Tourismus & Hospitality

The individual monetary valuation of online hotel ratings

Christos Evangelinos
Andy Obermeyer
Aaron Bartel
THE INDIVIDUAL VALUATION OF ONLINE HOTEL RATINGS

Christos Evangelinos
Andy Obermeyer
Aaron Bartel

Abstract:
We test the impact of online hotel ratings on the customers’ hotel choice using a binary choice experiment where the online rating score is one of the hotels’ attributes. Results show that online rating scores have a positive and significant impact on hotel choice. We also calculate the individual willingness to pay for higher rating scores and report the corresponding income and age elasticities.

Keywords:
Hotel rating, online review, discrete choice model, logit model, monetary valuation, eWOM.

JEL classification:
C25; L8; D12; Z31.
1. Introduction

The number of consumers searching for travel-related information online is steadily increasing (Morosan & Jeong, 2008), also leading to a drastically increased number of consumers booking online. A study for instance by the Association for Online Travel Sales (Verband Internet Reisevertrieb, 2017) found that the share of Germans who booked travel related services online has increased from 29% in 2010 to 43% in 2016.

Parallel to the rise of online bookings, the use of web 2.0, where users publish or search user-generated content (UGC), increased substantially as well (Sigala, 2009). Not only does UGC decrease the information asymmetry (Schegg & Fux, 2010; Zhang, Khan & Shih, 2015), the users also perceive the reviews as highly trustworthy (Del Chiappa, Lorenzo-Romero & Alarcón-del-Amo, 2015; Yoo, Lee, Gretzel & Fersenmaier, 2009). Moreover, UGC is the most crucial information source for travelers (Sidali, Schulze & Spiller, 2009).

From an economic perspective, the value travelers attach to such online hotel ratings is of fundamental interest. To determine travelers’ willingness to pay (WTP) for higher ratings, we examine how online ratings, along with other attributes such as prices, affect the travelers’ hotel choice. Based on these results we can also calculate demand elasticities. Such information may be of great importance for hotel management (pricing, assessment of investments in quality and service).

The empirical literature on the economic valuation of hotel reviews utilizes Lancaster’s (1966) and Rosen’s (1974) notion of hedonic prices. In this approach, the price of a good is a function of its characteristics. Several studies use this approach in the hotel sector and identify a positive and significant impact of online hotel ratings on room rates. Rosen’s (1974, p. 38, 46) approach relies on certain assumptions, in particular, complete information of buyers and sellers, the absence of transaction costs and perfect competition. One may argue that online hotel ratings eliminate information asymmetries, and booking portals reduce transaction costs so that the first two conditions are not critical. However, potential distortions may result from imperfect competition. Hotels may have a certain degree of spatial and temporal (high vs. low season) market power. Additionally, general econometric issues associated with revealed preferences, in particular, omitted variable bias have to be considered. Therefore, hedonic pricing may potentially derive distorted WTP estimates.

Analyzing stated preference data with discrete choice models offers an excellent opportunity to avoid such problems. In a well-defined experimental environment, respondents choose one of the presented alternatives associated with its attribute values. Such modelling approaches are conventional in the transportation literature, but to the best of our knowledge, have been rarely applied for the hospitality industry. Indeed, stated choice (SC) analyses show weaknesses concerning real-world forecasts, as they typically include only a selection of attributes. Furthermore, SC analyses may also be susceptible to strategic answers from the respondents. However, to estimate the willingness to pay, which is the primary objective of our analyses, they have proven to be one of the most suitable instruments across a wide range of economic applications.

This paper analyses travelers’ booking behavior in the presence of online ratings. We perform and evaluate discrete choice experiments with 250 individuals choosing between two different hotels characterized by specific attributes. Using this approach, we estimate the WTP for rating increases and demand elasticities.

Section 2 reviews the relevant literature. Section 3 describes the choice experiment and the data collected. Subsequently, section 4 presents the modelling approach, the estimation results accompanied by a discussion. Finally, section 5 summarizes the results.

2. Literature Review

Markets concerning travel and tourism, and in particular hotels, are characterized by significant information asymmetries (Cezar & Ögüt, 2016). A Hotel can signal quality by price, where customers perceive a higher room

---

1 For the factors contributing to trustworthiness see Filieri (2016). A review is provided by Cantallops & Salvi (2014).
rate as an indicator of higher quality (Chen & Schwartz, 2006), and its brand or in the context of the hospitality industry its chain affiliation (Israeli, 2002), which signals the reliability of the product and service quality (Keller, 2003). Furthermore, a very prominent and long-facilitated quality signal are hotel classification systems such as star ratings. Classification systems traditionally include the amenities available to the guest (World Tourism Organization, 2014) and are often introduced by governments to reduce information asymmetry (Šergo & Damijanić, 2015). A study by Callan (1998), carried out before the advent of the internet and thus unbiased by electronic word of mouth, showed that these star ratings are essential factors in the travelers’ decision process. Martin-Fuentes (2016) found that there is a statistically significant positive relationship between hotel star classification and a review score. Travelers are, therefore, on a large scale, agreeing with the official star classification.

The search for information about hotels is nowadays mostly done online (Xiang, Wang, O'Leary & Fesenmaier, 2015), where the traveler can use four different categories of information services: 1) search engines, 2) hotel metasearch websites, 3) online travel agencies (OTAs), and 4) the hotels website (Murphy & Chen, 2016). OTAs accounted for 22% of all European hotel bookings in 2015, where "Booking.com" was the most prominent platform with a share of 62% of all OTA-bookings (Schegg, 2016), indicating the importance of online intermediaries.

