The Effect of the Reviews in the Online Travel Market: an Empirical Investigation

Index

Introduction

Previous research has examined whether the electronic word of mouth (eWOM) has an effect on the theoretically highly efficient Internet markets. For the best of our knowledge, the literature on this field is mostly related to the theoretical description of the relations between the information asymmetries and the informational contents of the reviews with little attention to the price changes. However, much of the previous empirical work was focused on industries selling goods underling the presence of price dispersion and differentiation. There is lack of research literature on the empirical effects of the consumers' reviews in a service products market. In this study we investigate the effect of the consumers' reviews on a high-involvement service market: the online leisure travel market. In this framework, we are interested in quantifying the price changes relatively to both the quantity of the reviews and the relative users' grade. Performing this empirical analysis, we check whether the behaviour of the principal agents of this market, coherently with the literature, is influenced by the presence of information asymmetries and search costs. The online leisure travel market is a highly representative example of high-involvement market characterized by intermediaries between sellers and price-sensitive buyers. We use a software agent, emulating the preference of the leisure travellers, to create a tailored database of actual rooms, both from hotels and other accommodations, offered by a major corporate travel agencies. Later we conduct a descriptive investigation to present the database features and then we propose a multivariate linear regression analysis to find the price effects of the reviews controlling for all the quality variables and the hotel fixed effect. In this way we are able to make an inferential analysis to objectively identify the value of the reviews, all else being equal, in a scenario that closely matches how leisure travellers would use them. We expect that the price of the recommended rooms is affected from both the number and the quality of the reviews, with a higher influence of this latter. Moreover we expect to find empirical indications about the theoretically ambiguous effect, relatively to this market, of the competition between sellers.

1. The Internet Market

1.1. The Relevant Features

The impact of Internet on pricing and competition was actively debated by both practitioners and academics. The conventional wisdom regarding Internet competition is that their peculiar characteristics will bring about a nearly perfect market. This conception of the internet market derives mostly from the possibility of buyers to both compare the offering of sellers worldwide and take advantage of instantaneous circulation of the information. In the extreme version of this internet efficiency view, the characteristics of Internet will lead to a market where retailer location is irrelevant, consumers are fully informed of prices and product offering and all retailers make zero economic profit. With respect to the conventional retailer, with Internet market, consumers have the chance to price and compare features with ease (12). The information about a product sold can be classified in two broad categories. The seller-generated product information, available via its Web site or other traditional communication channels such as advertisements, or consumer-generated product information. This latter category may be available on forums, Web sites or even directly on the specific purchasing platform. The great heterogeneity of sellers and products on the internet market makes difficult a general comprehension of its features. Moreover, the internet-based electronic commerce is growing rapidly with the proliferation of commercial Web sites and the increasing acceptance of online transaction by consumers. Hence, the significant amount of information available on the internet market may misleads the consumers or at least requires some efforts to process it. There are empirical evidences that Internet may not be completely efficient (9). This expanding stream of research, even though mostly orientated on market for commodities, suggests that there is some observed variation in prices across retailers. Analogous result was proposed by the studies on different internet market segments. For example, Clemons, Hitt et al. (2002) found evidence of both price dispersion and price discrimination in the online travel market for airline tickets, suggesting the presence of information asymmetries. Hence the strength of Internet market is also its weakness as the presence of frictions in the online travel market, and in general on the Internet market, can be understood just taking into account the possibility of not fully informed consumers. Indeed, the internet sellers often propose non-transparent value products.

Consumer shopping online cannot touch or smell products, as would be possible in traditional retailer, so their purchasing decision must be based on the product information presented on the Web site itself. Online sellers seek to overcome this limitation by giving consumers the opportunity to share product evaluations on their platforms. Another important characteristic of the internet market is its bidirectionality (14). Through the Internet, not only can organizations reach audiences of unprecedented scale at low cost, but also individuals can make their personal thoughts, reactions, and opinions easily accessible to the global community of Internet users. The application on feedback mechanism in online marketplaces is particularly interesting, because many of the modern marketplaces would probably not have come into existence without them. According to Park, Lee et al. (2007), although also sellers provide product information, this differs from the kind of information provided by consumers reviews. When the sellers present product information, they tend to hide inferior aspects of a product and emphasize good aspects. Another difference is that the consumer-generated information is more understandable and familiar as it represents consumers' personal feelings. They suggest that the overall trustworthiness of the reviews is superior with respect to the seller-generated information. Furthermore, the informative effect of the reviews is also affected by other two critical dimensions: their quantity and their quality. Low-involvement consumers are affected mostly by the quantity while the high-involvement consumers are affected both by review quantity and by review quality.

1.2. The Online Leisure Travel Sector

Tourism is an information intensive industry whose organization rely on the communication with tourists through various channel to market their products and build customer relationship. Indeed, the internet has grown to be one of the most effective means for tourists to seek information and purchase tourism related products . Travel is a high involvement product and a large share of travel purchasers take advantage of Internet capability to share information for their online travel purchasing. Moreover, almost a third of these buyers said that consumer reviews helped with their purchase decision (19). The epithet leisure indicates a precise category of traveller. We can distinguish among two broad categories of voyagers: the time-sensitive travellers and the price-sensitive travellers. The former category includes the people who travel within a tight time window, as for example a business man, while, the latter category includes the tourists. Indeed, the tourist are ready to postpone the departure in exchange for a lower travel cost. Without having

a narrow time constraint, the leisure travellers are interested in gather the best deal and consequently conforming their purchasing decisions. Since in the internet market, as well as in other contexts, the surplus of a purchase is strictly related with the reduction of the information asymmetries (4), the leisure traveller will want to collect as much as possible information on the internet marketplace. As mentioned before the information source on Internet are essentially of two types. There are empirical evidences suggesting that main stream media, as for example specialist journal or persuasive advertising, play a little part on the purchasing decisions of the leisure travellers while the consumer-generated content are much more effective (22).

1.3. The Electronic Word of Mouth

Word of mouth (WOM) communication refers to interpersonal communication among consumers concerning their personal experience with a firm or a product. Concerning service products, because they are intangible and cannot be easily described, consumers tend to rely even more on word of mouth from an experienced source to reduce their perceived risk and uncertainty (26). WOM information search is greater in circumstances when a consumer is unfamiliar with a service provider, which is often the case in travel decisions. Different from traditional WOM, the word of mouth on the Internet is called electronic word of mouth (eWOM). Online word of mouth differs significantly from its offline form as it includes bilateral multiple communication between individuals who do not necessarily share any social relation but still preserving its informational content (10). As the use of Internet for travel panning is increasingly widespread, eWOM plays a key role in the online travel market. Not surprisingly the online travel sector is characterized by a massive use of internet generated contents.

Consumer reviews and ratings are the most accessible and prevalent forms of eWOM. During the purchasing process, consumers want product attribute and value information and recommendations from various sources. By acting as an informant and recommender, online consumers reviews have the capability of influencing the decision process of subsequent consumers. This would make the seller particularly attentive to deserve positive judgements from its customers, especially in a high-involvement market segment as for example the online travel market. Furthermore, there are empirical evidence pointing out that the reviews are perceived as more credible than information provided by marketers (19). There may be some limitation on the effectiveness of the consumers' reviews. The literature on the Cheap-Talk games suggest how

there are two main problems limiting the effectiveness of cheap talk among selfish rational agents. The first is the credibility as communication cannot work properly when there are incentives to lie and the second is the comprehensibility. The incentives to lie may arise if in a game (namely, a stylized situation) participants' interests are partly common and partly conflicting (16). Hence the leisure travellers on the online travel market who face the problem of finding the information about an unknown service product, reasonably in a different city from their own, must rely, even if within some limitations, in the eWOM.

2. Search Costs

If the consumers do not have information not only about price but also about product attributes then they will search for the best deal. Search costs are the costs related to the activities of acquiring information both in term of direct effort in searching job and in term of opportunity cost of the time employed in the research. As the consumers continue to search for product or services attributes until the marginal costs of searching match the marginal benefit, then the lower the search cost the more consumers search for the desired price. Hence a lower search cost leads to more intensive competition (4). At the mercy of the uncontrollable stream of information of the internet market the more price-sensitive travellers may experience considerable search costs. According to the literature on this field, buyers often face substantial search cost in order to obtain desired information about the prices and product offering of sellers in a non-transparent market. In the highly connected internet market, information systems can serve as intermediaries between the buyers and the sellers. A major impact of these electronic market system is that they typically reduce the search costs buyers must pay to obtain information about the commodities or the services available in the internet market. Because electronic market system generally reduces buyers' search cost in the process affecting the market power of buyers and sellers they ultimately increase the efficiency of the transactions. The sellers can still realize substantial profits as long as comparison shopping is costly for their customers (3). The eWOM and the presence of search costs are somehow related. We often observe the practice of reviewing in industries where consumers purchase goods or services infrequently, or where the product characteristics are difficult to assess. As a result, in these industries a typical consumer needs to conduct a costly search in order to find the products that match his tastes. In such industries, reviews can reduce consumer search costs and improve the quality of the transaction. The reviews affect consumers' search behaviour and

and consequently the firms' pricing behaviour as well (1). A commodity product bought from different sellers can differ only in its price. This effect may be even greater if the object of the deal is a service product. Hence both the intermediary ability to reduce the search costs and the consumers' reviews are likely to have a positive effect on product prices in the high-involvement online leisure travel market. The intermediary on the internet market serves also to improve its efficiency through referral fees. Intermediaries connect buyers to sellers who in return, if the deal is successfully completed, receive a commission for the creation of the match. This technique is prevalent in online markets. Without the presence of referral fees may emerge a problem of double marginalization on the market. The seller exercises market power against intermediaries who, in turn, exercise it against buyers (13). Overall there are two opposing tendency on the relatively immature internet market. On one hand the seller, often through an intermediary, encourages the use of eWOM form its customers while, on the other hand, the this would increase the competition. Thus, with a distinct lack of literature on this aspect, we are interested in empirically testing the net effect of the reviews on prices.

3. The Online Travel Agencies

3.1. The Online Travel Market Intermediaries

In physical retailing of commodities there is a market chain from the producer to distributor to retailer and finally to consumer, with the distributor and the retailer acting as intermediaries between the publisher and the customer. The role of the distributor is to aggregate products in a central warehouse and sell them to retailer, who in turn sell them to their customers. In a market without intermediaries the buyer would interact directly with the producer. In the internet world of electronic commerce, there are two additional possibilities of intermediaries, both as a software tool. It may be a search intermediary or a platform. The former is designed to help customers filter information from many retailers on the behalf of the customers while the latter allows retailer to take advantage of its great online visibility in exchange for referral fees. The online travel agencies (OTAs) are the online intermediaries on the travel sector. They are platform on which the seller of a specific structure can share its room(s) on Internet. The network effect of a product, is the users' payoff increase caused by the number of other users of that product. Agents using an OTA experience a positive network effect as it belongs to a particular case of network: the two-sided

market. Two-sided markets are characterized by several group of agents interacting via one, or more, platform. In this case the network externality experienced by two groups of the platform insiders depends on both the number of the other group agents, the cross-side network effects, or the number of the same group agents, the same-side network effect, that join the same platform (4). From the view point of the accommodating structures the cross-side effect in the online travel sector derives from the greater visibility of their rooms. Concerning the travellers, the cross-side effect arises from the increasing number of the purchasing alternatives. Furthermore, they benefit also from the same-side effect via the availability of the previous consumers' reviews. Hence the OTAs are ultimately an example of intermediaries in a high-involvement market characterized by the presence of both searching costs and eWOM. Because of their peculiarity the OTAs represent a suitable background to pursue our empirical analysis on the effect of the reviews.

3.2. OTA Business model

Online travel agencies provide a point of contact via Internet to enable consumers to search for appropriate accommodating solutions and prices and make a selection, which is then booked by the OTA. The operational process of an OTA is straightforward. It collects information from the customer, principally timing and location of the stay. The OTA then takes this request and some additional parameters set by the OTA and submit these to the computerized reservation system which recommend the relevant rooms from the collection available on its platform. The agent then takes the collection of rooms returned by the computerized reservation system, sorted by a specific algorithm, and present them to the traveller in pagination. The order of the selected recommendations depend upon a proprietary algorithm which selects the room on the basis of some opaque factors. The main factors considered by the algorithm are the level of the referral fees and the reviews of a structure. The seller can select the level of the referral fees in the signing of a new ad. In any case he cannot be sure of the exact result in term of visibility of a rise in the intermediary's commission, he just know that his room would be more visible. While the level of the reviews depends mostly the consumers' evaluation of the provided hospitality.

The other key feature to bear in mind continuing this analysis is the parity rate provision. The parity rate provision is an agreement between the OTA and the seller according to which the seller is free to select the price of its rooms but he cannot differentiate the price of a given accommodation across different OTAs. Punishment for an infringement of the parity rate provision

is the penalization in the order of recommendation committed by the algorithm. Even if it is actually difficult also for the OTA itself to monitor the extent of the online travel market, this clause is always included in the contract between the OTA and the seller. This provision plays a key role in the competition on this market segment. However the OTA can still influence the price both with special deals and with the small changes in the selling conditions. Furthermore in the submission of a new ad the OTA propose some advice on the seller's accommodation pricing on the base of the online travel market conditions. Onward on the exposition we will highlights how all these aspects may influence our observations, methodology and remark.

