The following publication Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. Tourism Management, 52, 498-506 is available at https://doi.org/10.1016/j.tourman.2015.07.018

ANALYSIS OF THE PERCEIVED VALUE OF ONLINE TOURISM REVIEWS: INFLUENCE OF READABILITY AND REVIEWER CHARACTERISTICS

Abstract

Online reviews provide additional product information to reduce uncertainty. Hence, consumers often rely on online reviews to make decisions. However, an explosion of online reviews brings the problem of information overload to individuals. Identifying valuable reviews from massive reviews becomes increasingly important to both consumers and companies, especially for experience products like attractions. Several online review platforms provide a function for voting a review as "helpful" to enable readers to rate whether the review is valuable. Unlike consumers, companies want to detect potential valuable reviews before it has been voted as valuable to avoid its negative influence or to promote its positive influence. Using online attraction review data retrieved from TripAdvisor, which is a famous travel product website, we conducted a two-level empirical analysis to explore factors that affect the value of reviews. We introduced a negative binomial regression model at the review level to explore the effects of the review itself, and then applied a Tobit regression model at the reviewer level to investigate the effect of reviewer characteristics. Finally, we identified two sets of factors that affect the perceived value of a review. One is text readability, and the other is the reviewer's sentiment. The characteristics of the reviewers are inferred from properties of historical rating distribution. These findings have direct implications for attraction managers to identify potential valuable reviews better.

Key words: Online review, review helpfulness, text readability, historical rating distribution

1. Introduction

As a form of user-generated content, online reviews are important information sources of consumer experience towards products. Online reviews do not only appear on product-selling websites, such as Amazon.com, but also on travel websites, such as Expedia and TripAdvisor. Research has been conducted to demonstrate the significant influence of reviews on consumer decision-making process for both search products and experience products in product-selling websites (Chevalier & Mayzlin, 2006; Duan, Gu, & Whinston, 2008; Forman, Ghose, & Wiesenfeld, 2008; Gu, Park, & Konana, 2012). As a typical experience product, performance of tourism-related products could also be influenced by online reviews. According to Collie (2014), 65% of leisure travelers will search online before deciding on a travel destination, and 69% of their plans are determined by online travel reviews. Prior research also claimed that travelers consider the reviews of past tourists in deciding on their trips (Gretzel & Yoo, 2008; Z. Liu & Park, 2015; Vermeulen & Seegers, 2009; Ye, Law, & Gu, 2009). Most reviews focus on hotels or restaurants, and pay little attention to attractions. Although they are all tourism-related experience products, they are not exactly the same. The overall quality of hotels can be inferred from their stars that are assessed by an official organization according to a unified standard, whereas attractions do not have a similar evaluation system. When comparing restaurants with attractions, consumers will face smaller losses if they choose a terrible restaurant compared with a disappointing attraction. Choosing a restaurant merely means a meal, whereas choosing an attraction needs an entire travelling plan including traffic, time, etc. Hence, consumer decision-making process in attraction selection would not be the same as in hotel and restaurant selection. Therefore, exploring the effect of online reviews on attraction decision is important.

Although online reviews provide convenience to consumers to have a comprehensive understanding of attractions and make decisions, the availability of hundreds of reviews creates a problem in information overload. For example, La Jolla Cove in California is a famous attraction, but potential travelers do not know whether it is worthy of its reputation, since it may be just a cove like other coves with nothing special. These potential travelers who are planning to go to La Jolla Cove want to find the answer from reviews on travel websites, such as TripAdvisor (http://www.tripadvisor.com/). However, TripAdvisor has more than 1,300 reviews about La Jolla Cove and reading all the reviews seems impossible. Therefore, TripAdvisor has designed a feature called "Was this review helpful?" to help travelers quickly identify the most helpful ones among the whole bunch of reviews. A "yes" button can be clicked by readers to rate the review. Through this function, travelers easily find the reviews voted most helpful by other travelers. This function is also meaningful to attraction managers because reviews serve as a tool not only for consumers to decide, but also for managers to improve their service quality. The reviews voted most helpful contain opinions most trusted by readers before or after traveling, have great influence, and thereby, are the most valuable reviews. Therefore, site managers should identify reviews that will potentially be voted most helpful and should fix the problems revealed in reviews before they could influence the decisions of potential customers.

The content is the most important factor that contributes to the value of a review, especially for attracting visitors, because information quality is critical in reducing uncertainty (Mudambi & Schuff, 2010). Although hotels and restaurants are both tourism-related experience products, their review contents do not play the same important role for decision making regarding attractions. Review content is less important for hotel reviews because hotel quality could be assessed mostly by its star and review scores for different aspects like cleaning, surroundings, etc. Review content in hotel reviews is more likely to provide details and support the review ratings. Restaurant reviews are also not substantially important because describing the exact taste of food is too difficult. Nevertheless, review content plays an important role in attraction reviews for two reasons. First, as attractions cannot be rated by some standard aspects like hotels, reviewers rely on review content more to state their experience. Moreover, potential travelers will also read review contents carefully for them to be acquainted with the attractions and to decide whether to go or not. Hence, the written style (readability) of a review, which represents how easily a review can be understood, would probably influence its value. Several studies have been conducted to explore the effect of review linguistic characteristics on review value (Ghose & Ipeirotis, 2011; Hao, Li, & Zou, 2009; Jeon, Croft, Lee, & Park, 2006; Kusumasondjaja, Shanka, & Marchegiani, 2012; J. Liu, Cao, Lin, Huang, & Zhou, 2007; Z. Liu & Park, 2015; Weimer & Gurevych, 2007), whether readability will affect perceived value of attraction reviews remains an open question.

