes logical sense and is shown to be a very important concept in understanding the consumers' information search. After Nelson proposed the concept of "product type" (1970), scholars have studied and linked *product type* to different forms of market competition (Kwark et al. 2014; Shapiro 1982; Villas-Boas 2004; Villas-Boas 2006), differences in marketing strategy (Hsieh et al. 2005; McWilliams and Siegel 2001; Park and Lee 2009), and consumer behavior (Huang et al. 2009; Klein 1998; Mudambi and Schuff 2010; Senecal and Nantel 2004; Weathers et al. 2007; Zhu and Zhang 2010). As an example, this stream of literature has shown that consumers tend to be more loyal and less price elastic for experience products, and demonstrated that advertising for search products should focus on characteristics. By contrast, advertising for experience products should focus on the "brand". Overall, the extensive literature on product type has suggested that product type is a key factor in understanding consumer search behaviors and in designing strategies.

The theoretical importance of "product type" is widely recognized; however, the measurement of this concept has been stalled. Notably, Nelson's original work (1970) proposed a measure using repair expenditure data because repair expenditure is a quality that is quite difficult to determine prior to purchase. Figure 1 shows this classification scheme as applied to a variety of products. Given the

significant attention on “product type” as an important construct, its measurement has not advanced. For example, researchers have used product price (Smith 1990), or surveys of consumer perception (Weathers et al. 2007) to measure product type. In most cases, researchers continue to use the original classification proposed by Nelson (Huang et al. 2009; Mudambi and Schuff 2010; Senecal and Nantel 2004) whose data is from the 1960s. The classification is used despite the recognition that representing product type on a spectrum from pure experience to pure search is more appropriate than using a dichotomous classification (Caves and Williamson 1985). In addition, the Internet, which significantly changes the way consumers acquire and process information, may have inadvertently changed the nature of some products, which raises concerns on the original classification made by Nelson. Therefore, although the theoretical foundation of the distinction between search and experience products is still relevant, the empirical development of a new measure is strongly needed to reflect better the changing nature of product type as perceived by consumers.

In this study, we show that the wealth of online product review data can be leveraged to infer consumer perception on products. The fundamental idea is that consumers process information differently when contending with either search or experience products. Based on this observation, we develop an alternative and continuous measure of product type as perceived by consumers according to the statistical properties of longitudinal data from online reviews. This data-driven measure has the capability to reflect the actual perception of consumers on a product and can accommodate the changing nature of search, which is especially relevant in online markets. These are advantageous properties that a subjective classification of search versus experience products (Weathers et al., 2007) or the repair expenditure-based measure proposed by Nelson (1970) does not possess.

Our measurement is motivated by a theoretical framework that is based on the research of the interactions among consumer reviews, consumer preferences, and product choice (Godes and Silva 2012; Hu et al. 2009; Kwark et al. 2014; Li and Hitt 2010; Sun 2012). In our theoretical framework, consumers make rational product choices based on information (e.g., existing product reviews), and report their post

consumption satisfaction in the form of product ratings³. This data generation process, where a consumer processes an *input* (i.e., existing product reviews) and produces an *output* (rating of the product based on his/her consumption experience), reflects how efficient product review information is being transferred among consumers. In particular, product reviews that are informative and indicative of product quality can lead a consumer to expect less uncertainty in their consumption utility, and therefore less “surprise” in his or her rating. Conversely, when product reviews do not inform quality or utility, consumers are more likely to disagree with prior ratings. To the extent that this data generation process reasonably characterizes consumer search and review behavior, the statistical properties of reviews will reflect the resolution of product uncertainty over time: reviews for products that are primarily chosen based on search attributes⁴ should exhibit less variation over time because of the process of averaging out random variation. The same situation is not applicable to products that are primarily composed of experience attributes, where the variation in reviews may be driven by systematic differences among consumers⁵ and is therefore less likely to be resolved. Text comments can also have an effect, and this effect, to some extent, can still be absorbed by numerical ratings. For example, when text comments are informative, then consumers who refer to these informative text comments will have less consumption uncertainty and less surprise, which will result in less variation in ratings over time. Thus, this data-driven method, which is based on the relationship between review variance and the sequence of reviews, provides an estimate of information transfer efficiency over time and offers insights into the relative importance of search versus experience attributes.

³ For instance, we do not consider review manipulation (e.g., Dellarocas 2006) in our theoretical framework. Although manipulation of reviews is likely present, it plays less important role in our analysis because we focus on deriving statistical inferences from large numbers of reviews where the fluctuations from these effects are at least partially averaged out. Empirically, we take potential review manipulation and self-selection influences on early reviews into consideration through robustness checks (such as dropping the first five and 10 reviews and repeating the analysis).

⁴ For example, a digital camera’s search attributes include price, maximum aperture, and sensor size; its experience attributes include durability, sharpness, and bokeh.

⁵ Other factors may also impact information transfer; for example, when consumers believe that significant fraud ratings exist, product reviews’ information transfer efficiency will be reduced, which will make products less search-oriented.

A clear advantage of this data-driven method is the continuous measurement that it provides and the applicability to category, product/merchant, and attribute levels of analysis. Category level classification is useful in the study of the role of product type or the effect of the inclusion of product type as a moderating variable. Product/merchant level classification is useful when there are significant variations in product characteristics at the product/merchant level (e.g., fine restaurants may be perceived more as experience products, whereas fast food chains are perceived as search products). Product/merchant level classification therefore allows manufacturers/merchants to understand better the perception of customers on their products and to refine/design strategies accordingly. Given that products often consist of both search and experience attributes, attribute level classification is useful in understanding whether each product attribute is perceived as search or experience. As mentioned previously, understanding the product type has strategic implications for marketing and advertisement strategies. We demonstrate the empirical application of the data-driven method proposed in this study using data sets from *Amazon* (retail product categories), *Yelp* (restaurants), and *Ctrip* (features/attributes of travel products). We validate this framework with user studies that compares the classifications derived from this method against the approach developed by Nelson. Overall, this review-based, data-driven measure is shown to be theoretically grounded, easy to implement, and more accurate in reflecting consumer perception on products than existing methods.

In Sections 2 to 4, we review the literature pertaining to product type and online product reviews, describe the theoretical model, discuss the empirical applications by detailing our data sets and econometric specifications and presenting the estimation results, and compare the relative performance of this review-based measure by reporting user studies conducted with *Amazon Mechanical Turk* (AMT). In Section 5, we discuss the implications of our findings and the limitations of this study as well as suggest a number of potentially fruitful avenues for future research.

2. Related Literature

2.1. Measuring Product Type

The concept of product type originates from search theory. Search theory forms the basis of the literature on product differentiation and price dispersion, as consumers need to search for information regarding the differences in products and prices to decide on their purchase. Search cost has been the unit of analysis in search theory. Nelson (1970) brought a new lens called product type to distinguish different ways in which consumers acquire information: from direct experience or from the search on product attributes (Nelson 1970, p. 317). When search costs are relatively high, consumers will rely more heavily on experience; otherwise, they will extensively search for product information. The majority of contemporary empirical studies that involve the construct of *product type* rely on this argument and the original classification of Nelson (1970). This reliance is observed despite the fact that studies using this approach have noted that the Internet is likely to change the traditional relationship between search and experience products (Huang et al. 2009) and that researchers may disagree on categorizations using this approach for products outside of Nelson's original list (Mudambi and Schuff 2010).