2. Data and methodology

Our analysis begins with a data set of actual rooms, both from hotels and other accommodations, offered by a major corporate travel agencies in the month of April 2017[1] in Italy. We ran our software agent six consecutive times, making requests for the corresponding six principal tourist destinations, collecting data from eight online travel agencies(OTAs). These cities, ordered by number of visits, are Rome, Milan, Venice, Florence, Turin and Naples. (*Source: elaboration ONT upon data of Bank of Italy, the international tourism of Italy – microdata distribution)*

2.1 Data Sources and Collection Methods[2]

2.1.1 Data scraping

Since there were not publicly available databases on the accommodating structures rooms recommended by the OTAs, we first needed to collect the observations on their prices and attributes. The key challenges in the construction of our database were both to be able to avoid exogenous influences on the observed prices and to control for the hotel fixed effect. Hence, we were interested in obtaining information on different OTAs contemporarily. Furthermore we were interested in imitating the behaviour of a typical leisure traveller. To perform this task, we used a software agent acting as a virtual consumer to collect the quantitative variables virtually eliminating the chance of price changes influencing the result. In order to emulate the preferences of a stylized leisure traveller we programmed it to make requests for a double room for the weekend (from 06/05/2017 to 07/05/2017) with thirty days advance. This ensures high quality data as we ran our software agent on every OTA contemporarily. The other quality variables were collected manually since OTAs have different standards (ex. On OTA 1 the zone is indicated in a district while on another OTA it may be indicated the distance of a room from downtown) as we will explain exhaustively in the following paragraph. Now it is sufficient to say that this is not an

[1] The reservation data only provide the input to the requests while it is the time advance and specification of the traveller's preference that allow us to emulate time-sensitive requests.

[2] For sake of exposition all the tables and graphs are presented after the outlier elimination except if explicitly indicated.

issue in our analysis because these variables do not change in time. For each accommodating structure the system collected names, prices, users' grades, number of reviews and whether the room was sold bundled with a service[3]. In our database these variables are called respectively Name, Price, Users' Grade and Reviews. Whereas there is little to say about the name of the structures, which just identifies the statistical units, there are some necessary clarifications about the other three variables. Our software agent collected prices shown in the recommendation pages of each OTA. Recent investigation of the European Commission (IP/17/844) pointed out how these may be ambiguous but, since we are focusing our attention on the behaviour of leisure traveller consumers, we will not consider this issue in our analysis. We can reasonably assume that consumers who decide to use an online travel agency base their choice on prices shown on the main pages. Concerning Grade, it may happen that an OTA presents more than one users' grade or even some opinions of doubtful origin about the zone or other room characteristics. We believe that this information is misleading and inconsistent among different travel agencies, thus we will only consider the grade assigned to the room by the former costumers. Depending on the OTA, variable Grade is expressed in different ways. We arbitrarily decided to adapt it as a scale of one to ten. The number of reviews is simply the one related to that room judgment. If travel agency is a *super-OTA*[4], our software agent collected also where does the price come from. Two of the eight OTAs provide users with the option to indicate a preference for either price or time. In this case software agent selected always price preference. The set of rooms alternative offered by OTAs are stored in eight different databases.

We applied three different decision rules to select rooms from the set of alternatives. First of all we emulated the preferences of a leisure traveller consumer selecting structures starting from the first page of results including special deals. The second decision rule specifies that the data collection ought to continue until the OTA gives back significant recommendations[5]. For this reason the total number of observation is limited by the less furnished OTA. In doing so our software agent stopped collecting data simultaneously on each of the eight OTAs. This is quite important in our analysis as the accommodating structures which are on the first pages are not only the ones often selected by consumers, but also homogeneous in their commission payments. The

[3] We select a bunch of quite homogeneous price services as specified after. This allow us to model this as a dichotomous variable which indicates whether a service is provided or not.

[4] A *super*-OTA is an OTA which works as comparator of other OTAs and which promised to recommend the lower price among them.

[5] As one proceeds through the pages of recommendation the results became ambiguous and irrelevant even though OTA claims thousands of available rooms.

third decision rule indicates that it should ignore sold out rooms which are often still recommended by OTAs, probably to encourage revisiting. Since it is technically difficult to impose a decision rule to select the right services among the several proposed, we let the system collect, if any, all of them. In order to model the variable Service as a dummy variable we treated ex post rough observations, consisting of text strings, on our statistical software. More precisely we imposed that it had to be substituted with 1 if the corresponding string indicated: included breakfast, airport shuttle service and free SPA access and 0 if it indicated free cancellation and the possibility not to pay upfront. To avoid discrepancy the same occurred for free wi-fi service which was sometimes explicitly indicated and some other times taken for granted. We assigned zero to them, assuming that their monetary value was negligible. Almost two thirds of the rooms of our sample were proposed with a relevant service included in their recommended prices whereas the other one third were not. In Table 1, and in the graph above it, it is possible to see how the distribution of this kind of room was not uniform across OTAs. The different propensity of OTAs in selling rooms with an included service is the first indicator of a vertically differentiated market that we found in our study.

Graph 1: Bar chart reporting the frequency of Service clustered for the online travel agencies

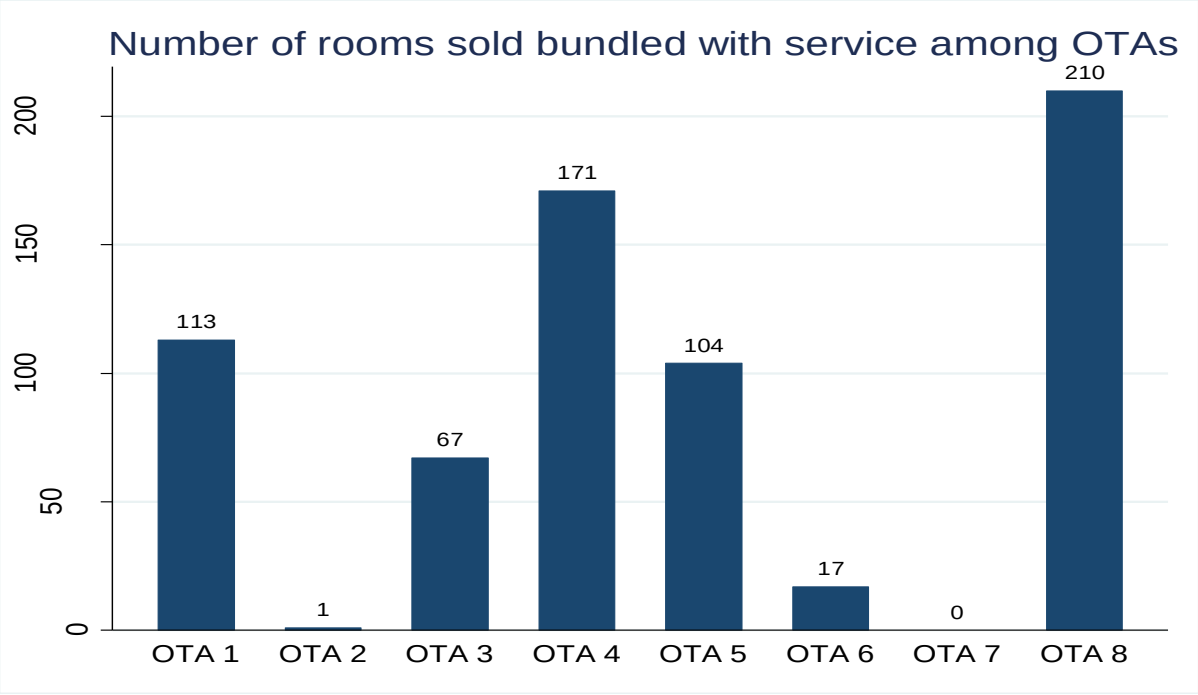


Table 1: Frequency of rooms sold with an included service for each OTAs

	ID OTA								
Service	OTA 1	OTA 2	OTA 3	OTA 4	OTA 5	OTA 6	OTA 7	OTA 8	Total
Not Provided	167	264	206	101	162	249	269	55	1473
Provided	113	1	67	171	104	17	0	210	683
Total	280	265	273	272	266	266	269	265	2156

2.1.2 Profile of information source

Concerning the possibility to include in the model the hotel fixed effect, we needed different observations of the same room. It was reasonable to assume that in a given time, an accommodating structure sold its room through different platforms contemporarily, thus we collected data on different OTAs. We selected the OTAs on the basis of their market shares. The two groups holdings the eight OTAs constituting our sample controlled the majority share of the online travel market in Italy. As stated before, the number of observation for each OTA was limited by the decreasing quality of the recommendations. Hence to reach a consistent level of observation we repeat our data scraping for each of the following cities: Rome, Milan, Venice, Florence, Turin and Naples. As well as for the OTAs selection we based our choose on the number of visits. Table 2 reports the two-way table of frequency of the observed rooms.

Table 2: Profile of information source

City							
ID OTA	ROME	MILAN	VENICE	FLORENCE	TURIN	NAPLES	Total
OTA 1	71	30	58	41	42	39	280
OTA 2	51	30	56	44	45	39	265
OTA 3	51	31	61	46	45	39	272
OTA 4	51	31	60	46	44	40	272
OTA 5	50	31	60	44	41	40	266
OTA 6	50	30	59	45	44	38	266
OTA 7	60	31	59	40	43	39	269
OTA 8	51	30	58	45	43	38	265
Total	435	244	471	351	347	312	2,156

2.2 Database Construction and Finish

2.2.1 Relevant Attributes of Room Variables

We were mainly interested in empirically testing the key drivers influencing the price of the accommodating structures from the leisure travellers' view point. In order to fit a robust regression model explaining rooms price as function of their relevant characteristics, we also need some quality variables describing them. Hence, we considered four more descriptive variables: Zone, Hotel, Stars and City. Practically these are observations of some of the characteristics of selected rooms which were collected and then modelled as several dichotomous variables[6]. Each of these variables assumes the value of 1 if the rooms have the corresponding stylized attributes. The rationale, for our purpose, was to catch possible features which had some monetary value for the tourist booking the room as well as we did before with the variable Service. Since we were considering 6 different cities, probably the most delicate matter was to design a criterion to assign 1 to the dichotomous variable Zone for each of the rooms. We decided to assign 1 to Zone if the accommodation structure which sold the room was in such a position that the guests were close to a destination of certain interest. If this was true the guests could save some transportation costs. Moreover, this fact often coincided with the prestige of their particular neighbourhood. Obviously this consisted in a conventionalization but, as we will see later, this variable fitted quite well with our model. Hotel was simply a variable which assumes value 1 if the seller of the room was an hotel and 0 otherwise. As one could imagine Stars as well was strictly related with the nature of the seller. It indicated the number of stars assigned by the regional authority[7] to each hotel. If the dummy variable Hotel assumed value 0 also the value of Stars was 0. Treating the rough observations with the statistical software this variable consisted in our final database of six (*n-1*, where *n* is the number of value which variable could assume) dichotomous variables indicating the attribute "number of stars received". In the below tables we will call them low, medium, good and luxury respectively for rooms with a star grade from two to five stars, while the remainder category identifies the other accommodations. Actually we got rid of outlier rooms with 1 star loosing just six observations and remaining with five star grade category. This variable was important because in addition to be a reliable proxy for quality it was also very similar to Users'

[6] According to econometric theory to represent a quality characteristic of a good which can assume *n* values we need just *n-1* binary variables. (Stock, James H. and Watson, Mark W., 2011. *Introduction to econometrics.*)

[7] Even though the six cities of our sample believe respectively to six different regions and so received stars from six different authority. We assume that the star grading system is reliable.

Grade and this allows us to make comparison between them testing some hypothesis of the eWOM theory. Our software agent took also track of the city in which the rooms were located as it represented another explanatory factor of their price. Also in this case, finishing our variable, we got five binary variables representing the six cities object of this study. As well as for cities, the system took track of the OTA of origin among the relevant attributes of the rooms. In Table 3 are reported the distributions of the hotel rooms, rooms sold with included service and stars categories for each OTA. The variable Other was the complementary of Hotel and it was created just for descriptive purposes. Almost all of the observations on it were concentrated on OTA 1 and OTA 7 while OTA 3 had a high number of luxury rooms. This table can be useful to better understand the following tables reporting average values as, for example, the high mean price of OTA3.

Table 3: Frequencies of room recommended by each OTA for relevant categorical groups

	Frequency						
Id Ota	Zone	Hotel	Other	Low	Medium	Good	Luxury
OTA 1	189	216	64	2	74	127	13
OTA 2	163	262	3	17	87	127	31
OTA 3	141	270	3	1	50	177	42
OTA 4	149	267	5	10	86	158	13
OTA 5	138	262	4	3	74	155	30
OTA 6	138	259	7	7	94	135	23
OTA 7	139	178	91	10	77	81	10
OTA 8	171	261	4	6	76	153	26
Total	1228	1975	181	56	618	1113	188

2.2.2 Outlier elimination

Studying the distribution of our variable we found out some outliers which may reduce the significance of our future analysis. Since the number of outlier observations was very limited we decided to keep them out of our database.