Moreover, how reviewer characteristics are inferred from his historical rating distribution influence as based on the perceived value of his review has not been answered. On TripAdvisor, readers could easily access the historical rating distribution of reviewers. Figure 1 is an example of a review on TripAdvisor. Historical rating distribution can reflect the personal preferences of the author such as rating criteria. According to the personal preferences inferred from historical rating distribution, readers can understand the meaning of the review more precisely. Hence, exploring the influence of historical rating distribution to perceived value of reviews is interesting.

****

Please Place Figure 1 Here

****

This study explores the factors that influence the perceived value of reviews. Using a dataset retrieved from TripAdvisor, we identified two sets of factors influencing review value, namely, review- and reviewer-related factors. Review-related factors are mainly about review text readability and rating, whereas the set of reviewer-related factors in this study includes whether the reviewer is positive and whether his mode rating is lower than mean rating.

Our empirical analysis yields three interesting findings. First, text readability exerts significant influence on the perceived helpfulness of reviews. Second, reviews would be perceived as more valuable when they express extreme sentiment. Third, the personal preference of reviewers is found to play an important role in influencing their perceived trustworthiness, and thereby affects the perceived value of their reviews. Specifically, readers are more likely to trust reviewers with higher mean historical ratings because such reviewers seem to be more positive. Although both positive and negative extreme sentiment will make reviews more valuable, reviews written by reviewers whose mode rating is lower than mean rating (positive skewness index) are more likely to gain more helpfulness votes.

The rest of the study is organized as follows. Section 2 presents related literature and hypothesis development. Section 3 includes econometric model specification and variable descriptions. Empirical results are presented in Section 4. Contributions and implications are discussed in Section 5. We conclude this study and present findings and limitations in Section

6.

2. Literature Review

In this study, we defined the value of a review as the helpfulness votes received or its perceived helpfulness. Hence, helpfulness and valuable are used interchangeably in the study. There are two main components influencing the value of reviews. The first component is the review itself. A review includes review content and rating. Many studies have been conducted to assess the value of a review by analyzing review content through natural language processing. Review length will influence perceived helpfulness significantly by exploring the effect of length according to machine learning approaches such as SVM (Jeon et al., 2006; Kim, Pantel, Chklovski, & Pennacchiotti, 2006; J. Liu et al., 2007; Weimer & Gurevych, 2007). Hao et al. (2009) explored the effect of review linguistic characteristics (length and subjective) on helpfulness in the movie industry. In the product review context, the extremity and depth of reviews affect perceived helpfulness (Mudambi & Schuff, 2010), especially when reviewers disclosed their identity (most reviewers do so). A negative review is deemed more credible than a positive review (Kusumasondjaja et al., 2012). Moreover, although both anxiety and anger are negative emotions, anxiety has more influence power than anger (Yin, Bond, & Zhang, 2014).

Besides the sentiment of the review, text readability is important to readers. Ghose, Ipeirotis, and Li (2012) proposed a new ranking system for hotel search engines by mining consumer reviews and by considering the readability of the review as one major factor in their system. In crowdfunding market, Burtch, Ghose, and Wattal (2013) found that a high readable project description would attract more contributors (investors). Specifically, helpful reviews that contain both pros and cons are expressed in clear writing and include product usage information and details. By contrast, unhelpful reviews are overly emotional or biased and lack of information (Connors, Mudambi, & Schuff, 2011; Wu, Heijden, & Korfiatis, 2011). Korfiatis, García-Bariocanal, and Sánchez-Alonso (2012) claimed that the readability of a review has greater influence than its length.

The other component is the reviewer. Hochmeister, Gretzel, and Werthner (2013) revealed that destination experts on TripAdvisor receive significantly more helpfulness votes from other community members than ordinary members. This finding indicates that the identity of the review generator has a significantly positive correlation with helpfulness of the review. Park, Xiang, Josiam, and Kim (2013) observed that self-disclosed personal profile information enables readers to evaluate whether the review generator is credible. We argue that trust in the reviewer results is trust in the review, that is, the identification of the review as helpful.

Other studies have explored the effect of review content and reviewer characteristics on review value (Ghose & Ipeirotis, 2011; Y. Liu, Huang, An, & Yu, 2008; Z. Liu & Park, 2015; Otterbacher, 2009). By mining IMDB reviews, Y. Liu et al. (2008) found that reviewer expertise and writing style influence review helpfulness received. Otterbacher (2009) further discovered that the number of reviews posted by a reviewer and the number of helpful votes the reviewer received will increase the helpfulness vote of the review in Amazon. Ghose and Ipeirotis (2011) combined the linguistic characteristics of reviews and reviewer characteristics to predict product review helpfulness vote on Amazon.com. They selected three product categories (audio and video players, digital cameras, and DVDs) and found that both review linguistic characteristics (measured by readability and spelling errors) and reviewer personal characteristics (measured by average helpfulness per review and personal information disclosure) do not consistently exert significant influence among these three categories. Z. Liu and Park (2015) investigated the effect of review and reviewer on the value of restaurant reviews in London and New York City. They found that review characteristics (rating, length, and readability) and reviewer characteristics (expertise, reputation) affect the perceived value of a review.