To avoid this criticism, researchers often used products that are seemingly located on the extreme ends of the search–experience scale. Senecal and Nantel (2004) contrasted calculators (an obvious search product) with wine (an obvious experience product) in studying the effect of online product recommendations on consumer choice. Alternative approaches have been employed such as surveys (Weathers et al. 2007), through which consumers were asked whether they believe that they have to use a product to make an informed decision. Similarly, product prices (Laband 1991) were used as a measure because higher-priced products have a greater search return. Lastly, although researchers have suggested the move from a dichotomous to a continuous search–experience classification (Klein 1998; Nelson 1981), for practical consideration, empirical implementation remains a challenge, and a significant number of the empirical studies have still relied on a dichotomous distinction between “search products” and “experience products”

using the original Nelson classification. Overall, the state of this stream of literature highlights the considerable value of developing an easily implementable continuous measure of *product type*.

2.2. Online Product Reviews

Online product reviews contain information that affects purchasing behavior. A substantial number of studies that link product reviews to market outcomes, such as product demand, for a variety of products sold online, have emerged. Chevalier and Mayzlin (2006) found a significant effect of review valence on book sales, whereas Liu (2006) found that review volume is a better predictor of future sales than review valence. Other studies have suggested that variance of ratings affect the sales of some products, such as books (Sun 2012) and beer (Clemons et al. 2006). Further research has indicated that the identity of the reviewer provides additional information because identity moderates the effect of product reviews on sales (Forman et al. 2008). Taken collectively, these findings prove that consumers find information in product reviews to be valuable in making product decisions, and that various statistics from online reviews such as mean, variance, and volume provide useful information about product quality and consumer preferences.

A related stream of literature examines the extent to which information in reviews may be biased and may not reflect true consumer utility. Dellarocas (2006) analyzed the potential of forum manipulation (firms pay reviewers to influence ratings), and concluded that, in equilibrium, this type of manipulation has a limited effect on review accuracy. An alternative argument for the inability of reviews to reflect utility is based on consumer heterogeneity, which states that consumers may be truthful, but not reflective of the “average” because reviews reflect both product attributes and consumer tastes (Li and Hitt 2008) or are shaped by the interaction with other reviewers (Hu et al. 2009; Moe and Schweidel 2012; Muchnik et al. 2013; Wang et al. 2009). These biases may be partially resolved by revealing the metadata on reviews such as those that pertain to the evaluation of reviewer characteristics (Forman et al. 2008) or by extracting the subjective attributes from review text (Ghose and Ipeirotsis 2010). We adopt the perspective that reviews are shaped by idiosyncratic preferences and avoid the potential bias-related issues by

focusing on longer-run ratings that are less subject to bias because a single review has limited influence on the overall effectiveness of reviews (Li and Hitt 2008).

2.3. Product Quality

A significant body of work on establishing a rigorous definition of product quality exists. Although scholars have offered different perspectives on the exact dimensions that define quality, the existence of objective (common component) and subjective (consumer-specific component) components of quality has been established⁶. This distinction is most clearly articulated by Garvin (1984), who proposed a separation of quality assessment into an objective “product-based approach” and a subjective “user-based approach” of quality. This view is consistent with that of the economic literature on product differentiation. Specifically, vertical differentiation is concerned with the gradation in quality that is perceived by all consumers, whereas horizontal differentiation focuses on consumer–product fit [e.g., Tirole (1988), which offers a textbook treatment of the standard models]. However, the literature has also recognized that as a result of the idiosyncratic consumer preference component, developing an objective quality measure that is independent of consumer preferences can be difficult (Rosen 1974; Theil 1971). Thus, instead of separating objective and subjective qualities, we take a utility-based view. Higher quality, irrespective of objective or subjective, leads to higher utility. Further, the relative weight of objective quality vs. subjective quality will be revealed by the heterogeneity of reviews among consumers because subjective quality is more difficult to transfer among consumers than objective quality. Therefore, we recognize the existence of different aspects of quality information contained in product reviews.

⁶ Tuchman (1980) suggests that quality is perceived excellence (“...a condition of excellence implying fine quality from poor quality...Quality is achieving or reaching for the highest standard as against being satisfied with the sloppy or fraudulent” (p. 38). Others have taken the more utilitarian approaches that are not objective such as “fitness for use” (Juran 1999; p. 242) and the argument that “the quality of a product depends on how well it fits patterns of consumer preferences” (Keuhn and Day 1962; p. 101).

3. Theoretical Model

3.1. Overview

Consistent with the literature on product differentiation (Nelson 1981; Tirole 1988), we adopt the perspective that a product is composed of a set of attributes. We divide these attributes into search attributes and experience attributes. Based on the literature (Nelson 1981), information for search attributes (e.g., color, size, outlook, general features, and use) can be unambiguously communicated to consumers, and therefore, this information helps educate consumers about the consumption utility that they can expect to receive from the product. By contrast, experience attributes (e.g., smell, taste, fit, special functionality, etc.) are more difficult to communicate because different consumers may perceive these attributes differently, and the information may not inform consumers of their expected consumption utility. A product usually comprises both search and experience attributes and is positioned on the continuum scale of a pure search product and a pure experience product.

We view online reviews as a reflection of the combination of objective reporting of quality related to search attributes (with some random noise) and a subjective component that highlights consumers' evaluation of the realized utility for experience attributes. Over time, information on these two components will evolve in different ways as more reviews accumulate. For search attributes, information can be easily communicated; thus, an increasing number of reviews will enhance the precision of the estimation of product quality as random noise is averaged out. The process helps in resolving the uncertainty of customers on consumption utility. Utility for experience attributes will not obtain the same properties of search attributes with regard to the increasing number of reviews for two reasons. First, experience attributes are difficult to describe through online reviews (e.g., the taste of a pumpkin soup offered by a particular restaurant). Second, due to the systematic differences in consumer preferences, the variation is less likely to be independent across consumers who have purchased a product. Additionally, more reviewers may indicate more heterogeneity in consumer preferences, which further increases the difficulty of transferring the information on experience attributes. Consequently, the ratings of a product

that primarily consists of search attributes are more likely to converge to an estimate of its true quality, whereas the ratings of goods that are largely composed of experience attributes may not exhibit such a converging behavior. In other words, the change in the variance of ratings *over time*, as more reviews are submitted, provides an estimate of information transfer efficiency and offers insights into consumer perception on products (whether or not information has helped them determine their utility). Therefore, this property can be used to infer product type. We formalize this data-driven approach in Section 3.2.

Figure 1. The Model: Identification Strategy

3.2. The Model

A product (indexed by j) is a bundle of search attributes and/or experience attributes. As noted earlier, consumers can form a reasonable expectation about the utility of search attributes by obtaining more second hand information (without direct usage), such as color, resolution, functionality, speed and quantity, among others. In other words, more information can lower the uncertainty regarding product utility. By contrast, experience attributes (e.g., taste) are difficult to ascertain through second hand information, and these usually provide idiosyncratic utilities to different consumers (Nelson 1981). Therefore, consumers have to experience the product directly to be certain of its suitability with their preferences (Hong and Pavlou 2014). Essentially, our identification between search attributes and experience attributes hinges on whether more review information helps reduce uncertainty related to consumption utility.