Online services enable travelers to evaluate hotels not only by its natural attributes (e.g. room size or amenities) but also by online ratings. Since this comparison often takes place on the same platform (e.g. OTA or meta-search website) on which the customer books the hotel, the influence of these online hotel ratings on the information- and evaluation stage of travelers (Varkaris & Neuhofer, 2017) is apparent.

Additionally, web 2.0 enables user interaction (O'Reilly, 2007). In contrast to traditional systems, such as travel guides or star classifications, the content (rating) is no longer created by a single author (individual, company, association). Instead, the content is created and continuously modified by all users (Kaplan & Haenlein, 2010). Web 2.0 as a platform forms the basis for social media (Yoo & Gretzel, 2012). The content created by end-users is publicly available and is referred to as user-generated content (UGC) (Kaplan & Haenlein, 2010). The UGC posted by travelers and influencing others is referred to as electronic word of mouth (eWOM) (Del Chiappa, et al., 2015). eWOM comprises “all informal communications directed at consumers through internet-based technology related to the usage or characteristics of particular goods and services, or their sellers” (Litvin, Goldsmith & Pan 2008). It shows substantial similarities with the traditional word of mouth (WOM), which is the informal communication between individuals about the quality and attributes of goods and services (Dichter, 1966). The informality of the communication and perceived independence of the information provider is the essential aspect of WOM (Litvin, et al., 2008).

In the hospitality industry, web 2.0 enables travelers to share their opinions easily with a broad audience (Xie, Zhang & Zhang, 2014), who use these reviews primarily in the information stage of the decision process (Cox, Burgess, Sellitto & Buultjens, 2009) to confirm or reject a pre-made choice of hotels (Anderson, 2012). Reviews and ratings in the context of eWOM have become the most influential source of information for booking decisions (Bore, Rutherford, Glasgow, Taheri & Antony, 2017; Sidali, et al., 2009). Empirical literature identifies a robust positive relationship between higher ratings and booking intentions of travelers (Mauri & Minazzi, 2013; Sparks & Browning, 2011; Vermeulen & Seegers, 2009; Ladhari & Michaud, 2015).

Regularly online consumer reviews result in accumulated ratings. We will focus on precisely these ratings in our analysis. We should mention, however, that travelers also use other forms of assessment, including written reviews. Potential customers assess the quality of these reviews and perceive them as a more valuable source of information than pure ratings (Filieri, 2015).

Liang, Ekinci, Occhiocupo and Whyatt (2013) note that reviews are perceived as a signal on the expected service level of the hotel. This relation is confirmed by Mauri and Minazzi (2013), who found a positive correlation between ratings and expected service level.

The findings of Zhao, Wang, Guo and Law (2015) on the influence of specific review aspects (e.g. usefulness, reviewer expertise, timeliness, volume, valence, and comprehensiveness) on booking intentions seem to be in line with the insights of the other literature mentioned here, in particular that higher ratings (also for single aspects) positively influence booking intentions. Stringam and Gerdes (2010) report that travelers assign higher ratings to...
well-located hotels, underlining the importance of considering spatial effects in the empirical analyses. In addition, with respect to usefulness Park & Nikolau (2017) showed that reviews are perceived in a non-linear (convex) fashion as useful and extreme reviews (in particular negative) are more useful than less extreme ones.

The increased booking intentions through higher ratings (Sparks & Browning, 2011) eventually result in increased sales volumes and higher revenues (Öğüt & Onur Taş, 2012). Öğüt and Onur Taş (2012) found that a one percent increase in rating leads to a sales increase per room of up to 2.68% in Paris and 2.62% in London. Since the authors did not find any impact of star ratings on sales volume, it is evident that consumers focus on online ratings rather than star ratings. Ye, Law, Gu and Chen (2011) also found an increase in sales volume of more than 5% if the hotel has a 10% higher rating.

The overall rating strongly influences the prices a hotel can charge. Anderson (2012) noticed that a rating increase by one point on a five-point scale enables the hotel to charge a price premium of 11.20% and corresponds to a 14.2% increase of the hotel choice probability. This finding is further supported by Torres, Singh and Robertson-Ring (2015), who found that each point on the “www.tripadvisor.com” five-point scale leads to a revenue increase of $280 per transaction.