Although a large numbers of works has explored and confirmed the effect of text readability on review value, they rarely paid attention on attraction reviews. As we demonstrated before, readers will have different attitudes towards reviews when they are reading attraction reviews to plan trips. Hence, this study fills this research gap by exploring the effect of text readability on attraction review value. Moreover, we seek answers on how reviewer characteristics inferred from historical rating distribution will influence the perceived value of the review. For example, when two reviewers both give the same rating of 4 to an attraction, but one gives all other reviewed attractions a rating of 2, and the other gives all others 5, these two ratings do not have the same recommendation intention even if the scores are the same. TripAdvisor provides a historical rating distribution of the reviewer on the review page. Therefore, investigating how reviewer characteristics inferred from historical distribution influence perceived value of the review is interesting.

2.1 The Review

The most important aspect of a review is the review content. To provide information effectively, the review should be precise or easy to understand without possible conflicts. As one of the quantifiable metrics of texts, readability, which is judged by its writing style, refers to how easily the text could be understood by readers (Klare, 1974). The readability of the text reflects the social status, education level, and social hierarchy of the author (Tausczik & Pennebaker, 2010). Therefore, written reviews with high readability would be treated as more reliable than written reviews with low readability, that is, the review source is more credible. As the review is precise or easy to understand, its meaning would spread to more people. Thus, the review would receive more helpfulness votes, provided that all else are equal. Therefore, we propose the following hypothesis:

Hypothesis 1a. *A more readable review will receive more helpfulness votes.*

Another aspect of a review is its rating, which is a brief overall evaluation of consumer experience. Readers can quickly identify the attitude and sentiment of the author based on the ratings. Li and Hitt (2008) demonstrated that the utility of each product is determined by its expected and perceived values. Hence, assuming that the rating of each review is composed of baseline rating, the expected and perceived value of the attraction is reasonable. In other words, the rating of a review for attraction j given by reviewer i can be seen as $rating_{ij} = b_i + p_{ij} - e_{ij}$, where b_i is the baseline rating of reviewer i for all reviews, e_{ij} is the expected value of attraction j by reviewer i, and p_{ij} is the perceived value of attraction j by reviewer i. Hence, it is not proper to directly use raw rating to indicate author's sentiment if the historical ratings are visible because the sentiment that the author wanted to express is $p_{ij} - e_{ij}$ part. The rating would be higher than the author's mean provided that the perceived value is higher than the expected value (i.e., the author is happy) and vice versa. Figure 1 shows that the historical rating distribution of the author is shown on the review page.

Thereby, readers could feel the author's sentiment more precisely. Obviously, the more a rating deviated from the mean rating (i.e., distance between expected value and perceived value is larger), the more extreme the sentiment expressed by the author. Such extreme sentiment could be either unexpectedly exciting or disappointing. Speeches with extreme sentiment will be more persuasive (Nabi, 1999). A more persuasive review means that it has a greater chance to be agreed upon by readers. In other words, readers would perceive the helpfulness of the review and vote for it. Therefore, we hypothesize the following:

Hypothesis 1b. *Reviews expressing more extreme sentiment will gain more helpfulness votes.*

2.2 The Reviewer

Figure 1 shows that readers can intuitively identify two properties from the historical rating distribution of the author. These properties are mean and skewness. The mean of the rating represents the author's baseline attitude towards the attractions reviewed. For instance, a reviewer tends to give a higher score in his reviews provided that he has a higher mean rating. Hence, readers could identify whether the author is positive or not. Positive attitude promotes trustworthiness in the online environment (Lu, Hayes, Yu, & Wang, 2010). Therefore, reviewers with higher mean ratings are more likely to be treated as trustworthy authors who provide credible information. As we argued before, trust in reviewers will lead to trust in reviews. Trust in reviews helps identify the perceived usefulness of reviews. Accordingly, we hypothesize:

Hypothesis 2a. *An author who wrote more reviews which stress the positive sides of an attraction is more likely to receive helpful votes than an author who stresses the negative sides of an attraction. .*

Skewness is a measure of distribution asymmetry. According to the skewness of rating distribution, we could find whether the author's rating habit in rating attractions. Unlike the mean skewness, which measures the relative position of mode and mean, is a measure of the distribution shape and is more intuitive. Figure 2 is an example of the three types of skewness. The one on the left is negative skewness, in which the author is more likely to give higher ratings (mode is 4). The middle one is neutral skewness, in which the rating distribution is almost normal. The one on the right is positive skewness, and the author is more likely to give lower ratings (mode is 2). All three distributions have similar means (left is 3.18, middle is 3, and right is 2.82). From the distributions, the readers could infer whether the author has a higher mode rating than mean rating (left one) or a lower mode rating than mean rating (right one). Consumers usually weigh negative information more heavily than positive information (Ahluwalia & Shiv, 1997; Kusumasondjaja et al., 2012), especially in the online environment (Basuroy, Chatterjee, & Ravid, 2003). According to prospect theory, pain brought by loss is a stronger emotion than happiness brought by success (Cenfetelli & Schwarz, 2011; Kahneman & Tversky, 1979). Hence, consumers are more willing to hear negative sounds to avoid losses when searching information online. Reviewers whose mean rating is higher than mode rating are more likely to give negative reviews. Thereby, they become credible information sources. Meanwhile, if a reviewer who usually posts negative reviews (i.e. mode rating is lower than mean rating) writes a positive review, readers will form various opinions, such as "Even such a reviewer highly recommends this attraction, this attraction must be quite interesting!" Thus, we conjecture that:

Hypothesis 2b. *An author whose mean rating is higher than his mode rating is more likely to receive helpfulness votes than other authors.*

****

Please Place Figure 2 Here

****

3. Methodology and Data

3.1. Econometric Model

In this study, we explored two sets of factors that influence the perceived usefulness of reviews. However, these two sets are not at the same level. One set included review-level factors, such as text readability and rating, whereas the other one includes reviewer level factors such as distribution properties. Not every reviewer has written the same number of reviews. This imbalance may raise the problem of biased estimation if we mix these two sets in one econometric model. Therefore, we proposed two econometric models to investigate the effect of these two sets. A review-level econometric model is used for testing H1a and H1b, while a reviewer-level econometric model is used for testing H2a and H2b.

3.1.1. Review Level Model

One model (Research model I) is a review-level model, in which the number of helpfulness votes received for each review is the dependent variable. Majority of reviews did not receive any helpfulness vote. Hence, we first summarized the distribution of helpfulness votes (Table 1). Obviously, the distribution is not normal even if we exclude the zeros. Hence, a linear regression model is not proper in this study. As the helpfulness votes is a count variable, this study clearly adopted a count data model. One typical count data model is the Poisson regression model, which assumes that the dependent variable is drawn by a Poisson process. However, the Poisson process requires the mean to be equal to the variance. In our study, the mean of helpfulness votes was smaller than the variance (mean is 0.400, standard variance is 1.182). This over-dispersion problem required us to apply extended models of Poisson regression. Therefore, we introduced negative binomial regression model as Equation (1), which relaxed the Poisson assumption that the mean should equal the variance (Greene, 2011). In Equation (1), $\boldsymbol{x_i}$ represents a vector of independent variables, and $\boldsymbol{\beta}$ is a vector of parameters to be estimated.

$$\mathrm{P}(\mathrm{Y} = y_i | \boldsymbol{x_i}) = \frac{\Gamma(\theta + y_i)}{\Gamma(1 + y_i)\Gamma(\theta)} r_i^{y_i} (1 - r_i)^{\theta}, \quad (1)$$

$$\lambda_i = exp(\boldsymbol{x}_i'\boldsymbol{\beta}), r_i = {\lambda_i}/{(\theta + \lambda_i)}$$

****

Please Place Table 1 Here

****

3.1.2. Reviewer Level Model

The other model (Research model II) is the reviewer level model. As different reviewers wrote varied numbers of reviews in our dataset, we took the average number of helpfulness votes for each reviewer in this data set as the dependent variable. This dependent variable was also a non-negative variable, but it is not a count variable like helpfulness votes in the review level model. As this dependent variable is clearly a continuous variable, it would be more proper to apply the type II Tobit model for the reviewer level research model as Equation (2).

Similar to Equation (1), $\boldsymbol{x}_i$ represents a vector of independent variables in Equation (2), and $\boldsymbol{\beta}$ is a vector of parameters to be estimated.

$$y_i = \begin{cases} y_i^* & if\ y_i^* > 0 \\ 0 & if\ y_i^* \leq 0 \end{cases}, \text{ where } y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i \tag{2}$$

3.2. Data

We obtained data from the leading traveler community, TripAdvisor. TripAdvisor provides a platform for travelers to post their opinions of attractions. When selecting the target attractions, we selected attractions that are not too famous or too rare. Too famous attractions, such as the Grand Canyon, are too familiar to tourists, and they do not need reviews to help them decide. Too rare attractions usually have a few visiting tourists. Thereby, reviews about such attractions probably would not gain attention to receive helpfulness votes. Hence, we chose New Orleans as our target city. New Orleans is a city with plenty of natural and man-made landscapes, but it is not as famous as popular travel cities like Los Angeles, New York City, or cities in Florida. Therefore, New Orleans was an ideal city for our study.

We retrieved reviews for attractions in New Orleans on October 30, 2014, as well as reviewer profiles. The data contain 41,061 reviews for 106 attractions. Since TripAdvisor only provides rating distribution for reviewers who have already posted at least three reviews, we obtained 19,674 reviewers with historical rating distribution in our dataset. The review data included review rating, review date, and helpfulness votes received. The reviewer data included historical rating distribution, total reviews published, and total helpfulness votes received. We collected the total review number and photo number for each attraction as well.

3.3. Variables

3.3.1. Dependent Variable

Total helpfulness votes of a review received is the dependent variable in our study for research model I. The dependent variable was the only metric used to measure the perceived helpfulness of a review. The average helpfulness vote of the reviews of each author for New Orleans attractions is the dependent variable for research model II.