We build upon the theoretical literature of product type, and we show that the dynamics of the longitudinal ratings for search and experience products are systematically different. This difference forms a basis for the empirical inference of product type. In other words, we do not intend to specify a product or its attributes, but analyze user-generated data on product rating using this model to inform us about the perception of consumers on each product. If a product is perceived as a search product, certain rating dynamics (convergence) should be observed; whereas if a product is perceived as an experience product, different rating dynamics (non-convergence) should be observed. We set up our model as follows:

A consumer (indexed by i) derives utility from consuming a product (j). Without any consumer reviews, the utility of product j received by consumer i can be formulated as follows:

$$U_{ij} = w_j q_{ij}^s + (1 - w_j) q_{ij}^e + \varepsilon_{ij} - p_j; q_{ij}^s \sim N(q_j, \sigma_s^2), q_{ij}^e \sim N(0, \sigma_e^2), \varepsilon_{ij} \sim N(0, \sigma_f^2). [1]$$

In Equation [1], q_{ij}^s is the utility (or quality) that consumer i obtains from the search attributes of product j , which is independently and normally distributed with mean q_j and variance σ_s^2 . Essentially, we can consider q_j as the true objective quality of product j , and therefore σ_s^2 reflects the random error around the true quality of product j . σ_s^2 may be caused by imperfect distribution, imperfect manufacturing process, or other unknown random factors that lead to the random deviation in the quality (e.g., a consumer may receive a defective or inferior quality product that falls short of product descriptions).

q_{ij}^e captures the utility (or quality) associated with consumers' tastes, preferences, or fit on the experience attributes of product j , which can be positive (fit) or negative (misfit) in relation to the objective quality.

In other words, a product that fits consumer preferences will result in an additional utility beyond the objective quality, whereas a misfit one will reduce the utility. The consumer faces uncertainty regarding his or her own fit with the product, and this uncertainty (the extent to which consumers suffer from misfit) may differ from one consumer to another. Therefore, σ_e^2 can also be dependent on the consumer.

ε_{ij} captures random individual-level noise, such as in the case where a consumer may have a bad mood

and therefore experience a lower utility regardless of quality. We assume that this variable is independently and identically distributed.

In order to examine theoretically how each of these three components of utilities affects the sequential dynamics of longitudinal ratings, we assume that the three components of utilities are independent of each other. Note however, complete independence is not required for empirical analyses. Although high correlation among these components will make it difficult to distinguish different product types empirically, absolute independence is not required in most regression analyses to obtain a consistent estimation, as long as the correlation is not strong enough to make multi-collinearity a serious issue.

Furthermore, w_j is the weight of search attributes for product j in determining the utility of consumers; therefore, $(1-w_j)$ is the weight of experience attributes for product j , $w_j \in [0,1]$. For a pure search product, $w_j = 1$; whereas for a pure experience product, $w_j = 0$. Lastly, p_j is the price for product j .

Based on this model set-up, we discuss how online product reviews affect consumers' utility level.

3.2.1. Pure Search products ($w_j=1$)

First, we focus on the case of pure search products. The information on the product attributes of search products, by definition, is transferrable. Therefore, we can consider a consumer's rating as a data point of the realized quality of the product, and thus, the set of reviews become a sample from which consumers can infer the true quality q of the product. The average value of the ratings, that is, the sample mean, provides a point estimate of the true quality, and consumers form confidence (or uncertainty) levels around the sample mean that reflect the true quality. The law of large numbers and the central limit theorem state that the *variance* of the sample mean decreases as sample size increases. For a given product j , the variance for each data point can be obtained by the following:

$$Var(q_{ij}) = \sigma_{sj}^2 \cdot [2]$$

The variance of the sample mean of n ratings will therefore become $Var(\bar{q}_j) = \frac{\sigma_{sj}^2}{n}$ [3], which measures

the approximate distance of the sample mean (the average rating) from the true utility q_j , from which one can form the confidence interval of average rating as an estimate of true quality.

Therefore, consumers become more confident that the sample average (average rating) is the true quality when more reviews are available. In other words, consumer uncertainty decreases as the number of information through which consumers can draw upon increases. When n reviews are available, consumers

form the expectation of quality: $E(q_{ij}^s) = \bar{q}_{ij}^s \sim N\left(q_j, \frac{\sigma_{sj}^2}{n}\right)$.

Assume that consumers' ratings reveal their realized utility from consuming the products, where realized utility can be described as the expected utility conditional on available reviews plus deviation around the expected utility (i.e., individual random noise). Then conditional on having n prior reviews, the rating of consumer i can be written as follows:

$$r_{ij} = U_{ij}|n = \bar{q}_{ij}^s + \varepsilon_{ij} - p_j; \quad \bar{q}_{ij}^s \sim N\left(q_j, \frac{\sigma_{sj}^2}{n}\right), \quad \varepsilon_{ij} \sim N(0, \sigma_j^2). \quad [4]$$

Based on the abovementioned equation, the variance of the n^{th} rating for product j can be constructed as follows:

$$Var(r_{nj}) = \frac{\sigma_{sj}^2}{n} + \sigma_j^2. \quad [5]$$

Equation [5] shows that the variance of the ratings of search products evidently decreases as the number of ratings increases.

3.2.2. Pure Experience products ($w_j=0$)

In the case of pure experience products, we use a set up similar to that in Section 3.2.1, where consumer reports a rating based on realized utility, which is the expected utility conditional on the number of prior reviews and the individual-level random noise. Specifically, for product j , conditional on n prior reviews, the rating of consumer i of product j can be written as:

$$r_{ij} = U_{ij}|n = q_{ij}^e + \varepsilon_{ij} - p_j ; q_{ij}^e \sim N(0, \sigma_{e_{ij}}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_j^2). \quad [6]$$

$$Var(r_{ij}) = (\sigma_{e_{ij}}^2 + \sigma_j^2). \quad [7]$$

A consumer's uncertainty with regard to fit will not necessarily be reduced because of prior ratings because quality information on experience attributes may not be transferrable among consumers. For example, the fact that consumer A likes the taste of brand X canned tuna does not mean that consumer B will enjoy it as much as consumer A can exemplify the idea. Statistically, consumers' tastes and preferences are endowed and not likely change even if many consumers like the product. Therefore, the law of large numbers will not be applicable, and as a result, uncertainty is not necessarily reduced.

As previously noted, uncertainty about the fit (or the sensitivity of consumer utility to fit/misfit) can be consumer dependent. When every consumer has the same degree of uncertainty or sensitivity to fit/misfit, that is, when $\sigma_{e_{ij}}^2$ is independently drawn from the same distribution with variance $\sigma_{e_j}^2$, we can further simplify Equation [7] as:

$$Var(r_{ij}) = (\sigma_{e_j}^2 + \sigma_j^2). \quad [8]$$

Equation [8] suggests that the sample variance of the ratings of pure experience products will *not* decrease as the number of ratings increases. Comparison of Equation [5] and [8] brings us to our first proposition:

Proposition 1. If the variance (or standard deviation) of the ratings does not decrease with the increase in the number of reviews, then a product has at least one experience attribute.