The literature mentioned above underlines the importance travelers assign to eWOM. Given the impact of online ratings on booking intentions and hotel revenues, several researchers became interested in the monetary value that travelers attach to online ratings. Table 1 provides a selection of research contributions on the estimation of WTP for online hotel rating increases.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Topic</th>
<th>Method/Approach</th>
<th>Data Source</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Öğüt and Onur Taş (2012)</td>
<td>Impact of star classification and online ratings on hotel room sales and prices.</td>
<td>Hedonic Pricing</td>
<td>388 hotels in London (UK), 562 in Paris (France) on Booking.com</td>
<td>1% higher rating increases sales by 2.68% in Paris and 2.62% in London</td>
</tr>
<tr>
<td>De la Pena et al. (2016)</td>
<td>Impact of innovative activity in the hotel industry on the willingness to pay.</td>
<td>Hedonic Pricing</td>
<td>176 Cuban hotels</td>
<td>Positive impact of higher ratings on prices.</td>
</tr>
<tr>
<td>Zhang et al. (2011)</td>
<td>Influence of hotel attributes on room rates</td>
<td>Hedonic Pricing</td>
<td>243 hotels in New York City on TripAdvisor</td>
<td>12.4% price increase per point in tripadvisor.</td>
</tr>
<tr>
<td>Xie et al. (2016)</td>
<td>Consumer preferences among three types of online distribution channels: regular full information sales channels, and opaque sales channels.</td>
<td>Discrete Choice Experiment</td>
<td>531 participants, 10 choice scenarios each.</td>
<td>Utility increases with the rating. But: results not significant (p &gt; 0.10)</td>
</tr>
<tr>
<td>Schamel (2012)</td>
<td>Differentiation between single vs. double rooms, and weekdays vs. weekends.</td>
<td>Hedonic pricing</td>
<td>911 Observations from trivago N.V. in the vicinity of Bolzano (Italy).</td>
<td>Depending on the model WTP for one point on a 100 point scale between €0.54 and €2.07.</td>
</tr>
<tr>
<td>Abrate et al. (2011)</td>
<td>Star ratings as a function of hotel attributes.</td>
<td>Hedonic pricing; simultaneous log-linear equations</td>
<td>145 hotels in the Piedmont region (Italy).</td>
<td>Strong positive and significant effects of star ratings and other quality signaling variables on room rates.</td>
</tr>
</tbody>
</table>

Table 1: Willingness to pay for online ratings – selected empirical literature. Source: Own depiction.
As Table 1 shows, the majority of researchers employ hedonic regression in order to quantify the WTP for online hotel ratings. Notwithstanding the different scales of measurement, all papers in Table 1 found a positive and significant impact of online reviews on hotel prices. Note that some of the hedonic studies in Table 1 employ a log-linear modelling approach, meaning that the estimated coefficient for online reviews corresponds to a semi-elasticity. The approach by Xie et al. (2016) is to the best of our knowledge the only study, which employs choice modeling for WTP estimation in this context. However, the authors of this study did not obtain significant results.

3. Choice experiment and data

We focus on online reviews but not on the star ratings. We restrict our experiment to eWOM ratings within the same star category to keep the experiment simple for the participants. We constructed a choice experiment that allows choosing between two hotels described by specific attributes (see below). As usual in such experiments, participants were informed that the alternatives do not differ across the remaining but unmentioned attributes that could be relevant for hotel selection. Hence, these attributes are irrelevant for the persons’ choice, and their absence does not distort the results.

We have chosen the city of Düsseldorf, Germany as the reference market for hotels in our analysis. Using a particular location makes the experiment more realistic for the participants. Düsseldorf is a destination for both leisure and business travelers, with a steadily increasing number of guests, and with a large proportion of international guests (Brabant & Schwendner, 2014). Therefore, the location is known to both German and international respondents. However, selecting Düsseldorf as an illustration for the experiment is not expected to distort the choices of the respondents.

Furthermore, we decided on the reference hotel star rating (fixed in our analysis) based on information about the Düsseldorf hotel market. According to the Hotelstars Union, approximately 55% of the available hotel capacity (by number of guest rooms) in Düsseldorf has four stars (Brabant & Schwendner, 2014). Thus, we fixed the number of stars according to the star classification system to four stars in our experiment description.

Having defined the framework of the choice experiment, we proceed with a description of the hotel attributes:

- Rating score: Being the primary component of interest, eWOM enters the choice experiment in the form of a scale from 0 (worst) to 10 (best). The attribute values comprise eight levels, increasing in 0.3 increments and ranging from 7.4 to 9.5. The scale and its range are based on the ratings of the majority of four-star hotels in Düsseldorf on the website of the OTA “Booking.com”.

- Chain affiliation: As the majority of studies (using hedonic pricing) shows, chain affiliation influences room rates. Hotel guests are quite willing to pay an additional amount of money if a hotel belongs to a chain. Theoretically, chain affiliation may be regarded as a signal of quality (Israeli, 2002). We consider preferences for chain affiliation in our experiment by stating whether the alternative belongs to a chain or not. Chain affiliation was specified by three examples of chains in the four-star category, namely Best Western, Mercure, and NH Hotels, below the choice situations. In our model, the chain affiliation is captured by a binary variable indicating chain affiliation by 1 and no chain affiliation by 0.

- Travel time to the final destination: Hotel guests usually book a room not to stay in the hotel but to pursue various private and business-related activities in the city. For that reason, people usually have a particular point of interest in the city (e.g. sightseeing destination, conference center). The distance between the hotel and the person’s point of interest probably influences its booking decision. Therefore, we included the hotel’s distance from the point of interest in the experiment by specifying the walking time. This attribute has four levels (categories), namely 0 to 5 minutes, 6 to 10 minutes, 11 to 15 minutes, and 16 to 20 minutes.

Note that Table 1 does not contain the whole literature on hedonic approaches in the hotel sector. Several further studies have been performed, e.g., by Aguiló, Alegre and Riera, 2001; Alegre, Caldera and Sard, 2013; Espinet, Saez, Coenders and Fluvià, 2003; Papatheodorou, 2002; Thrane, 2005 utilizing information on package tour prices, or by Carvell and Herrin, 1990; Thrane, 2007; Hamilton, 2007; Monty and Skidmore, 2003; Rigall-i-Torrent and Fluvià, 2011 using room rates.
Price: The room rate is one of the most critical variables for potential customers when making booking decisions. This attribute comprises nine levels, from €80 to €120 in increments of €5. The lower and upper limits were chosen on the basis of an average hotel room rate for four-star hotels in Düsseldorf of €104 as in 2014 (Brabant & Schwendner, 2014) and a realistic price volatility of 20% around the mean.