3.3.2. Independent Variables of Interest

There are two sets of independent variables of interest. One set comprised review-related variables. The first one is review length. As prior research suggest (Hao et al., 2009; Jeon et al., 2006; Kim et al., 2006; J. Liu et al., 2007; Weimer & Gurevych, 2007), review length has a positive effect on the perceived helpfulness of a review. The second one is review text readability. We used the Gunning-FOG index (FOG) developed by Gunning (1969) to measure the syntax style complexity of the review. FOG is one of the most popular readability measurements (Ghose & Ipeirotis, 2011). Based on Stajner, Evans, Orasan, and Mitkov (2012), the FOG index is calculated as Equation (3).

$$\text{FOG} = 0.4 \times (average\ sentence\ length + Hard\ words) \tag{3}$$

Hard words refer to the number of words containing more than two syllables for each 100 words of a document. The FOG index gives years of formal education required to understand the text on first reading, i.e., the higher the FOG index, the more difficult it is to understand the review text. In research model II, these two variables are averaged according to author.

The third one is absolute distance from the rating to the author's mean rating, which represents the sentiment that the author wanted to express in the review. It is designed to validate the effect of extreme sentiment on the perceived usefulness of the review. Given that prior studies on extremity used quadratic rating to explore the U curve effect of rating (Z. Liu & Park, 2015; Mudambi & Schuff, 2010), we introduced a similar variable for robust results. We took the mean rating of the author from the review ratings and included this term and its quadratic term in research model I to comprise the robust test. These variables were also averaged in research model II.

The other set of independent variables contained two variables related to reviewer rating distribution. One is the distance between the mean rating of the author and the rating of the review. The other is the skewness of the historical rating distribution, which is calculated as $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^{3/2}}$.

3.3.3. Control Variables

The time elapsed from the review published is obviously correlated with helpfulness votes received. Usually, the longer time that has elapsed, the more helpfulness votes were received by a review. This variable was also averaged in research model II.

To control further the attraction effect, we included the attraction ranking in TripAdvisor. This value referred to the number of total reviews for an attraction and the number of total photos for the said attraction. Attraction ranking represented the popularity of the attraction. As we demonstrated before, the popularity of an attraction would affect the helpfulness of its reviews. Hence, we included ranking as a control variable. We also introduced the number of total reviews because the chance of receiving helpfulness vote for a review would be reduced if it were among massive reviews. Photos are always more persuasive than words. Therefore, the number of photos for the attraction was included as a control variable.

As Hochmeister et al. (2013) stated, experts received more acknowledgements than others. Whether a reviewer is an expert could be inferred partially from the number of total reviews he published. Therefore, the number of total reviews a reviewer published was considered. We also introduced average helpfulness votes received of a reviewer following Ghose and Ipeirotis (2011).

Tables 2 and 3 describe the variables in this study for models I and II. Table 4 shows the corresponding descriptive statistics.

****

Please Place Table 2 Here

****

****

Please Place Table 3 Here

****

****

Please Place Table 4 Here

****

4. Results

4.1. Review Model

We first ran a negative binomial regression model at review level for research model I. Column 1 of Table 5 reports these results. When the hypothesis of dispersion parameter α is equal to zero, we rejected it because the likelihood-ratio χ^2 test is 5502.48 with p-value <

0.001. This result indicated that the dependent variable distribution is more likely to be a negative binomial distribution, but not a Poisson distribution. Hence, our model selection is supported.

****

Please Place Table 5 Here

****

Hypothesis 1a investigated whether a more readable review will receive more helpfulness votes. This hypothesis is supported because *FOG* is negative and significant (coef. = -0.009, p-value<0.01). The negative coefficient of *FOG* indicated that when the review text requires fewer years of formal education to be understood, the review would receive more helpfulness votes. A review that can be easily comprehended by readers who have fewer years of formal education implied that, the text is more readable because it has less difficult vocabulary. Therefore, our H1a was supported.

Hypothesis 1b, which states that reviews expressing more extreme sentiment would receive more helpfulness votes, was supported (coef. = 0.236, p-value<0.01). Hence, a review is thought to be useful when the rating is far from the mean rating of the author, either much higher or much lower. To illustrate this effect further, we replaced *RatingDis* by *RatingDif* and its quadratic term and ran the model again. The result is shown in column 2 of Table 5. The positive significance of $RatingDif^2$ confirms the previous results.

4.2. Reviewer Model

We then ran a type II Tobit model at reviewer level for research model II. The results are presented in Table 6. Hypothesis 2a, which investigated the effect of the author's mean rating on perceived helpfulness is supported because *AuthorMean* is significant (p-value < 0.01) and positive (coef. = 0.096). As we mentioned before, higher mean rating means the author is positive, whereas lower mean rating means the author is pessimistic. Therefore, the result indicated that reviews written by positive authors were more likely to receive helpfulness votes than reviews written by pessimistic authors.

****

Please Place Table 6 Here

****

Hypothesis 2b on the effect of historical rating distribution skewness was supported (coef. = 0.027, p-value < 0.01). The positive coefficient indicated that a review will receive more helpfulness votes when the skewness of the historical rating distribution of its author was larger. Note that the more a distribution skews to the right side (i.e., the mode is larger), the smaller the skewness index is. Hence, this result indicated that reviewers with lower mode rating than mean rating were more likely to receive helpfulness votes.