3.2.3. Mixed Attributes with Risk Averse Consumers

Although it is reasonable to assume that $\sigma_{e_{ij}}^2$ is independently drawn from the same distribution, an implicit assumption behind Equation [8] is that the time of purchase (and giving reviews) is independent of $\sigma_{e_{ij}}^2$. The assumption may be invalid based on the evidence from the literature that proves the dependence of time of purchase on $\sigma_{e_{ij}}^2$. Specifically, the literature has suggested the existence of a self-

selection bias in experience products because early reviewers may behave systematically different from late reviewers (Li and Hitt 2008). With consumer risk aversion, one possible pattern is that those with smaller uncertainties (i.e., lower $\sigma_{e\ ij}^2$) may enter the market earlier than those with higher uncertainties. For example, some consumers may have a better idea about a clothing brand (e.g., fabric, size) or an author's books (e.g., writing style). Therefore, these consumers face lower uncertainty when they purchase clothes with this brand or a book from the author, and these consumers tend to be early buyers of these products. Similarly, product diffusion literature on information systems (IS) and marketing also supports the view that early adopters tend to be those who are more familiar with the product and/or those who are less risk-averse (i.e., less sensitive to fit/misfit). In this case, these patterns will suggest that $\sigma_{e\ ij}^2$ can increase over time along the review sequence. In other words, the variance of ratings for experience products will likely increase as the number of reviews increases.

Based on the abovementioned discussions, we derive the following relationships for general products that possess a combination of search and experience attributes:

$$r_{ij} = U_{ij}|n = w\bar{q}_{ij}^s + (1 - w_j)q_{ij}^e + \varepsilon_{ij} - p_j; \quad \bar{q}_{ij}^s \sim N\left(q, \frac{\sigma_{sj}^2}{n}\right), \quad q_{ij}^e \sim N\left(0, \sigma_{e\ ij}^2\right), \quad \varepsilon_{ij} \sim N(0, \sigma_j^2) \quad [9]$$

$$Var(r_{nj}) = \left(\frac{w_j^2 \sigma_{sj}^2}{n} + (1 - w_j)^2 \sigma_{e\ nj}^2 + \sigma_j^2\right). \quad [10]$$

In Equation [10], we assume $\sigma_{e\ nj}^2$ weakly increases with n , to be consistent with the practical observation that later consumers have higher uncertainty than early consumers.

For comparison purposes and to gain deeper insights, we adopt a general function for $\sigma_{e\ nj}^2$, $\sigma_{e\ j}^2(1 + n^a)$, which increases in number of reviews n , where $\sigma_{e\ j}^2$ can be considered as the base level variance related to the product fit uncertainty on the experience attributes for product j , and a is a parameter influencing how the fit uncertainty (variances) may increase with the number of consumers over time to accommodate the fact that as time goes by, more heterogeneous consumers enter the market. With this general functional

form, we derive the following proposition:

Proposition 2: *$\text{Var}(r_{nj})$ increases as $\frac{w_j}{1-w_j}$ decreases and as $\frac{\sigma_{sj}^2}{\sigma_{ej}^2}$ decreases.*

This proposition suggests that, as experience attributes become more important (i.e., $1 - w_j$ increases) to the determination of utility, or as the general uncertainty related to experience attributes is higher compared with that of search attributes, the variance of ratings is likely to increase as the number of reviews increases.

Therefore, the contrapositive of Proposition 2 provides a foundation for inferring the degree to which experience attributes dominates product utility, that is: ***as the number of reviews increases, the lesser the increase in the variance (or standard deviation) of ratings, the more likely that search attributes dominate this product.***

4. Empirical Applications

In the previous section, we provided a theoretical foundation to show that a data-driven method based on the sequential dynamics of longitudinal reviews can be used to distinguish product types. In this section, we demonstrate the application of this data-driven method to classify products empirically at three different levels: category (e.g., clothing, computers), product/merchant (e.g., Red Lobster Restaurant in Philadelphia), and attribute (e.g., location, ease of use, taste). Category level classification is useful when researchers study product type or include product type as a moderating variable. Product/merchant level classification is useful when a significant variation exists in the product characteristics at the product/merchant level (e.g., a fine dining French restaurant may be perceived to be more experience-dominated, while a fast food chain restaurant may be perceived to be more search-dominated). Given that products are defined as a set of attributes, attribute level classification is useful in understanding the perception on each product attribute as either search-based or experience-based. We utilize the theoretical

framework described in the previous section to identify product type in these three levels. As a performance comparison, we rely on user studies to compare our classification with the original classification of Nelson (1970).

4.1. Data

Two types of data were collected for this study: archival data from three websites and primary data from a user study on AMT. In this part, we discuss the collection of archival website data. We present the collection of data and the results of the study on AMT in Section 4.6.

Using a combination of automated web crawlers and manual data crosscheck, we collected archival website data from three sources: *Amazon*, *Yelp*, and *Ctrip*. We utilized the data from *Amazon* for the category-level application, *Yelp* for the service/merchant-level application, and *Ctrip* for attribute-level application of travel products. Data from *Amazon* cover the period from July 2004 to February 2012, data from *Yelp* cover the period from October 2006 to October 2011, and data from *Ctrip* cover the period from May 2003 to May 2010. For each review, we collected product identification information and category along with detailed information from each review (time and rating) within the timeframes of our data. The details for each data source are described below.

Data from *Amazon* cover 15 pre-defined product categories (Table 1), which have been commonly studied in the literature and include apps, video games, and digital cameras, among others. The abovementioned review information of all products within these categories were collected for the sampling window. The data were expected to enable the conduct of our estimations and the classification of experience versus search products at the category level.

--Insert Table 1 about here--

To obtain data on services, we collected all restaurant reviews from the 50 largest cities in the US⁷ over a five-year period (October 2006 to October 2011). This procedure yielded 249,199 reviews for 5,557 restaurants.

To obtain data on attributes, we collected information on travel products from *Ctrip.com*, which is a large online information intermediary and travel agent in China that collects consumer ratings for travel products such as hotels. *Ctrip* provides an overall rating, as well as attribute ratings, for location and service. We obtained data on 5,439 hotels with a total of 901,482 ratings.

To avoid potential unwanted effects of review bias, we retained products that have at least 25 reviews, which is a common threshold for defining a “large” sample in a statistical sense (Stock and Watson 2003). To avoid biases from self-selection or forum manipulation (Dellarocas 2006; Li and Hitt 2008), we dropped the first five reviews⁸. We retained the original *Amazon* pre-defined product categories for the category-level analysis, although this could introduce some variance in our results if consumers perceive categories differently from how these are classified on *Amazon*.

4.2. Model Specification

Using the time stamps and the rating values of the reviews, we computed the cumulative standard deviation (STD_i) up to the n^{th} review ($n = 1, 2, \dots N$):

$$STD_{nj} = \left\{ \frac{1}{n} \sum_i^n (r_{ij} - \mu_{nj})^2 \right\}^{1/2}. \quad [11]$$

In Equation [11], r_{ij} is the i^{th} consumer’s posted rating for product j , and μ_{nj} is the average of the first n ratings for product/hotel attribute j . The standard deviation will be related to the position of the review in sequence ($sequence(n) = n$). After re-arranging our analytical results (Equation [10]), we obtained the

⁷ The choice of cities is based on the size using population as selection criteria. We referred to the US Census Bureau for the population estimates. Source: U.S. Census Bureau. Web: www.census.gov.

⁸ We repeated all of our analyses using the full set of data and found that the ordering of our results remains unchanged.

following:

$$STD_{nj} = STD(r_{nj}) = Var(r_{nj})^{1/2} = \left(\frac{w_j^2 \sigma_s^2}{n} + (1 - w_j)^2 \sigma_e^2 + \sigma_j^2 \right)^{1/2}, \quad \frac{d\sigma_{e_{nj}}^2}{dn} \geq 0. \quad [12]$$

We estimated the relative importance of experience attributes versus search attributes based on reduced-form panel data estimations of cumulative variance with reference to the review sequence n .