Table 2 summarizes the attributes and levels used in the experimental design.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating score</td>
<td>7.4 – 9.5 in steps of 0.3</td>
</tr>
<tr>
<td>Chain affiliation</td>
<td>Yes – No</td>
</tr>
<tr>
<td>Price</td>
<td>€80 – €120 in steps of €5</td>
</tr>
<tr>
<td>Travel time to the final destination</td>
<td>0 – 5; 6 – 10; 11 – 15; 16 – 20 (minutes)</td>
</tr>
</tbody>
</table>

Table 2: Attributes and levels for the choice experiment.

In the experiment, each respondent was asked to choose between two hotels whereby the attributes varied according to the values presented above. The choice experiment was repeated five times for each participant with different combinations of the attribute levels. The combinations were assigned at random from the set of all possible combinations (see e.g. Brownston, Bunch & Train, 2000; Hollander, 2006; Train & Wilson, 2008). Since all choice alternatives are characterized by the same attributes with the same levels, the names “Hotel A” and “Hotel B” are not inherently specific. This kind of choice experiment is denoted as unlabeled in the literature (Rose & Bliemer, 2009).

Before the choice situations were presented to the participants, the setting of the experiment was described. Persons were asked to imagine planning a private overnight stay in a double bedroom in a four-star hotel in Düsseldorf whereby the room has to be paid for by themselves. This information was also made available in every choice situation to ensure that each participant was informed adequately at any point in time. Figure 1 shows an example of a choice situation.

Please choose spontaneously the hotel that is most appealing to you, and that you would consider booking.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Hotel A</th>
<th>Hotel B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price per night</td>
<td>€115</td>
<td>€90</td>
</tr>
<tr>
<td>Does the hotel belong to a chain?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Distance (minutes) to final destination</td>
<td>11 – 15</td>
<td>0 – 5</td>
</tr>
<tr>
<td>Rating score</td>
<td>8.6</td>
<td>7.7</td>
</tr>
</tbody>
</table>

Table 3: A typical choice situation in the experiment.

Following the experiments, participants were asked to provide socio-demographic information such as gender, age, education level, occupation, and net monthly income.

We distributed the questionnaire via social media and encouraged participants to disseminate it further. This snowball sampling allows us to efficiently obtain a sufficiently large and heterogeneous sample. We believe this sampling method is appropriate for our research purpose because choosing between travel options is a common task for the vast majority of people. The proportion of Germans, for instance, who travel for more than five days per year was 77% in 2016, and only 12% of Germans indicated that they have not traveled in the preceding three years. (Verband Internet Reisevertrieb, 2017).
The survey period ran from 13th January to 9th February 2017. During this period, 300 responses were recorded, 250 of which were usable for data analysis, resulting in 1250 observations (5 choices per individual). About half of the respondents were female. The average age in the sample is 30 years with a standard deviation of roughly nine years and a median of 27 years. Therefore, focusing on the German population, the sample is representative concerning gender but not concerning age since we do not have enough observations for high age groups (Statistisches Bundesamt, 2016). To encourage respondents to indicate their net income honestly, we used five income categories in the questionnaire rather than asking to state the income directly. Using the middle of each income class, we calculated an average monthly net income of €1848 with a standard deviation of €1337. This seems to be very close to the overall average net monthly income of €1888 in Germany for 2017 (Statistisches Bundesamt, 2018).

4. Model

We analyze the data using choice models, especially logit models, to determine the impact of attributes on the persons’ choices. Logit models are conventional in the literature on stated or revealed choices. Therefore, we do not provide a detailed explanation of the theoretical principles. A comprehensive overview of discrete choice models can be found, for example, in Train (2009) or Ben-Akiva and Lerman (1994). The choice process and the influencing attributes are related to the latent utility a person receives from choosing a particular alternative (e.g. “Hotel A”) from a given set of alternatives (e.g. “Hotels A and B”). It is assumed that decision makers always choose the option with the highest utility ($U$), which is decomposed into a deterministic ($V$) and a stochastic ($\varepsilon$) component. In the following, we consider the utility difference between “Hotel A” and “Hotel B”, because this determines the choice decision:

$$\Delta U = \Delta V + \Delta \varepsilon$$

(1)

The deterministic utility component is a function of the differences in travel time to the final destination ($T$), chain affiliation ($C$), rating score ($R$) and price ($P$):

$$\Delta V = \beta_T \Delta T + \beta_C \Delta C + \beta_R \Delta R + \beta_P \Delta P$$

(2)

In the questionnaire, travel time was presented to the respondents in categories with a minimum and a maximum walking time (cf. Section 3). However, for model estimation travel time is represented by a variable indicating the maximum walking time to the final destination. Accordingly, travel time is treated as a continuous variable in the model. Using a continuous variable is advantageous in the sense that the model is much more parsimonious and marginal rates of substitutions can be determined. For a description of the remaining variables see Section 3.