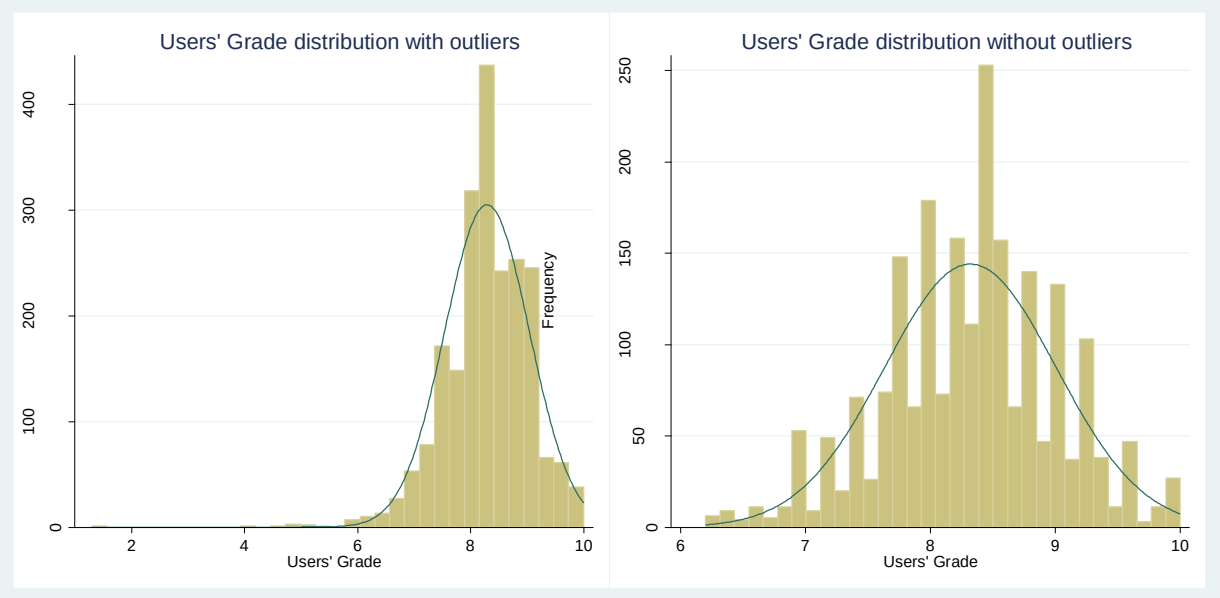
In Table 4 is reported the percentiles of the users' grade. In this case the first percentile, which is the value below which 1% of the observations may be found, corresponds to 6.2. Namely just one observation every a hundred is characterized by a users' grade lower than 6.2. This suggested us that the distribution of users' grade may be negatively skewed which was indeed -0.9433.

Table 4: Users' grade percentile distribution

Percentage of observation	1%	5%	10%	25%	50%	75%	90%	95%	99%
smallest	1,3	4,1	4,6	4,8					
percentiles	6,2	7	7,4	7,9	8,3	8,8	9,2	9,4	10
largest						10	10	10	10

As it possible to see in the below Graph 2 eliminating the observation with users' grade lower than 6.2 we get a much more normal distribution at the cost of just 20 observations. We also eliminated observations which had stars rating lower than 1 losing just 6 observations. Hence after outlier elimination we passed from 2182 to 2156 observations.

Graph 2: Table of graphs comparing the distribution of Users' Reviews before (left) the outlier elimination and its distribution after it (right), reporting the plot of the normal distribution.



2.2.3 Creation of Variables and Merge

By then we have all relevant attributes of available rooms in eight different data sets expressed due to the variables listed before (Name, Users' Grade, Star, Review, Service, Zone, Stars, Hotel, City, Ota). To go forward with our analysis, we merged the eight different data sets with our data elaboration software checking whether a structure sold the same room using different OTAs. This was needed to end up with a database usable to provide meaningful information through descriptive statistics and to run regressions empirically testing the effect of the reviews on price. The command merge combines datasets horizontally adding variables to the existing observations. Actually to run the merge operations the software does not require that the two data sets have all the observations on the same statistical units. If occurred that there are cross-sample correspondences between statistical units, then the observation of the two data sets are considered as two observations on a single statistical unit, which is exactly what we need to make cross-OTA matching. [8]. Since our statistical software could match structures only if the Name was identical on two or more data sets, we programmed it to get rid of the miscellaneous components of Name. Actually this variable was less uniform than expected. Indeed it often happened that the miscellaneous parts of the name, for example *hotel*, *Rome* and so on, were mixed up around the relevant part of the name or it might be used *&* instead of *and*. Possible explanations are the difficulty of the sellers to provide always the same name to different websites or the attempt of both sellers and OTAs to disincentive automatic comparison among different platforms. Once the name was corrected, we asked statistical software to make seven consecutives merging with the eight data sets taking track of the cross-OTA correspondences with the variable match[9]. Further more we imposed to it to consider two statistical units identical only if they had not only the same name but also the same city, the same zone and the same star grade[10]. After the merge, our database consisted of 1058 observations each one with eight times the original variables (every variable coming from each of the data sets) mostly with a null value. Thus, we reshaped it from wide to long obtaining a database constituted by 2156 observations and we substituted the name of each

[8] As we will see later, there are some differences among prices of the same room on different online platforms despite the parity rate provisions. Anyway some OTAs claim in their terms and conditions that if an accommodation structure sold a room at low price through another platform it will be penalized in visibility.
[9] Every time that two datasets were merged and a room was matched between them variable Match was increased by one.
[10] This because eliminating the miscellaneous parts of the name, some rooms were ambiguously identified.

room with an ID number if the room has the same name, the same city, the same zone and the same star grade so to uniquely identified identical rooms.

Following, the existent variables were finished and the database was reshaped, we generated some other variables needed for our analysis on the data. First of all we created the variable Match, which could assume value from zero to eight excluding 1. The value 1 is excluded because if a room compared just on one OTA it did not has this attribute. Then, Match equal to 0 was assigned by our statistical software to a room if its purified name appeared only in one of the eight OTAs, while it assumed 2 if it appeared in two and so on up to the value of 8. In Table 5 are presented the number of rooms sold on one or more OTAs contemporarily.

Table 5: Frequencies of rooms sold on more than one OTA for nature of the structure

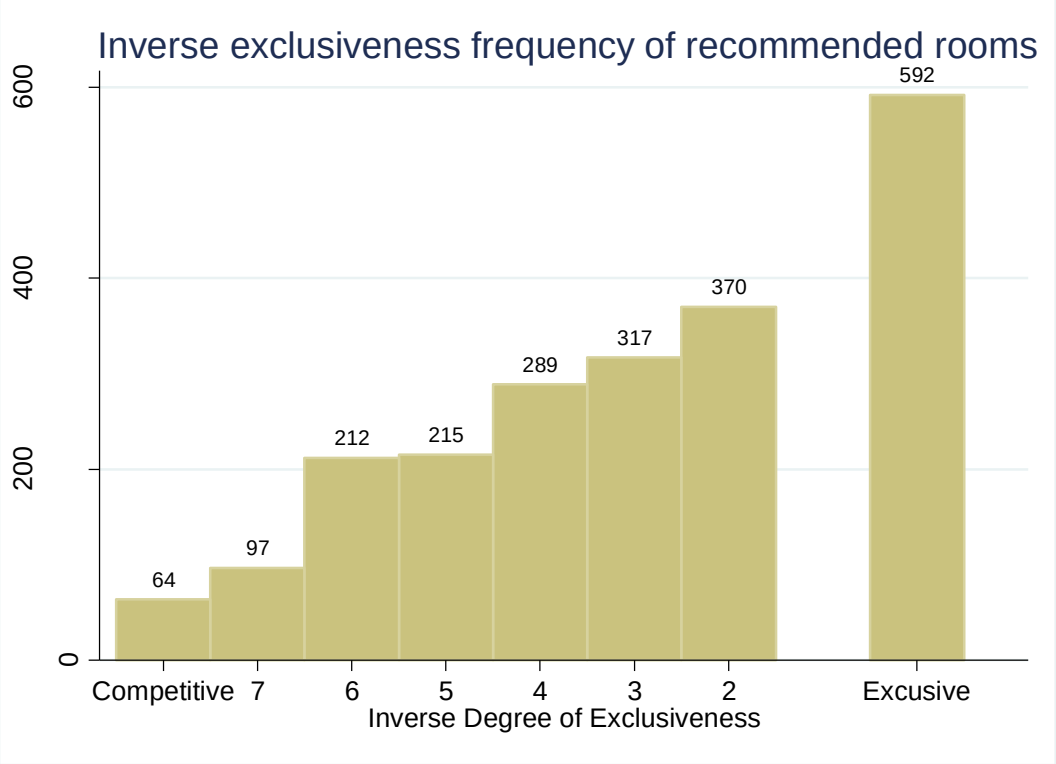
		Inverse degree of exclusiveness							
Matched OTAs	Competitive	7	6	5	4	3	2	Exclusive	Total
Hotel	64	97	212	215	289	315	354	429	1975
Accommodation	0	0	0	0	0	2	16	163	181
Frequency	64	97	212	215	289	317	370	592	2156
Percentage	9,70%	4,5%	9,83%	9,97%	13,40%	14,70%	17,16%	27,46%	100%

The 27.46% of the rooms of our sample were exclusively sold by one OTA while the other 72.54% were sold on at least two online travel agencies. This latter category was composed mostly by hotel rooms as only 16 other accommodation structures were recommended on two OTAs and just 2 on three OTAs. In general, the number of rooms belonging to one of these categories must be divisible for the number indicating the category. Despite the fact that this is not always true, as consequence of outlier elimination, a set of matched rooms which lost one or more observations still preserve this attribute. To better understand this point, let us take as example the fully competitive rooms. The number 64 indicates that there were 8 hotel structures[11] which sold their rooms on all of the eight observed OTAs and were not affected by outlier elimination whereas the rooms of the fourth category were. Since higher was the Match value, higher was the number of OTAs competing in the sale of a specific room, this variable could be interpreted as the inverse degree of exclusiveness

[11] $Unique\ rooms_{Match\ i} = {}^{Frequency_{Match\ i}}/_{Match_i}$

(Graph 3). We expected that the price of a set more competitive rooms would have a lower price respect to a less competitive one. In the following paragraph we will see whether the mean price and its standard deviation varies among different categories.

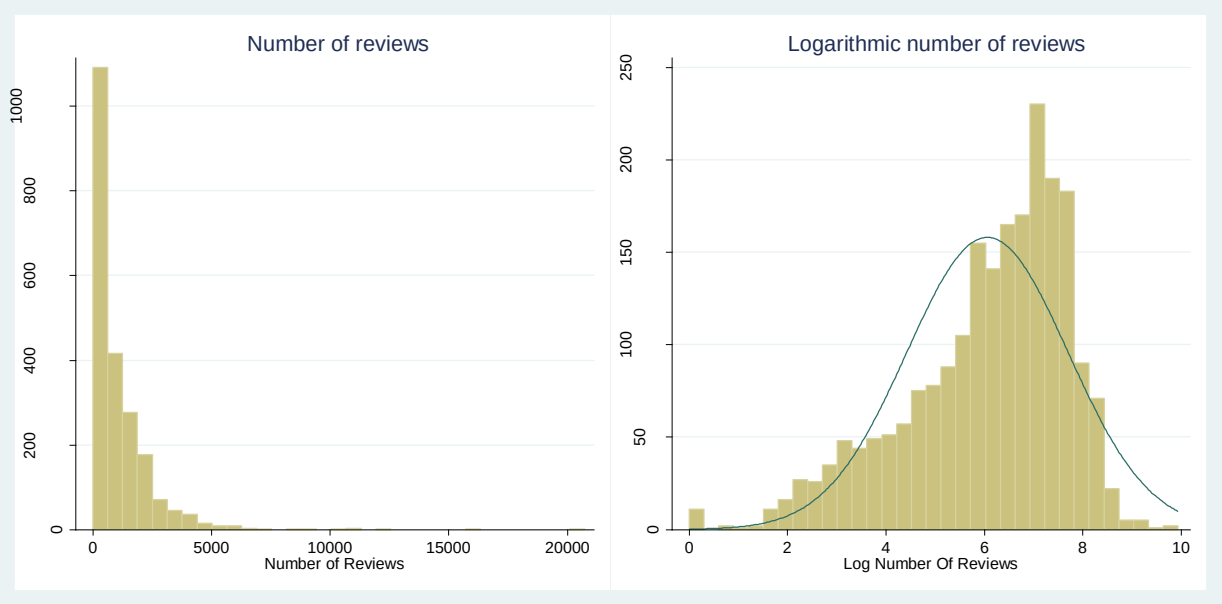
Graph 3: Histogram of the inverse degree exclusiveness



Reviews is characterized by high variability as it took value from 1 to 20730. If the distribution of a variable is characterized by high skewness[12], taking a natural logarithm of the variable sometimes helps fitting the variable into a model.

[12] The Skewness is a measurement of the degree of asymmetry of a distribution

Graph 4: Table of graphs comparing the distribution of number of reviews before (left) the logarithmic transformation and its distribution after it (right) reporting the plot of the normal distribution.



Since a variable with such a positively skewed distribution was not suitable as explanatory variable in our regression model, we transformed it in a logarithmic variable (logReviews) as it made skewed distribution more normal (Graph 4). Moreover, when a change in the dependent variable is related with percentage change in an independent variable, or vice versa, the relationship is better modelled by taking the natural logarithm of either or both of the variables. In our case the independent variable, but not the dependent variable was logged thus one percent change in the independent variable was associated with 1/100 times the coefficient change in the dependent variable. This means one percent increase in the number of reviews is associated with an increase in the price of one percent of the coefficient associated to logReviews.

