“Only the Interested Learn”
– A Model of Proactive Learning of Product Reviews

Liye Ma

October 2015

1 Liye Ma (email: liyema@rhsmith.umd.edu) is an Assistant Professor of Marketing at the Robert H. Smith School of Business at the University of Maryland. The author would like to thank the data sponsor which provided the data used in this study. All errors are my own.
Abstract

We develop a sequential learning model to analyze how consumers proactively acquire product quality information through reading online product reviews. Product reviews are often treated as exogenous quality signals in the Internet word-of-mouth literature. Review reading, however, is a costly activity undertaken in a deliberate manner. Our model casts consumers’ review reading process in a rational framework. In our model, whether a consumer reads reviews, and of which products, depend on the consumer’s information set and expectations. We estimate the model using a rich dataset from a restaurant review website which contains information on both browsing and purchase. We find strong evidence of consumers consciously seeking product reviews, and parameter estimates reveal distinct types of information acquisition behaviors. Comparison with alternative models shows that taking reviews as exogenous signals leads to biased estimates on quality levels and signal precisions. Counterfactual analysis further shows that the reviews a consumer sees earlier is more consequential, due partly to their impact on subsequent search actions, and that sorting reviews from lowest to highest ratings leads to highest differentiation across products. Our study is the first to analyze consumer’s review reading process and to analyze the effect of reviews in this dynamic context.

Keywords: Internet WOM, online product reviews, learning models, dynamic programming, sequential learning, choice models
1. Introduction

In this study, we develop a sequential, proactive learning model where consumers rationally seek out product reviews to learn about product qualities, and we evaluate the effect of reviews while accounting for such endogenous search behavior of consumers. The proliferation of online product reviews has attracted great interest of the research community. Studies on Internet word-of-mouth (WOM) show that product reviews have significant impacts on consumer purchases and other marketing outcomes (Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Duan et al. 2008, Godes and Mayzlin 2009, Chintagunta et al. 2010). As product reviews typically contain information on existing consumers’ experiences with the products, they convey vital information pertaining to product quality and effectively facilitate consumer-to-consumer learning.

Early studies on the effect of Internet WOM typically used aggregate measures of product reviews such as the average review rating or the variance of ratings as explanatory variables, instead of analyzing reviews at individual level. While confirming the significant impacts of product reviews, by staying at aggregate level, these studies provide limited insights on the process through which product reviews enter consumers’ decisions. Recent studies have started investigating consumers’ decision process using individual level models. Zhao et al. (2013) analyzed the effect of reviews on consumer purchases of experiential products using a Bayesian learning model, while accounting for both product quality and review credibility. Wu et al. (2015) developed a differential learning model by incorporating reviewer characteristics and preference heterogeneity. These studies with individual level analysis provide more refined knowledge on the effect of product reviews. However, while differing in approach, existing studies usually treat product reviews as quality signals that are exogenously given to consumers, i.e. consumers are assumed to acquire the information from reviews in a costless manner. This assumption of exogeneity is not just made in the studies on product reviews, but is often adopted in studies on consumer learning in general. The learning literature typically models consumers updating quality belief either through direct purchase and experience, or through external signals such as advertising or product reviews (Ching et al. 2013). While purchase is inherently an endogenous decision, hence analyses of consumer purchases, particularly in a forward-looking
framework, automatically account for the endogenous acquisition of product quality information through purchase, studies on learning from external signals routinely take such signals as exogenously given (e.g. Erdem and Keane 1996, Ackerberg 2003), with only a few exceptions.²

Reading reviews, however, is a time-consuming activity. While the consumer search literature has recognized that looking up price and product characteristic information is costly (Weitzman 1979, Kim et al. 2011), poring through review texts to extract important quality information is equally time consuming and costly. Given the cost involved, consumers are expected to be judicious in deciding what reviews to read. For example, if the first few reviews a consumer sees about a product are all negative, the consumer may simply give up and search for other more promising products instead – it may not worthwhile to invest additional time to learn more about a product that is unlikely to be good. In contrast, if those first few reviews are positive, the consumer may develop a higher interest in the product, and decide to read more reviews to learn more about the product. As another example, after reading a number of reviews, a consumer may believe that she has enough information about the product and decide to stop, even if there are many more reviews available.

The acquisition of quality information from product reviews, therefore, is a dynamic and endogenous process. In this process, the information a consumer already acquired not only has a direct impact on purchase through shaping the consumer’s view on the product quality, but also has an indirect impact through determining whether the consumer will decide to acquire more information about the same or other products. Taking product reviews as exogenous signals, however, would recognize only the direct impact, but not the indirect impact on the information set formation process. To get an accurate understanding of how product reviews affect consumer decisions, therefore, it is necessary to first understand how consumers decide to read product reviews. This, however, has not been done in the existing literature. An in-depth

² Chintagunta et al. (2012) controls for the endogeneity of detailing behaviors using a rational expectation approach. Erdem et al. (2005) account for consumers’ decision on which source of information to seek, which is not product specific. Our model differs from both these approaches. Both studies are discussed in more detail in the literature review section.
understanding of this information acquisition process is also crucial for product review platforms, which should present product reviews to consumers in an efficient manner.

Motivated by this gap in the literature, we study in this paper how consumers decide whether and what product reviews to read in a sequential manner, and we analyze the effect of product reviews in this dynamic context. To do so, we develop a dynamic model in which consumers rationally acquire product reviews to build up their information sets. We call this a “proactive learning” model, as consumers proactively decide what to learn rather than passively receive quality signals. In our model, a consumer solves a dynamic programming problem to build up their optimal information set on product qualities, while minimizing the information acquisition cost. Whether a consumer will proceed to acquire more product review information, and for what product to acquire such information, depend on the current information set of the consumer, and on the anticipation of how additional information on certain products will alter the information set. This information acquisition process is also intermingled with the purchase decisions, where in each purchase incidence a consumer makes decisions based on the quality perceptions according to their information set. The model thus both endogenizes consumers’ learning actions, and enables the analysis of the effect of product reviews on purchases with such proactive learning behaviors accounted for.

We estimate the model using a unique dataset obtained from a popular restaurant review website in an Asian country. The dataset contains information on product reviews and consumer purchases. More importantly, it contains the information on consumers’ browsing activities at the website, at the clickstream level. This detailed browsing information enables the identification of the proactive learning model. Through reduced form analyses of the data, we first show that there is strong model-free evidence of consumers consciously seeking product review information. For example, the ratings of early reviews a consumer sees about a product are correlated with the number of additional reviews the consumer subsequently reads about the product, lending support to the model developed in our study. We then estimate the structural parameters of the proactive learning model using the dataset. The parameter estimates reveal different behavioral patterns of learning from product reviews, where one consumer segment is significantly more quality conscious than
the other segment, is more risk averse, updates quality beliefs more slowly, and incurs higher learning cost. The estimates also show that the quality level indicated by a 4-star review or a 5-star review is significantly higher than that of a 2-star or a 3-star review, while the latter two indicate somewhat similar quality levels. These estimates provide a refined understanding on both consumer’s learning behavior and the quality implications of product reviews.

We also compare our model with a benchmark model where product reviews are taken as exogenous quality signals, as is commonly done in the literature. The comparison shows that if we fail to recognize that review reading is an endogenous process but instead take product reviews as exogenous quality signals, the quality differences among reviews of different ratings will be overestimated. The speed of learning and the extent of risk aversion also get biased estimates in this exogenous learning model. Accounting for the proactive learning process, therefore, is necessary for accurately evaluating the effect of reviews. Furthermore, an important implication of the proactive learning process is that the impact of a product review will depend on where the review is located and how early a consumer sees it. Through simulation, we show that a product review on earlier pages is much more consequential than a similar one on later pages. Specifically, a positive review on the first page has about three times the impact of a review of the same rating on the next page, on both how it influences subsequent searches and purchases. Finally, we also evaluate the effects of different review sorting criteria through simulation, and show that showing reviews with the lowest ratings first is the most diagnostically efficient criterion, in the sense that it maximizes the difference in purchase probabilities across products with different review ratings.

We contribute to the literature by developing a dynamic and rational model to understand consumers’ process of learning from product reviews. Since reading reviews is a costly and deliberate activity, yet extant literature often treat reviews as exogenous signals, the proactive learning model developed in this study fills an important gap in the Internet WOM literature, and more broadly contribute to the learning literature as well. That an exogenous learning model leads to biased estimates further confirms the importance of accounting for this proactive learning process. Substantively, by analyzing consumers’
information acquisition process in our framework, we provide a detailed understanding of the different behavioral profiles of searching product reviews; we show that the effect of reviews crucially depend on where the reviews are located, as earlier reviews would also determine whether consumers would decide to read later reviews at all; we show that different review sorting criteria have different efficiency implications. All these add to our understanding of how product reviews affect consumer decisions, and inform industry managers on both the impact of reviews and the design of review platforms.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature. Following that, we describe the data used in our study and discuss model-free analysis in Section 3. We then develop the formal model in Section 4 and discuss the identification and estimation. In Section 5, we discuss the estimation results, followed by simulation analysis in Section 6. Finally, we conclude in Section 7.

