Online Review Impacts on Hotel Online Booking Decision

Yingxiao Yu[1, a], Xueyan Guo[2, 3, b*], Ying Zhang[4, c] and Huanyu Zhao[2, 3, d*]

[1]Hebei HYJK Information Technology Co. Ltd, Shijiazhuang, China

[2]SJZ JKSS Technology Co. Ltd, Shijiazhuang, China

[3]Institute of Applied Mathematics, Hebei Academy of Sciences, Shijiazhuang, China

[4]Hebei Academy of Sciences, Shijiazhuang, China

[a]kappel89@163.com, [b]guoxueyan1989@126.com, [c]spicegirl_2000@163.com,

[d]zhaohuanyu@163.com

Keywords: Online review; Hotel online booking decision; Regression analysis; Significance test.

Abstract. With the development of the economy and the maturity of the network market, more and more people book hotels and tickets online. When the customer book hotels online, online reviews play an important role in consumer booking decision. Despite researcher's attempts on examining different aspects of online reviews, its impact on hotel sales remains largely unknown in the existing literature. *Objective:* In this paper, we give one method to investigate the impact of online reviews on consumer's hotel online booking decision. *Method:* Utilizing data collected from the so jump website in China, linear regression analysis is used to find out which aspects of online reviews have impacts on the consumer's hotel online booking decision. *Result:* Our results indicate a significant relationship between online reviews and consumer hotel online booking decision. Six factors of online reviews influencing consumers in hotel online booking were found. *Conclusion:* In this paper, we have successfully indicated that the linear regression model analysis can be used to investigate the relationship between online reviews and hotel online booking decision.

Introduction

All manuscripts must be in English, also the table and figure texts; otherwise we cannot publish your paper.

With the development of the economy and the maturity of the network market, the Internet has been permeated to the crowd. According to the CNNIC (China Internet Network Information Center) report, by December 2014, scale of Chinese Internet users had reached 649 million, increased by 2.1% than the end of 2013. But along with the continuous improvement of the network infrastructure and the continuous expansion of Internet users, the network shopping market has been into the rapid development, online shopping increase rapidly. Besides, online shopping has huge potential market. The report pointed out that more than half of Internet users said they will increase online shopping in the future. The Internet users who do not have online shopping experience has strong online shopping intentions and they will shop online in the future, indicating that online shopping market has strong growth potential.

Online shopping brings consumers a lot of convenience and economy, so is the travel industry. According to China's online travel market research, the Internet users who book travel online has reached 133 million. More and more people book hotel and ticket online. Hotel online booking model

Have mainly two forms. One is booking directly on the official web site. On one hand, this form can reduce the cost paid for the third party web site. On the other hand, the hotel's management can directly see the consumer evaluation of the hotel and improve their service. This strengthens the interaction between the managers and the consumers greatly. Another form is reservations through the third party sites. Consumers can make comments freely and make comparison between different hotels. This form is used by most hotel consumers for it is more reliable than the official web site.

Online reviews are made by consumers and most are anonymous, thus it has highly credibility. Online reviews have become one important reference when people book hotels online. It has become

a habit for consumers to make comments online. Online review is an important resource for both consumers and enterprises. For hotel managers and decision makers, online reviews reflect the consumer's opinion and suggestion of the hotel. Thus research of the relationship between online reviews and consumers purchase decision is of great significance

Early researches about relationship between online reviews and consumers purchase decision is whether online reviews can promote product sales. The research about relationship between online reviews and product sales has been mature. Chevalier and Mayzlin investigate influence of online book reviews on book sales in the amazon web site as a platform [1]. Their research shows that the higher the online review scores, the higher of the book sales. Clemons et al. point out those beer sales can be predicted through the consumer online reviews [2]. The more of the online reviews, the higher of the beer sales. Senecal and Nantel indicate that subjects who consulted product recommendations selected recommended products twice as often as subjects who did not consult recommendations [3]. Henning-Thurau and Walsh demonstrate that the negative comments play more important roles than the positive ones in consumer purchase decision [4]. Ahluwalia et al. also demonstrated that negative reviews are more likely to cause the attention of consumers [5].