Recapitulating our final database was constituted by 2156 observation on 1069 actual rooms, both from Italians hotels and other accommodations, offered by a major OTAs in the month of April 2017. All the relevant attributes, indicated on the recommendation pages, were described by a total of 33 variables. Among these, 23 are dichotomous variables representing the 6 variables corresponding to: the star grade (6), the city (6), the OTA of origin[13] (8) and whether if the room was an hotel (1), if it was sold bundled with a service (1) and if it was located in one of the selected

[13] Even though for Star Grade, OTA and City we actually needed just n-1 dummy variables, we none the less created n of it. This to be able to select the more appropriate benchmark, namely the one not included in the regression, and to descriptive statistical purposes.

zone (1). The only two quantitative variables were the price and the users' grade while the name and the ID of the hotel served to uniquely identified each of the rooms.

2.3 Data Characteristics and descriptive evidences

In this paragraph we will present some descriptive tables and graphs to highlight the principals properties of the data constituting our sample. We were dealing with a unique database tailored for our study about the effect of the reviews on the online leisure traveller market. Because of the peculiarity of the market segment and the data sources, our database was mostly composed by variables of qualitative nature. After the data overview of the previous section, we expected that the OTA of origin of a specific room constituted one of its relevant attributes affecting price and, consequently, that prices of a specific room is not the same among different OTAs. Moreover, accordingly with the widely acknowledged negative effect of competition on price, we presume that the price decreases as the number of OTAs selling the same room increases. Hence, before going through to the inferential analysis of the data we had to understand precisely which were the features of this dataset, in order to design an appropriate regression model. During the descriptive analysis we will check whether these results will coincide with our expectation, suggesting case by case possible explanations.

2.3.1 Price, Users' Grade and Reviews

The unit of measure of the quantitative variables are different as the price is indicated in Euro, the users' grade in a continuous scale from one to ten and the number of reviews are counted in units making impossible relative comparison among them. We found out that the most appropriate approach to present them was to consider the means and the standard deviations of the three qualitative variables in relation to each of the six group categories. In this way we could analyse the intra group changes of price, the users' grade and the number of reviews to check our prior hypothesis and to better understand the regressions results. Table 6 shows some basic statics of our sample. Overall we had a total of 2156 observations with an average price of €210 for a stylized leisure traveller request. The standard deviation in quite high with respect to mean price as it is equal to €162, as the price ranges from €31 to €1600. This is not surprising since the significant

difference in quality of recommended rooms. Relatively to the other two quantitative variables collected, the average users' grade[14] is nearly 8.3 with a standard deviation of 0.69 taking value from 6.2 up to the full score of 10, and the average numbers of reviews related to these valuations are 1034 ranging from 1 to 20730. Partially, the low standard deviation of Users' Grade derives from the outlier elimination discussed before which was in any case equal to 0.75.

Graph 5: Table of graphs reporting mean and standard deviation of Price, Users' Grade and Reviews

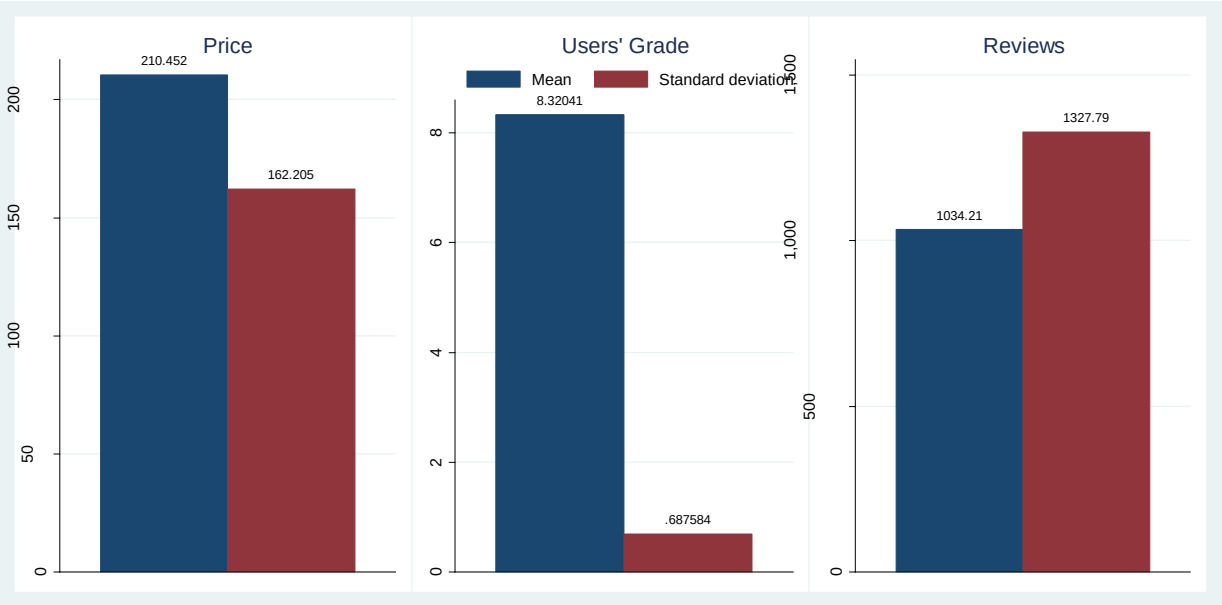


Table 6: Summary of basic statistic of qualitative variables

VARIABLES	(1) Number of obs.	(2) Mean	(3) Standard deviation	(4) min	(5) MAX	(6) Coefficient of variation
Price	2156	210,45	162,205	31	1600	0,770
Users' Grade	2156	8,32	0,69	6,2	10	0.083
Number of Reviews	2156	1034,21	1327,79	1	20730	1.283

Let us now understand how the quality variables average values changed in relation with each of the categorical variables. As it is possible to see in Table 7, the average price of one of the 1228 rooms of our sample, situated in an accommodation structure in good position, is almost €244 while, for less attractive location, the average price of the 928 rooms declines to €166. The

[14] Even though in the following regression analysis we will use the logarithmic transformation of this variable, in this section it results much more useful to discuss taking in consideration the observed users' grade.

important price difference between the two categories corroborated not only the validity of the zone as an explicative variable of the rooms price, but also that our selection criterion was quite accurate. This difference is not extended to the mean grade assigned to these kinds of rooms, which are respectively 8.36 upon an average number of reviews of 1070, and 8.19 with 977 reviews.

Graph 6: Table of graphs reporting mean and standard deviation of Price, Users' Grade and Reviews depending on whether the rooms are situated in an actrative zone.

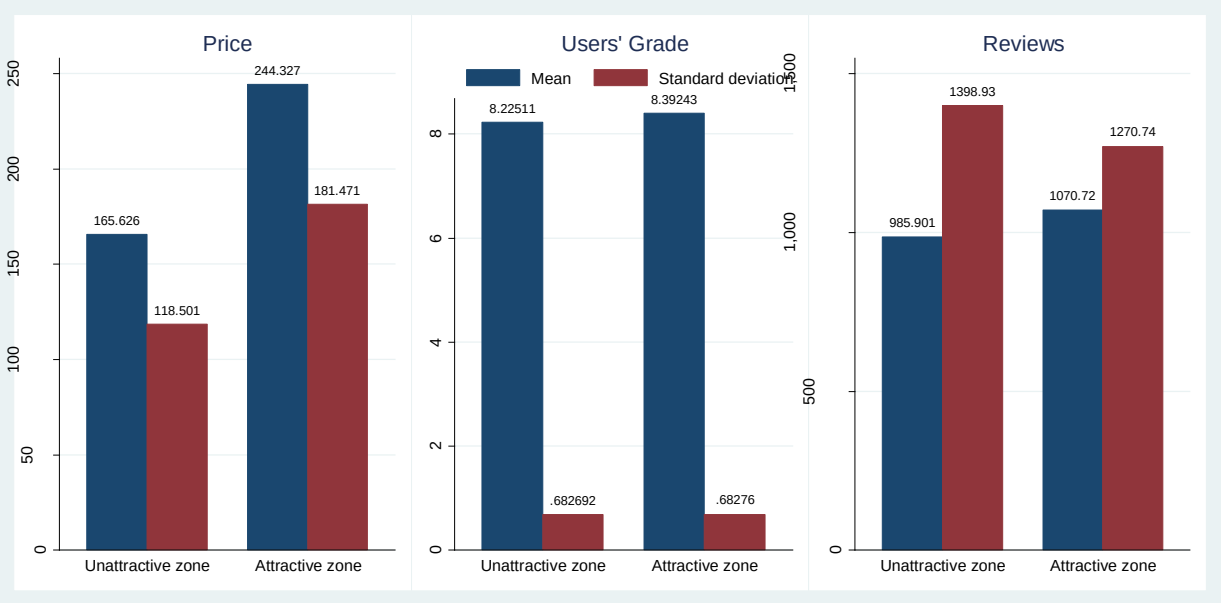


Table 7: mean of quantitative variables grouped for zone attributes

	Mean values			
Zone	Price	Users' Grade	Number of reviews	Frequency
Unattractive (0)	165.63	8.22	958.90	928
Attractive (1)	244.33	8.39	1070.72	1228

Table 8 summarizes the mean value of Price, Grade and Reviews relatively to the nature of recommended rooms. In our sample there were 1975 hotel rooms and 181 in different structures. On average, the price of a room in a hotel is equal to €216 with a judgment of 8.28 based on 1070 reviews. The mean values for other accommodations correspond to a price of nearly €154 with relatively higher grade of 8.39 deriving from a lower average amount of reviews equal to 480. As well as before there is significant price difference between the two categories but, in this case, this is also true for the number of reviews as the hotel has more than twice as much the reviews of the other structures. This is not unforeseen as one considers that often the hotels have much more rooms respect to, for example, a bed and breakfast.

Graph 7: Table of graphs reporting mean and standard deviation of Price, Users' Grade and Reviews depending on whether the room was sold in a hotel or in an others accomodation.

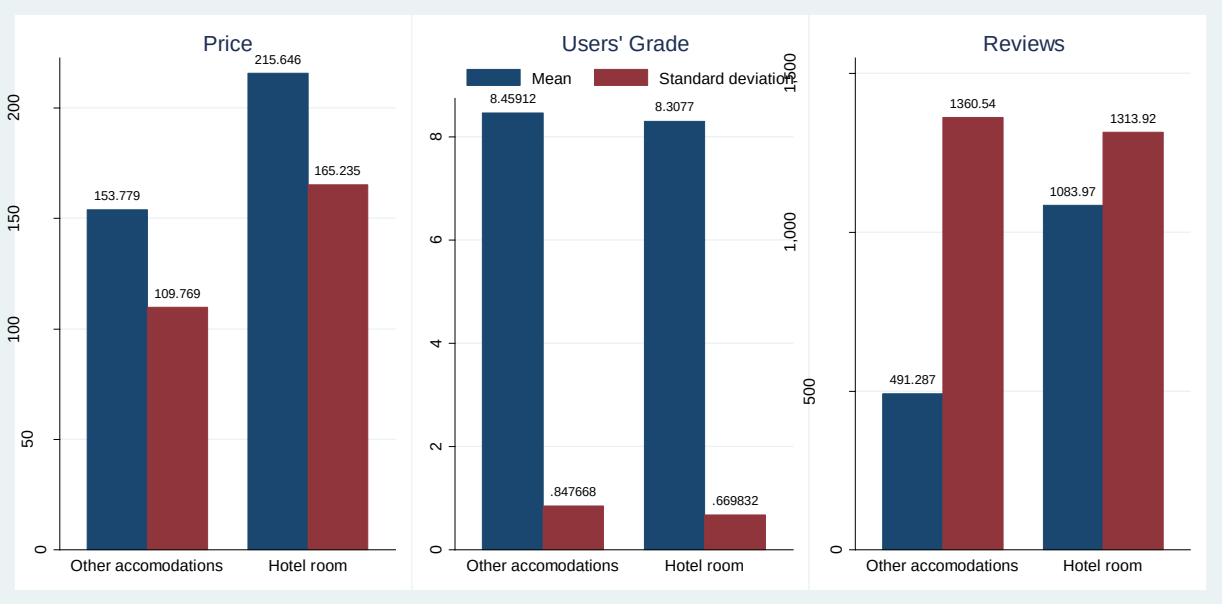


Table 8: mean of quantitative variables grouped for hotel attributes

	Mean values			
Hotel	Price	Users' Grade	Number of reviews	Frequency
Other accommodation (0)	153.78	8.46	491.29	181
Hotel (1)	215.65	8.31	1083.97	1975

In the following table are reported the mean Price, Grade and Reviews for each possible value of Star. In our sample there are respectively 57, 628, 1121 and 188 rooms in hotels with two to five assigned stars and 182 rooms in different kind of accommodating structures, which then have 0 stars. Not surprisingly, the mean price became higher as the star grade increased, with a notable increase for five-stars hotels which whereas suffer a relatively low number of reviews. In Table 9 emerges how, on average, a room which is not in an hotel is sold at the same price of a room in a three-stars hotel, but the mean grade assigned to these rooms is comparable with four-stars hotel rooms. Nevertheless, there seem to be a positive correlation between Users' Grade and Stars.

Graph 8: Table of graphs reporting mean and standard deviation of Price, Users' Grade and Reviews depending on the star grade assingned to each group of rooms. (OA*: other accomodation).