5. Discussion and Implications

5.1. Theoretical Implications

This paper provides three main theoretical contributions. First, our study sheds lights to future tourism management studies on online review by exploring the effect of text readability. In this study, we investigated the effect of text readability on perceived value of attraction reviews, whereas most literature on text readability or review value/helpfulness mainly focused on product reviews (Ghose & Ipeirotis, 2011; Y. Liu et al., 2008; Otterbacher, 2009). As such, it is of paramount importance for tourism management research. Unlike the search products investigated in prior research [e.g. audio and video players, digital cameras, and DVDs in Ghose and Ipeirotis (2011)], attraction is a kind of experience product. Consumers cannot infer the quality of an attraction by some objective parameters, which can be used to describe a search product. The only thing consumers can rely on is the description provided by reviewers. Some studies were conducted to explore the effects of text readability on reviews for experience products such as restaurants (Z. Liu & Park, 2015), but attraction reviews are not exactly the same as restaurant reviews. Consumers would read reviews more carefully to make trip plans because choosing wrong attractions will result in more losses than in choosing wrong restaurants. Hence, the text readability of a review should be a more important factor than other linguistic characteristics. Our empirical results supported this point, i.e., more readable reviews would receive more helpfulness votes. This finding provides a new insight for future research on eWOM of attractions. When exploring the effect of eWOM on attraction, such as online reviews, researchers should not equally treat all reviews with the same influence power. More readable attraction reviews would play a more important role than the rest.

Second, our study has made a contribution to improve prediction models and causality/correlate relationship research models of online tourism reviews. To the best of our knowledge, this study is the first one to investigate the effect of reviewer historical rating distribution properties on online tourism reviews. We explored the impact of reviewer historical rating distribution on the perceived value of their reviews, whereas prior research only paid attention to the summary statistics of reviewers (Ghose & Ipeirotis, 2011; Z. Liu & Park, 2015), such as number of reviews, number of helpfulness votes received, etc. Compared with the summary statistics of the historical ratings of a review, the distribution of historical ratings could more comprehensively reflect the personal characteristics of a reviewer. In addition, the empirical result shows that historical rating distribution properties influence the perceived value of an online attraction review. This finding has implications for both prediction problem research and causality/correlate relationship that explores problem research. Research on predictive problems of online tourism reviews should incorporate the visible historical records of the author into the predictive model to reach higher predictive accuracy. For causality/correlation relationship problem, researchers should include variables about the visible historical records of the authors in research model to control the individual effect caused by review authors when they are performing online tourism review research.

Last, from the methodology perspective, we introduced a negative binomial regression model to investigate the factors influencing helpfulness votes of reviews. Most prior studies applied linear regression model using OLS estimation or logistic model. Z. Liu and Park (2015) introduced the Tobit model, which would fit the data better. However, according to our statement in the research model section (Section 3.1), the application of a count model would be more proper to avoid biased estimation when the number of helpfulness votes is the dependent variable. The negative binomial regression model fits data better in this research context. The negative binomial regression model could be widely applied in tourism management research, since plenty of tourism studies might use count variable as a dependent variable. For example, both 'the number of days spent in a destination' and 'number of visitors of an attraction' are typical count variables. Therefore, the introduction of a count model (i.e. negative binomial regression model) in this study opens a new venue for future tourism management research in methodology.

5.2. Managerial Implications

Our results yield managerial implications for tourism management in practice, especially for attraction managers. First, this paper would aid attraction managers in detecting possible influential online reviews. Helpful reviews are more favored by readers, and thereby they are more influential on decision. As a result, these comments and reviews have to be considered and analyzed carefully. Attraction managers should quickly identify potential helpful reviews to cope with both positive aspects and negative aspects in reviews. Attraction managers should prepare for an increase in number of visitors if the review is positive, or they need to take remedial actions if the review shows some complains. According to our findings, managers should mainly focus on precise or easy to understand reviews because they would be more influential than obscure reviews.

Second, this paper also shows that readers consider the author's mean rating when they read reviews. Based on the findings, the reviewers rated differently than their average ratings about a tourism spot would attract more helpful votes. Specifically, a review would receive helpfulness votes if the rating is relatively lower than the mean rating of the author's historical reviews even if this rating is not a bad rating. Therefore, when attraction managers want to know the potential influence of a review, they should not only consider the raw rating but also the relative rating to mean rating of the author.

Third, our study also suggests attraction managers should pay attention to authors of the reviews. According to the findings, reviews written by positive reviewers (those having higher mean ratings) are worthy of attention because positive reviewers are more likely to receive helpfulness votes. Positive individuals tend to accept their situation and are less likely to be escapists (Scheier & Carver, 1987). Lu et al. (2010) found that the impact of positive attitude on web trustworthiness is very strong. This finding may be the reason why their views sound more reliable and helpful. Besides positive reviewers, reviewers whose mode rating is lower than mean rating are also important reviewers who will generate influential reviews. Identifying positive reviewers or reviewers whose mode rating is lower than mean rating, and taking their views into consideration would be more logical for the attraction professionals to develop and improve their tourism product (e.g. attractions) further.

6. Conclusion and Limitations

Two sets of main findings are indicated in this study. The first comprises review-related findings. First, the readability of a review text is correlated with perceived helpfulness of the reviews. Reviews with precise or easy to understand writing styles will receive more helpfulness votes. Second, reviews expressing extreme sentiment would be considered as valuable.