The stochastic component of the utility for each hotel is assumed to be iid Gumbel distributed and, therefore, the difference between the two stochastic components ($\varepsilon$ for “Hotel A” and “Hotel B”) is distributed logistically. Applying the corresponding cumulative distribution function to the calculation of the probability $P_A$ for choosing “Hotel A” results in the following formula:

$$P_A = \Pr(\Delta U > 0) = \Pr(-\Delta \varepsilon < \Delta V) = \frac{1}{1 + e^{-\Delta V}}$$

(3)

Since the probabilities for “Hotel A” and “Hotel B” must add up to one, the probability $P_B$ is one minus $P_A$. The optimal parameters are determined by maximizing the log-likelihood function (4).

$$\Lambda = \sum_{n \in N} \sum_{h \in H} y_{nh} \ln P_{nh}$$

(4)

Set $H$ contains the available alternatives, in this case “Hotel A” and “Hotel B”, $N$ is the set of observations and the variable $y_{nh}$ equals one if and only if alternative $h$ was chosen in the $n$-th observation.
In addition to this Base Model (BM), we use more complex models, especially utility functions, to analyze the influence of age and income on the choices. We hypothesize (1) that persons of a younger age react more sensitive to upward changes in the hotel rating and (2) that people with higher income are less sensitive to price increases. The reason for the former is that younger people could be better acquainted with online reviews. The latter hypothesis is based on the expectation of a diminishing marginal utility of income. The influence of income on price sensitivity is modeled in two distinct ways. In a first model, denoted as Income Category Model (ICM), we estimate price parameters for each of the five income categories available in the questionnaire (cf. Section 3). In another model, we construct a continuous income variable based on the income categories. For each income category, we use its mid-point to approximate a person’s income. Using the continuous variable instead of the categories makes the model much more parsimonious, but it can no longer capture non-monotonic effects concerning income. Furthermore, one should be cautious in interpreting the resulting income elasticity ($\xi_I$) of the WTP, because the data provides only five different income levels, which in turn are only approximations. In the following, we refer to this model as the Income Elasticity Model (IEM). A similar model structure with respect to income has been used, for example, by Obermeyer, Treiber and Evangelinos (2015) and Axhausen et al. (2008) in the context of the value of travel time savings. The variable age is continuous by definition and, therefore, its influence on the rating score parameter is modelled by means of an elasticity ($\xi_A$) in both ICM and IEM. Eq. (5) and Eq. (6) show the corresponding deterministic utility differences for the ICM and the IEM. The binary variable $I_j$ indicates whether a person has been assigned to income category $j$. Consequently, the parameter $\beta_{Pj}$ denotes the price sensitivity of persons in the respective income category. For the ease of interpretation, the models include the mean-scaled age ($A_s$) and in the case of the IEM also the mean-scaled income ($I_s$). Hence, the parameters $\beta_R$ and $\beta_P$ represent marginal utilities for a person with mean age and mean income, respectively. 

\[
\Delta V = \beta_T \Delta T + \beta_C \Delta C + \beta_R A_s^{\xi_A} \Delta R + \left( \sum_{j=1}^{5} \beta_{Pj} I_j \right) \Delta P 
\]

(5)

\[
\Delta V = \beta_T \Delta T + \beta_C \Delta C + \beta_R A_s^{\xi_A} \Delta R + \beta_P I_s^{\xi_I} \Delta P 
\]

(6)

The interpretation of the elasticity parameters is described in the next section together with the derivation of the WTP measure. The use of both models is advantageous for interpretation. The ICM explicitly provides the price sensitivity of the respective income category, while the IEM immediately shows the price sensitivity of an average agent and the influence of income on the price sensitivity.
5. Results

We used BIOGEME (Bierlaire, 2016) to estimate the models. Table 3 summarizes the estimation results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>BM*</th>
<th>ICM*</th>
<th>IEM*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel time</td>
<td>-0.1300*</td>
<td>-0.1310*</td>
<td>-0.1310*</td>
</tr>
<tr>
<td></td>
<td>(0.0108)</td>
<td>(0.0109)</td>
<td>(0.0109)</td>
</tr>
<tr>
<td>Chain affiliation</td>
<td>0.3160*</td>
<td>0.2970*</td>
<td>0.2970*</td>
</tr>
<tr>
<td></td>
<td>(0.1050)</td>
<td>(0.1050)</td>
<td>(0.1050)</td>
</tr>
<tr>
<td>Rating score</td>
<td>1.0800*</td>
<td>1.0700*</td>
<td>1.0600*</td>
</tr>
<tr>
<td></td>
<td>(0.0886)</td>
<td>(0.0919)</td>
<td>(0.0917)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.0748*</td>
<td>-</td>
<td>-0.0718*</td>
</tr>
<tr>
<td></td>
<td>(0.0050)</td>
<td></td>
<td>(0.0053)</td>
</tr>
<tr>
<td>Price (IC 1)*</td>
<td>-</td>
<td>0.0856*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0087)</td>
<td></td>
</tr>
<tr>
<td>Price (IC 2)</td>
<td>-</td>
<td>0.0815*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0090)</td>
<td></td>
</tr>
<tr>
<td>Price (IC 3)</td>
<td>-</td>
<td>0.0702*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0104)</td>
<td></td>
</tr>
<tr>
<td>Price (IC 4)</td>
<td>-</td>
<td>0.0652*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0156)</td>
<td></td>
</tr>
<tr>
<td>Price (IC 5)</td>
<td>-</td>
<td>0.0513*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0119)</td>
<td></td>
</tr>
<tr>
<td>Income Elasticity</td>
<td>-</td>
<td></td>
<td>0.1620*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0734)</td>
</tr>
<tr>
<td>Age elasticity</td>
<td>-</td>
<td>0.588*</td>
<td>-0.6150*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3010)</td>
<td>(0.3030)</td>
</tr>
<tr>
<td>Null-LL</td>
<td>-866.434</td>
<td>-866.434</td>
<td>-866.434</td>
</tr>
<tr>
<td>Final-LL</td>
<td>-571.181</td>
<td>-566.211</td>
<td>-567.059</td>
</tr>
<tr>
<td>Adj. rho-squared</td>
<td>0.3362</td>
<td>0.3361</td>
<td>0.3386</td>
</tr>
<tr>
<td>AIC</td>
<td>1150.363</td>
<td>1150.421</td>
<td>1146.118</td>
</tr>
<tr>
<td>BIC</td>
<td>1170.886</td>
<td>1196.599</td>
<td>1176.904</td>
</tr>
</tbody>
</table>