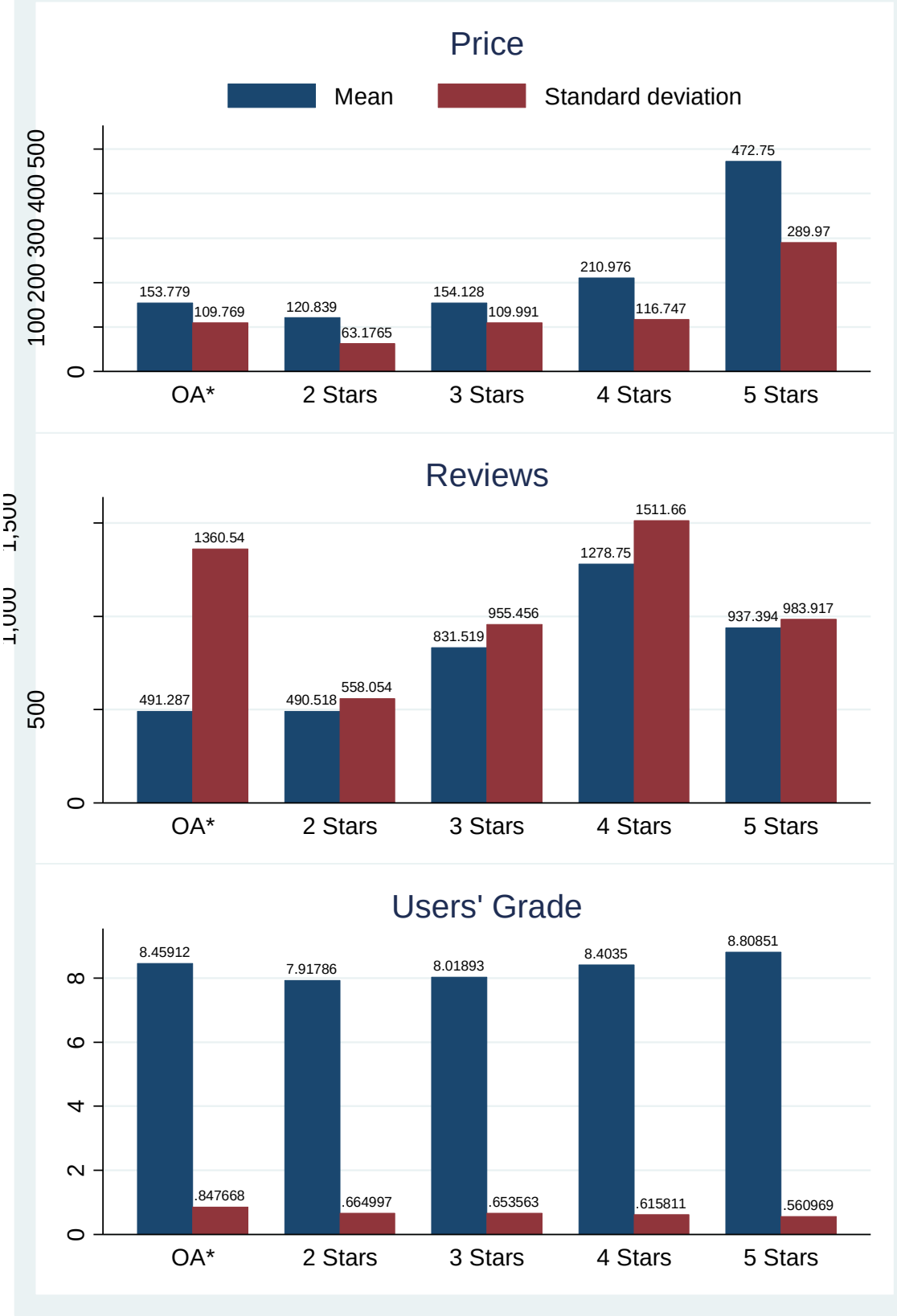


Table 9: mean of quantitative variables grouped for star grade attributed

Star Grade	Price	Users' Grade	Number of reviews	Frequency
0	153.78	8.46	491.29	181
2	120.84	7.92	490.52	56
3	154.13	8.02	831.52	618
4	210.98	8.40	1278.75	1113
5	472.75	8.81	937.39	188

As we can see in Graph 9 the average price varies across the eight OTAs object of our study. Visiting them, one fundamental difference suddenly caught the eyes: their different graphic appearances. Indeed, many of them are really refined and user friendly, while other are bare and tiring to use. We arbitrarily decided to divide the OTAs, except for *super-OTAs*, in two categories reflecting this feature. From now on we will refer intuitively at these kinds of interface as *state of the art* OTAs and *basic* OTA[15]. Among the OTAs in our data set, Ota_1 and Ota_6 are *state of the art* online travel agencies, Ota_2, Ota_4 and Ota_7 are *basic* and Ota_3, Ota_5 and Ota_8 are *super-OTA*. In Table 10 it can be noticed how *basic* OTAs have the lowest average prices (from €148 to €201), while *super-OTA* have the highest (from nearly €280 to €215), closely followed by the *state of the art* OTA (from €213 to €216). According to the literature in this field the information systems, as the online travel agencies, can serve as intermediaries between the buyers and the sellers typically reducing the search cost[16]. Hence we can imagine that a better graphic interface improves the consumers' search effort increasing their willingness to pay. Anyway, these talking was not enough to exhaustively justify cross OTA price variation. Whereas the users' grade seems to be stable as usual among different groups, the average number of reviews follows the price pattern.

[15] We will not reveal the name of the OTAs which constitute our database and we will provide only aggregate information of them for commercial reasons; although, we still know the corresponding OTA to each identification number.

[16] For an exhaustive treatise of the argument see for example *J.Y. Bakos : A Strategic Analysis of Electronic Marketplaces, university of California, 1991, Irvine.*

Graph 9: Table of graphs reporting mean and standard deviation of Price depending on the OTA of origin

Table 10: mean of the qualitative variables depending on the OTA of origin

	Mean values			
Online Travel Agency	Price	Users' Grade	Reviews	Frequency
OTA 1	217.17	8.40	1864.61	280
OTA 2	202.28	8.34	810.88	265
OTA 3	215.41	8.39	1752.91	273
OTA 4	184.99	8.19	280.05	272
OTA 5	219.69	8.15	1620.60	266
OTA 6	213.85	8.37	496.10	266
OTA 7	149.38	8.09	184.52	269
OTA 8	281.10	8.61	1227.81	265

We can see the three *super-OTAs* ranging from 1127 to 1752 with the higher number of reviews, followed by the *state of the art* OTAs with an average of 1180 reviews; to close there are the *basic* OTAs, which span from 184 to 810.

By then, our prior hypotheses are not unambiguously confirmed, but so far we can state that the OTA of origin is an explanatory variable for the price to the extent that it influences rooms price with both their ability to reduce search costs and their reviews capital.

Let us now compare the sample of rooms sold in a single platform with the ones sold by many agents. In Table 11 are reported the average values of our quantitative variables in relation with

the number of OTA competing in selling a specific room. Our sample consists of 592 rooms which are recommended just by a single OTA whereas the remaining 1564 are proposed by different agents contemporarily. As it is clear from Graph 10, it is not possible to identify a pattern for Price and Grade in relation with Match, hence further investigations are needed to shed light on this aspect. We tried to repeat the analysis sub clustering for some pertinent qualitative variables, in order to get across group homogeneous quality rooms to isolate the competitive effect on price. Also in this way, we got the same results. Indeed, the rooms price seems to be indeterminately affected from the inverse degree of competitiveness.

Graph 10: Table of graphs reporting mean and standard deviation of Price depending on the number of OTAs compeating in selling a specific room.

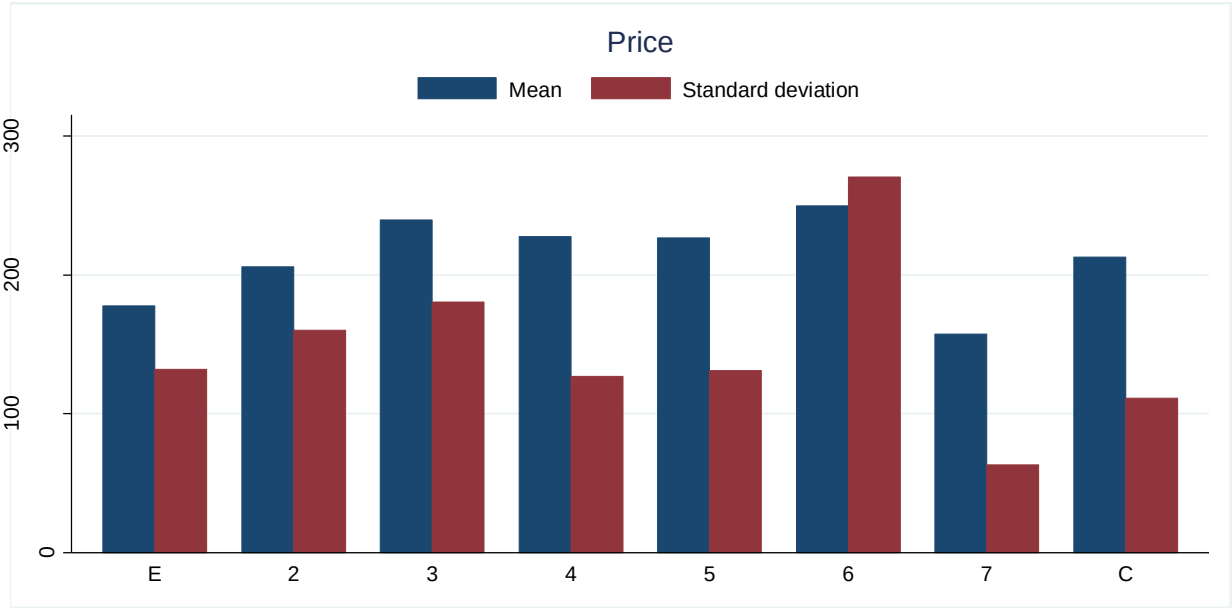


Table 11: mean of quantitative variables grouped for the inverse degree of exclusiveness

Inverse degree of exclusiveness	(mean) Price	(standard deviation) Price	(mean) Users' Grade	(mean) Reviews	Frequency
Exclusive	177.55	132.13	8.27	771.69	592
2	206.14	160.21	8.26	960.51	370
3	239.17	180.29	8.33	925.10	317
4	227.70	126.95	8.34	1203.45	289
5	226.62	131.11	8.36	1417.30	215
6	249.68	270.69	8.23	1192.75	212
7	157.31	63.09	8.50	1514.05	97
Competitive	212.76	111.29	8.72	1125.18	64

Possible explanations to this puzzling result relies on the fact that the price was selected by selling structures, which might spare effort applying the same price on each OTA with the only purpose to increase rooms visibility. Thus, our prior hypothesis is likely to be rejected.

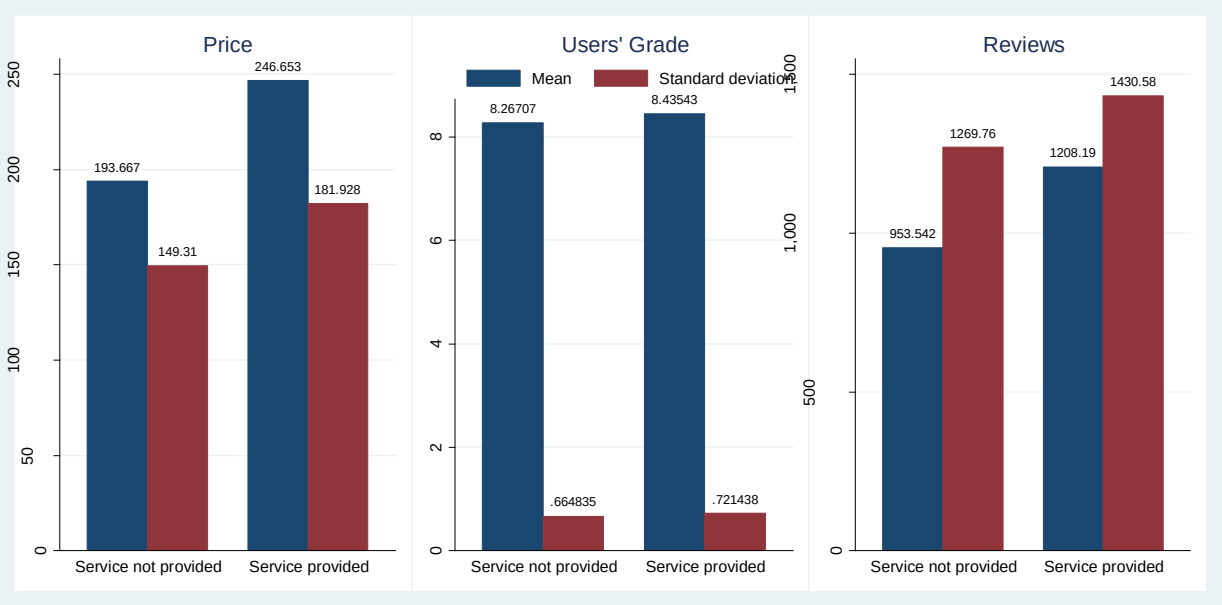
2.3.2 Price Discrimination

The across-OTA average price differences and the absence of competitive effect on the online leisure travel market suggest the presence of vertical differentiation. Since we had cross-OTA observations we were able to check whether the same room was sold at the same price on every OTA. We imposed our statistical software to count observations with same name and positive value of Match, so to uniquely identify all the specific rooms sold on several platforms, which had the same price in the correspondent number of matched OTAs. We found out that any room had this characteristic, namely any room was sold at the same price contemporarily on each of the OTAs in which it was observed. Overall these suggest the presence of price discrimination in the online leisure travel sector but, since the parity rate provision between the selling structure and the platforms, this price difference must found different explanation. A room sold bundled with a service has an average price of €246 with a users' grade of 8.42 based upon a mean number of 1205 reviews, while the average price falls to nearly €193 for a without-service-room. Same happens to mean grade and number of reviews as they are correspondingly equal to 8.27 and 953 (Table 12). Hence if a room was sold with an included service its price, on average, rises significantly both confirming the validity of the decision rule imposed on our software agent in data collection and suggesting it as a tool to price differentiate in the parity rate framework.