2. Literature

Our study falls within the literature on the Internet word-of-mouth (WOM). The proliferation of online product review websites and discussion forums has led to a large number of studies on Internet WOM. Both the creation and the effect of WOM have been investigated in many different contexts. Studies on WOM creation have revealed a diverse set of motivating factors underlying the generation of WOM, including self-enhancement, emotion, social considerations, images, economic incentives, etc. (Anderson 1998, Hennig-Thurau et al. 2004, Schlosser 2005, Albuquerque et al. 2012, Lovett et al. 2013, Toubia and Stephen 2013). These studies also showed that factors such as product and brand characteristics, quality, and content type, etc. affect WOM generation (Berger and Schwartz 2011, Lovett et al. 2013, Dellarocas and Narayan 2006, Berger and Milkman 2012). Several studies also investigated the dynamic aspect of WOM creation (Li and Hitt 2008, Wu and Huberman 2008, Godes and Silva 2012, Moe and Schweidel 2012). Meanwhile, the effect of Internet WOM has also been studied on a wide ranges of products, including books (Chevalier and Mayzlin 2006), movies (Liu 2006, Duan et al. 2008, Chintagunta et al. 2010), TV ratings (Godes and Mayzlin 2004), restaurants (Godes and Mayzlin 2009), micro-lending deals (Stephen and Galak 2012), stock returns (Tirunillai and Tellis 2012), and social network memberships (Trusov et al. 2009), among others. While results
vary across studies, they together show that Internet WOM has a significant impact on sales and other marketing outcomes. By investigating how consumers seek out product review information to form their quality perceptions, our study is more closely related to this latter stream of literature on the effect of Internet WOM. In this stream of literature, most early studies used aggregate review statistics such as the number of reviews, average review ratings, and variances of reviews as explanatory variables, instead of performing “micro” level analyses. Recent studies using individual review level analysis provided more refined understanding on how reviews affect consumer decisions. Zhao et al. (2013) developed a Bayesian learning model to analyze the effect of online product reviews on consumer purchases of experiential products, while accounting for both product quality and review credibility factors. Wu et al. (2015) developed a differential learning model which incorporates reviewer characteristics and preference heterogeneity. Similar to these studies, our study also investigate consumer learning, and we extend from the literature by investigating how consumers proactively seek out product reviews before learning from such reviews.

More broadly, our study is related to the literature on consumer learning (Ching et al. 2013). An extensive literature now exists which investigates how consumers form and update their quality beliefs of products, typically employing the Bayesian learning framework (Roberts and Urban 1988, Erdem and Keane 1996, Ackerberg 2003, Crawford and Shum 2005, Narayanan et al. 2005, Erdem et al. 2008, Ching 2010, Zhang 2010, Zhao et al. 2013; see Ching et al. 2013 for a review of the literature). The learning can come either from consumer’s own purchase and experience, or from external quality signals (e.g. Erdem and Keane 1996). The handling of these two types of quality signals in the literature, though, has been somewhat different. Purchase decisions are inherently endogenous, so studies of forward looking consumer behaviors automatically take into account the endogeneity of the acquisition of such quality signals, as they can be considered as byproducts of purchases. In contrast, most studies on consumer learnings from external quality signals, such as advertisement (Ackerberg 2003), price (Erdem et al. 2008), detailing (Narayanan et al. 2005), or product review (Zhao et al. 2013), etc. treat such quality signals as exogenously given to consumers. However, just as purchasing a product costs money and likely involves deliberate decisions, learning from information sources takes time, and the acquisition and processing of such information also likely involve
deliberate decisions. Treating these quality signals as exogenous is thus a simplification that may preclude a more accurate understanding of how they affect consumers. Several existing studies have sought to overcome this limitation. For example, in Chintagunta et al. (2012), physicians are modeled as rationally expecting the subsequent detailing efforts from which they would learn about drug qualities, thus addressing the endogeneity concern. In Erdem et al. (2005), consumers rationally decide on which information source they will seek out to learn about computer products in a forward looking framework. Our study is similar to Erdem et al. (2005) in that we also model consumers rationally seeking out product quality information. However, rather than deciding which information source to consult with as in Erdem et al. (2005), in our study consumers decide for what product to acquire information. This distinction is crucial, as consciously acquiring product specific information can significantly alter the competitive landscape among products. By looking at how consumers decide on what reviews to read, which is inherently product specific, our study thus also contributes to the learning literature.

More distantly, our study is also related to the search literature (Weitzman 1979, Ratchford and Srinivasan 1993, Mehta et al. 2003, Zwick et al. 2003, Kim et al. 2011, De Los Santos et al. 2012, etc), which investigates how consumers optimally search for information on product prices and characteristics. Typically, uncertainty regarding a specific product is resolved upon searching the product in this literature, and there is no need to repeatedly search for the same product. In contrast in our study, which is more closely related to the learning literature in general, consumers develop more precise understanding of product qualities after processing product review information, but the uncertainty on quality is not fully resolved. Repeatedly acquiring quality information of the same products is a key aspect of our study.

3. Data and Reduced-Form Analysis

3.1 Data Overview

Our data is obtained from a major restaurant review web site from an Asian country. The web site, which shall remain anonymous in this study, is similar to Yelp.com in the United States, and is the market leader in
its country of operation. The website hosts consumer submitted product reviews of thousands of restaurants in each major city of the country, and consistently attracts millions of views each month.

Our dataset covers all restaurants in a district of the country’s largest city. The dataset consists of three parts. First, for each restaurant, the dataset contains all reviews of the restaurant since inception. Each review consists of a star rating, which is an integer from 1 to 5, where 5-star is the highest rating, and a paragraph of text describing the reviewer’s experience at and the evaluation of the restaurant. Second, the dataset contains the clickstream level browsing records of a set of randomly selected consumers over a 31-day period, in the month of March 2012. There are three types of browsing records: accessing the main restaurant page, accessing a restaurant review page, and accessing a restaurant photo page. The main restaurant page lists the name, address, and other factual information about the restaurant. It also shows the overall average review rating. On this main page, the five most recent reviews of the restaurant are also displayed. Subsequently, each restaurant review page shows twenty additional consumer reviews of the restaurant in chronological order, from newer to older ones. Each photo page shows fifteen photos about the restaurant, usually pictures of dishes served at the restaurant. For each restaurant, a viewer will start from the main page, and may subsequently proceed to browse one or more review pages and photo pages. The browsing of a restaurant may span across multiple days, intermingled with browsing of other restaurants. Both review and photo pages are similar to web pages at Yelp.com. This detailed step-by-step record of user’s browsing behavior enables an in-depth analysis of consumers’ information acquisition process.

In addition to review and browsing information, the third part of the dataset contains information on the actual restaurant visitations of the same set of consumers for whom the browsing information is available. The website offers a “checkin” functionality through consumers’ mobile phones. Upon entering a restaurant, a consumer can check in using her mobile phone to register her visitation at the website, and retrieve

---

3 This was later increased to ten reviews on the main page, a change which occurred after the period covered by our data.
4 Due to non-disclosure agreement we cannot show screenshots of the actual webpages.
customized promotion information about the restaurant. Mobile phone users in the country of the website routinely use this feature, and the checkins are reliable proxies of consumers’ actual purchase incidences.\textsuperscript{5}

The entire history of restaurant reviews is included in the dataset, together with consumers’ browsing and checkin actions over the month of March, 2012. The browsing and checkin information provided by the company covers a shorter period of time than does the review information, both due to the more sensitive nature of the data and the much larger size.

### 3.2 Descriptive Statistics

The descriptive statistics of the consumer activities is reported in Table 1. There are 6,518 consumers in the dataset. Over the course of the 31 days covered by the dataset, a consumer on average browsed through 4.92 restaurants. As is typical, the data is positively skewed, with a standard deviation of 10.4 and max of 483. On average, a consumer visited 1.15 review pages and 3.03 photo pages, with standard deviations of 3.51 and 10.6, respectively. The data show that some browsing actions stopped at the restaurant main page level, while others proceeded to the subsequent review and photo pages. The average per-consumer checkin is 0.48, with a standard deviation of 2.25.

[Insert Table 1 about Here]

For each of the 2,367 restaurants in the dataset, the entire history of consumer reviews is available. Each review contains a star rating – an integer ranging from 1 to 5, where higher star rating indicates better qualities – together with a paragraph of textual evaluations. The restaurant-level descriptive statistics of these reviews are reported in Table 2; the total numbers of reviews of each star rating are plotted in Figure 2; and the histogram of restaurant-level average star ratings is plotted in Figure 3. Consistent with existing literature (e.g. Chevalier and Mayzlin 2006), product reviews are in general positive. There are considerably more 4-star and 5-star reviews than 1-star and 2-star ones. The average restaurant-level review rating is 3.59. The

\textsuperscript{5} It is possible that a consumer goes to a restaurant without using the checkin feature. To address this limitation, the model accounts for purchase choice probability conditional on a purchase being made. This is discussed in more detail in Section 4.2 and the accompanied footnotes.
descriptive statistics also show that there is a great amount of review information for these restaurants, where a restaurant on average has more than 70 reviews.