Subsequently, we could infer the product type as a continuous measure based on the relative value of this coefficient. Given the *Amazon* category-level data set pools multiple products (same with the *Ctrip* data where multiple hotels are pooled together), we included a fixed effect α_x for each product/hotel x in the analysis. For example, given a category (e.g., books) j on *Amazon* ($j = 1, 2, \dots, 15$), the estimating equation for every product x in this category j ($x = 1 \dots X$) is:

$$STD_{nj} = \beta_j * Sequence_{nxj} + \alpha_{xj} + \varepsilon_{nxj}, \quad [13-1]$$

where $Sequence_{nxj}$ denotes the n th review for product x in category j , α_{xj} is the fixed effect for product x in category j , and ε_{nxj} is the random noise associated with the n th review for product x in category j .

Since we intended to estimate one product (restaurant) at a time for product-level analysis (*Yelp*), the estimation equation for each product (restaurant) j is:

$$STD_{nj} = \beta_j * sequence_{nj} + \varepsilon_{nj}. \quad [13-2]$$

In Equation [13-2], $Sequence_{nj}$ indicates the n th review of product j , and ε_{ij} captures individual random errors for the n^{th} review of product j .

In both equations, β_j captures the relationship between the variance and the sequence of ratings for category/product/attribute j , and as such, it is the parameter we seek to estimate. If this coefficient is large, then experience attributes become more important in determining utility (i.e., w_j is smaller and/or uncertainty related to experience attributes is relatively large). Although the dependent variable is not independently distributed, any standard OLS χ^2 test of the null hypothesis can be calculated and

interpreted in the usual way in the presence of only one trend parameter (Hamilton 1994). In addition, the test is asymptotically valid for any hypothesis on any subset of the parameters (p. 472). Therefore the estimates are consistent, unbiased, and efficient for large N .

4.3. Category Level Estimations (*Amazon* Categories and Products)

Table 2 shows the estimations of Equation [13-1] for each *Amazon* category ordered from largest to smallest in terms of the value of the variance trend estimate (β_j). As noted earlier, a larger value of β_j implies that the products within this category comprise more experience attributes (smaller w_j and/or $\frac{\sigma_s^2 j}{\sigma_e^2 j}$, or both), whereas a smaller (especially negative) value of β_j suggests that the products within this category are dominated by search attributes (larger w_j and/or $\frac{\sigma_s^2 j}{\sigma_e^2 j}$). The ordering can be regarded as intuitively reasonable, at least on the boundaries, with memory cards being the most “search” related and beauty products being the most “experience” related. Based on these estimates, we suggest several cut-off points for the 15 product categories obtained from *Amazon*. The first cut-off point is labeled as **dominant experience** products (statistically significant coefficient with effect size $> +0.002$), which exhibit strongly increasing variance in the review sequence: *Beauty products, TVs, Musical Instruments, Video Games, Camera Lenses, Keyboards, Books, Android Apps, and Software* are in this category. The second batch of products is categorized as **experience** products, where the trend is slight, but non-decreasing (statistically significant coefficient with effect size > 0): *DVDs and Digital Cameras* are in this group. The third category of products did not show significant changes (either increase or decrease) over the increase in the sequence of reviews. We categorize these as **search** products (statistically significant coefficient with effect size < 0): *Hard Drives and Laptops* fall in this category. The last group of products exhibits significant decrease over the sequence of reviews, such as *Portable GPS* and *Memory Cards*. We categorize them as **dominant search** products. The grouping of products can be affected by the heterogeneity within the group product, which could reduce the ability to discriminate among the trends for different categories. Therefore, the category-level analysis is only meaningful when the category can

be clearly defined and the variations within the category are not large. In the event that this is a concern, more fine-grained categories can be chosen. This procedure is demonstrated below for product (restaurants)-level analysis, and discussed further in the conclusion.

--Insert Table 2 about here--

4.4. Product Level Estimations (*Yelp* Restaurant Reviews)

Restaurants are generally considered as experience products. This observation is confirmed by our model-free sanity check, as presented in Figure 3. The figure shows that in aggregate, variance increases with sample size per restaurant for our *Yelp* restaurant review data after omitting the first five reviews as discussed earlier. A regression analysis (Table 3) for the aggregate category shows a significant positive coefficient of the number of ratings on review variance, which is consistent with the result in Figure 3. Nonetheless, there are significant variations among restaurants, e.g., certain restaurants are essentially standardized by design (e.g., McDonald’s) while fine-dining restaurants are likely to be highly idiosyncratic, therefore, deriving finer-grain distinctions between different restaurants within the “restaurant category” would be useful. This step is possible with our method. We show that the same theoretical model applied at the category level could also be applied at the product level (individual restaurant). The product level analysis can be estimated with Equation [13-2].

Figure 3. Variance of Ratings versus Number of Ratings per Restaurant

--Insert Table 3 about here--

Table 4 shows the results from different subgroups of the restaurant category that are reasonably well defined. Fast food has a significantly negative coefficient; therefore it is mostly a “search” product. Pizza shop has a coefficient not significantly different from zero, which suggests that it is a mix of search and experience. The Fine Dining category has a coefficient that is significantly positive, which confirms our intuition that it is generally experience-based. These results can be extended down to the individual facility. For instance, we find similar results when narrowing down our analysis to, for example, the 123-Eatery Restaurant in Minneapolis. Overall, our results suggest that our model can be used effectively to understand product type and classify each product or sub-group of products as search or experience product in a continuous scale. The results also suggest that merchants/manufactures can employ our data-driven method to understand better the perception of consumers on their products/services and to design strategies accordingly. For example, if a particular pizza shop is perceived as more search than experience, then the advertising strategy should focus more on product attributes. Moreover, given that experience-dominant products/services are more likely to command price premiums (e.g., Gucci bag), owner or manager can also use our data-driven measure to evaluate whether a particular strategy to change perception works. For example, some luxury brand manager can test whether a particular advertising strategy successfully make their products more “experience” based.

--Insert Table 4 about here--

4.4. Attribute Level Estimations (*Ctrip* Data)

The *Ctrip* data set contains an aggregate review (overall rating) that is consistent with our earlier data, as well as separate ratings for service (likely an experience attribute) and location (which should be a search attribute because location information can be communicated relatively easily). Table 5 shows the estimations obtained through Equation [13-1]. Once again, the results are consistent with our expectations that the coefficient of review variance trend is significantly positive for service, significantly negative for

location, and essentially zero for the aggregate. The results demonstrate the effectiveness of our model as a method in classifying products/services.

--Insert Table 5 about here--

Figure 4. Empirical Observations of the Regression Line for Different Attributes

4.6. Relative Performance in Comparison with the model of Nelson (1970)

Although the empirical applications of our data-driven method have demonstrated its effectiveness in classifying product type and confirmed our expectations of product types accurately, a benchmark comparison would be useful to achieve a more objective measure of the performance of our method. To date, the most widely used product classification method with regard to experience versus search products is the measure developed by Nelson (1974). To determine the performance of our measure relative to that of Nelson, we employed users on AMT⁹ to conduct a classification task. They were asked to rate 35 product classes according to: (a) the ability to evaluate the product based on second-hand information, (b)

⁹ Our use of AMT is consistent with prior literature (e.g., Archak et al. 2011, Ghose et al. 2012). AMT has been shown to be a reliable respondent pool for simple information tasks such as content coding (Buhrmester et al. 2011). The survey was administered in September 2013. Twenty respondents were paid \$10 to categorize 35 questions. On average, respondents took 27 seconds to do each category (an effective wage of \$13 per hour).

whether a consumer can be certain of product attributes before purchase, and (c) whether a consumer needs to use the product to ascertain its attributes before purchase. A sample question for one category can be seen in Figure. The first question was based on our definition of product type, and the second and third questions were adapted from Weathers et al. (2007) and Hong and Pavlou (2014)¹⁰.

