doi: 10.5367/te.2015.0464

Research note: CHOICE BEHAVIOUR IN ONLINE HOTEL BOOKING

Lorenzo Masiero (*)
School of Hotel and Tourism Management
The Hong Kong Polytechnic University
17 Science Museum Road
TST East, Kowloon, Hong Kong
E-mail: lorenzo.masiero@polyu.edu.hk

Juan L. Nicolau
Dpt. of Marketing
Faculty of Economics
University of Alicante
Spain
E-mail: jl.nicolau@ua.es

(*) corresponding author

Acknowledgement: We thank the destination management organization of Locarno and Ascona for their collaboration in the data collection.

Research note: CHOICE BEHAVIOUR IN ONLINE HOTEL BOOKING

Keywords: online booking; choice behavior; nested decision process; nested logit

ABSTRACT

The objective of this study is to analyse how people process information and make decisions with regard to booking hotel rooms via online booking systems. The authors propose a nested decision process and compare it with the approach of previous literature in analysing the determinants in choosing a hotel. This research attempts to model online hotel bookings in the real market using discrete choice modelling. The methodology is based on the estimates of nested logit models, and the results show that tourists choose a hotel by going through a number of staged decision structures, which is in line with Associative Network Theory and the Cybernetic decision-making model.

INTRODUCTION

Choosing a hotel is a fundamental decision-making process whose determinants have been studied profusely in related literature (Kim & Perdue, 2013). The process of choosing a hotel itself is potentially complex (Lockyer, 2005) and most studies fail to recognize that travel might follow a nested process (Jeng & Fesenmaier, 2002). In particular, although individuals can easily make a reservation with just one click with little search cost (Boffa & Sucurro, 2012; Guo, Ling, Dong & Liang, 2013), they cannot ascertain the best hotel among a wide

selection of alternatives. They may simultaneously consider all the available hotels and then select one of them based on certain criteria, or they may group these hotels based on a specific criterion or by consciously or unconsciously using some heuristics. Accordingly, we state the hypothesis that: H.1.- Tourists choose a hotel through nested decision structures. To test this hypothesis we use the simplest non-single decision process, that is, a two-stage process (see figure 1). Given that we cannot foresee the sequence of decisions, we test a priori sequence (based on the star ratings of the hotels) and several empirically derived sequences. Therefore, this research intends to address theoretical implications in the context of tourist decision process for hotel selection.

FIGURE 1 - ABOUT HERE

RESEARCH DESIGN

Real market data on a tourist's choice of hotel provides valuable information for the investigation of the tourists' behavior in booking hotel rooms online. The research design involves specific assumptions and treatments that are aimed to prepare an appropriate data set for this study's analysis. Our sample comprises the online bookings for three- and four-star hotels in Locarno-Ascona near Lake Maggiore, which is the most important tourist area in Canton Ticino, Switzerland. These bookings are registered during the peak months of July to October 2011. We only focus on three- and four-star hotels to reduce the choice set and given the similarities between these two hotel categories. Within this context, our data refers to the online bookings for double rooms in three- and four-star hotels that are offered in Booking.com and are simultaneously connected to the booking engine of the local destination management organization (DMO).

The local DMO's booking engine has recorded details of each booking that are made on the OTA platform, as well as the rates of the different rooms available for each hotel that are

quoted on the OTA on a daily basis. We select for further screening 31 hotels that offer double bedrooms and are available during the analyzed time period. To guarantee the stability of our econometric results, a hotel must have a minimum of 15 bookings to be included in our final sample. Five hotels that have satisfied our predefined sample requirement are selected to represent the deterministically identified choice set. The cut-off point was arbitrary decided and set to a minimum of 5% of market share, resulting in a final data set of 320 online bookings. Table 1 shows the descriptive statistics of the sample. The choice of the hotels represents the dependent variable while three hotel's critical dimensions, namely, price, location, and user-generated ratings and two control variables (i.e. lake view and swimming pool) are used as independent variables.