[Insert Table 2 about Here]

[Insert Figures 1 and 2 about Here]

3.3 Preliminary Analysis

We now perform initial, reduced-form analysis related to our research question. The key to consumers’ optimal information acquisition is that consumers would decide whether to acquire additional information, and for what product to acquire such information, based on what they already know in their past searches. In the context of our dataset, consumers acquire restaurant quality information through reading product reviews or viewing photos, in order to make informed decisions on what restaurants to go to. Since processing information is costly, consumers are expected to seek out additional information of a restaurant only when they believe doing so is worthwhile. Intuitively, if the first several reviews a consumer sees about a restaurant have low ratings, the consumer may decide that the restaurant is unlikely to be good, hence stop spending more time on this restaurant and instead search for other restaurants. In contrast, if the first few reviews a consumer sees have high ratings, the consumer may be intrigued and decide to find out more about the restaurant. After the information search, a consumer would be more likely to choose a restaurant with higher review ratings than one with lower ratings. Equally importantly, a consumer would be more likely to choose a restaurant about which she knows more than one about which she knows less. Evidence of consumers performing such rational information acquisition actions should thus be reflected from the relationship between the initial reviews a consumer sees and the subsequent browsing and purchase, i.e. checkin, actions.

Figure 3 plots the average numbers of browsing and checkin activities for each consumer and each restaurant corresponding to the restaurant’s initial review ratings, i.e. the average ratings of the first five reviews displayed on the restaurant’s main page that the consumer sees. As the figure shows, as the initial review rating increases, the number of subsequent browsing activities (left Y axis) also increases in general. This is evidence that a negative initial impression indeed discourages further information search of the
restaurant. The number of browsing activities stabilizes when the initial rating reaches 3.5 stars, and drops slightly for those with initial rating higher than 4.5 stars. This slight dip suggests certain consumers may be convinced that the restaurant is good enough once seeing the overwhelmingly positive reviews at first. Meanwhile, the figure shows that as the initial review rating increases, the likelihood of the consumer visiting the restaurant also increases (right Y axis). Two factors may exist here. A positive initial impression may lead the consumer to form a positive quality view, which itself would increase the chance of purchase. Meanwhile, it also encourages the consumer to seek out more information, as discussed earlier on the increased search activities. This additional information reduces consumer’s uncertainty about the quality, also increasing the chance of purchase. In contrast with browsing which stabilizes after the initial ratings reach 3.5 stars, the checkin activities continues to increase after that, suggesting there is no saturation in the quality implications of the star ratings, i.e. a 5-star review is considered as indicating higher quality than a 4-star review.

Furthermore, Figure 4 plots the number of checkins of a restaurant with respect to the average review ratings of the restaurant, which shows that higher overall average review ratings corresponds to higher numbers of checkins. This suggests that in general consumers perceive product reviews as legitimate quality signals, and their decisions are influenced by them. This is also consistent with the existing literature, which shows that reviews are informative of qualities. Finally, Figure 5 plots the number of checkins with respect to the number of browsing activities. As the figure shows, the more browsing activities a consumer has on a specific restaurant, the higher her likelihood of visiting the restaurant. More searches of a restaurant would typically reflect a consumer’s heightened interest and likely positive quality perception of the restaurant, which should correspond to higher chance of purchase. This is another evidence of the close relationship between the consumers’ actions of acquiring quality information and their subsequent purchases. Taken together, these figures show strong evidence that consumers make their information acquisition decisions dynamically and consciously.
3.3.1. Reduced-form Regression Analysis

The evidence of consumers’ proactive learning behavior can also be shown in reduced-form regressions. Two regression analyses are performed. In the first regression, we explore the association between the average review rating a consumer sees on the main restaurant page and the number of times she subsequently viewed the information of that restaurant. We analyze this through the following Poisson regression:

\[
\text{View}_{ij} \sim \text{Poisson}(\lambda_{ij})
\]

Where

\[
\lambda_{ij} = \exp(\alpha + \beta \times \text{AvgStar}_{ij})
\]

In this regression, \( \text{View}_{ij} \) is the number of times consumer \( i \) viewed photo and review pages of restaurant \( j \), and \( \text{AvgStar}_{ij} \) is the average star rating of the first five reviews the consumer saw at the restaurant’s main page. The regression result, reported in Table 3, shows that the coefficient of the average initial star rating is positive and statistically significant at .001 level. This echoes the first pattern discussed above as shown in Figure 3, and it shows that the relationship between the initial review ratings and the subsequent searches is not just visually apparent, but also statistically significant.

[Insert Tables 3 and 4 about Here]

In the second regression, we explore the association between consumers’ checkins at restaurants and the information they acquire through viewing reviews and photos of those restaurants. We analyze this through a logistic regression:

\[
\text{logit}(\text{Checkin}_{ij}) = \alpha + \beta_1 \times \text{AvgStar}_{ij} + \beta_2 \times \text{AvgStarAll}_{ij} + \beta_3 \times \text{View}_{ij} + \epsilon_{ij}
\]

In this regression, \( \text{Checkin}_{ij} \) is the binary indicator of whether consumer \( i \) visited restaurant \( j \); \( \text{AvgStarAll}_{ij} \) is the average star rating of all reviews of the restaurant \( j \); \( \text{AvgStar}_{ij} \) and \( \text{View}_{ij} \) are the same as in the previous regression. The regression result is reported in Table 4. The coefficients show that the
average initial review ratings, the overall average review ratings, and the number of pages browsed are all positively related to the eventual checkin decisions, and the effects are all statistically significant. These results again echo the patterns in Figures 3-5 which are discussed above, and show that these relationships are also statistically significant. That the effect of average initial star rating is statistically significant even after controlling for the overall average ratings is particularly noteworthy, as it indicates higher impact of reviews a consumer sees earlier on, and underscores the importance of in-depth analyses of consumers’ sequential information acquisition behavior.

We also performed regression analysis of consumers’ information search and checkin activities after they view the first review page and after they view the first photo page. The results are similar to those discussed above. Overall, these reduced-form analyses reveal strong evidence that consumers acquire product review and photo information in a calculated manner, thus supporting our modeling of these activities using an optimization framework.

4. Model

We now set up our model. Certain aspects of our model are chosen to match the empirical setting of our study – the restaurant review website. However, the modeling framework is intended to be general enough to apply to a broader context where consumers proactively seek out product information to form quality beliefs and make purchase decisions. Therefore, in this section we use generic terms such as “product” and “purchase”, instead of “restaurant” and “checkin” which are specific to the empirical setting of our study.

There are $I$ consumers, each indexed by $i, i = 1, \ldots, I$. Time is discrete and is indexed by $t, t = 1, \ldots, T$. Consumers go to a product review web site to seek out information about one or more products, and they purchase products over time. There are $J$ products in the market, each indexed by $j, j = 1, \ldots, J$.

4.1 General Structure and Timeline

We first lay out the overall structure of consumer activities. Consumers engage in two types of activities: information acquisition and purchase. First, in each time period, a consumer may actively acquire product
quality information at the product review web site. If the consumer decides to acquire product information, she would take a sequence of information search actions. This sequence of actions performed in a time period is called a *session*. In each step of a session, the consumer may read additional product reviews of a product which she already knows (i.e. she has acquired information about the product before), view additional photos of a product which she already knows, or search for a new product. The consumer may stop searching after any step, and may decide not to start a session at all in a time period. Through searching products, reading reviews and viewing photos over time, a consumer gains knowledge about a set of products. For each product, the consumer forms the expectation about the product quality based on the information she acquires. This knowledge of the set of products constitutes her *information set*. We denote $I_{itk}$ as consumer $i$’s information set at time $t$ after the $k$-th search action.\(^6\)

The second type of activity is purchase. In each time period, a consumer may decide to purchase a product. For purchase, she will choose one product from her information set.\(^7\) For model simplicity, we assume that the information acquisition activity precedes product purchase in each time period. The timeline is illustrated in Figure 6.

[Insert Figure 6 about Here]

### 4.2 Utility and Consumer Learning

We first discuss a consumer’s utility and information learning. The indirect utility of product $j$ to consumer $i$ at time $t$ after the $k$-th search action, i.e. based on information set $I_{itk}$, is:

\[
U_{ijtk} = \theta_i(q_{ijtk} - r_i q_{ijtk}^2) + \epsilon_{ijtk}
\]

(1)

In equation (1), $\theta_i$ is consumer $i$’s valuation of quality. $q_{ijtk}$ is consumer $i$’s perceived quality of product $j$ at time $t$ after the $k$-th search. This is a subjective quality from consumer $i$’s perspective, based on

\(^6\) When the context is clear, we also use $I_{itk}$ to denote the set of products that are contained in the consumer’s information set.

\(^7\) For brevity, we refer to the set of products contained in a consumer's information set simply as the consumer's information set, when doing so would not cause confusion.
the information she has acquired to that point, contained in information set \( I_{itk} \). \( r_i \) is the risk coefficient, which indicates the extent of the consumer’s risk aversion \( (r_i > 0) \). \( e_{ijtk} \) is a random error which we assume follows a Type I Extreme Value distribution, which yields closed-form solution of choice probabilities at each purchase incidence.