This relationship of tourism product has its own characteristics. These researches mainly focus on two aspects of scenic spots and hotel online reviews. Qing Ye and Rob Law demonstrate that there is significant relationship between hotel online reviews and sales [6]. Beverley and Victoria verify that the positive reviews can improve customer booking intention [7].

This paper focus on online hotel reservation consumers and potential consumers. Aim at investigating the relationship between the online reviews and consumers booking decision and validate the influence factors of consumers booking decision.

The rest of this paper is structured as follows. Our method is introduced in Section 2 which includes the two main parts of this paper: the linear regression analysis and the significance test method. The experiment result is presented in Section 3. At the end of this paper, we briefly address our works and the next jobs we will do.

All the experiments are implemented in SAS (Statistical Analysis System).

Method

Our research framework is given in Section 2.1. Section 2.2 describes the data set used in our experiments. The linear regression analysis and the significance test method are introduced in Section 2.3.

Research Framework and the Hypotheses. Based on the preceding review of the literature we postulate that the following six factors have significant impacts on consumers in hotel online booking: content of the comment, comment whether detailed, prestige of the reviewer, comment score, timeliness of the commentate number of useful comment. We thus formulate the following general hypotheses.

H1. The more of the comment content, the greater the impact on consumer booking decision;
H2. The more specific of the comment, the greater the impact on consumer booking decision;
H3. The higher of the reviewer prestige, the greater the impact on consumer booking decision;
H4. The higher of the comment score, the greater the impact on consumer booking decision;
H5. The more timely of the comment, the greater the impact on consumer booking decision;
H6. The more of the useful comment, the greater the impact on consumer booking decision.
The dependent variable is consumer booking decision.

Data Acquisition. We designed a questionnaire and post it on the sojump website. A sample of 265 subjects was generated on the basis of 265 questionnaires obtained from sojump. The questionnaire contains six questions about whether the six factors impact the subject in hotel online booking. There are five options: 5, 4, 3, 2, 1. Option 5 represents there will be a very important influence. Option 1 represents there will be little influence. The other eight questions are about the influence of online review on consumers booking decision: the comment provides some information I have never focused on in the past, the comment makes my purchase decision clearer, the comment mentions some of the things I have never considered, etc. There is also five options scale from

"strongly agree" to "strongly disagree". Option 5 represents strongly agree. Option 1 represents strongly disagree.

The majority of the subjects are between the age of 20 and 39 (90%). Fifty-two percent are female, 58% are working full time; 3% of the subjects are full-time students and another 39% are part-time workers and the other. Most people (94%) have a monthly income of less than 5000.

Data Analysis

The linear regression analysis and the significance test method are introduced in this section.

Correlation Analysis. Correlation analysis is a kind of statistical research to measure the relevance between variables. It is mainly used to judge whether there is relevant relationship between different variables. Simple correlation coefficient is widely used to measure linear relationship between two variables. Through correlation analysis, we can find which variables are more related to the dependent variable among all the variables. If the correlation coefficient between two variables is positive, there is a positive correlation relationship between the two variables. On the contrary, there is a negative correlation relationship between the two variables.

Giving two vectors x, y $\in$ R*. The correlation coefficient between x and y is defined as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{1}$$

Where $\bar{x}$ and $\bar{y}$ is the mean value of x and y.

The correlation between influence factors and the consumer booking decision is analyzed here.

Significance Test of Regression Model and Coefficients. Generally, one linear regression model can be written as

$$Y = \beta_0 + \beta_1 X_1 + \dots \beta_p X_p + \varepsilon. \tag{2}$$

Where $\beta_0, \beta_1, \dots, \beta_p$ are the unknown constant parameters, ε is the unaccounted for random error, meet $E(\varepsilon) = 0$ and $Var(\varepsilon_i) = \sigma^2 > 0$.