TABLE 1 - ABOUT HERE

Price. To perform a discrete choice analysis, each observation in the data set must consider the attributes of both chosen and non-chosen alternatives. Therefore, for hotels that offer two or more types of rooms (hotels C, D, and E), the price of the non-chosen alternative must be accordingly selected among the prices of the different types of rooms (see Table 1). In this context, for the non-chosen hotels C and D, the price of the standard room is considered because it represents the typical layout within the sampled hotels. A further step must be undertaken with regard to the price of the non-chosen hotel E, as the three available rooms either have or do not have a lake view. The price of a standard room with a garden view is selected whenever Hotel A or Hotel B is chosen, the price of a standard room with a lake view is selected whenever a small room with a lake view from hotels C or D is chosen, and the price of a superior room with a lake view is selected whenever a standard room with a lake view from hotels C or D is chosen.

Location. The location of the hotel is measured based on its proximity to the city center (being it either Locarno or Ascona). Table 1 distinguishes the hotels near the city center (hotels A, B, and E) from the hotels in a natural setting (hotels C and D).

User-generated ratings. We use the widely popular TripAdvisor travel site with regard to the user-generated ratings. The overall rating of each hotel at the time of booking is included in the data set. Based on the five-point scale of TripAdvisor (from "terrible" to "excellent"), the hotels in the sample have obtained a remarkable average rating of 4.4, which is between "very good" and "excellent."

Control variables. The four-star hotel provides the option to choose between rooms with lake view and garden view allowing for the distinction of the marginal utility associated with the specific feature "lake view" compared to "garden view". We further notice that two hotels (hotels C and E) offer a swimming pool among their amenities.

We rely on a multinomial logit (MNL) model for the single-stage process, whereas we rely on a nested logit (NL) model for the two-stage process. To validate the findings and their robustness, two model specifications are estimated while distinguishing the types of stay, namely, weekend (or non-weekend) stays and up to (or more than) three nights stays. The utility that is associated with person n and alternative j (U_{nj}) is assumed to be the sum of the systematic part (V_{nj}) and the error term (ε_{nj}). In this context, the systematic part of the five hotels' utility functions in the first model specification is described as follows:

$$M1\begin{cases} V_{n(A,B)} = ASC_{(A,B)} + \beta_{P_{WE}} * Price_{WE} + \beta_{P_{\neq WE}} * Price_{\neq WE} + \beta_{D_{WEnt}} * Dist_{WEnt} + \beta_{D_{\neq WEnt}} * Dist_{\neq WEnt} \\ \quad + \beta_{Pool_{nt}} * Pool_{nt} + \beta_{Rating} * Rating \\ V_{n(C,D)} = ASC_{(C,D)} + \beta_{P_{WE}} * Price_{WE} + \beta_{P_{\neq WE}} * Price_{\neq WE} + \beta_{DNS_{WEnt}} * Dist_{WEnt} + \beta_{DNS_{\neq WEnt}} * Dist_{\neq WEnt} \\ \quad + \beta_{Pool_{nt}} * Pool_{nt} + \beta_{Rating} * Rating \\ V_{nE} = \beta_{P4_{WE}} * Price_{WE} + \beta_{P4_{\neq WE}} * Price_{\neq WE} + \beta_{D_{WEnt}} * Dist_{WEnt} + \beta_{D_{\neq WEnt}} * Dist_{\neq WEnt} \\ \quad + \beta_{Pool_{nt}} * Pool_{nt} + \beta_{Rating} * Rating + \beta_{Lake_{View}} * Lake_{View} \end{cases} \quad (1)$$

where ASC(.) refers to the alternative specific constants that are estimated for the four three-star hotels (i.e., normalized in respect to the four-star hotel), β_{PWE} and β_{P4WE} ($\beta_{P\neq WE}$ and $\beta_{P4\neq WE}$) are the price coefficients that are associated with the weekend (non-weekend) bookings for the three- and four-star hotels, respectively, β_{DWEnt} and $\beta_{DNSWEnt}$ ($\beta_{D\neq WEnt}$ and $\beta_{DNS\neq WEnt}$) are the location coefficients that are associated with the weekend (non-weekend) bookings interacted by the number of nights for hotels that are close to the city center and for hotels in a natural setting, respectively, β_{Poolnt} is the coefficient that is associated with the availability of a swimming pool interacted by the number of nights, β_{Rating} is the coefficient that is associated with the overall rating in TripAdvisor, and $\beta_{LakeView}$ is the coefficient capturing the marginal utility associated with the room attribute "lake view" with respect to "garden view" in the four-star hotel.