A consumer does not know a product’s quality with certainty. Instead, she holds a belief about the quality, and updates her belief based on the information acquired at the web site. We model this as a Bayesian learning process. Specifically, a consumer starts with a prior on product quality that is common across all products.\(^8\)

\[
q_{ij0} \sim N(\mu_{i0}, \sigma_{i0}^2)
\]

As the consumer gains more information about the product, her perceived quality of the product evolves to:

\[
q_{ijtk} \sim N(\mu_{ijtk}, \sigma_{ijtk}^2)
\]

The information can be acquired through reading reviews of the product, viewing photos of the product, or seeing the product’s average review rating on the first page, i.e. the product’s main web page. A product review consists of a star rating, ranging from 1 to 5, and a paragraph of text describing the reviewer’s experience of the product. We assume that the quality signal a consumer takes from a product review follows a normal distribution:

\[
A_{ijtk}^r \sim N(\mu_{ijtk}^r, (\sigma_{ijtk}^r)^2)
\]

---

\(^8\) Note that the prior is assumed to be common across products but not common across consumers. The former is reasonable because, as will be discussed later, we explicitly account for the quality update based on the average review rating of a product the consumer sees to begin with, which would lead to different quality perceptions afterward. We do not assume it to be common across consumers, because the prior is also a proxy for the way a consumer gets a new product which may differ across consumers (but for which we do not have data). A consumer who sets a filter of, say, average rating above 3.5 stars, would have a higher initial quality prior than would another consumer who searches all products. In our empirical context, though, there is little evidence of consumers setting such filters, as the average star rating of restaurants consumers actually searched is 3.69, very similar to the overall average rating of all restaurants, which is 3.59.
In equation (4), $s_{ijtk}$ is the star rating of the review consumer $i$ reads in step $k$ of time $t$ about product $j$. That is, we model the information contained in a review as coming from a normal distribution, where the mean of the distribution depends on the star rating of that review. A review with higher star ratings carries a more positive quality message about the product, i.e. $\mu_j^r$ increases with $s$ (Zhao et al. 2013). However, all reviews with the same star rating may not indicate exactly the same quality, and even when reading the same review, different consumers may form different impressions. Therefore, the quality signal is modeled as a noisy one. In equation (4), $(\sigma_j^r)^2$ is the variance of this quality signal, where the superscript $r$ indicates the signal is derived from a product review. A larger variance indicates higher noise in the signal from the consumer’s perspective, hence lower speed of learning. We make this variance consumer-specific as people may process the same review with different levels of attention, hence getting signals of different precisions.

Similarly, we model the quality signal a consumer takes from a product photo as following a normal distribution:

$$ A_{ijtk}^p \sim N(\mu_j^p, (\sigma_j^p)^2) $$

In equation (5), the superscript $p$ indicates the signal is derived from viewing a photo, and $(\sigma_j^p)^2$ is the variance of the quality signal. Unlike reviews, photos do not contain star rating information. Instead, we model the mean of the quality signal as a linear function of the overall rating of the restaurant:

$$ \mu_j^p = \omega_0^p + \omega_1^p S_{j,ijtk} $$

In equation (6), $S_{j,ijtk}$ is the average star rating of product $j$ across all its reviews that are available at the time of consumer $i$’s $k$-th search action at time $t$ (i.e. all reviews posted at the website prior to the viewing action of the consumer). In modeling this way, we assume that a photo conveys information about product quality that is on average proportional to the product’s overall quality ratings. We expect $\omega_1^p > 0$, i.e. that photos of a higher quality product would indeed signal higher quality.
Furthermore, we model the quality signal a consumer takes from the average review ratings on the first page also as following a normal distribution:

\[ A_{ijtk}^f \sim N(\mu_{ijtk}^f, (\sigma_{ijtk}^f)^2) \]

In equation (7), the superscript \( f \) indicates the signal is derived from viewing the average review rating on the first page, and \( (\sigma_{ijtk}^f)^2 \) is the variance of the quality signal. Similar to modeling photos, we model the mean of this quality signal as a linear function of the overall rating of the restaurant at the time of this consumer search:

\[ \mu_{ijtk}^f = \omega_{ijtf}^f + \omega_{ijtS}^f S_{i,j,t,k} \]

The learning process follows the standard Bayesian update procedure. Prior to viewing a product review, the consumer’s quality belief is \( q_{ijt,k-1} \sim N(\mu_{ijt,k-1}, \sigma_{ijt,k-1}^2) \). After processing the quality signal \( A_{ijtk}^r \), the quality belief becomes:

\[ \mu_{ijtk} = \frac{\mu_{ijt,k-1} \sigma_{ijt,k-1}^2}{\sigma_{ijt,k-1}^2 + (\sigma_t^f)^2} + \frac{A_{ijtk}^r}{\sigma_t^f} \]

And

\[ \sigma_{ijtk}^2 = \frac{1}{\sigma_{ijt,k-1}^2 + (\sigma_t^f)^2} \]

Similarly, after processing the quality signal \( A_{ijtk}^P \), the quality belief becomes:

\[ \mu_{ijtk} = \frac{\mu_{ijt,k-1} \sigma_{ijt,k-1}^2}{\sigma_{ijt,k-1}^2 + (\sigma_t^P)^2} + \frac{A_{ijtk}^p}{\sigma_t^P} \]

And
(12) \[ \sigma_{ijtk}^2 = \frac{1}{\frac{1}{\sigma_{ijtk-1}^2} + \frac{1}{(\sigma_i^2)^2}} \]

Also similarly, after processing the quality signal \( A_{ijtk}^f \), the quality belief becomes:

(13) \[ \mu_{ijtk} = \frac{\mu_{ijtk-1} + A_{ijtk}^f}{\frac{1}{\sigma_{ijtk-1}^2} + \frac{1}{(\sigma_i^2)^2}} \]

And

(14) \[ \sigma_{ijtk}^2 = \frac{1}{\frac{1}{\sigma_{ijtk-1}^2} + \frac{1}{(\sigma_i^2)^2}} \]

Conditional on the consumer’s information set, the expected utility of product \( j \) to consumer \( i \), based on her information set after search step \( k \) at time \( t \) is:

(15) \[ E[U_{ijtk}|I_{ltk}] = \theta_i (\mu_{ijtk} - r_i \mu_{ijtk}^2 - r_i \sigma_{ijtk}^2) \]

In a purchase incidence of the consumer with this information set, assuming the utility of each product also contains the purchase incidence specific error term which follows a type I extreme value distribution, the probability of choosing product \( j \) by consumer \( i \) at time \( t \), conditional on a purchase being made, is:

(16) \[
\Pr(j|I_{ltk}) = \frac{e^{\theta_i (\mu_{ijtk} - r_i \mu_{ijtk}^2 - r_i \sigma_{ijtk}^2)}}{\sum_{j \in \Omega(I_{ltk})} e^{\theta_i (\mu_{ij'tk} - r_i \mu_{ij'tk}^2 - r_i \sigma_{ij'tk}^2)}}
\]

4.3 Proactive Learning

The setup of information learning and product choice discussed in the previous section is consistent with standard setups in the extant learning literature (e.g., Erdem and Keane 1996, Erdem et al. 2005, Erdem et al. 9 Considering the data we have, we model the choice probability conditional on a purchase being made, i.e. we model the choice aspect but not incidence aspect of the purchase decision. This way we do not have to make the assumption that checkins capture all purchase incidences. In another setting, e.g. using scanner panel data, the purchase incidence can easily be modeled by including an outside option.
We now discuss the modeling of consumers’ proactive acquisition of product quality information from reading reviews.

4.3.1. Search Actions

In each time period, a consumer may start a search session, which consists of a sequence of information acquisition actions. For each action in this sequence, the consumer may decide to read more review or view more photo of a product that she already knows, or retrieve information about a new product. Acquiring and processing such information are costly, so the consumer makes conscious decision on which type of information to acquire on what product. If the few reviews the consumer has read about a product are all very negative, for example, the consumer may not spend time to read additional reviews. In contrast, if the consumer has just found a new product and the initial information is highly positive, then she may be more likely to read more reviews to find out more about the product. If a lot of information has been gathered, then the consumer would be unlikely to search further about the product, since additional information, even though available, would be of marginal value.

We model each information search session as a dynamic programming problem, where the consumer performs guided searches to maximize her expected discounted utility. The actions available to consumer \(i\) at the \(k\)-th step at time \(t\) are to read a set of reviews of an existing product in her information set, to view a set of photos of an existing product in her information set, to retrieve a new product, and to stop the search session:\(^{10}\)

\[
a_{itk} = \begin{cases} 
R_j & \text{get a new set of reviews of product } j, j \in I_{it,k-1} \\
P_j & \text{get a new set of photos of product } j, j \in I_{it,k-1} \\
N & \text{get a new product} \\
S & \text{stop searching}
\end{cases}
\]

Searching for information is a costly activity. The cost of each search activity is denoted as:

---

\(^{10}\) Web sites typically show a set of product reviews or a set of photos in one web page, hence the model is set up as retrieving a set of reviews or photos, instead of a single review or photo, in each step. This setup includes retrieving a single review or photo in each step as a special case (the size of the set is one). The difference is at operational, not conceptual, level.
In the equation, $c_{ir}$ is the cost of obtaining and reading a set of reviews by consumer $i$; $c_{ip}$ is the cost of retrieving and viewing a set of photos; $c_{in}$ is the cost of retrieving and processing the information of a new product. The cost of stopping is set to zero.