Table 12: mean of qualitative variables grouped for service attribute

	Mean variable values			
Service	Price	Users' Grade	Reviews	Frequency
Service not provided (0)	193.68	8.27	953.54	1473
Provided service (1)	246.65	8.43	1208.19	683

Graph 11: Table of graphs reporting mean and standard deviation of Price, Users' Grade and Reviews depending on provided serice.



In conclusion, in this section we found three main descriptive evidence. The OTA of origin is an explanatory variable for the price to the extent that it influences rooms price with both its reviews capital and its ability to reduce search costs. From the descriptive analysis, the rooms price seems to be indeterminately affected from the inverse degree of competitiveness. In the following chapter, by using a base of actual available rooms, we will be able to make inferential analysis to objectively quantify our findings about online travel market under a scenario that closely matches how it would be used by leisure travellers.

3. Empirical Analysis and Results

What we found in the previous chapter was substantially aligned with the literature on the information asymmetries on the online travel market. However much of the previous work has a qualitative approach relatively to the effect of the online consumers' reviews. Now we are interested in quantifying the effect of these reviews on room prices.

3.1 Econometric Models

3.1.1 Linear Regression Model (No Hotel Fixed Effect)

The key challenge in empirically testing our predictions is to properly model the interdependence between room characteristics, which are the explanatory variables, and the price, that is the explained variable. First of all, we assumed that the relationship between these variables was linear. Practically this assumption can virtually never be confirmed, however multiple regression procedures are not greatly affected by minor deviations from this assumption. The critical dimensions that affect the room prices are the city, the location, whether the service is provided or not, the nature of the structure, the quality (caught by both star rating and users' grade) and the OTA which sells the room to the extent that it reduces the searching costs. As discussed before, all these relevant attributes are described by both the quantitative and the qualitative variables constituting our tailored database. We began with a linear regression model (1) on our cross-sectional data defined as follows.

$$\begin{aligned} Room\ Price_r = \alpha + \beta_1 Users'Grade_r + \beta_2 logReviews_r \\ + \gamma_{City_{r,c}} + \gamma_1 Zone_r \\ + \delta_{Star\ Grade_{r,s}} + \delta_1 Service_r + \varepsilon_r \end{aligned}$$

Where γ_{City_r} and $\delta_{Star\ Grade_r}$ represent respectively the city fixed effect and the star grade fixed effect and r indexes the room.

The underlying assumptions of this formulation are the standard ones for the ordinary least squares (OLS) model. The first assumption is that other factors affecting the price, contained in the error term ε_r, are uncorrelated to the explanatory variables specified in the model. This means that, on average, these other factors do not affect our results. Whereas there is little to say about the first assumption, which just describes an ideal condition, there are some caveats about the following two. The second assumption is about how the sample is drawn. It postulates that the observations of the sample are independently and identically distributed (i.i.d.) across observation. Namely, in our case, that if a given sample of rooms is drawn from its population, then it necessarily has the same distribution. We relaxed this assumption, because of spatial correlation[17], clustering our sample for the city categories. While we allowed the units within each cluster to be correlated, we expected independence throughout the clusters. Namely, including six clusters in our regression model, we assumed across city i.i.d. observations. The prices of the accommodations within a city may be correlated because of common characteristics of the rooms within that city or because of the features of the city itself (such as the high maintenance cost in Venice). The third assumption is that large outliers are unlikely. About this issue one may remember the discussion in the previous chapter about the outlier elimination. In order to let our model likely to respect this latter condition we eliminated outliers from the observation on users' grade while we applied a logarithmic transformation to the number of reviews. The last assumption is that there is not a perfect collinearity between the explanatory variables. This assumption is always violated including in the model *n* categorical variables describing a variable which can assume exactly *n* values. Indeed, such an explanatory variable can be described with just *n-1* dummy variables as the *n-th* category will result from all the others assuming a value equal 0. In this case the excluded binary variable serves as benchmark to evaluate the effect of the other categories on the explained variable. From now on we will refer to it as the base category. If a perfect collinearity occurred between two explanatory variables our statistical software automatically dropped one of the two to avoid the collapse of the variance covariance matrix and proceeded with the coefficients estimation. As we will show below this will not necessarily constitute a problem in our analysis.

In multiple linear regression, the size of the coefficient of each of the independent variables indicate the magnitude of the effect of that variable on the dependent variable, and the sign on the coefficient simply reflects the direction of that effect. This type of correlation is also referred to as a partial correlation. Our interest lies in the unknown parameters β, γ and δ which represent

[17] This situation arises when dealing with geographical units in which the observations are not truly independent.

the independent contributions of each independent variable to the prediction of the dependent variable; respectively: the partial effect of the reviews, the location and the intrinsic attributes of a room. Note that we arbitrarily adopted different Greek letters to indicate different groups of the variable coefficients for exposition purposes. In interpreting these results it is important to bear in mind the units in which each of the variables is measured. Let us now understand the meaning of these coefficients. The more straightforward coefficient is β_1 which can be interpreted as the price variation due to a one unit variation of the explanatory variable $\mathrm{Users'Grade}$. All other coefficients need a different interpretation. Concerning the logged dependent variable $logReviews_r$ a one percent change in the independent variable was associated with one percent variation multiplied by the coefficient change in the dependent variable. Thus, the coefficient β_2, which is attached to a logarithmic transformation of the number of reviews, indicates that for a, let's say 1% increase in the number of reviews, the difference in the expected mean price will be always $\beta_2/100$. In other words, one percent increase in the number of reviews is associated with an increase in the price of one percent of the coefficient associated to variable $logReviews_r$. As long as its percentage increase is fixed, we will see the same difference in the price, regardless where the baseline number of reviews is. Regarding the other coefficients attached to a binary variable, for example γ_1 which is related to $Zone_r$, partial effects are computed at different settings of that dichotomous variable. Thus, these coefficients indicate the absolute measure of the price increase related to the particular attribute described by the correspondent variable. In the example above, if the room is located in an attractive zone, then, given that the coefficient is positive, its price will be γ_1 greater than the price of a room situated in an unattractive zone. The interpretation is even different for a group of binary variables describing an explanatory variable which is not dichotomous as the zone above. This occurs, for example, for $\gamma_{City_{r,c}}$ (with $c=1,\ldots,6$) that indicates price increase with respect to the baseline, omitted, city category. Hence, the price of a room in a structure of the $City_c$ will be, for a positive value of the coefficient, $\gamma_{City_{r,c}}$ greater than a room located in the city identified by the base category.

Another way to think about the partial effect, described by the regression coefficients, is to consider it as a measure of the correlation of an explanatory variable with the independent variable, after controlling for all the other independent variables. A formulation of this kind is appropriate in this framework because of the high number of categorical variables. Indeed, each of them control for a certain room attribute. In the preceding chapter we found out that the OTA of origin of a room may be one of its relevant attribute. For this reason we are interested in comparing our first linear regression model (2) with another similar but which controls also for the OTAs.

$$Room\ Price_r = \alpha + \beta_1 Users'Grade_r + \beta_2 logReviews_r + \gamma_{City_{r,i}} + \gamma_1 Zone_r + \delta_{Star\ Grade_{r,j}} + \delta_1 Service_r + \varphi_{OTA_{r,m}} + \varepsilon_r$$

Here the φ_{OTA_r} represents the OTA fixed effect which catch the price effect of been sold by the OTA_o. We expected to find positive value $\varphi_{OTA_{r,o}}$ by choosing the cheaper OTA as baseline category.

3.1.2 Linear Regression Model (With Hotel Fixed Effect)

An explanatory variable is said to be endogenous if it is correlated with the error term ε_r. It might be useful to specify the meaning of endogenous in this discussion since it differ from other branches of economics, namely determined within the model formulation. Here instead it is related to any situation where an explanatory variable is correlated with the unobserved part of the price. One of the main causes because it usually arises is the omitted variable bias. An omitted variable appears when one would like to control for one or more additional variables but, usually because of data unavailability, he cannot include them in a regression model. If a certain variable is unobserved we can still estimate the dependent variable as a linear function of the other explanatory variables but this is not necessarily significant if the unobserved variable is correlated with the others. If the unobservable variable and one (or more) of the independent variable is correlated then it is said to be endogenous. The correlation of explanatory variable with the unobservable is often to self-selection: agents choose the value of the former which might depends on the latter, that is unobservable.

It is reasonable to assume that the leisure travellers deciding whether to write a review and choosing the relative evaluation are influenced by the specific hospitality of a given accommodating structure. More precisely travellers decide whether to review and the grade to assign depending on the specific hospitality of a given accommodating structure, that is unobservable. If this is true, $Users'Grade_r$ and $logReviews_r$ may be endogenous and consequently our estimation biased because of the omitted variable bias. Hence, since our database is suitable to control for cross-OTA observations, we propose a third model (3) considering also

the hotel fixed effect (i.e. the specific ability of a structure). In this way we are able to catch the actual effect of the reviews on price all else been equal.

$$Room\ Price_r = \alpha + \beta_1 Users'Grade_r + \beta_2 logReviews_r + \gamma_{City_{r,c}} + \gamma_1 Zone_r + \delta_{Star\ Grade_{r,s}} + \delta_1 Service_r + \varphi_{OTA_{r,o}} + \vartheta_{Hotel_{r,h}} + \varepsilon_r$$

Where ϑ_{Hotel_r}, represents the hotel fixed effect.

3.2 Regression Analysis

3.2.1 Linear Regression Model Results

Table 1 presents estimations (with standard errors in parentheses) of the equations specified above. Each column reports coefficients, with the corresponding significance levels, from the different specifications that use both qualitative and quantitative variables as regressors of the price of a room. In the bottom part of the table are also reported the fixed effects considered in each regression.

Table 1: Regression results (Dependent variable Price)

VARIABLES	(1) Base Linear Regression	(2) Linear Regression with OTA Fixed Effect	(3) Linear Regression with Hotel Fixed Effect
Users' Grade	54.0541***	50.8340***	5.1488**
	(11.1779)	(10.0861)	(1.8036)
Log Number of Reviews	4.4832**	1.5168	1.7324**
	(1.5074)	(1.4326)	(0.5824)
Included Service	19.5211**	6.2215	-1.4037
	(6.4177)	(5.4025)	(1.5097)
Zone	63.8119**	62.6936**	
	(19.7247)	(19.7674)	
Low	-24.9181	-21.9934	
	(30.5392)	(20.2305)	
Medium	40.8322**	42.8649***	
	(10.9761)	(5.8229)	
Good	63.3641***	68.2829***	
	(2.7948)	(8.7968)	
Luxury	270.4941***	278.2366***	
	(31.4725)	(37.4667)	
Observations	2,156	2,156	1564
Database	Full	Full	Hotel fixed effect subsample
R-squared	0.5765	0.5863	0.9897
City Fixed Effect	YES	YES	YES
Star Grade Fixed Effect	YES	YES	YES
OTA Fixed Effect	NO	NO	YES
Hotel Fixed Effect	NO	NO	YES

With standard errors in parentheses.
+ p < 0.15. * p < 0.1. ** p < 0.05. *** p < 0.01.

Even if biased the first two model provide useful insights about the market framework of this study, more precisely the effect of the OTA on price. The first column depicts the coefficients taking into account only the city and the star grade fixed effects. An additional point in

$Users'Grade_r$[18] worth for the seller of a specific room, on average, 54€ while the value of an additional one percent of reviews is 0.045€[19]. Since this regression do not includes the ability of a specific accommodating structure these results are abnormally high. This in is not surprisingly considering how the leisure travellers' evaluations may be actually influenced by the hospitality. Indeed, as we will see later, after controlling for the hotel fixed effect the value of these coefficients will considerably drop. In the second column are reported the coefficient including also the OTA fixed effects. We expected that OTA of origin would affect the price to the extent that it influenced rooms price with both its reviews capital and its ability to reduce search costs. The main difference in the coefficients controlling for the OTA fixed effect is that both $logReviews_r$ and $Service_r$ on price, besides the reduction in magnitude, lost their significance. Hence the specification of the model without considering the OTA fixed effect suffer for the omitted variable bias. Considering our descriptive findings and comparing the results of the two model we can reasonably corroborate our previous hypothesis about the effect of the OTA of origin. On one hand, we can interpret this variation as if the reviews capital of the OTAs displaces part of the effects of the number of reviews because of their partial collinearity. While, on the other hand, it suggests that the provision of a service is somehow related with the specific OTA recommending the room. Hence, to the extent that this is true, we can point out the presence of vertical product differentiation[20] in the online travel market. In both the first and the second column of Table 1 the coefficients of the star rating are positive as expected, except for one of them. The negative coefficient of Low, the room attribute of having two stars, indicates that a room with two stars has a price of almost 25€ lower with respect to the base category, which is other accommodation.