Instructions:

Strongly positive	I can sufficiently evaluate the product's characteristics by reading second hand information from other consumers (such as online product reviews).
Positive	I can evaluate the product's characteristics by reading second hand information from other consumers (such as online product reviews).
Neutral	I can evaluate the product's characteristics by reading second hand information from other consumers (such as online product reviews) to some extent.
Negative	I can not sufficiently evaluate the product's characteristics by reading second hand information from other consumers (such as online product reviews).
Strongly negative	It is not possible to evaluate the product's characteristics by reading second hand information from other consumers (such as online product reviews).

Judge the sentiment expressed by the following item toward: Ability to value the product based on second hand information.

Jewelry

Strongly negative Negative Neutral Positive Strongly positive

Figure 6. Sample Page for Data Collection on AMT

The coded data from 20 respondents show acceptable inter-rater reliability (Krippendorff 2004) and internal consistency of the three questions (Cronbach 1951); thus, we averaged the responses from this five-point Likert-type scale¹¹. From the rank of the products on a straight line ranging from experience to search, the categorization of Nelson is only 30% correlated with the result of the user survey; however, our proposed measure achieved a categorization that is 70% correlated with the result of the user survey. The scatter plots show that the user survey results fit more precisely with our proposed measure than with that of Nelson (1970). This result indicates the advantage of our proposed product review-based measure, in capturing the view of consumers. Figures 6a and 6b show this relationship.

¹⁰ The three questions are: (a) I can evaluate the product characteristics based on reading second hand information from other consumers (such as online product reviews), (b) It's important for me to see/touch/hear (whichever applies) this product to evaluate its attributes, and (c) I can evaluate the quality of this product simply by reading descriptions about the product.

¹¹ Krippendorff's alpha > 0.8 and Cronbach's Alphas > 0.7.

Figure 6. Scatter Plots of Ranks by Product Review, Nelson (1970) and User Studies

5. Discussions and Implications

5.1 Key Findings

We leverage the statistical properties of online reviews to construct an objective measure of product type. On the one hand, our theoretical framework suggests that products principally chosen based on search attributes will have a downward trend in review variance as more reviews are received because information on search attributes can be efficiently transferred among consumers. Therefore, more information decreases the uncertainty of product quality. On the other hand, experience attributes are more difficult to communicate and are less transferrable among consumers because of the differences in consumer preferences. Therefore, ratings for products that place more importance on experience attributes will not show the same pattern as those that are more concerned with search attributes. This insight leads to a simple measure of product type that can be easily implemented using publicly available longitudinal online review data. Our results suggest that this measure provides intuitively plausible results and outperforms existing methods, such as the Nelson repair-cost classification method. We also show that this approach can be applied at multiple levels of aggregation from product category to product attribute.

5.2. Contribution and Implications

This research offers several distinct contributions and practical implications. First, it significantly advances the empirical measure of product type. Although an extensive stream of research employs product type, particularly in IS, economics, and marketing, the widely adopted approach to measure product type developed by Nelson (1970) has not been updated for 50 years. Scholars have expressed that despite the applicability of the theoretical model on consumer information search, using the measure developed about 50 years ago may limit the confidence of stakeholders in drawing accurate theoretical implications and actionable findings (Lucas Jr 1976) because the influence of the reduced search cost (Bakos 1997) on different products are not the same. For example, to avoid ambiguity, many studies collected sample products based on extreme examples of search or experience products (Senecal and Nantel 2004), whereas others acknowledged that the Nelson measure could be potentially problematic (Caves and Williamson 1985; Huang et al. 2009), specifically for products that are outside of the original list of Nelson (1970) (Mudambi and Schuff 2010). Without a better measure of product type in place, some scholars continued to use Nelson's original classification despite their open acknowledgement that the "Internet is likely to change the traditional relationship between search and experience products" (Huang et al. 2009, p. 56). Our study contributes to this aspect by leveraging large amounts of free data generated by consumers to inform scholarly research. We also contribute to an ongoing research stream in IS that demonstrates the generation of useful insights from embedded information in online reviews for IS and economic research. Specifically, we show that the dynamics of longitudinal reviews can provide an estimate of information transfer efficiency. This can assist in the observation of this dynamics from a practical perspective, which has not been achieved successfully considering that economists have defined experience and search attributes only on the theoretical level. Nowadays, the unprecedented large-scale review data provide a vehicle to measure and empirically classify products as perceived by consumers, offering the opportunity to close the gap between theory and measurement of product type. Our data-driven method to discern product type provides a continuous measure, which is more practical than the

dichotomous view of product type, and is easy to implement and most importantly, it can be applied at different levels. Our data-driven measure can also reflect the potential changing perception of consumers on the products, as a result of the changing marketplace and information available to consumers.

Therefore, our review-based measure can be used by academic researchers in such disciplines as IS, marketing, economics, and psychology that aim at studying the direct or moderating effects of product types, an endeavor that requires proper tools.

Second, this study highlights the design implications for the literature on online product reviews. To make reviews helpful for the decision making, we clarify that for different types of products, different online product review systems should be created. The informational role of the distributional characteristics of product reviews also indicate that consumer text comments for pure search products may not add much value on top of numerical ratings. People generally have the capacity to interpret statistical properties of everyday events (such as political polls) based on heuristics (Darke et al. 1998; Nisbett et al. 1983), especially when they are within a scenario that motivates them to require high accuracy (e.g., online shopping task)¹². Thus, consumers no longer need to read many text comments because the distributional characteristics already convey what they need to know including the average rating for expectation of product quality and variance of rating for product quality uncertainty.

On the other hand, for products that are chosen based on experience attributes, additional supplemental information (such as reviewer metadata or text comments) may be especially important for extracting the useful part of the review signal for the individual utility when average rating is not as meaningful and review variance does not decrease over time. The current review systems (e.g., *Amazon* and *Yelp*) allow consumers to post a rating (an integer between 1 and 5) and a text comment that complement the numerical rating. We maintain that expanding the uni-dimensional review system into multiple dimensions is useful in separating the main search attributes and experience attributes to facilitate

¹² We argue that online shopping indeed is a task that motivates consumers to achieve high decision accuracy because they are clearly driven to maximize utility in this task.

matching consumer preference with the attributes of products. A laptop computer, for example, can have dimensions such as build quality, processing efficiency, aesthetics, and so on. In addition, consumers are bounded by rationality (Simon 1982); thus, excessive unstructured information may be detrimental to decision making (Jones et al. 2004). As a result, the information role of online product reviews is undermined. By correctly identifying the attributes (experience versus search), online product review systems could be directly designed for higher consumer search efficacy. For example, a pure search product would benefit from a uni-dimensional numeric rating, whereas products with dominant experience attributes would benefit from more structured text comments, discussions, and even product forums, which allow later consumers to identify “consumer–product fit” from heuristics.