### 4.3.2 Consumer’s Optimization Problem

We model each search session as a dynamic programming problem. In each step of the search process in the session, the consumer makes search decision to maximize her expected discounted utility.

\begin{equation}
V_i(l_{itk-1}) = \max_{a_{itk}, a_{itk+1}, \ldots} E \left[ \sum_{k'=k}^{\infty} \beta^{k'-k} u_i(a_{itk}|l_{itk'-1}) \right]
\end{equation}

In equation (19), $u_i(a_{itk}|l_{itk'-1})$ is the utility derived from each step $k'$, depending on the action she takes and the information set. When the search is ongoing, the consumer incurs search cost but does not derive any flow utility. When the consumer stops searching, the consumer derives utility from the information set that is accumulated, through expected future consumptions of the product:

\begin{equation}
u_i(a_{itk}|l_{it,k-1}) = \bar{u}_i(a_{itk}|l_{it,k-1}) + \varepsilon_{itk}(a_{itk}) = \begin{cases} -c_i(a_{itk}) + \varepsilon_{itk}(a_{itk}) & a_{itk} \neq S \\ TV_i(l_{itk}) + \varepsilon_{itk}(a_{itk}) & a_{itk} = S \end{cases}
\end{equation}

In the equation, $\varepsilon_{itk}(a_{itk})$ is an action-specific random term that follows the type I Extreme Value distribution, which yields closed-form solutions for the choice probabilities of search actions based on the value function. $TV_i(l_{itk})$ is the terminal value of the information set to the consumer. Since the information set

\footnote{Note that the dynamic optimization is assumed at the session level. In other words, we assume that a consumer optimally plots out search actions within each specific session, but not that the consumer optimally forward look into indefinite future for subsequent search sessions (instead of subsequent actions within the same search session). While a “fully optimal” consumer who anticipates and plans search sessions days or weeks later can be modeled without altering the framework much, we believe that is a much stronger optimality assumption which is unnecessary for the purpose of this study.}
assists the consumer in her subsequent purchase, this terminal value quantifies the subsequent utilities derived from purchase and consumption. Given the utility functions discussed in the previous section, the expected per-purchase utility is simply the logit inclusive value according to the information set (Gowrisankaran and Rysman 2012):

\[
IV_i(I_{itk}) = \ln(\sum_{j \in L_{itk}} e^{E[U_{ijtk}|I_{itk}]})
\]

Let \( \delta_i \) be the discount rate on purchase utilities, the terminal value of the information set is then:

\[
TV_i(I_{itk}) = \frac{1}{1-\delta_i} IV_i(I_{itk}) = \frac{1}{1-\delta_i} \ln(\sum_{j \in L_{itk}} e^{E[U_{ijtk}|I_{itk}]})
\]

The action-specific value function is:

\[
V_i(a_{itk}|I_{it,k-1}) = \bar{v}_i(a_{itk}|I_{it,k-1}) + \epsilon_{itk}(a_{itk})
\]

\[
= \bar{v}_i(a_{itk}|I_{it,k-1}) + \beta E[V_i(I_{itk}(a_{itk}|I_{it,k-1}))] + \epsilon_{itk}(a_{itk})
\]

And the choice probability for each search action is:

\[
Pr(a_{itk}|I_{it,k-1}) = \frac{e^{V_i(a_{itk}|I_{it,k-1})}}{\sum_{a' \in A(I_{it,k-1})} e^{V_i(a'|I_{it,k-1})}}
\]

### 4.4 Heterogeneity

Different consumers may have different initial quality beliefs, and may have different capacities or preferences on processing reviews and photos as quality signals. They may also differ in their costs of processing different types of information, and their degrees of risk aversion. In our study, heterogeneity among consumers is modeled using latent class (Kamakura and Russell 1989). Specifically, let there be \( M \) type of consumers, each of which accounts for \( \pi_m \) portion of the population. Each consumer type has different quality valuation, risk aversion, quality learning, and cost parameters.
4.5 Estimation

The parameters to be estimated are:

\[
\left\{ \{\pi_m, r_m, \{\mu_{m0}, \sigma_{m0}^2\}, \{(\sigma_m^p)^2, (\sigma_m^f)^2\}, \{c_{mr}, c_{mp}, c_{mn}, \delta_m\}\}_{m=1,\ldots,M}, \{\mu_s^P\}_{s=1,\ldots,S}, \omega_0^P, \omega_1^P, \omega_0^f, \omega_1^f, \beta \right\}
\]

The model can be estimated using Simulated Maximum Likelihood. The likelihood is constructed based on Equation (16) for purchase decisions and Equation (24) for search actions. Action-specific value function is needed to compute the likelihood according to equation (24). To solve for the value function, the solution to the dynamic programming problem of equation (19) is needed. We use the method developed in Keane and Wolpin (1994) to solve the dynamic programming problem through backward induction with interpolation. Since quality signals are not observed by the econometrician, the likelihood is calculated through simulated draws. The standard errors are estimated through bootstrapping. The detail of the estimation procedure is discussed in Appendix 1.

4.6 Identification and Normalization

The quality valuation and risk aversion parameters and the mean quality rating parameters \(\{\theta_m, r_m, \{\mu_{m0}, \sigma_{m0}^2\}, \omega_0^P, \omega_1^P\}\) are identified through information search actions depending on the information they have acquired, as well as through consumers’ purchase decisions. The prior quality belief and signal variance parameters \(\mu_{m0}, \sigma_{m0}^2, (\sigma_m^p)^2, (\sigma_m^f)^2\) are identified through the evolution of purchase decisions over time depending on the additional information consumers acquire along the way, and the evolution of search actions depending on the information they have acquired. The search cost parameters \(c_{mr}, c_{mp}\) and \(c_{mn}\) are identified through the consumers’ information search actions depending on the information they have acquired.

All search actions in each session happen in a relatively short period of time. Consistent with the prior literature (Kim et al. 2011), we set the discount rate parameter of the dynamic programming problem to \(\beta=1\), i.e. no discount within a search session. The discount rate on future purchase utilities, \(\delta_m\), is not identified, as for any change of \(\delta_m\), corresponding changes of search cost parameters can be made to yield
the same value function. Therefore, we normalize $\delta_m = 0$. This means that the search cost should be interpreted as the cost corresponding to each incidence of purchase (see Equation 22). Furthermore, a proportional change of $\theta_m$ can be compensated by corresponding changes of the quality parameters, the risk aversion parameters, and the signal variance parameters. We thus normalize this parameter for the first consumer type: $\theta_1 = 1$. Finally, learning related parameters also need normalization: the initial variance is normalized to 1 (Mehta et al. 2003); the signal quality variance of the average review rating on the first page is also normalized to 1, since a consumer will get this signal once and only once for each restaurant and it cannot be identified through the endogenous information acquisition process; the quality level of 3-star rating is normalized to 0, since review ratings signal qualities only on a relative scale.

5. Results

5.1 Empirical Specification

We first clarify a few specifications that cast our model to the institutional details of the data. At the restaurant review website, each review page contains twenty individual reviews. Therefore, the cost parameter in equation (18) corresponds to the cost of retrieving and reading twenty reviews. Similarly, each photo page contains fifteen photos. Both the quality signal parameters in equations (5) and (6) and the corresponding cost parameter in equation (18) thus represent the information available at, and the cost of retrieving and viewing, a page of fifteen photos.

Our dataset contains more browsing activities than checkins. Although browsing activities are the key to identifying the proactive learning model parameters, checkins are needed to compare our model with existing models that do not account for the endogenous information acquisition decisions but use only purchase information to identify product qualities. Considering this, we use the set of consumers who have at least two checkins over the time period. This subset contains 416 consumers. This subset of consumers obviously has on average more checkin activities, and also has on average more search activities. Importantly, though, these consumers have very similar reduced-form patterns as discussed in Section 3. The reduced-form regressions using only this subset of consumers are discussed in Appendix 2. The similar coefficient
estimates suggest that from the perspective of proactive learning, which is the focus of the study, this subset is qualitatively similar to the whole set of consumers.

Our dataset covers a 31-day period. We use the first 20 days for model calibration, and the remaining 11 days as holdout sample.

5.2 Parameter Estimates

Table 5 reports the parameter estimates of our proposed model. Two consumer segments are estimated which reveal distinct characteristics. To begin, a consumer in the second segment has considerably higher valuation of quality than the first segment ($\theta_2 = 6.005$ vs. $\theta_1$ which is normalized to 1). The consumers from both segments have similar initial quality perceptions ($\mu_{10} = 0.793$ and $\mu_{20} = 0.669$). Consumers in the second segment are more risk averse than consumers in the first segment ($r_2 = 0.064$ vs. $r_1 = 0.011$). About two-third of consumers belong to the first segment, while the remaining one-third belongs to the second segment.

The learning patterns also differ across the two segments. Consumers in both segments get quality signals of similar precision, hence updating quality beliefs at similar speed, from viewing photos ($(\sigma^p_1)^2 = 5.003$ and $(\sigma^p_2)^2 = 4.098$). Consumers in the second segment update beliefs from reading reviews more slowly than do consumers in the first segment ($(\sigma^r_1)^2 = 1.115$ and $(\sigma^r_2)^2 = 5.479$). For both segments, the variance parameters for photo are similar to or higher than those for review, suggesting that the information contained in one page of photos is similar to or lower than that in one review (recall that the data is precise to only page level for photos, but the star rating is known for each individual product review).