Similarly, the systematic part of the five hotels' utility functions in the second model specification is described as follows:

$$M2\begin{cases} V_{i(A,B)} = ASC_{(A,B)} + \beta_{P_{n13}} * Price_{n13} + \beta_{P_{n4+}} * Price_{n4+} + \beta_{D_{n13nt}} * Dist_{n12nt} + \beta_{D_{n4+nt}} * Dist_{n4+nt} \\ +\beta_{Pool_{nt}} * Pool_{nt} + \beta_{Rating} * Rating \\ V_{i(C,D)} = ASC_{(C,D)} + \beta_{P_{n13}} * Price_{n13} + \beta_{P_{n4+}} * Price_{n4+} + \beta_{DNS_{n13nt}} * Dist_{n13nt} + \beta_{DNS_{n4+nt}} * Dist_{n4+nt} \\ +\beta_{Pool_{nt}} * Pool_{nt} + \beta_{Rating} * Rating \\ V_{iE} = \beta_{P4_{n13}} * Price_{n13} + \beta_{P4_{n4+}} * Price_{n4+} + \beta_{D_{n13nt}} * Dist_{n13nt} + \beta_{D_{n4+nt}} * Dist_{n4+nt} \\ +\beta_{Pool_{nt}} * Pool_{nt} + \beta_{Rating} * Rating + \beta_{Lake_{View}} * Lake_{View} \end{cases} \quad (2)$$

where β_{Pn13} and β_{P4n13} (β_{Pn4+} and β_{P4n4+}) are the price coefficients that are associated with the bookings for up to three (more than three) nights in three- and four-star hotels, respectively, and β_{Dn13nt} and $\beta_{DNSn13nt}$ (β_{Dn4+nt} and $\beta_{DNSn4+nt}$) are the location coefficients associated with the bookings for up to three (more than three) nights interacted by the number of nights for hotels that are close to the city center and for hotels in a natural setting, respectively. The rest of the coefficients are explained in Equation (1).

The two model specifications in Equations (1) and (2) are estimated for both the single-stage (MNL model) and two-stage (NL model) process assumptions, which amounts to a total of four models. A higher performance associated with the NL model (compared to the MNL model) would provide supporting evidence for a nested decision making process over a single-stage process.

RESULTS AND CONCLUSIONS

We rely on the likelihood function and the Akaike Information Criterion (AIC) to identify the best model. Tables 2, 3 and 4 show the comparison among the three models, where the empirically derived two-stage model significantly ($p < 0.05$) outperforms the single-stage model and the a priori-defined two-stage model.

TABLE 2 - ABOUT HERE

TABLE 3 - ABOUT HERE

TABLE 4 - ABOUT HERE

This result implies that, first, the two-stage models better represent the decision-making process of a tourist in choosing a hotel, which confirms our hypothesis that tourists choose their hotels through nested decision structures, that is in line with the Associative Network Theory (Collins & Loftus, 1975) (activating of information through nested links) and the Cybernetic decision-making model (Steinbruner, 2002) (nested choice processes to reduce the decision-making process' complexity). Second, individuals consider other characteristics aside from star ratings when forming groups or nests of hotels. Hotel C, which is located in a natural setting, possesses unique characteristics (particularly its panoramic view) that separate it from the other nests. Star ratings effectively help guests realize their expectations as these ratings can be seen as an explicit promise of quality service that consequently raises

a potential guest's service expectations (Zeithaml, Bitner & Gremler, 2006). Therefore, a hotel with a higher star rating is expected to provide better service (Ariffin & Maghzi, 2012). However, according to this result, other relevant attributes may also emerge when individuals use their perceptions to form groups of hotels.