As can be seen from the above equation, the discrepancy of each observed value of the $Y_i(i = 1, 2\dots, n)$ is caused by two aspects: one is through the linear model $\beta_0 + \beta_1 X_1 + \dots \beta_p X_p$, different $X_1, X_2, \dots, X_n$ cause different Y; the other is due to the unaccounted for random factors. If the variation of different Y_i can be decomposed into the above two aspects, significance of the linear regression model can be analyzed by comparing the two aspects [8, 9, 10].

Make $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, the variation of Y_i relative to the center $\bar{Y}$ is

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2. \tag{3}$$

Call SST as Total sum of squares.

$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}(i = 1, 2, \dots, n)$ Is the estimated value of linear regression model? For $\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n}(Y_i - Y_i) = 0, \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$. The variation of Y_i relative to the center $\frac{1}{n}\sum_{i=1}^{n} Y_i$ is

$$SSR = \sum_{i=1}^{n}(Y_i - \bar{Y})^2. \tag{4}$$

Call SSR as Regression sum of squares.

It can be proved that $SST = SSR + SSE$. Define

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \quad (5)$$

The larger of R2, the more significant of the linear regression model. Call R2 as coefficient of multiple determinations.

The regression relationship significance test is to test the hypothesis

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0.$$

If H_0 is true, the regression model becomes $Y = \beta_0 + \varepsilon$. That is $X_1, X_2, \ldots, X_P$ not significant to Y, it is meaningless to construct the linear regression model between Y and $X_1, X_2, \ldots, X_P$. Otherwise at least one X_i has effect on Y in the linear regression model.

The linear regression model under the hypothesis H_0 becomes

$$\mathbf{Y} = \mathbf{X}\beta_0 + \varepsilon, \varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n).$$

Where $\mathbf{X} = \mathbf{1}$, thus the least square estimation of β_0 is

$$\beta_0 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y}.$$

For all the $i = 1, 2, \ldots, n, Y_i = \overline{Y}, SSE(H_0) = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = SST$, the test statistic is

$$F = \frac{SSR/p}{SSE/(n-p-1)} = \frac{MSR}{MSE}. \quad (6)$$

If H_0 is true, $F \sim F(p, n-p-1)$, the test value p is

$$p_0 = P_{H_0}(\mathrm{F} > f) = P(F(p, n-p-1) > f). \quad (7)$$

Where f is the observed value of F. For the given significance value α, if $p_0 < \alpha$, the linear regression model is significant; otherwise the linear regression model is meaningless.

Experimental Result

The experiment results are presented in this section.

Correlation Analysis Result. The correlation analysis result between the consumer booking decision and the six affecting factors is presented in Table 1. It can be seen that all the six affecting factors have significant impact on consumer booking decision.

Table 1. The correlation analysis results

Y	X1	X2	X3	X4	X5	X6	
Y	1.00						
X1	0.60	1.00					
X2	0.57	0.49	1.00				
X3	0.44	0.46	0.56	1.00			
X4	0.27	0.31	0.37	0.47	1.00		
X5	0.36	0.29	0.41	0.53	0.67	1.00	
X6	0.29	0.37	0.27	0.41	0.37	0.45	1.00

Regression Model Significance Test Result. Using multiple regression analysis in SAS (Statistical Analysis System), the linear regression model significance test result is presented in Table 2.

Table 2 Variance analysis table

source of variance	degree of freedom	sum square	mean square	F value	P value
regression	5	25.48	25.48	100.63	0.001
residuals	259	65.58	0.253		
total	264	91.06			

The test value $p_0 = 0.001$, so the linear regression model is highly significant.

Regression Coefficients Significance Test Result. The significance test result of X_1 is in Table 3. The significance level p is close to zero. The coefficient is 3.02. Without considering other factors, comment content has a positive effect on purchase decision. The hypothesis H1 established.