[Insert Tables 5 and 6 about Here]

The two segments also differ much in their costs of searching for information. For the first segment, retrieving a new restaurant has lower cost ($c_{1n} = 0.542$) than either viewing another page of photos or reading another page of reviews of an existing restaurant ($c_{1p} = 3.035$ and $c_{1r} = 3.488$), while the cost of viewing another page of photo and that of reading another page of reviews are statistically similar. For the
second segment, in contrast, the cost of retrieving a new restaurant is higher than that of viewing a new page of photos \( (c_{2n} = 2.895 \text{ and } c_{2p} = 2.096) \), while the cost of reading another page of reviews is higher \( (c_{2r} = 4.875) \). Other things equal, these estimates suggest that the first segment would tend to get information of new restaurants, while the second segment would favor either getting new restaurants or viewing photos. Between the segments, the cost of retrieving a new restaurant and that of reading reviews are lower for the first segment than for the segment, while the cost of viewing photos is the opposite.

Comparing the two segments in general, the second segment has much higher valuation of quality and is more risk averse. Furthermore, the second consumer segment incurs higher cost of reading reviews. These suggest that the second segment is on the balance more “serious” than the first segment. Consistent with this are the behavioral profiles of the two segments reported in Table 6, summarized by assigning each consumer to the more probable segment based on the likelihood. The table shows that consumers in the first segment on average searched 8.41 restaurants, while those in the second segment searched a similar 8.65 restaurants. However, consumers in the second segment on average read more review pages (2.52 vs. 1.98) and viewed many more photo pages (22.23 vs. 3.32). Consumers in the second segment thus on average search more than those in the first segment do, likely driven by their higher quality valuation and risk aversion, suggesting that they are more serious consumers.

Moving on to the mean quality parameters of reviews and photos, we can see that the mean quality level of a 1-star review is much lower than that of a 2-star review, which is somewhat close to that of a 3-star review \( (\mu_1^r = -1.465 \text{ and } \mu_2^r = -0.552, \text{ vs. } \mu_3^r \text{ normalized to 0}) \). The mean quality levels of a 4-star review and a 5-star review are much higher than that of a 3-star review \( (\mu_4^r = 1.733 \text{ and } \mu_5^r = 2.185) \). This suggests that reviews of 1-star are very detrimental to a consumer’s perception of the restaurant’s quality, while a 4-star or 5-star review helps significantly. The initial priors of both segments \( (\mu_{10} = 0.793 \text{ and } \mu_{20} = 0.669) \) are between the qualities implied by 3-star and 4-star reviews. Recall that the mean average review rating across all restaurants is 3.59 (Table 2), this suggests that consumers’ initial quality perception is fairly accurate, i.e. close to the actual average qualities of the restaurants. The quality signals of photos and average review
ratings have slopes of 1.532 and 1.172, respectively, similar to the scales as suggested by the quality levels of reviews of different stars.

We also estimated the model without consumer heterogeneity, i.e. with a single segment. The single segment model has noticeably worse likelihood than that of the two segment version discussed above ($-LL = 10713$ for the single segment version, vs. $-LL = 9924$ for the two-segment version discussed above). This shows that accounting for consumer heterogeneity is important.

### 5.3 Comparison with the Exogenous Learning Model

Reviews are often taken as exogenous quality signals in the extant literature. We now compare our proactive learning model with one such alternative model. In this alternative model, which we call the “exogenous learning” model, a consumer also updates her quality belief based on the product reviews and photo she actually reads or views. However, it is assumed that the consumer is exposed to these photos and reviews exogenously instead of proactively acquiring such quality signals, i.e. deciding to search a new restaurant, view another page of photos, or read another page of reviews. The utility and quality learning process of this exogenous learning model is identical to those of the proactive learning model, as specified in Section 4.2. However, since the endogenous search actions are not considered, the optimization process as specified in Section 4.3 and the related cost parameters do not apply to this exogenous learning model.

[Insert Table 7 about Here]

The parameter estimates for the exogenous learning model are reported in Table 7. Comparing this with the estimates of the proactive learning model shows clearly that ignoring the endogenous learning process leads to biased estimates on several fronts. First, the exogenous learning model overestimates the quality difference across the star ratings. While the quality levels estimated from the proposed model range from -1.465 for a 1-star review to 2.185 for a 5-star one, those estimated from the exogenous learning model span across a much wider range, from -3.497 for a 1-star review to 7.021 for a 5-star review. This difference is noteworthy: the reviews a consumer sees not only has a direct impact on her quality belief of the restaurant
and hence purchase probability, but also has an indirect effect in that it affects whether the consumer decides to seek out more reviews of the restaurant. For example, unfavorable reviews may discourage the consumer from acquiring more information of the restaurant, which would further reduce the attractiveness of the restaurant comparatively, given that consumers prefer more accurate understandings of the product quality. Since the exogenous learning model takes review reading actions as exogenous, it ignores this indirect link. The parameter estimates suggest that the effect of this indirect effect may have been loaded onto the parameters for the direct effect, hence overestimating the quality differences among reviews of different star ratings. Furthermore, the signal variances of product reviews and photos are mostly overestimated, i.e. the speed of learning from product reviews and photos are underestimated. This may have compensated for the overestimation of the quality differences among reviews of different ratings. Finally, the risk aversion estimates are also biased, with that for the first segment overestimated while that for the second segment underestimated. Taken together, these show that accounting for the endogenous nature of consumers’ information acquisition process is necessary for an accurate understanding of the effect of product reviews and the related consumer learning.

6. Simulation

The key point highlighted by the proactive learning model is that consumers determine whether and what quality signals to acquire conditional on their existing information sets. Consequently, the sequence in which consumers acquire different quality signals is important. In this section, we perform two simulations to investigate this. In the first simulation, we calculate the incremental effect of reviews of different star ratings, and see the difference in such effect when the same reviews show up in different places. In the second simulation, we calculate the effect of sorting the reviews of each restaurant based on different criteria.

6.1 The “Value” of a Review

In the first simulation, we evaluate the incremental effect of a product review of a certain star rating. Since search action is endogenous and history dependent, the effect will depend on the place where this review shows up. We therefore evaluate the effect of a product review which is displayed when a user first retrieves a
restaurant, and when a user loads the next page of product reviews. We perform the simulation as follows.

We use the actual dataset of consumer search actions. To evaluate the effect of an $x$-star review ($x$ from 1 to 5) when it is displayed when a user first retrieves a restaurant page (or when the user loads the next page of reviews), for each actual action of retrieving a new restaurant in the dataset (or retrieving the next page of reviews), we simulate the subsequent search and purchase activities by replacing the review with the lowest star rating in that page with a review of $x$-star.

[Insert Table 8 about Here]

The result of the simulation is reported in Table 8. Both search and purchase activities are summarized for each star rating, for both the restaurant’s main page (“first page” column in the table) and the next review page (“second page” column in the table), where 1-star is used as baseline with which other star ratings are compared. The table clearly shows the higher value of a review with higher star rating: compared with a 1-star review, a 4-star review and a 5-star review on the first page increase the amount of search activities by 3.66% and 4.53%, respectively. The increase in purchase probability is even more striking: compared with a 1-star review, a 4-star review and a 5-star review increase the purchase probability by 13.44% and 17.41%, respectively. That one single review can have such large impact on the purchase probability demonstrates the importance of product reviews, especially those which consumers see early on.

This impact contains both the direct effect, where the review affects the consumer’s quality perception and hence purchase probability, and the indirect effect, where the review also changes the subsequent search activity which then also affects purchase probability. Not accounting for the effect on subsequent search activity (by holding the subsequent search unchanged in the simulation), a 4-star review and a 5-star review would increase the purchase probability by 8.00% and 10.17%, respectively. The direct effect thus roughly accounts for 60% of the change in purchase, while the indirect effect through changing subsequent search accounts for about 40% of the impact. Comparatively, a 4-star review and a 5-star review on the second page increase the amount of search activities by 1.08% and 1.28%, respectively, and the purchase probabilities by 4.63% and 5.37%, respectively. All these are also positive, although the effect sizes are significantly lower than
their counterparts for the first page. The value of a review, therefore, crucially depends on where the review is located, in addition to its star rating. These results highlight the importance of understanding the sequential nature of the process of consumers learning from reviews. Note that the probabilities for the second page are calculated conditional on the consumer having actually visited that page. The actual difference between the impact of a review on the first page and that on the second page is therefore even higher, as many consumers may not visit the second page at all. Industry managers would benefit from this detailed account of the value of a review with its nuanced aspects, as it may have important business implications. For example, certain existing studies have shown that review ratings decline over time (Li and Hitt 2008, Wu and Huberman 2008, Godes and Silva 2012). When product review web sites display the most recent reviews first, then, the first few reviews a consumer sees may have lower ratings. Industry managers thus need to be especially cognizant about the outsized impact of these early reviews on the consumer’s purchase decisions.