(•) Standard error.
* Significant on 1% level.
# Significant on 5% level.
a Base Model (BM), Income Category Model (ICM), Income Elasticity Model (IEM).
b Price parameter for the listed income category (IC).
c AIC: Akaike Information Criterion, BIC: Bayesian Information Criterion.

Table 4: Estimation results of BM, ICM, and IEM. Source: own estimations.

Considering the outcome of a likelihood-ratio (LR) test in Table 4 (see e.g. Train, 2009, p. 70) we conclude that the IEM is superior to the BM. Therefore, taking into account the influence of age and income improves the model’s ability to explain people’s behavior. Both the IEM and the ICM are using the information on income and age. Since these models are non-nested, we use a test developed by Horowitz (see e.g. Ben-Akiva & Lerman, 1994, p. 172; Horowitz, 1983, p. 336) to compare them. As Table 4 shows, the null hypothesis that the ICM is the true specification cannot be rejected. However, according to the AIC and the BIC, the IEM is superior to the ICM.

---

3 When travel time enters the models as a categorical variable (cf. Section 3 and 4), the estimated coefficients of the remaining attributes change only marginally. Furthermore, the models with a continuous travel time variable offer a better fit quality than the models with time categories. Therefore, we refrain from reporting the results of these models.

4 When time enters the model as a categorical variable, the p-values in Table 4 change only marginally.
We also validated the ability of our models to reproduce the observed choices. Using the coefficients and attribute values, we simulated a decision for each observation in the data. The simulation procedure is structured as follows: Using the deterministic utilities we calculate the probability according to formula (3). In the next step, we generate a uniformly distributed random number from the interval [0,1]. If the random number is less than the calculated probability, we define “Hotel A” as the simulated choice. Based on the simulated and observed (real) choices, we calculate a hit ratio that provides the proportion of decisions that were determined correctly. This procedure is repeated 2500 times to derive an average hit ratio. The results show that on average about 70% of the decisions are correctly predicted by each of the three models.

The estimated parameters are robust across the three models in terms of sign and size. The ICM and the IEM produce significant elasticity parameters on a 5% level, while the remaining parameters are highly significant across all three models. The parameters show the expected sign: An increase in walking time or hotel price decreases utility, while a rating score increase raises the utility level. Chain affiliation also has a positive impact on the traveler’s utility. The elasticity parameters take effect in the expected direction: Younger people are c. p. more sensitive to the reviews and higher income people c. p. care less about the hotel price. The latter is also supported by the ICM price coefficient pattern, which shows a continuously decreasing price sensitivity for higher income categories. However, only the difference between the income categories 1 and 5 is significant on a 5% level.

A comparison of the chain affiliation and rating score parameters reveals that the utility gain induced by a rating increase of one is more than three times as high as the additional benefit that a consumer would receive from choosing a chain affiliated instead of a non-chain hotel. Therefore, online ratings seem to be a much more influential variable in decision making than chain affiliation. In the next step, we calculate the customers’ monetary valuation of attribute changes. The emphasis is on the rating score, while the attributes chain affiliation and time are only briefly addressed.

The willingness to pay for a marginal rating increase \( WTP_R \) is defined as the marginal rate of substitution \( MRS_{R,P} \) between the rating score and the price attribute. Eq. (7) shows the corresponding formula for the IEM.

\[
WTP_R = MRS_{R,P} = \frac{\partial V / \partial R}{\partial V / \partial P} = -\frac{\beta_R A_s \xi_R}{\beta_P A_s \xi_I}.
\]

The corresponding formula for the BM results from setting the elasticities in Eq. (7) to zero. The WTP for the ICM is calculated in the same way as in the other models, but the calculations must be performed separately for each income category. Using Eq. (7), we clarify the meaning of the elasticity parameters \( \xi_A \) and \( \xi_I \). Eq. (8) and (9) show the calculation of the income and age elasticity of WTP.

\[
\Xi_{I,R} WTP_R = \frac{\partial WTP_R}{\partial I_s} = \xi_I \frac{\beta_R A_s \xi_R}{\beta_P A_s \xi_I} = \xi_I.
\]

\[
\Xi_{A,R} WTP_R = \frac{\partial WTP_R}{\partial A_s} = \xi_A \frac{\beta_R A_s \xi_R}{\beta_P A_s \xi_I} = \xi_A.
\]