The second finding centers on reviewer characteristics inferred from historical rating distribution. This distribution, which is the summary of all review ratings given by a reviewer, will significantly influence perceived helpfulness of his reviews. Specifically, the mean rating of the historical ratings of an author could be used to infer the baseline attitude towards travelling reviews (positive or negative). Positive reviewers (reviewers with higher mean) will receive more helpfulness votes. Moreover, the skewness (a measure of the distribution asymmetry) of the historical rating distribution, which represents whether a reviewer's mode rating is higher than mean rating, is another factor that influences the perceived helpfulness of a review. Reviews written by reviewers who have historical distributions with positive skewness are more likely to receive helpfulness votes.

This study has several limitations. First, we only used reviews for attractions in New Orleans. Although the city is ideal for conducting our research, using a broader range of attractions would be better to understand how the reviews gain helpfulness votes. Second, although we partially controlled the attraction effect and the author's individual effect by including related variables, these factors may not be enough to control all unobservable effects. One possible approach is to apply a fixed effect model by retrieving more data at different periods.

Table 1. Distribution of review helpfulness votes

Votes Value	Frequency	Proportion
0	31,045	75.607%
1	7,131	17.367%
2	1,762	4.291%

3	514	1.252%
4	220	0.536%
5	111	0.270%
6~10	201	0.490%
11~15	40	0.097%
16~20	19	0.046%
21~25	9	0.022%
26~30	6	0.015%
>30	3	0.007%
Total	41,061	100%

Table 2. Variable Description for Model I

Variable	Description
Helpful	Helpfulness votes received by a review
FOG	FOG readability index of the review text
Length	Total number of words of a review
RatingDis	Distance between rating and the reviewer’s mean rating
RatingDif	Rating minus the reviewer’s mean rating
ReviewDate	Number of days elapsed from the day the review was published
Ranking	Ranking of the attraction
AttReview	Number of total reviews for the attraction
AttPhoto	Number of total photos for the attraction

Table 3. Variable Description for Model II

Variable	Description
AvgHelpful	Average helpfulness votes received by the author's reviews in New Orleans
AvgFOG	Average of FOG readability indexes of the review texts
AvgLength	Average of total number of words of the author's reviews
AvgReviewDate	Average number of days elapsed from the publication day of the author's reviews
Skewness	Skewness of the reviewer's historical rating distribution
AuthorMean	Mean of the reviewer's historical ratings
ReviewNum	Number of total reviews published by the reviewer
AvgHelpfulAll	Average helpfulness votes received of the reviewer's all reviews
AvgRatingDif	Average of ratings subtracted from the reviewer's mean rating

Table 4. Variable Descriptive Statistic

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Helpful	41,061	0.400	1.182	0	62
FOG	41,061	10.223	3.686	0.34	73.35
Length	41,061	62.155	58.307	5	3,204
RatingDis	39,950	0.587	0.465	0	3.75
RatingDif	39,950	0.196	0.723	-3.75	2.667
ReviewDate	41,061	577.541	422.285	18	4870
Ranking	41,061	18.164	17.166	1	106
AttReview	41,061	2,907.352	2,154.697	2	6,134
AttPhoto	41,061	480.448	459.652	0	1,475
AvgHelpful	19,674	0.387	1.025	0	38
AvgFOG	19,674	10.242	3.279	0.34	73.35
AvgLength	19,674	64.185	53.384	6	969.5
AvgReviewDate	19,674	592.993	407.727	42	4422
Skewness	19,674	-1.028	0.814	-7.417	3.175
AuthorMean	19,674	4.280	0.408	1	5
ReviewNum	19,674	41.094	61.200	3	2450
AvgHelpfulAll	19,674	0.598	0.641	0	26.667

Table 5. Model Results

Helpful	(1) **Coef. (Std. Err)**	(2) **Coef. (Std. Err)**
FOG	-0.009 (0.0030)***	-0.009 (0.0030)***
Length	0.007 (0.0002)***	0.007 (0.0002)***
RatingDis	0.236 (0.0199)***	
RatingDif		0.256 (0.0189)*
RatingDif^2		0.117 (0.0092)***
ReviewDate	0.001 (0.0002)***	0.007 (0.0002)***
Ln(AttReview+1)	-0.047 (0.0153)***	-0.048 (0.0154)***
Ln(AttPhoto+1)	-0.240 (0.0139)***	-0.239 (0.0139)***
AttRanking	0.002 (0.0007)***	0.002 (0.0008)***
Cons.	-0.459 (0.0924)***	-0.384 (0.0930)***
alpha	1.108 (0.0334)	1.100 (0.0332)
Log Likelihood	-29,017.383	-28,994.364
LR Chi2	5,315.63	5,361.67
Pseudo R2	0.0839(p=0.000)	0.0846(p=0.000)
Likelihood-ratio test of alpha=0	5,502.48(p=0.000)	5,502.84(p=0.000)
Number of Obs.	39,950	39,950

* p<0.1, ** p<0.05, ***p<0.01, standard errors are reported in parenthesis

Table 6. Model Results

AvgHelpful	**Coef.(Std. Err)**
AuthorMean	0.096 (0.0183)***
Skewness	0.027 (0.0082)***
ReviewNum	0.0001 (0.0001)
AvgHelpfulAll	0.550 (0.0101)***
AvgFOG	-0.006 (0.0019)***
AvgLength	0.004 (0.0001)***
AvgRatingDif	-0.055 (0.0103)***
$AvgRatingDif^2$	0.065 (0.0077)***
AvgReviewDate	0.001 (0.00001)***
Cons.	-0.874 (0.0821)***
Log Likelihood	-25,018.186
LR Chi2	6,776.75(p=0.000)
Pseudo R2	0.1193
Number of Obs.	19,674