6.2 Sorting Reviews

Although initially product review sites simply ordered review chronologically with most recent review displayed first (as is done at the web site for this study), many subsequently added functionalities which allow users to sort reviews in certain orders, or to specify sorting preferences so reviews will be sorted automatically based on the user preferences. The ordering of the reviews is expected to have significant impact given the endogenous information acquisition process. In the second simulation, we evaluate the effect of different sorting criteria of product reviews. We evaluate three alternative scenarios, each corresponding to a specific sorting criterion: 1) best-first, where the reviews of the restaurant are sorted from highest star ratings to the lowest; that is, a consumer will be shown all 5-star reviews before shown all 4-star reviews, and so on; 2) worst-first, where reviews are sorted from those with lowest ratings to those with highest ratings; 3) even distribution, where reviews are sorted such that a review of each of the five star ratings is shown in turn until it exhausts all reviews of a certain star rating, followed by the same procedure for the remaining stars, and so on. Each scenario is compared with the original one, where reviews are sorted temporally from most recent to the oldest. We sorted the reviews based on these criteria for all restaurants in the dataset, and performed forward simulation of search and purchase given the newly sorted reviews.
The effect of such sorting criteria on purchases is reported in Table 9. Since the sorting criteria will affect restaurants with different qualities differently, the result is summarized according to the average review ratings of the restaurant. As expected, across all sorting criteria, restaurants with higher star ratings get more purchases than those with lower star ratings. Comparing across the sorting criteria, we can see that restaurants with highest average ratings (4-star and above) get the highest purchase amount under the worst-first sorting criterion, while those restaurants get the lowest purchase amount under the best-first sorting criteria. Considering that the average ratings are in general positive (3.59 stars in our data), restaurants would usually have more 4-star and 5-star reviews than 1-star and 2-star ones. The best-first sorting criterion, therefore, would not show consumers much difference across restaurants in the first one or two pages of reviews. This results in relatively closer quality perceptions across restaurants in consumers’ information sets, when consumers stop reading further after one or two pages due to search costs, hence smaller differences in purchase probabilities. In contrast, the worst-first sorting criterion would highlight the difference across restaurants – a very high quality restaurant may have only a few 1-star reviews while an average restaurant may have many more, and this will all be shown on the first page so consumers can easily see. The effects of the current chronological sorting and the even distribution sorting criteria are between those of the best-first and worst-first, where the purchase amount for the highest rated restaurants is slightly higher under the current sorting criterion than under the even distribution sorting criterion. By having an equal representation of each star rating as much as possible, the even distribution sorting criterion leads to smaller difference (based on the first few pages of reviews) among restaurants than chronological sorting does, which likely would show more reviews close to the restaurants’ average rating to begin with. Since product reviews are quality signals, a diagnostically efficient system should see higher quality products and lower quality products clearly differentiated. Considering this, the simulation result suggests that the worst-first sorting criterion has the highest diagnostic efficiency across the sorting criteria, while the best-first sorting criterion is the least diagnostically efficient.
7. Conclusion

Online product reviews have become a major component of Internet commerce, greatly facilitating consumer-to-consumer knowledge transfer. Given the ubiquity and prominence of product reviews, researchers have naturally devoted much effort to understanding their impact. Aggregate statistics such as average review ratings have been shown to significantly impact sales and other marketing outcomes in many different settings. Recent studies have also analyzed the effect of reviews at individual level, using learning models to analyze the information consumers extract from individual reviews. Existing studies on the effect of WOM usually treat product reviews as exogenous quality signals. However, product reviews do not get pushed into consumers’ eyes automatically. Instead, reading review is a time consuming process, and it takes an interested consumer to decide to do that consciously. Whether a consumer would decide to read a review of a product depends on what the consumer already knows in general, what she already knows about the product in specific, and what she expects to learn from additional reviews. Treating product reviews as exogenous signals ignores the dynamics in this sequential process, and will likely lead to biased understandings.

To accurately understand how product reviews affect consumer decisions, it is necessary to first understand how consumers read product reviews. By endogenizing consumers’ review reading behavior in a rational framework, the proactive learning model developed in this study fills this gap in the literature. The unique dataset with consumer browsing behavior used in this study enables the model estimation, and sheds light on the intricacies of consumers’ information acquisition process. The dataset shows strong evidence of consumers consciously deciding what to read, and of the significant impact of such information acquisition actions on sales. Positive initial reviews are shown to lead consumers to read further, and they lead to higher purchase probabilities even after the average review ratings are controlled for. Parameter estimates of the proactive learning model further reveal different behavioral profiles of acquiring information from product reviews, with consumers differing in their valuation of quality, risk aversion, as well as the cost of reading reviews and the speed with which they learn from them. The parameter estimates also show the quality differences as suggested by reviews of different ratings, with 4-star and 5-star reviews indicating significantly
higher quality levels than other reviews, while 2-star and 3-star reviews indicating somewhat similar quality levels. Equally importantly, the analysis shows that if we fail to recognize that review reading is a guided and deliberate process, but instead treat reviews as exogenous quality signals, we will get biased estimates on the difference between reviews of different star ratings, and on consumers’ speed of learning from reviews.

Our analysis has significant managerial implications. Constant on the mind of industry managers is the question of the value of product reviews. In addition to giving a more accurate measure of the quality differences between reviews of different ratings than those measured using an exogenous model, our analysis also shows that the value of a product review crucially depends on where the review is located. The impact of a review on the first page on subsequent searches and purchases is shown to be about three times higher than that of a review on just the next page. The earlier reviews a consumer sees, in other words, are much more consequential than later ones. That the same review can have different effect when placed at different pages is an important takeaway for managers, and suggest that sorting product reviews in different ways can lead to different results. Indeed, through simulation we show that sorting reviews from lowest ratings to highest ratings is the most efficient in terms of maximizing the difference across products of different qualities, while sorting reviews in the opposite order is the least efficient, and chronological sorting and even sorting lead to results in between those two. All these findings inform managers on the effect of reviews and the design of Internet WOM platforms.

Several limitations of our study call for future research. First, existing research has shown that reviews not only differ in ratings, but also differ in credibility (Zhao et al. 2013). Furthermore, reviews written by different reviewers may also have different implications (Wu et al. 2015). While our study goes deeper on the dynamic aspect of the information acquisition process of reading reviews than has been done in the extant literature, the dynamic framework also limits our ability to account for such horizontal differentiation of reviews and reviewers. Incorporating such richer heterogeneity is left for future work. Second, similar to reviews and reviewers, products can also be horizontally differentiated. In fact, the competitive landscape among products can be much more complex. Restaurants clustered together in location, for example, will see
fiercer competition than restaurants farther apart. Incorporating horizontal differentiation and competition among firms will further improve our understanding of the effect of product reviews. Finally, our study focuses on the reading of reviews but does not account for their creation. However, review creation is the logical next step after purchase and consumption, and is also endogenous. If reading positive initial review leads a consumer to read more reviews and become more likely to purchase, it will also increase the probability of the consumer writing a review later on, which is an additional quality signal to subsequent consumers. The effect of a review, therefore, is likely even higher once this subsequent creation of additional reviews is accounted for. Such a “close-loop” analysis has the potential to provide a complete picture around online product reviews, and is an exciting topic for future research.

References


Appendix

A.1 Estimation

We estimate our model using the Simulated Maximum Likelihood (SMLE) approach. The model parameters to be estimated are (certain parameters are normalized as discussed in Section 4.6):

$$\Theta = \left\{ \{\pi_m, \{\theta_m, r_m\}, \{\mu_{m0}, \sigma^2_m\}, \{\sigma^2_{m}, (\sigma^2_{m})^2, (\sigma^2_{m})^3\}, \{c_{mr}, c_{mp}, c_{mn}, \delta_m\}\right\}_{m=1,...,M'} \{\mu_s\}_{s=1,...,S}, \omega_0, \omega_1, \omega_2, \omega_3, \beta \right\}$$

Central to the model estimation is the task of solving the dynamic programming problem of equation (19). The solution to this dynamic programming problem is the set of value functions $V_m(1|\Theta), m = 1, ..., M$. Each value function corresponds to a type of consumers in the latent class setup. State is a consumer’s information set, which consists of the set of products the consumer is aware of. Each product in the information set is represented by the mean and variance of the consumer’s quality belief. To solve for the value function, we use the backward-induction with interpolation method developed in Keane and Wolpin (1994). The regressors for the interpolation regression are the inclusive value of the information set, the number of products, and the variances of quality perception of the three products with the highest mean utility. From the estimated value function, the estimated action-specific value function (equation 23) is calculated through simple average of 30 simulated draws. We denote the estimated value function and action-specific value function as $\hat{V}_m(1|\Theta)$ and $\hat{V}_m(a,1|\Theta)$. The estimated value functions allow us to compute the likelihood of search actions based on equation (24).