The results show that the parameters \( \xi_A \) and \( \xi_I \) correspond to the respective elasticities. Therefore, the interpretation of the parameter \( \xi_I \) is as follows: For a 1% increase in income the willingness to pay for a rating increase rises by \( \xi_I \) percent. The interpretation of \( \xi_A \) is analogous.
The WTP for a decrease in travel time (maximum walking time) \( WTP_T \) is, as shown in Eq. (10) for the IEM, determined similarly. However, the chain affiliation attribute is a binary variable with the values zero or one. To determine the corresponding WTP, we calculated the absolute price increase for a chain affiliated hotel for which the person’s utility remains constant. For the utility function chosen in the IEM, the WTP for chain affiliation \( WTP_C \) is calculated according to Eq. (11). The adjustments required for the BM and the ICM are similar to those applied to the WTP for a rating score increase.

\[
WTP_T = \frac{\beta_T}{\beta_P} I_s \eta_i
\]

\[
WTP_C = -\frac{\beta_C}{\beta_P} I_s \eta_i
\]

Table 5 shows the point estimates of the WTP.\(^5\) For the ICM and IEM, the table presents the WTP for a person with a mean age and a mean age and income, respectively. The percentage change in WTP for a one percent change in income or age results from the elasticities. Note that the age elasticity only applies to \( WTP_R \).

<table>
<thead>
<tr>
<th>WTP</th>
<th>BM(^a)</th>
<th>ICM(^a)</th>
<th>IEM(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel Time(^b)</td>
<td>1.74</td>
<td>-</td>
<td>1.82</td>
</tr>
<tr>
<td>Chain affiliation</td>
<td>4.22</td>
<td>-</td>
<td>4.14</td>
</tr>
<tr>
<td>Rating score</td>
<td>14.44</td>
<td>-</td>
<td>14.76</td>
</tr>
<tr>
<td>Travel time (IC 1)(^c)</td>
<td>-</td>
<td>1.53</td>
<td>-</td>
</tr>
<tr>
<td>Travel time (IC 5)(^c)</td>
<td>-</td>
<td>2.55</td>
<td>-</td>
</tr>
<tr>
<td>Chain affiliation (IC 1)(^c)</td>
<td>-</td>
<td>3.47</td>
<td>-</td>
</tr>
<tr>
<td>Chain affiliation (IC 5)(^c)</td>
<td>-</td>
<td>5.79</td>
<td>-</td>
</tr>
<tr>
<td>Rating score (IC 1)(^c)</td>
<td>-</td>
<td>12.50</td>
<td>-</td>
</tr>
<tr>
<td>Rating score (IC 5)(^c)</td>
<td>-</td>
<td>20.86</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^a\) Base Model (BM), Income Category Model (ICM), Income Elasticity Model (IEM).

\(^b\) WTP per minute.

\(^c\) WTP for the listed income category (IC).

Table 6: WTP point estimates in euros for different attributes and models. Source: own calculations.

The results show that the WTP point estimate for a reduction in travel time (maximum walking time) varies between 1.53 and 2.55 euros per minute, depending on the model and income class. These are rather high values compared to typical values of travel time savings (see e.g. Wardman, Chintakayala & de Jong, 2016). Walking time, however, is often valued twice as high as the in-vehicle time (Small 2012, p. 5; Wardman, 2004, p. 376). In addition, hotel guests may consider walking this distance more than once during their stay. The WTP for chain affiliation varies from 3.47 to 5.79 euros. Compared to the WTP for a rating increase of one, which varies from 12.50 to 20.86 euros, the WTP for chain affiliation is relatively low. The positive income elasticity indicates an increase in WTP with income level, while the negative age elasticity shows a decrease of the WTP for a rating score increase with age. We will discuss the monetary valuation of online ratings in more detail below.

The WTP point estimate is not a constant but a random variable. However, the probability density function of the WTP is unknown a priori since it is the ratio of two normally distributed parameters (Armstrong, Garrido & de Dios Ortúzar, 2001, p. 144). Nevertheless, there are several methods to determine the confidence interval of the WTP. We use a method called asymptotic t-test (see e.g. Armstrong et al. 2001, p. 145). The necessary inputs are the estimated coefficients, t-statistics and the coefficient of correlation between the two parameters. We showed the

\(^5\) The point estimate of WTP is typically smaller than its mean (Armstrong et al. 2001, p. 158). However, the difference between the two values is small for a large number of observations. We do not present the mean WTP, but later give a confidence interval of the point estimate.
estimated coefficients in Table 3. The t-statistic for a parameter is calculated by dividing the estimated coefficient by its standard error (Wooldridge 2009, p. 130). Table 6 summarizes the correlation of coefficients.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>BM</th>
<th>ICM</th>
<th>IEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating score - Price</td>
<td>-0.403</td>
<td>-</td>
<td>-0.355</td>
</tr>
<tr>
<td>Rating score - Price (IC 1)</td>
<td>-</td>
<td>-0.256</td>
<td>-</td>
</tr>
<tr>
<td>Rating score - Price (IC 2)</td>
<td>-</td>
<td>-0.217</td>
<td>-</td>
</tr>
<tr>
<td>Rating score - Price (IC 3)</td>
<td>-</td>
<td>-0.200</td>
<td>-</td>
</tr>
<tr>
<td>Rating score - Price (IC 4)</td>
<td>-</td>
<td>-0.144</td>
<td>-</td>
</tr>
<tr>
<td>Rating score - Price (IC 5)</td>
<td>-</td>
<td>-0.076</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^{a}\) Base Model (BM), Income Category Model (ICM), Income Elasticity Model (IEM).