The significance test result of X_2, X_3, X_4, X_5, X_6 is in Table 4, 5, 6, 7, 8 respectively.

Table 3 The significance test result of X1

	estimated value	Standard deviation	t value	p value
constant	3.02	0.15	20.30	<0.0001
X1	0.19	0.04	4.9	<0.0001

Table 4 The significance test result of X2

	estimated value	Standard deviation	t value	p value
constant	2.7	0.17	16.3	<0.0001
X2	0.26	0.04	6.36	<0.0001

Table 5 The significance test result of X3

	estimated value	Standard deviation	t value	p value
constant	3.03	0.16	19.21	<0.0001
X3	0.19	0.04	4.53	<0.0001

Table 6 The significance test result of X4

	estimated value	Standard deviation	t value	p value
constant	2.52	0.17	15.1	<0.0001
X4	0.31	0.04	7.35	<0.0001

Table 7 The significance test result of X5

	estimated value	Standard deviation	t value	p value
constant	2.59	0.18	14.62	<0.0001
X5	0.29	0.04	6.51	<0.0001

Table 8 The significance test result of X6

	estimated value	Standard deviation	t value	p value
constant	2.38	0.17	14.19	<0.0001
X6	0.34	0.04	8.19	<0.0001

Conclusion

In this paper, we have successfully indicated that the linear regression model analysis can be used to investigate the relationship between online reviews and hotel online booking decision. We fist establish the linear regression model between the six factors of online reviews and consumer booking decision. Then correlation analysis is used in our approach to measure the correlation between the six influence factors and consumer booking decision. Significance test of regression model is used to test the significance of regression model we have established. The influence factors of consumer booking decision can be found out by regression coefficient significance test. Our results suggest that online comments have an important impact on hotel online booking. Therefore, hotel managers should seriously consider online reviews.

A sample of only 265 subjects was generated on the basis of 265 questionnaires obtained from sojump, the sample size is not big enough. Besides, there are many factors which have impacts on hotel online booking. Only six aspects of online reviews is investigated in our paper. More subjects should be obtained and more aspects of online reviews should be considered. We will do these works next.

References

[1] J. Chevalier and D. Mayzlin, 2006 "The effect of word of mouth on sales: Online book reviews," Journal of Marketing Research, pp. 345–354.

[2] E. K. Clemons and L. M. Hitt, 2006 "When online reviews meet hyper differentiation: A study of the craft beer industry," Journal of Management Information Systems, vol. 23, no .2, pp. 149–171.

[3] S.Senecal and J Nantel 2004"The influence of online product recommendations on consumers online choices,"Journal of Retailing, vol. 80, no. 80, pp.159–169.

[4] T. Hennig-Thurau and G. Walsh,2003."Electronic word of mouth: Motives for and consequences of reading customer articulations on the internet". International Journal of Electronic Commerce, vol. 8, no. 2, pp. 51–74.

[5] R. Ahluwalia.R. E. Burnkrant and H. R Unnava2000,"Consumer response to negative publicity: The moderating role of commitment," Journal of marketing research, vol. 37, no. 2, pp. 203–214.

[6] Q. Ye, R. Law, and B. Gu, 2009."The impact of online user reviews on hotel room sales,"International Journal of Hospitality Management, vol. 28, no. 1, pp. 180–182.

[7] B. A. Sparks and V. Browning, 2011"The impact of online reviews on hotel booking intentions and perception of trust," Tourism Management,vol. 32, no. 6, pp. 1310–1323.

[8] T. Hastie, R. Tibshirani,and J. Friedman, 2009."The elements of statistical learning.second edition"Springer,Berlin.

[9] J. O. Rawlings, S. G. Pantula and D.A. Dickey,1998."Applied regression analysis: a research tool," Journal of the American Statistical Association, vol. 85, no. 41.

[10]G.A.F Seber and A.J.Lee, 2012"Linear regression analysis,"John Wiley & Sons.