TABLE 5 - ABOUT HERE

Table 5 shows the results of the empirically derived two-stage model. Regarding the model's variables, we find that price has an expected negative influence on the choice of hotel. Surprisingly, the negative effect of prices decreases when choosing a four-star hotel (despite its higher prices) rather than three-star hotels. Not only does this prove that the expected superior quality of service and the intangible part of tourism consumption may counteract the negative effect of prices, but this also confirms our previous reasoning in which individuals may not resort exclusively to star ratings when forming groups of alternatives from which they can choose their preferred hotel. Location has different effects on the choice of hotel, which depends on the tourists' length of stay. Individuals who stay for up to three days prefer hotels that are closer to the city center. In line with the previous literature, online hotel user-generated ratings can increase the tendency for consumers to book a room in the rated hotel.

The theoretical implications of this study are as follows: first, although the mere booking and choice of a hotel is not considered to be a purchase with the highest involvement among tourism products (Sanchez et al., 2008), the results show the individuals' tendency to use a nested decision process. Therefore, if the analyst wants to imitate an individual's decision-making process as much as possible, he or she must consider more than one stage as it can accurately reflect an individual's mental processing of information. Second, although a priori-defined groups of hotels can be readily obtained, the inclusion of a hotel to a specific group seems to be more of an empirical issue than a theoretical one as the included hotel

must comply with the prevalent hotel attributes in each group. Third, the differentiation of the price's effect depending on the hotel's star rating supports Lockyer's (2005) findings, in which the expected negative effect of prices can be influenced by its connection with other "trigger points." Accordingly, the effect of price must not be modeled in isolation.

Finally, as the results prove that a two-stage decision process is favored -compared to a single decision process- it means that a variety of nested decision processes could be proposed and analyzed in the context of hotel online bookings with real market data, both at the hotel level (as the one tested here) or at specific services of a hotel (e.g. location-hotel-room type).

REFERENCES

Ariffin, A.A.M., & Maghzi, A. (2012). A preliminary study on customer expectations of hotel hospitality: Influences of personal and hotel factors. *International Journal of Hospitality Management*, 31, 1, 191-198.

Boffa, F., & Succurro, M. (2012). The impact of search cost reduction on seasonality. *Annals of Tourism Research*, 39, 2, 1176-1198.

Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 6, 407-428.

Guo, X., Ling, L., Dong, Y., & Liang, L. (2013). Cooperation contract in tourism supply chains: the optimal pricing strategy of hotels for cooperative third party strategic websites. *Annals of Tourism Research*, 41, 20-41.

Jeng, J., & Fesenmaier, D. R. (2002). Conceptualizing the travel decision-making hierarchy: A review of recent developments. *Tourism Analysis, 7*(1), 15–32.

Kim, D., & Perdue, R.R. (2013). The effects of cognitive, affective, and sensory attributes on hotel choice. *International Journal of Hospitality Management*, 35, 246-257.

Lockyer, T. (2005). The perceived importance of price as one hotel selection dimension. *Tourism Management*, 26, 529–537.

Sanchez, J.; Callarisa, L.; Rodriguez, R.M. and Moliner, M.A. (2006), "Perceived value of the purchase of a tourism product", *Tourism Management*, 27, 394–409.

Steinbruner, J. (2002). *The cybernetic theory of decision*. Princeton University Press, Princeton.

Zeithaml, V.A., Bitner, M.J., & Gremler, D.D. (2006). *Services Marketing*. Integrating Customer Focus Across the Firm, 4th ed. McGraw-Hill, New York.

Tables

Table 1. Descriptive statistics.

MARKET SHARE				
Hotel A (3-star)	15%			
Hotel B (3-star)	14%			
Hotel C (3-star)	8%			
Hotel D (3-star)	17%			
Hotel E (4-star)	46%			
TYPE OF STAY				
Weekend stays (2 nights)	15.3%			
Short stays (1-3 nights)	86.6%			
TYPE OF DOUBLE ROOM				
Hotel A (three-star)	Standard without a lake view			
Hotel B (three-star)	Standard without a lake view			
Hotel C (three-star) – with pool	Small with a lake view			
	Standard with a lake view			
Hotel D (three-star)	Small with a lake view			
	Standard with a lake view			
Hotel E (four-star) – with pool	Standard with a garden view			
	Standard with a lake view			
	Superior with a lake view			
HOTELS	Mean	Std. Dev.	Min	Max
Price CHF[(a)]	239.9	55.4	136.0	380.0
Length of stay (nights)	2.2	1.2	1.0	7.0
Distance from city center (km)				
Hotels located in the city (A, B, E)	1.1	0.9	0.2	2.0
Hotels located in the nature (C, D)	7.3	0.1	7.2	7.3
TripAdvisor overall rating	4.4	0.2	4.0	4.7

[(a)] 1 CHF = approx. 0.8 EUR

Table 2. Single-stage model vs. empirically derived two-stage model.