In the third column are reported the coefficients of the model with the hotel fixed effect which constitute the main findings of our empirical analysis. In this way, the value of the coefficients is purified from the omitted variable bias related to the ability of a specific accommodating structure. The number of observations in this regression model drop from 2156 to 1564 because this formulation considers only the matched observations (namely, the rooms contemporarily sold on several OTAs). The value of an additional point in $Users'Grade_r$ is now, on average, nearly 5.15€. As we can see from Table 1 this value significantly falls of almost fifty euro. This means that the users' grade is mostly function of the hospitality of the structure which sells the room.

[18] As indicated in the previous chapter, we arbitrarily adapted the users' grade in a scale from one to ten for all the OTAs. Hence the result related to a one point increase in the users' grade must be consider relatively to this scale.
[19] $\Delta Room\ Price_r = (\beta_2/100)\Delta_{\%} logReviews_r$
[20] The vertical differentiation occurs when different price-quality combinations are offered to target different consumer.

From the leisure traveller's point of view the informational content of the reviews is mostly about the hospitality rather than the objective quality of the proposed room. Indeed, the objective quality attributes of the specific accommodating structure are already explained by the star rating. In the third column the coefficients for these variables are dropped by our statistical software. This because of the perfect collinearity with the hotel fixed effect. Indeed all the observations on a same room share the same star grade, as well as for the other variables save for the reviews, the users' grade and the OTA of origin. Anyway, we can consider their value as reported in the second column. On average a leisure traveller's willingness to pay for a three-star room (i.e. Medium) is nearly 43€ higher with respect to the base category. In other world, the consumer is willing to pays fortythree additional euro for the higher quality of a three-star room respect to another accommodation. The additional willingness to pay reach the considerable value of 278€ for a luxury room. As expected in a high-involvement sector, the effect of the quantity of reviews is relatively lower. This amount to 0,017€ for each additional percentage increase of the number of reviews. The reminder parts of both the effects of the reviews suggest that there are some objectives qualitative features of the recommended rooms which are not included in the star rating.

3.2.2. Matched Rooms Subsample

In this section we will present the results of the investigation about the effect of the competition between the OTAs in selling a room. We conclude, from the descriptive analysis, that it is not possible to identify a clear pattern for the price and users' grade in relation with the inverse degree of competitiveness. Our partial explanation of this puzzling result relies on the possibility of inefficiencies in the sellers' pricing behaviour. Table 2 compare the previous results with the ones relative to the subsample analysis, respectively on columns (1), (2) and (5) and columns (3) and (4). The results in the fifth column deserve some additional explanations. As can be notice it reports the same results as the third column of the previous table even though it is referred to a different regression. This occurs because the competition fixed effect is already included in the hotel fixed effect or, in other words, they are perfectly correlated.

Considering the matched rooms subsample, the resulting coefficients of the significant variables are higher than in the analysis with the full database. Comparing the first two columns with the third and the fourth columns it is possible to see how, taking into account only the matched rooms, the effects of the variables describing the quality of the rooms is greater respect to the full database.

Table 2: Matched rooms subsample results (Dependent variable Price)

VARIABLES	(1) Base Linear Regression	(2) Linear Regression with OTA Fixed Effect	(3) Linear Regression with OTA fixed Effect	(4) Linear Regression with Competition Effect	(5) Linear Regression with Hotel Fixed Effect
Users' Grade	54.0541***	50.8340***	58.1837***	52.6514***	5.1488**
	(11.1779)	(10.0861)	(13.8073)	(9.2637)	(1.5360)
Log Number of Reviews	4.4832**	1.5168	1.2412	1.1723	1.7324**
	(1.5074)	(1.4326)	(1.8128)	(2.1370)	(0.4960)
Included Service	19.5211**	6.2215	5.0715	7.2315	-1.4037
	(6.4177)	(5.4025)	(5.9044)	(5.0397)	(1.2857)
Zone	63.8119**	62.6936**	67.6315**	62.6901**	
	(19.7247)	(19.7674)	(23.3180)	(16.1546)	
Low	-24.9181	-21.9934	13.2420	-23.0986	
	(30.5392)	(20.2305)	(15.0360)	(19.5904)	
Medium	40.8322**	42.8649***	72.8131***	36.6727***	
	(10.9761)	(5.8229)	(14.1408)	(4.5405)	
Good	63.3641***	68.2829***	89.9519***	63.2768***	
	(2.7948)	(8.7968)	(17.5998)	(5.3251)	
Luxury	270.4941***	278.2366***	298.5531***	266.9425***	
	(31.4725)	(37.4667)	(48.9569)	(29.8708)	
Observations	2,156	2,156	1,564	1564	1,564
Database	Full	Full	Matched Room Subsample	Matched Room Subsample	Matched Room Subsample
R-squared	0.5765	0.5863	0.6003	0.5924	0.9871
City Fixed Effect	YES	YES	YES	YES	YES
Stars Grade Fixed Effect	YES	YES	YES	YES	YES
OTA Fixed Effect	NO	YES	YES	YES	YES
Competition Effect	NO	NO	NO	YES	YES
Hotel Fixed Effect	NO	NO	NO	NO	YES

With standard errors in parentheses.
+ p < 0.15. * p < 0.1. ** p < 0.05. *** p < 0.01.

Moreover, after controlling for the competition fixed effect these coefficients come back close to the ones of the full database. Ignoring for a while all the others possible distortions, this indicates how, in a more competitive framework, the qualitative indicators have a greater relevance. One of the effects of the competition in a service product market is to bring the price of a service near to its objective value reducing possible price distortions. Hence in a more competitive framework the other partial effect merge within the more objective qualitative variables. This can be interpreted

as a reduction of the price distortion caused by the information asymmetries. However there may be other explanations to these simple observations. Since these formulations still suffer from the omitted variable bias relative to the hospitality of each accommodating structure, we cannot be sure of the validity of this latter result. The weak evidences about the relationship between the rooms price and the inverse degree of competitiveness are not sufficient to confirm or reject our prior hypothesis on it. The limitations of our dataset do not allow us to unambiguously identify the effect of the competition among OTAs in the leisure travel sector as it needs further investigation and it may constitute the basis of future researchs.

Conclusion

Our interest lies in the empirical measurement of the price changes relatively to reviews in the online leisure travel market. We find that the value of an additional point in the users' grade, all else being equal, on average, is nearly 5.15€. While, the effect of the quantity of reviews on price amount to 0,017€ for each additional percentage increase of the number of reviews. This finding has important implications from both sector marketers' selling rooms online and leisure travellers using the OTA for their travel planning.

From the sector marketers' prospective this constitutes and important indication for future pricing decisions. Once the sellers know the value of the consumer-generated content relative to their room they are able improve their pricing strategy. For example looking on the market at a quality homogeneous rooms, save for a the of users' grade and the number of reviews, which are successfully sold on the market at given price, they can infer whether they are efficiently pricing their rooms. From the leisure travellers' point of view it constitutes a further confirmation about the goodness of the peer-generated contents. Moreover, the great reduction of the endogeneity due to the hotel fixed effect is a good indicator of how the reviews are an effective means to overcome the information asymmetries in a service product market. Indeed they are able to catch the omitted information of more traditional quality indicators. However the leisure travellers' ought to bear in mind how the informational contents of the reviews are mostly about the hospitality rather than the objective quality of the proposed room. We are likely to relax our previous position about the low trustworthiness of the market-generated content. Even if assign by the regional authorities, not by the marketers, we see how the star grade is a reliable quality indicator even if actually there are some objectives qualitative features of the recommended rooms which are not included in the star rating. The reminder parts of both the effects of the reviews suggest, in line with the literature, that they explain just a small fraction of the quality of a room, probably in relation with the problematic information omitted by the sellers.

Checking whether the behaviour of the intermediaries acting on online travel market is influenced by the presence of information asymmetries and search cost, we find just weak evidences. For this reason we are able to make just some "educated observation". Concerning the presence of price distortions we found just descriptive evidences about the OTA behaviour. More precisely we find that a room sold with an included service has an average price significantly higher than if it is not included. We also argue, from the regression analysis, that the provision of a service is related

with the specific OTA recommending the room. Overall this suggests that OTA are managing the services included in the deal trying to vertically differentiate their market. This is even more reasonable considering also the parity rate provision which avoids any form of horizontal differentiation. Overall our analysis on the OTAs indicates that they explain a small portion of the rooms price. A possible explanation to this result is that they are able to affect the leisure travellers' willingness to pay via both their review capital and their ability to reduce search costs. Anyway we are not able to provide quantitative measures of these influences. Regarding the puzzling effect of the OTAs competing in selling a same room we do not propose a possible explanation. Indeed there are not sufficient elements to confirm or reject our prior hypothesis on this relation. Both from the descriptive and inferential analysis of the relation between the rooms price and the inverse degree of competitiveness it seems to be indeterminate. The limitations of our dataset do not allow us to unambiguously identify the effect of the competition among OTAs in the leisure travel sector but they may constitute the basis for further investigation in future researches.

.

References

1. Arbatskaya, Maria and Konishi, Hideo, 2012. *Referrals in search markets*. USA. International Journal of Industrial Organization 30: 89-101.
2. Arndt, Johan, 1967. *Role of product-related conversations in the diffusion of a new product*. USA. Journal of Marketing Research 4 (3): 291-295.
3. Bakos, Yannis J., 1991. *A strategic analysis of elcectronic marketplaces.* Irvine, California. Journal of MIS.
4. Belleflamme, Paul and Peitz, Martin, 2010. *Industrial organization markets and strategies*. Cambridge, New York, USA. Cambridge University Press
5. Bickart, Barbara and Schindler, Robert M., 2001. *Internet forums as influential sources of consumer information*. Camden, New Jersey. Journal of Interactive Marketing 15 (3): 31-39.
6. Bone, Paula F., 1995. *Word-of-mouth effects on short-term and long-term product judgments*. New York. Journal of Business Research 32: 213-223.
7. Bonn, Mark A. and Furr H. Leslie and Susskind, Alex M., 1999. *Predicting a behavioural profile for pleasure travellers on the basis of internet use segmentation.* USA. Journal of Travel Research 37: 333-340.
8. Brown, Jacqueline J. and Reingen, Peter H., 1987. *Social ties and word-of-mouth referral behaviour*. USA. Journal of Consumer Research 14: 350-362.
9. Brynjolfsson, Erik and Smith, Michael D., 2000. *Frictionless commerce? A comparison of internet and conventional retailers*. Massachusetts, USA. Management Science 46 (4): 563-585.
10. Chatterjee, Patrali, 2001. *Online reviews: do consumers use them?* Rutgers University, USA. Advances in Consumer Research 28: 129-133.
11. Clemons, Eric K. and Hann, Il-Horn and Hitt, Lorin M., 2002. *Price dispersion and differentiation in online travel: an empirical investigation*. Philadelphia, USA. Management Science 48 (4): 534-549.
12. Combes, G. C. and Patel, J. J., 1997. *Creating lifelong customer relationships: why the race for customer acquisition on the internet is so strategically important*. USA: iword, Hambrecht & Quist 2 (4).
13. Condorelli, Daniele and Galeotti, Andrea and Skreta, Vasiliki, 2014. *Selling through referrals.* USA.

14. Dellarocas, Chrysanthos, 2003. *The digitization of word of mouth: promise and challenges of online feedback mechanisms.* Massachusetts, USA. Management Science 49 (10): 1407-1424.
15. European commission, 2017. *Booking your holidays online: Commission and consumer protection authorities act on misleading travel booking websites*. Brussels, Belgium. European commission- press release IP/17/844.
16. Farrell, Joseph, 1993. *Meaning and credibility in cheap-talk games.* California, USA. Games and economic behaviour 5: 514-531.
17. Fodness, Dale and Murray, Brian, 1997. *Tourist information search*. USA. Annals of Tourism Research 24 (3): 503-523.
18. Galenianos, Manolis, 2013. *Learning about match quality and the use of referrals.* Pennsylvania, USA. Review of Economic Dynamics 16: 668-690.
19. Gretzel, Ulrike and Yoo, Kyung Hyan, 2008. *Use and impact of online travel reviews*. Texas, USA. Conference Paper.
20. Gursoy, Dogan and McCleary, Ken W., 2004. *An integrative model of tourists' information search and behavior*. Great Britain. Annals of Tourism Research 31 (2): 353-373.
21. Hadj, Héla Ali and Nauges, Céline, 2007. *The pricing of experience goods: the example of "en primeur" wine.* USA. American Journal of Agricoltural Economics 89 (1): 91-103.
22. Hanlan, Janet and Kelly, Stephen, 2004. *Image formation, information sources and an iconic Australian tourist destination*. Australia. Journal of Vacation Marketing 11 (2): 163-177.
23. Hennig- Thurau, Thorsten and Gwinner, Kevin P. and others, 2004. *Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?* Journal of Interactive Marketing 18 (1): 38-52.
24. Inderst, Roman and Ottaviani, Marco, 2012. *Competition through commissions and kickbacks.* USA. The American Economic Review 102 (2): 780-809.
25. Murray, Keith B. and Schlacter, John L., 1990. *The impact of services versus goods on consumers' assessment of perceived risk and variability*. USA. Journal of the Academy of Marketing Science 18 (1): 51-65.
26. Murray, Keith B., 1991. *A test of services marketing theory: consumer information acquisition activities.* USA. Journal of Marketing 55: 10-25.