To calculate the likelihood of both search actions (equation 24) and purchase actions (equation 16), we need the consumer’s information set. This information set is not fully observed by the econometrician, as only the search actions are known, but not the exact quality signals $A^S_{itk}, A^P_{itk}$ and $A^T_{itk}$. We handle this through numerical simulation. For each consumer $i$, we generate $N$ sequence of draws $\bar{a}_n^i, n = 1, ..., N$. Each element in each draw is used to construct the equality signal for the corresponding search action. The log-likelihood is then computed as:

$$LL(D|\Theta) = \sum_{i=1}^I \ln \left( \sum_{m=1}^M \pi_m \left( \frac{1}{N} \sum_{n=1}^N L_m(D_i|\Theta, \bar{V}_m, \bar{a}_n^i) \right) \right)$$

In the equation, $D$ represents the whole dataset, i.e. the search and purchase actions of all consumers, while $D_i$ represents the sequence of search and purchase actions of consumer $i$. The function $L_m(\cdot)$ is the likelihood of the data for consumer of type $m$:

$$L_m(D_i|\Theta, \bar{V}_m, \bar{a}_n^i) = \prod_{t=1}^t \prod_{k=1}^{K_{it}} \Pr(a_{itk}|\Theta, \bar{V}_m, \bar{a}_n^i) \prod_{t=1}^t \Pr(c_{it}|\Theta, \bar{a}_n^i, p_{it} = 1)$$

The first component of the function represents the likelihood of the search actions, while the second component represents the likelihood of the purchase choice actions conditional on the purchase incidences ($c_{it}$ denote the choice of consumer $i$ at time $t$, while $p_{it}$ is the binary indicator of whether the consumer made a purchase at the time).

The simulated likelihood is then maximized using standard numerical optimization methods.
The standard error of parameter estimates are computed through bootstrapping, by resampling the data with replacement 20 times.

A.2 Reduced Form Regression using the Subset of Consumers
As stated in Section 5.1, in order to enable comparison with existing models which treat reviews as exogenous quality signals and thus rely on purchase for identification, we estimated the model using the subset of consumers with at least two checkins in the dataset. To ensure that this subset of consumers demonstrate similar reduced form patterns as discussed in Section 3.3 of the main paper, which is critical for the identification of the proactive learning behavior, we performed the same regression analyses on this subset.

Table A2-1 reports the result of the first regression using this subset. This corresponds to Table 3 of the main paper. As the result shows, the coefficient of the initial review rating is positive and statistically significant at .001 level, the same as discussed in Section 3.3. Table A2-2 reports the result of the second regression using this subset. This corresponds to Table 4 of the main paper. The coefficients for initial review rating, overall average rating, and number of searches are all positive and statistically significant, the same as discussed in Section 3.3. These show that the proactive learning behavior is qualitatively similar between this subset of consumers and the whole set.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.109(***)</td>
</tr>
<tr>
<td>AvgStar</td>
<td>0.289(***)</td>
</tr>
</tbody>
</table>

Dependent Variable: Number of Views
Link Function: Poisson

Table A2-2: Reduced-form Regression Using Consumer Subset – Checkin

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.636 (***)</td>
</tr>
<tr>
<td>AvgStar</td>
<td>0.197 (*)</td>
</tr>
<tr>
<td>AvgStarAll</td>
<td>0.275 (*)</td>
</tr>
<tr>
<td>NumView</td>
<td>0.132 (***)</td>
</tr>
</tbody>
</table>

Dependent Variable: Checkin
Link Function: Binomial
Significance Level: *** < 0.001 < ** < 0.01
Tables and Figures

Table 1: Descriptive Statistics – Consumer Actions

<table>
<thead>
<tr>
<th>Per-User Statistics</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Restaurants Browsed</td>
<td>4.92</td>
<td>10.4</td>
<td>1</td>
<td>483</td>
</tr>
<tr>
<td>Number of Review Pages Read</td>
<td>1.15</td>
<td>3.51</td>
<td>0</td>
<td>121</td>
</tr>
<tr>
<td>Number of Photo Pages Viewed</td>
<td>3.03</td>
<td>10.6</td>
<td>0</td>
<td>244</td>
</tr>
<tr>
<td>Number of Checkins</td>
<td>0.48</td>
<td>2.25</td>
<td>0</td>
<td>123</td>
</tr>
<tr>
<td>Number of Consumers</td>
<td>6518</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Restaurants</td>
<td>2367</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Days</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Descriptive Statistics – Restaurant Reviews

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Star Rating</td>
<td>3.59</td>
<td>0.55</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Number of Reviews</td>
<td>77.81</td>
<td>190.98</td>
<td>1</td>
<td>2764</td>
</tr>
<tr>
<td>Number of Restaurants</td>
<td>2367</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Reduced-form Regression – Number of Views

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.822 (***)</td>
</tr>
<tr>
<td>AvgStar</td>
<td>0.194 (***)</td>
</tr>
</tbody>
</table>

Dependent Variable: Number of Views
Link Function: Poisson
***: significant at 0.001 level

Table 4: Reduced-form Regression – Checkin

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-4.479 (***)</td>
</tr>
<tr>
<td>AvgStar</td>
<td>0.217 (***)</td>
</tr>
<tr>
<td>AvgStarAll</td>
<td>0.229 (**)</td>
</tr>
<tr>
<td>NumView</td>
<td>0.088 (***)</td>
</tr>
</tbody>
</table>

Dependent Variable: Checkin
Link Function: Binomial
Significance Level: *** < 0.001 < ** < 0.01
### Table 5: Parameter Estimates – Proposed Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>6.005 (0.743)</td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.011 (0.012)</td>
<td>0.064 (0.030)</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>0.793 (0.122)</td>
<td>0.669 (0.125)</td>
</tr>
<tr>
<td>$(\sigma^p)^2$</td>
<td>5.003 (2.254)</td>
<td>4.098 (0.478)</td>
</tr>
<tr>
<td>$(\sigma^r)^2$</td>
<td>1.115 (0.193)</td>
<td>5.479 (0.627)</td>
</tr>
<tr>
<td>$c_n$</td>
<td>0.542 (0.036)</td>
<td>2.895 (0.226)</td>
</tr>
<tr>
<td>$c_p$</td>
<td>3.035 (0.110)</td>
<td>2.096 (0.039)</td>
</tr>
<tr>
<td>$c_r$</td>
<td>3.488 (0.045)</td>
<td>4.875 (0.143)</td>
</tr>
<tr>
<td><strong>Segment Proportion</strong></td>
<td>71.30%</td>
<td>28.70%</td>
</tr>
<tr>
<td><strong>Quality Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Review</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1^r$</td>
<td>-1.465 (0.086)</td>
<td></td>
</tr>
<tr>
<td>$\mu_2^r$</td>
<td>-0.552 (0.380)</td>
<td></td>
</tr>
<tr>
<td>$\mu_3^r$</td>
<td>normalized to 0</td>
<td></td>
</tr>
<tr>
<td>$\mu_4^r$</td>
<td>1.733 (0.136)</td>
<td></td>
</tr>
<tr>
<td>$\mu_5^r$</td>
<td>2.185 (0.101)</td>
<td></td>
</tr>
<tr>
<td><strong>Photo</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_1^p$</td>
<td>1.532 (0.020)</td>
<td></td>
</tr>
<tr>
<td><strong>Average Rating</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_1^f$</td>
<td>1.172 (0.128)</td>
<td></td>
</tr>
<tr>
<td><strong>-LL (In-sample)</strong></td>
<td>9924</td>
<td></td>
</tr>
<tr>
<td><strong>-LL (Holdout)</strong></td>
<td>7644</td>
<td></td>
</tr>
</tbody>
</table>

Numbers in parentheses are standard errors

### Table 6: Consumer Behavioral Patterns

<table>
<thead>
<tr>
<th>Average Number of</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurants Browsed</td>
<td>8.41</td>
<td>8.65</td>
</tr>
<tr>
<td>Review Pages Read</td>
<td>1.98</td>
<td>2.52</td>
</tr>
<tr>
<td>Photo Pages Viewed</td>
<td>3.32</td>
<td>22.23</td>
</tr>
</tbody>
</table>

### Table 7: Parameter Estimates – Exogenous Learning

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td></td>
<td>115.08 (20.156)</td>
</tr>
</tbody>
</table>
Table 8: Simulation – Change in Activities when Changing One Review

<table>
<thead>
<tr>
<th>Star Rating</th>
<th>Change in Search Activities</th>
<th>Change in Purchases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Page</td>
<td>Second Page</td>
</tr>
<tr>
<td>1-Star</td>
<td>baseline</td>
<td></td>
</tr>
<tr>
<td>2-Star</td>
<td>0.30%</td>
<td>0.17%</td>
</tr>
<tr>
<td>3-Star</td>
<td>0.76%</td>
<td>0.32%</td>
</tr>
<tr>
<td>4-Star</td>
<td>3.66%</td>
<td>1.08%</td>
</tr>
<tr>
<td>5-Star</td>
<td>4.53%</td>
<td>1.28%</td>
</tr>
</tbody>
</table>

Table 9: Simulation – Number of Purchases per-Restaurant by Sorting Criteria

<table>
<thead>
<tr>
<th>Restaurant Star Rating</th>
<th>Sorting Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>&lt;2</td>
<td>1.67</td>
</tr>
<tr>
<td>2-3</td>
<td>3.97</td>
</tr>
<tr>
<td>3-4</td>
<td>10.28</td>
</tr>
<tr>
<td>4-5</td>
<td>21.92</td>
</tr>
</tbody>
</table>
Figure 1: Number of Reviews by Star Rating

Figure 2: Histogram of Restaurant Average Star Ratings

Figure 3: Search and Checkin Activity by Initial Review Ratings Viewed
Figure 4: Checkin Activity by Overall Review Ratings

Figure 5: Checkin Activity by Number of Searches
Figure 6: Model Timeline

Search action: read review, view photo, or search a new product
Checkin: proxy for purchase