Log-likelihood t-test	Estimate 1 (weekend)	Estimate 2 (length of stay)
MNL	-358.73	-355.42
NL ‘‘empirically derived’’	-355.68	-352.34
Statistic	6.100	6.160
p-value	0.014	0.013
AIC - MNL	2.34	2.32
AIC - NL ‘‘empirically derived’’	2.32	2.30

Table 3. Single-stage model vs. “stars” two-stage model.

Log-likelihood t-test	Estimate 1 (weekend)	Estimate 2 (length of stay)
MNL	-358.73	-355.42
NL ‘‘stars”	-358.381	-355.065
Statistic	0.698	0.710
p-value	0.40	0.39
AIC - MNL	2.34	2.32
AIC - NL ‘‘stars”	2.34	2.32

Table 4. A priori-defined vs. empirically derived two-stage models.

Log-likelihood t-test	Estimate 1 (weekend)	Estimate 2 (length of stay)
NL ‘‘stars’’	-358.381	-355.065
NL ‘‘empirically derived’’	-355.680	-352.340
AIC - NL ‘‘stars’’	2.34	2.32
AIC - NL ‘‘empirically derived’’	2.32	2.30

Table 5. Estimates of the empirically derived two-stage model.

	Equation 1		Equation 2	
	Weekend		Length of stay	
	Parameter	**t-ratio**	**Parameter**	**t-ratio**
ASC A	11.532***	(6.25)	10.148***	(5.49)
ASC B	5.396***	(3.50)	4.267***	(2.68)
ASC C	4.742	(1.11)	6.534*	(1.72)
ASC D	6.263***	(4.37)	4.931***	(3.37)
Price WE three-star	-0.084***	(-8.35)	-	-
Price N-WE three-star	-0.094***	(-9.73)	-	-
Price n123 three-star	-	-	-0.091***	(-9.50)
Price n4+ three-star	-	-	-0.085***	(-7.78)
Price WE four-star	-0.036***	(-6.12)	-	-
Price N-WE four-star	-0.045***	(-9.20)	-	-
Price n123 four-star	-	-	-0.045***	(-9.34)
Price n45+ four-star	-	-	-0.028***	(-3.67)
Distance WE	-0.701**	(-2.24)	-	-
Distance N-WE	-0.296**	(-1.97)	-	-
Distance n123	-	-	-0.353	(-1.58)
Distance n4+	-	-	-0.452**	(-2.40)
Distance WE (N.S.)	-0.189***	(-3.55)	-	-
Distance N-WE (N.S.)	-0.052*	(-1.90)	-	-
Distance n123 (N.S.)	-	-	-0.070*	(-1.75)
Distance n4+ (N.S.)	-	-	-0.047	(-1.59)
Pool*Nights	0.490**	(2.46)	0.071	(0.26)
Rating (TripAdvisor)	6.108***	(5.15)	5.745***	(4.97)
LAKE view four-star	3.139***	(7.90)	3.067***	(7.76)
Scale Parameters				
A (A,B,D,E)	0.40**	(2.10)	0.46**	(2.45)
B (C)	*1.000*	*(fixed)*	*1.000*	*(fixed)*
Model Fits				

Number of observations	320	
Log-likelihood restricted	-630.764	
Log-likelihood (ASCs)	-453.44	
Log-likelihood	-355.69	-352.35
Number of parameters	16	16
McFadden pseudo ρ^2	0.436	0.441
AIC normalized	2.323	2.302

Note: *** = prob < 1%; ** = prob<5%; * = prob < 10%.

Figures

Figure 1. Nested vs. non-nested decision structures.