27. Olshavsky, Richard W. and Granbois, Donald H., 1979. *Consumer decision making—Fact or fiction?* Indiana, USA. Journal of Consumer Research 6: 93-99.
28. Park, Do-Hyung and Kim, Sara. 2008. *The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews*. Chicago, USA. Electronic Commerce and Applications 7: 399-410.
29. Park, Do-Hyung and Lee, Jumin and Han, Ingoo, 2007. *The effect of on-line consumer reviews on consumer purchasing intention: the moderating role of involvement.* USA. International Journal of Electronic Commerce 11 (4): 125-148.
30. Shavell, Steven, 1994. *Acquisition and disclosure of information prior to sale.* USA. The RAND Journal of Economics 25 (1): 20-36.
31. Stock, James H. and Watson, Mark W., 2011. *Introduction to econometrics*. USA. Pearson.
32. Valletti, Tommaso M., 2000. *Price discrimination and price dispersion in a duopoly*. Venezia, Italia. Research in Economics 54: 351-374.
33. Woolridge, Jeffrey M., 2010. *Econometric analysis of cross sections and panel data.* Massachusetts, USA. MIT Press.

The Effect of the Reviews in the Online Travel Market: an Empirical Investigation

Abstract

Previous research has examined whether the electronic word of mouth (eWOM) has an effect on the theoretically highly efficient Internet markets. For the best of our knowledge, the literature on this field is mostly related to the theoretical description of the relations between the information asymmetries and the informational contents of the reviews with little attention to the price changes. However, much of the previous empirical work was focused on industries selling goods underling the presence of price dispersion and differentiation. There is lack of research literature on the empirical effects of the consumers' reviews in a service products market. In this study we investigate the effect of the consumers' reviews on a high-involvement service market: the online leisure travel market. In this framework, we are interested in quantifying the price changes relatively to both the quantity of the reviews and the relative users' grade. Performing this empirical analysis, we check whether the behaviour of the principal agents of this market, coherently with the literature, is influenced by the presence of information asymmetries and search costs. The online leisure travel market is a highly representative example of high-involvement market characterized by intermediaries between sellers and price-sensitive buyers. We use a software agent, emulating the preference of the leisure travellers, to create a tailored database of actual rooms, both from hotels and other accommodations, offered by a major corporate travel agencies. Later we conduct a descriptive investigation to present the database features and then we propose a multivariate linear regression analysis to find the price effects of the reviews controlling for all the quality variables and the hotel fixed effect. In this way we are able to make an inferential analysis to objectively identify the value of the reviews, all else being equal, in a scenario that closely matches how leisure travellers would use them. We expect that the price of the recommended rooms is affected from both the number and the quality of the reviews, with a higher influence of this latter. Moreover we expect to find empirical indications about the theoretically ambiguous effect, relatively to this market, of the competition between sellers.

Our analysis begins with a data set of actual rooms, both from hotels and other accommodations, offered by a major corporate travel agencies in the month of April 2017 in Italy. We ran our

software agent six consecutive times, making requests for the corresponding six principal tourist destinations, contemporarily collecting data from eight online travel agencies (OTAs). These cities, ordered by number of visits, are Rome, Milan, Venice, Florence, Turin and Naples. Since there were not publicly available databases on the accommodating structures rooms recommended by the OTAs, we first needed to collect the observations on their prices and attributes. The key challenges in the construction of our database were both to be able to avoid exogenous influences on the observed prices and to control for the hotel fixed effect. Hence, we were interested in obtaining information on different OTAs contemporarily. Furthermore we were interested in imitating the behaviour of a typical leisure traveller. To perform this task, we used a software agent acting as a virtual consumer to collect the quantitative variables virtually eliminating the chance of price changes influencing the result. In order to emulate the preferences of a stylized leisure traveller we programmed it to make requests for a double room for the weekend (from 06/05/2017 to 07/05/2017) with thirty days advance. This ensures high quality data as we ran our software agent on every OTA contemporarily. Concerning the possibility to include in the model the hotel fixed effect, we needed different observations of the same room. It was reasonable to assume that in a given time, an accommodating structure sold its room through different platforms contemporarily, thus we collected data on different OTAs. We selected the OTAs on the basis of their market shares. The two groups holdings the eight OTAs constituting our sample controlled the majority share of the online travel market in Italy. The number of observation for each OTA was limited by the decreasing quality of the recommendations proceeding through the result pages. Hence to reach a consistent level of observation we repeat our data scraping for each of the cities (Rome, Milan, Venice, Florence, Turin and Naples). The collection for each city was not simultaneous but it took place in few hours. Our final database was constituted by 2156 observation on 1069 actual rooms, both from Italians hotels and other accommodations, offered by a major OTAs. All the relevant attributes, indicated on the recommendation pages, were described by a total of 33 variables. Among these, 23 are dichotomous variables representing the 6 variables corresponding to: the star grade (6), the city (6), the OTA of origin (8) and whether if the room was an hotel (1), if it was sold bundled with a service (1) and if it was located in one of the selected zone (1). The only two quantitative variables were the price and the users' grade while the name and the ID of the hotel served to uniquely identified each of the rooms.

After the database construction we presented some descriptive tables and graphs to highlight the principals properties of the data constituting our sample. We were dealing with a unique database tailored for our study about the effect of the reviews on the online leisure traveller market. Because

of the peculiarity of the market segment and the data sources, our database was mostly composed by variables of qualitative nature. We expected, coherently with the literature, that the OTA of origin of a specific room constituted one of its relevant attributes affecting price and, consequently, that prices of a specific room is not the same among different OTAs. Moreover, accordingly with the widely acknowledged negative effect of competition on price, we presume that the price decreases as the number of OTAs selling the same room increases. Hence, before going through to the inferential analysis of the data we tried to understand precisely which were the features of this dataset, in order to design an appropriate regression model. In this section we found three main descriptive evidence. The OTA of origin is an explanatory variable for the price to the extent that it influences rooms price with both its reviews capital and its ability to reduce search costs. From the descriptive analysis, the rooms price seems to be indeterminately affected from the inverse degree of competitiveness.

We were interested in quantifying the effect of these reviews on room prices. The key challenge in empirically testing our predictions is to properly model the interdependence between room characteristics, which are the explanatory variables, and the price, that is the explained variable. First of all, we assumed that the relationship between these variables was linear. Practically this assumption can virtually never be confirmed, however multiple regression procedures are not greatly affected by minor deviations from this assumption. The critical dimensions that affect the room prices are the city, the location, whether the service is provided or not, the nature of the structure, the quality (caught by both star rating and users' grade) and the OTA which sells the room to the extent that it reduces the searching costs. As discussed before, all these relevant attributes are described by both the quantitative and the qualitative variables constituting our tailored database. We proposed three linear regression models with an increasing number of explanatory variables among which the more general was the following (3).

$$
\begin{aligned}
Room\ Price_r = {} & \alpha + \beta_1 Users'Grade_r + \beta_2 logReviews_r \\
& + \gamma_{City_{r,c}} + \gamma_1 Zone_r \\
& + \delta_{Star\ Grade_{r,s}} + \delta_1 Service_r \\
& + \varphi_{OTA_{r,o}} \\
& + \vartheta_{Hotel_{r,h}} + \varepsilon_r
\end{aligned}
$$

Where γ_{City_r} represents the city fixed effect, $\delta_{Star\ Grade_r}$ represents the star grade fixed effect, φ_{OTA_r} represents the OTA fixed effect and ϑ_{Hotel_r} represents the hotel fixed effect and r indexes the rooms.

The underlying assumptions of this formulation are the standard ones for the ordinary least squares (OLS) model. The first assumption is that other factors affecting the price, contained in the error term ε_r, are uncorrelated to the explanatory variables specified in the model. This means that, on average, these other factors do not affect our results. Whereas there is little to say about the first assumption, which just describes an ideal condition, there are some caveats about the following two. The second assumption is about how the sample is drawn. It postulates that the observations of the sample are independently and identically distributed (i.i.d.) across observation. We relaxed this assumption, because of spatial correlation, clustering our sample for the city categories. While we allowed the units within each cluster to be correlated, we expected independence throughout the clusters (namely we assumed across city i.i.d. observations). The prices of the accommodations within a city may be correlated because of common characteristics of the rooms within that city or because of the features of the city itself (such as the high maintenance cost in Venice). The third assumption is that large outliers are unlikely. In order to let our model likely to respect this latter condition we eliminated outliers from the observation on users' grade while we applied a logarithmic transformation to the number of reviews. The last assumption is that there is not a perfect collinearity between the explanatory variables. If a perfect collinearity occurred between two explanatory variables our statistical software automatically dropped one of the two to avoid the collapse of the variance covariance matrix and proceeded with the coefficients estimation.

Concerning the hotel fixed effect it is reasonable to assume that the leisure travellers deciding whether to write a review and choosing the relative evaluation are influenced by the specific hospitality of a given accommodating structure. If this is true our estimation is biased because of the omitted variable bias. In our case $Users'Grade_r$ and $logReviews_r$ are correlated with the unobservable because of self-selection: travellers decide whether to review and the grade to assign depending on the specific hospitality of a given accommodating structure, that is unobservable.

In the following table are reported the main results of our regression analysis controlling for different fixed effects as indicated in the bottom part of the table.

Table: Regression results (Dependent variable Price)

VARIABLES	(1) Base Linear Regression	(2) Linear Regression with OTA Fixed Effect	(3) Linear Regression with Hotel Fixed Effect
Users' Grade	54.0541***	50.8340***	5.1488**
	(11.1779)	(10.0861)	(1.8036)
Log Number of Reviews	4.4832**	1.5168	1.7324**
	(1.5074)	(1.4326)	(0.5824)
Included Service	19.5211**	6.2215	-1.4037
	(6.4177)	(5.4025)	(1.5097)
Zone	63.8119**	62.6936**	
	(19.7247)	(19.7674)	
Low	-24.9181	-21.9934	
	(30.5392)	(20.2305)	
Medium	40.8322**	42.8649***	
	(10.9761)	(5.8229)	
Good	63.3641***	68.2829***	
	(2.7948)	(8.7968)	
Luxury	270.4941***	278.2366***	
	(31.4725)	(37.4667)	
Observations	2,156	2,156	1564
Database	Full	Full	Hotel fixed effect subsample
R-squared	0.5765	0.5863	0.9897
City Fixed Effect	YES	YES	YES
Star Grade Fixed Effect	YES	YES	YES
OTA Fixed Effect	NO	NO	YES
Hotel Fixed Effect	NO	NO	YES

With standard errors in parentheses.
+ p < 0.15. * p < 0.1. ** p < 0.05. *** p < 0.01.

We find that the value of an additional point in the users' grade, all else being equal, on average, is nearly 5.15€. While, the effect of the quantity of reviews on price amount to 0,017€ for each additional percentage increase of the number of reviews. this finding has important implications from both sector marketers' selling rooms online and leisure travellers using the OTA for their travel planning.

From the sector marketers' prospective this constitutes and important indication for future pricing decisions. Once the sellers know the value of the consumer-generated content relative to their room

they are able improve their pricing strategy. For example looking on the market at a quality homogeneous rooms, save for a the of users' grade and the number of reviews, which are successfully sold on the market at given price, they can infer whether they are efficiently pricing their rooms. From the leisure travellers' point of view it constitutes a further confirmation about the goodness of the peer-generated contents. Moreover the great reduction of the endogeneity due to the hotel fixed effect is a good indicator of how the reviews are an effective means to overcome the information asymmetries in a service product market. Indeed they are able to catch the omitted information of more traditional quality indicators. However the leisure travellers' ought to bear in mind how the informational contents of the reviews are mostly about the hospitality rather than the objective quality of the proposed room. We are likely to relax our previous position about the low trustworthiness of the market-generated content. Even if assign by the regional authorities, not by the marketers, we see how the star grade is a reliable quality indicator even if actually there are some objectives qualitative features of the recommended rooms which are not included in the star rating. The reminder parts of both the effects of the reviews suggest, in line with the literature, that they explain just a small fraction of the quality of a room, probably in relation with the problematic information omitted by the sellers.

Checking whether the behaviour of the intermediaries acting on online travel market is influenced by the presence of information asymmetries and search cost, we find just weak evidences. For this reason we are able to make just some "educated observation". Concerning the presence of price distortions we found just descriptive evidences about the OTA behaviour. More precisely we find that a room sold with an included service has an average price significantly higher than if it is not included. We also argue, from the regression analysis, that the provision of a service is related with the specific OTA recommending the room. Overall this suggests that OTA are managing the services included in the deal trying to vertically differentiate their market. This is even more reasonable considering also the parity rate provision which avoids any form of horizontal differentiation. Overall our analysis on the OTAs indicates that they explain a small portion of the rooms price. A possible explanation to this result is that they are able to affect the leisure travellers' willingness to pay via both their review capital and their ability to reduce search costs. Anyway we are not able to provide quantitative measures of these influences. Regarding the puzzling effect of the OTAs competing in selling a same room we do not propose a possible explanation. Indeed there are not sufficient elements to confirm or reject our prior hypothesis on this relation. Both from the descriptive and inferential analysis of the relation between the rooms price and the inverse degree of competitiveness it seems to be indeterminate. The limitations of our dataset do not allow

us to unambiguously identify the effect of the competition among OTAs in the leisure travel sector but they may constitute the basis for further investigation in future researches.