* p<0.1, ** p<0.05, ***p<0.01; standard errors are reported in parenthesis

References

Ahluwalia, R., & Shiv, B. (1997). Special session summary: The effects of negative information in the political and marketing arenas: exceptions to the negativity effect. *Advances in Consumer Research, 24*, 222

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing, 67*(4), 103-117

Burtch, G., Ghose, A., & Wattal, S. (2013). An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research, 24*(3), 499-519

Cenfetelli, R. T., & Schwarz, A. (2011). Identifying and testing the inhibitors of technology usage intentions. *Information Systems Research, 22*(4), 808-823

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research, XLIII*(August), 345-354

Collie, M. (2014). Internet leads travel. *Geton.com.* Retrieved from http://www.geton.com/blog/geton/internet-leads-travel

Connors, L., Mudambi, S. M., & Schuff, D. (2011). Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness. *Hawaii International Conference on System Sciences*. In (Eds.). Hawaii. IEEE. pp. 1-10.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems, 45*(4), 1007-1016

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

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering, 23*(10), 1496-1512

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science, 31*(3), 493-520

Greene, W. H. (2011). *Econometrics analysis* (7 ed.). Upper Saddle River, NJ: Prentice Hall.

Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. *Information and Communication Technologies in Tourism 2008*. In P. O'Connor, W. Höpken & U. Gretzel (Eds.). ENTER 2008. Austria. New York: Springer-Verlag. pp. 35-46.

Gu, B., Park, J., & Konana, P. (2012). Research Note--The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research, 23*(1), 182-196

Gunning, R. (1969). The FOG index after twenty years. *Journal of Business Communication, 6*(2), 3-13

Hao, Y., Li, Y., & Zou, P. (2009). Why some online product reviews have no usefulness rating. In *Pacific Aisa Conference on Information Systems*, Hyderabad, India. pp. 100.

Hochmeister, M., Gretzel, U., & Werthner, H. (2013). Destination expertise in online travel communities. *Information and Communication Techonologies in Tourism 2013*. In C.

Lorenzo & X. Zheng (Eds.). ENTER 2013 Proceedings. Austria. Berlin Heidelberg: Springer-Verlag. pp. 218-229.

Jeon, J., Croft, W. B., Lee, J. H., & Park, S. (2006). A framework to predict the quality of answers with non-textual features. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. In (Eds.). Seattle, WA, USA. New York. pp. 228-235.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*(2), 263-292

Kim, S.-M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. In (Eds.). Sydney, Australia. pp. 423-430.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 10*(1), 62-102

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications, 11*(3), 205-217

Kusumasondjaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. *Journal of Vacation Marketing, 18*(3), 185-195

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research, 19*(4), 456-474

Liu, J., Cao, Y., Lin, C.-Y., Huang, Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. In (Eds.). Prague, Czech Republic. pp. 334-342.

Liu, Y., Huang, X., An, A., & Yu, X. (2008). Modeling and predicting the helpfulness of online reviews. In *Eighth IEEE International Conference on Data Mining*, Pisa, Italy. pp. 443-452.

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management, 47*, 140-151

Lu, J., Hayes, L., Yu, C.-S., & Wang, L.-Z. (2010). Technology readiness, trustworthiness, and use of e-auctions. In *Southwest Decision Sciences Institute Conference* Dallas, TX. pp.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly, 34*(1), 185-200

Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude Change, and recall. *Communication Theory, 9*(3), 292-320

Otterbacher, J. (2009). “Helpfulness” in online communities: a measure of message quality. In *CHI*, Boston, MA, USA. pp. 955-964.

Park, H. L., Xiang, Z., Josiam, B., & Kim, H. M. (2013). Personal profile information as cues of credibility in online travel reviews. *Information and Communication Techonologies in Tourism 2013*. In C. Lorenzo & X. Zheng (Eds.). ENTER 2013 Proceedings. Austria. Berlin Heidelberg: Springer-Verlag. pp. 230-241.

Scheier, M. F., & Carver, C. S. (1987). Dispositional optimism a physical well-being: The influence of generalized outcome expectancies on health. *Journal of Personality, 55*(2), 169-210

Stajner, S., Evans, R., Orasan, C., & Mitkov, R. (2012). What can readability measures really tell us about text complexity? *Proceedings of Workshop on Natural Language Processing for Improving Textual Accessibility*. In L. Rello & H. Saggion (Eds.). Istambul. pp. 14-22.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24-54

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management, 30*(1), 123-127

Weimer, M., & Gurevych, I. (2007). Predicting the perceived quality of web forum posts. *Proceedings of the Conference on Recent Advances in Natural Language Processing*. In (Eds.). Borovets, Bulgaria. pp. 1-6.

Wu, P. F., Heijden, H. v. d., & Korfiatis, N. (2011). The influences of negativity and review quality on the helpfulness of online reviews. In *International Conference on Information Systems*, Shanghai. pp. 1-10.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management, 28*(1), 180-182

Yin, D., Bond, S., & Zhang, H. (2014). Anxious or Angry Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews. *MIS Quarterly, 38*(2), 539-560