Finally, this study also provides practical implications for managers. As Villas-Boas (2004) has observed, consumers can gain further information regarding the fit of experience products to their preferences only by experiencing it after purchase. Therefore, for manufacturers and retailers, educating consumers about product attributes could lower pre-purchase uncertainty and enhance performance. Online retailers can adopt state-of-the-art technologies such as “3D virtual experience” (Jiang and Benbasat 2007) to simulate “touch and feel” experiences for potential consumers. Full understanding of product type is crucial because adopting such technologies is costly. For experience attributes that cannot be easily transferred, advertising the brand rather than the specific product attributes may be beneficial. Another implication for managers of online or offline shops is that the variance trend of their products with product review data can be used to determine whether the consumers view their products using the same standard (searchable attributes) or different standards (preferences). In the restaurant sample, 68% of all restaurants have experience attributes (and therefore any restaurant is more likely to have experience attributes); however, 32% are dominated by search attributes. Let us consider two restaurants: one only serves fast food such as KFC (search) and the other is an exotic French restaurant (experience). To boost the sales the sales of the former, taking a vertical differentiation strategy is optimal (e.g., the restaurant can increase volume per meal, or lower its prices). In the case of the French restaurant that features a number of experience

attributes, only people who like the taste matter. Thus, the French restaurant may not be required to cater to the general taste of the public; the restaurant's efforts should instead be focused on facilitating "matching".

5.3. Limitations and Opportunities for Future Research

Our method of classifying product types is based on information transfer efficiency. The dynamics of longitudinal ratings provide a lens to observe how prior ratings are communicated and consumed by a later consumer (from his or her rating). Our theoretical model assumes that a consumer rates a product based on his or her consumption utility (accommodating quality variances, differences in tastes, and other random factors). The extent to which reviews can be manipulated purposely can potentially decrease the accuracy of our measure. For example, if the amount of fraudulent reviews is significant, consumers will not trust reviews as much. This phenomenon makes information transfer through ratings less efficient, which in turn will make products more experience oriented. However, as long as fraud or manipulation is not so serious that make ALL reviews not trustable, our model remains applicable because the fundamental idea that ratings for search and experience products are processed differently remains. Fraudulent reviews can also be believed to have a limited influence on our results pursuant to the efforts of review platforms to combat review fraud. In addition, this study focuses solely on the quantitative review component, which likely could leave out significant, potentially useful information in the review text. However, as long as ratings reflect the processing of information (presumably including the textual information) in the aggregate level, our results remain unbiased. For example, informative textual reviews may help consumers learn about various product attributes. If this is true, that is, if textual information helps reduce uncertainty, then we should observe a decrease of rating variances over time, which is consistent with the view that the product is more search oriented. Therefore, ratings would encompass the effectiveness of text comments in transferring product quality information, and longitudinal rating variance would still provide a proxy whether information can be used to reduce uncertainty (the key difference between search and experience products). That being said, effectively incorporating semi- or

unstructured review text information to fine-tune our classification method remains a potentially fruitful avenue of future work. Moreover, following prior literature (Nelson 1981), utilities from search were assumed to be taken independently from experience attributes in our theoretical model in order to understand each of their effects. Empirically, such absolute independence is not required as long as the dependence is not too high. If the two utilities are highly dependent to each other, it would suggest that it is not necessary to separate the utilities, because they are essentially of the same type (i.e. only search attributes or only experience attributes), and empirically it would be impossible to discern them. However, we believe that in reality, some experiences are just not as easy to communicate than some others, and this is supported by our empirical studies as we are able to reasonably distinguish between different product types, and our approach also outperforms that of Nelson based on user survey.

5.4. Conclusion

Data-driven analytics has attracted significant attention from both academics and practitioners. The digital revolution and digital innovation (Brynjolfsson and McAfee 2011) have brought about data that researchers have only dreamed of a decade ago. One challenge that remains is how to make sense of and leverage the freely available user-generated data to inform research and practice by building theoretically grounded and empirically validated analytical tools. This challenge has already caught the attention of many scholars in the field of IS¹³. As an important type of consumer-generated content, online product reviews are prime examples of micro-level data that are free and readily available in digital format. A better understanding of the types of information that can be extracted from these data has considerable value. Our proposed data-driven method shows that the dynamics of longitudinal ratings provides an effective way to understand information transfer efficiency, which could be used to infer product type.

¹³The *Workshop on Statistical Challenges in E-commerce Research* (SCECR) was organized as an effort to encourage IS and marketing scholars to take advantage of the newly available data to inform research and practice.

References

1. Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Science*. **57**(8):1485-1509.
2. Bakos JY (1997) Reducing buyer search costs: Implications for electronic marketplaces. *Management science*. **43**(12):1676-1692.
3. Brynjolfsson E, McAfee A (2011) Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy.
4. Caves RE, Williamson PJ (1985) What is product differentiation, really? *The Journal of Industrial Economics*. 113-132.
5. Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*. **54**(3):477-491.
6. Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*. **43**(3):345-354.
7. Clemons EK, Gao GG, Hitt LM (2006) When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*. **23**(2):149-171.
8. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*. **16**(3):297-334.
9. Darke PR, Chaiken S, Bohner G, Einwiller S, Erb HP, Hazlewood JD (1998) Accuracy motivation, consensus information, and the law of large numbers: Effects on attitude judgment in the absence of argumentation. *Personality and Social Psychology Bulletin*. **24**(11):1205-1215.
10. Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*. **49**(10):1407-1424.
11. Dellarocas C (2006) Strategic manipulation of Internet opinion forums: implications for consumers and firms. *Management Science*. **52**(10):1577-1593.
12. Dellarocas C, Zhang XM, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*. **21**(4):23-45.
13. Duan W, Gu B, Whinston AB (2008) The dynamics of online word-of-mouth and product sales--An empirical investigation of the movie industry. *Journal of Retailing*. **84**(2):233-242.
14. Forman C, Ghose A, Wiesenfeld B (2008) Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*. **19**(3):291-313.
15. Garvin DA (1984) What does "product quality" really mean? *Sloan management review*. **26**(1):25-43.
16. Ghose A, Ipeirotis P (2010) Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Knowledge and Data Engineering, IEEE Transactions on*. (99):1-1.
17. Godes D, Silva JC (2012) Sequential and temporal dynamics of online opinion. *Marketing Science*. **31**(3):448-473.
18. Hamilton JD. 1994. *Time series analysis*. Princeton university press Princeton.
19. Hong Y, Pavlou PA (2014) Product Fit Uncertainty: Nature, Effects and Antecedents. *Information Systems Research*. **Forthcoming**.
20. Hsieh Y-C, Chiu H-C, Chiang M-Y (2005) Maintaining a committed online customer: a study across search-experience-credence products. *Journal of Retailing*. **81**(1):75-82.
21. Hu N, Zhang J, Pavlou PA (2009) Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*. **52**(10):144-147.
22. Huang P, Lurie NH, Mitra S (2009) Searching for experience on the web: an empirical examination of consumer behavior for search and experience products. *Journal of Marketing*. **73**(2):55-69.
23. Jiang Z, Benbasat I (2007) The effects of presentation formats and task complexity on online consumers' product understanding. *MIS Quarterly*. **31**(3):475-500.