\(^{b}\) Price parameter for the listed income category (IC).

Table 7: Correlation of coefficients for BM, ICM, and IEM. Source: own calculations.

Using these numbers and a critical t-value of 1.96, we calculated the lower and upper bounds of the 95% confidence intervals. Figure 2 illustrates the resulting confidence intervals for BM and IEM. The results for the IEM continue to apply to a person of mean age and income.

![Figure 1: Confidence intervals (95%) of the rating score WTP for the Base Model (BM) and the Income Elasticity Model (IEM).](image)

The confidence intervals for the BM and IEM are quite similar, although the upper bound for the IEM is a little higher. Furthermore, we conclude that in 95% of cases the WTP point estimate of a person with mean age and income is approximately between 12 and 18 euros. Relative to the average price level of €100 in the data (cf. Section 3), this is a mark-up of 12 to 18 percent. We applied the same method as for BM and IEM to the ICM, but calculated confidence intervals for five income categories. The results shown in Figure 3 are for a person with mean age.
Figure 2: Confidence intervals (95%) of the rating score WTP for the Income Category Model (ICM). Source: own depiction.

The results show that lower and upper bounds of the confidence intervals increase with income level, except for the lower bound in the income category 4 (€3001 to €4000). Furthermore, the confidence intervals are wider for higher income categories. This result is not surprising as there are fewer observations for categories of higher income. Also, the confidence intervals overlap. The confidence interval of income category 2 (€1000 to €2000), which includes the average income used for the other two models, is similar to that of BM and IEM, indicating the robustness of the results across different specifications. Depending on a person’s income category the WTP point estimate of a person with mean age is between 10 and 39 euros. To verify the robustness of our results we also calculated the confidence intervals according to the multivariate normal simulation method (Armstrong et al. 2001, p. 146). We have again taken the necessary inputs for this procedure from table 4 and table 7. We determined the confidence intervals as follows: Based on a bivariate normal distribution, a rating score and a price parameter are generated simultaneously in a first step. In a second step, the WTP is calculated using these two parameters. We repeated steps one and two 100000 times. From the resulting 100000 WTP values, we excluded the upper and lower 2.5% of the realizations to determine the bounds of the 95% confidence interval. The results are close to those of the asymptotic t-test, which is in line with the literature on these two methods (see Armstrong et al. 2001, p. 157). Therefore, we refrain from presenting the figures of the multivariate normal simulation method.

6. Conclusions

As insights from the literature show, online hotel ratings are an important factor influencing the customers’ hotel choice. Concerning the monetary valuation of online hotel ratings, several hedonic price studies show significant and positive figures. In this paper, we used a discrete choice experiment to determine the individual monetary valuation of online hotel ratings in Germany. Our results show, in line with the previous literature, a positive and significant effect of online rating scores on the hotel choice. For a rating increase of one point on a scale from 0 (worst) to 10 (best), we computed that the willingness to pay (point estimate) of a person with mean age and income ranges from 12 to 18 euros. Taking into account different income categories, the range increases to 10 to 39 euros.

Furthermore, results show that the rating score WTP increases with income but decreases with age. We also computed the willingness to pay for chain affiliation as well as travel time savings. Compared to the values of
travel time savings reported in the literature we determined rather high values. The WTP for chain affiliation is relatively low compared to the monetary valuation of a rating score increase. Thus, potential customers seem to rely more on guest reviews than on chain affiliation to obtain information about the hotel’s quality.

These results underline the importance of online reviews for hotel management. Under normal conditions in competition, a hotel, which does not match the guests’ expectations receives at least in the long-run lower ratings. Given the customers’ WTP for higher ratings, the room rate will eventually drop, which will reduce the hotel’s revenues. In other words, if a hotel manages to increase its rating score, it may charge higher prices. Certainly, these additional profits should exceed the costs of implementing the corresponding improvement. In the short-run, it may seem attractive to reduce costs at the expense of quality. However, an associated increase in profits would be a short-term effect. The quality decline will be reflected in the customers’ reviews, and a resulting deterioration in the rating score is associated with loss of revenue. But it may take a long time and require serious efforts on the part of the hotel management to achieve a rating increase again. In the present paper, however, we have not discussed all these aspects but limited our consideration to the individual monetary valuation of rating score increases.

Finally, we would like to point out a rather special problem with rating scores. Online platforms provide the infrastructure and customers the data to calculate these ratings. No official organization is regulating nor standardizing this user-generated content. Although the platform providers will endeavor in their own interest to increase the credibility of the information they provide, manipulation cannot be ruled out. The incentive is obvious if the costs of an artificial increase in the rating (e.g. through falsified positive reviews) are far below the actual costs of adequate quality improvements. Since the manipulation of the rating becomes more expensive with the actual number of reviews already received, platform providers usually report the rating score together with the corresponding number of reviews. First research to this indicates that travelers pay attention to the number of reviews in particular for positive ones (Galivan, Avello & Martinez-Navarro, 2018). A more fundamental discussion of this issue, however, goes beyond the scope of this paper and should be addressed in future research.
References