24. Jones Q, Ravid G, Rafaeli S (2004) Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information systems research*. **15**(2):194-210.
25. Klein LR (1998) Evaluating the potential of interactive media through a new lens: Search versus experience products. *Journal of business research*. **41**(3):195-203.
26. Krippendorff K (2004) Reliability in content analysis. *Human Communication Research*. **30**(3):411-433.
27. Kwark Y, Chen J, Raghunathan S (2014) Online Product Reviews: Implications for Retailers and Competing Manufacturers. *Information Systems Research*. **Forthcoming**.
28. Laband DN (1991) An objective measure of search versus experience products. *Economic Inquiry*. **29**(3):497-509.
29. Li X, Hitt L (2010) Price effects in online product reviews: an analytical model and empirical analysis. *MIS Quarterly*, Vol. 34, No. 4, pp. 809-831.
30. Li X, Hitt LM (2008) Self selection and information role of online product reviews. *Information Systems Research*. **19**(4):456-474.
31. Liu Y (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*. **70**(3):74-89.
32. Lucas Jr RE. 1976. Econometric policy evaluation: A critique *Carnegie-Rochester conference series on public policy*. Elsevier, 19-46.
33. McWilliams A, Siegel D (2001) Corporate social responsibility: A theory of the firm perspective. *Academy of management review*. **26**(1):117-127.
34. Moe WW, Schweidel DA (2012) Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*. **31**(3):372-386.
35. Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science*. **341**(6146):647-651.
36. Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS Quarterly*. **34**(1):185-200.
37. Nelson P (1970) Information and consumer behavior. *Journal of Political Economy*. **78**(2):311-329.
38. Nelson P (1974) Advertising as information. *Journal of Political Economy*. **82**(4):729-754.
39. Nelson P (1981) Consumer information and advertising. *Economics of information*. 42-82.
40. Nisbett RE, Krantz DH, Jepson C, Kunda Z (1983) The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*. **90**(4):339.
41. Park C, Lee TM (2009) Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research*. **62**(1):61-67.
42. Pavlou PA, Dimoka A (2006) The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*. **17**(4):392-414.
43. Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy*. **82**(1):34-55.
44. Senecal S, Nantel J (2004) The influence of online product recommendations on consumers' online choices. *Journal of Retailing*. **80**(2):159-169.
45. Shapiro C (1982) Consumer information, product quality, and seller reputation. *The Bell Journal of Economics*. 20-35.
46. Simon HA. 1982. *Models of bounded rationality: Empirically grounded economic reason*. MIT press.
47. Smith DB (1990) The economics of information: An empirical approach to Nelson's search-experience framework. *Journal of Public Policy & Marketing*. 111-128.
48. Stock JH, Watson MW. 2003. *Introduction to econometrics*. Addison Wesley Boston.
49. Sun M (2012) How Does Variance of Product Ratings Matter? *Management Science*. **Forthcoming**.
50. Theil H. 1971. *Principles of econometrics*. Wiley New York.
51. Tirole J. 1988. *The theory of industrial organization*. MIT press.

52. Villas-Boas JM (2004) Consumer learning, brand loyalty, and competition. *Marketing Science*. **23**(1):134-145.
53. Villas - Boas JM (2006) Dynamic competition with experience products. *Journal of Economics & Management Strategy*. **15**(1):37-66.
54. Wang C, Zhang X, Hann I-H. 2009. Social bias in online product ratings: a quasi-experimental analysis *The Workshop on Information Systems and Economics*.
55. Wang J, Ghose A, Ipeirotis P. 2012. Bonus, Disclosure, and Choice: What Motivates the Creation of High-Quality Paid Reviews? *ICIS*.
56. Weathers D, Sharma S, Wood SL (2007) Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience products. *Journal of Retailing*. **83**(4):393-401.
57. Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*. **74**(2):133-148.

TABLES

Table 1. Summary Statistics for *Amazon* Product Samples

#	Category	Number of Products	Total Observations
1	Android Apps	880	48400
2	Video Game	648	28751
3	TV	552	24918
4	Digital Camera	473	20285
5	Music Instrument	395	16575
6	Books	384	15889
7	Laptop	275	8089
8	DVD	176	9153
9	Camera Lens	152	7146
10	Memory Card	151	6726
11	Keyboard	137	6900
12	Software	133	6367
13	GPS	98	4945
14	Hard Drive	85	3891
15	Beauty Products	70	2742

Table 2. Estimation Results of *Amazon* Data (DV = STD_{ij})

Category	Coefficient	Std. Err.	Within R ²	F Statistic	Classification
Beauty Product	0.0043***	0.0016	0.07	7.32	Dominant Experience
TV	0.0038***	0.0004	0.1	88.73	Dominant Experience
Music Instrument	0.0037***	0.0005	0.09	53.17	Dominant Experience
Video Game	0.0034***	0.0003	0.11	113.30	Dominant Experience
Camera Lens	0.0033***	0.0006	0.09	24.03	Dominant Experience
Keyboard	0.0032***	0.0005	0.12	30.12	Dominant Experience
Book	0.0026***	0.0005	0.05	27.28	Dominant Experience
Android App	0.0023***	0.0002	0.06	98.90	Dominant Experience
Software	0.0022***	0.0007	0.05	9.52	Dominant Experience
DVD	0.0016***	0.0004	0.04	14.57	Experience
Digital Camera	0.0013***	0.0003	0.02	16.61	Experience
Hard Drive	0.0004	0.001	0.002	0.12	Search
Laptop	-0.0002	0.0005	0.002	0.06	Search
Portable GPS	-0.0011***	0.0003	0.04	9.35	Dominant Search
Memory Card	-0.0013*	0.0007	0.02	2.94	Dominant Search

Notes: 1. Estimation includes product level fixed effects.

2. Cluster robust standard errors are in parentheses, and standard errors are adjusted for all clusters (Groups).

3. Coefficients are significant at *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 3. Fixed Effects Estimations in *Yelp* Data (Restaurant Sample, DV=STD)

Estimation Method:	(1) OLS	(2) FE
<i>Sequence</i>	0.003*** (0.000)	0.003*** (0.000)
Constant	0.813***(0.002)	0.813***(0.003)
Observations	170,586	170,586
Restaurant FE	No	Yes
R-squared	0.029	0.130
Number of Merchants	--	4,374

Notes: 1. For the Fixed Effects Model-Cluster robust standard errors are in parentheses, and standard errors are adjusted for all groups.

2. Coefficients are significant at *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 4. Estimations for Individual Restaurants at *Yelp.com* (DV=STD)

	(1) Fast Food ¹⁴	(2) Fine Dining ¹⁵	(3) Pizza ¹⁶
<i>Sequence</i>	-0.0003*** (0.000)	0.0003*** (0.000)	-0.000 (0.000)
Constant	1.314*** (0.005)	0.643*** (0.005)	1.110***(0.015)
Observations	1,007	1,151	629
R-squared	0.760	0.811	0.01

Notes: 1. Robust standard errors are in parentheses.

2. Coefficients are significant at *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 5. Fixed Effects Estimation Results in *Ctrip* Data (DV = STD_{ij})

VARIABLES	(1) Service Rating	(2) Location Rating	(3) Overall Rating
<i>Sequence</i>	0.0001*** (0.000)	-0.0001*** (0.000)	0.000 (0.000)
Constant	0.882*** (0.002)	0.749*** (0.002)	0.803***(0.002)
Observations	299,259	299,275	299,275
Hotel FE	Yes	Yes	Yes
Number of product ID	5,439	5,439	5,439

Notes: 1. Cluster robust standard errors are in parentheses..

2. Coefficients are significant at *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

¹⁴ The *Yelp* page is <http://www.Yelp.com/biz/cin%C3%A9bistro-miami-2>

¹⁵ The *Yelp* page is <http://www.Yelp.com/biz/112-eatery-minneapolis>

¹⁶ The *Yelp* page is <http://www.Yelp.com/biz/galactic-pizza-minneapolis>