Substitution threat between Airbnb and hotels: myth or reality?

Ruggero Sainaghi\textsuperscript{a}, Rodolfo Baggio\textsuperscript{b,c}

\textsuperscript{a} Università IULM, Dipartimento di Business, Law, Economics, and Consumer Behaviour, Via Carlo Bo, 1, I-20143 Milan, Italy – ruggero.sainaghi@iulm.it

\textsuperscript{b} Bocconi University, Master in Economics and Tourism, Via Röntgen, 1, I-20136 Milan, Italy – rodolfo.baggio@unibocconi.it

\textsuperscript{c} National Research Tomsk Polytechnic University, 30 Lenin Avenue, Tomsk 634050, Russian Federation

ABSTRACT

The rise of peer-to-peer accommodation platforms in the tourism and hospitality industry has created an interesting and growing debate around the threats of substitution between them and traditional hotels. Previous studies have provided contradictory findings. Here we address the issue by analyzing the degree of synchronization between the daily occupancy of hotels and that of Airbnb listings in Milan, Italy, over a period of four years. The findings show that the two series are widely desynchronized during the week, on workdays and trade-fair days, when hotels work prevalently within the business segment, and when Airbnb listings mainly accommodate leisure guests. By contrast, a partial synchronization (and therefore a potential substitution threat) is revealed during weekends and holidays.

KEYWORDS

Peer-to-peer accommodation platforms; Airbnb; mutual information; substitution threats

INTRODUCTION

A few years ago, peer-to-peer accommodation platforms were identified as the hottest trend for tourism and hospitality (Pizam, 2014). Some authors revealed the paucity of studies (Koh & King, 2017; Tussyadiah & Pesonen, 2018), mainly represented by so-called “grey literature” (conference proceedings, commissioned research, and newspaper articles) (Cheng, 2016; Dredge & Gyimóthy, 2015; Oskam & Boswijk, 2016). However, the situation has quickly changed. In fact, the exponential growth of Airbnb and other peer-to-peer accommodation platforms has attracted the interest of researchers around the world (Altinay & Taheri, 2019; Dolnicar, 2019). This body of literature has explored many topics; one of these is the substitution threat generated by these platforms towards traditional hotels (Prayag & Ozanne, 2018). Despite the relevancy of this central question, there are
few empirical studies and, of those, the results are contradictory. For example, Zervas, Proserpio, and Byers (2017), based on a longitudinal study of 7,361 Airbnb listings and 4,006 hotels, affirmed that an increase in 1% of Airbnb listings resulted in a 0.05% reduction in hotel revenue in Texas, with low-end hotels and those not specializing in business guests being most affected. Apparently, the substitution threat is very small; however, if Airbnb listings rise by 20%–30% the effect is relevant. This study was later criticized by Heo, Blal, and Choi (2019) because the research did not consider either location or seasonal patterns. Xie and Kwok (2017) identified potential substitution for the demand for hotels, with a consequent reduction in revenue per available room. However, in the study of Guttentag (2015), built around the disruptive innovation theory, three arguments limit the impact generated by Airbnbs on hotels: i) Airbnbs are a niche product; ii) they are not a perfect substitute for traditional accommodation firms; and iii) there are not enough Airbnbs to impact upon hotels. In other papers, the concentration of Airbnb listings in some zones is evaluated as considerably high (Boros et al., 2018; Brauckmann, 2017; Gutiérrez et al., 2017) and, therefore, the third consideration expressed by Guttentag (2015) is at least nuanced (Oskam & Boswijk, 2016). Some recent studies, then, support the absence of (e.g., Ginindza & Tichaawa, 2019) or limited competition threat (Mhlanga, 2019) between Airbnb and traditional hotels. Moreover, using different methods (an agent-based model), Bruno and Faggini (2020) conclude that low substitutability exists and is mainly determined by demand levels, rather than by prices or competition within and between markets.

The central question about the competitive threat generated by Airbnb towards hotels remains largely a black box. Previous papers have mainly researched quantitative evidence about the effects generated by Airbnb on hotel metrics, such as revenue per available room, average daily rate, or occupancy, or, more rarely, on financial indicators (Aznar et al., 2016, 2017; Blal, Singal, & Templin, 2018; Choi et al., 2015). Surprisingly, very few studies have analyzed the substitution threat focusing on the comparison of seasonal patterns regarding peer-to-peer accommodation platform and hotels (Heo, Blal, & Choi, 2019; Martín et al., 2018). This paper contributes towards filling this gap. The main research question below is later operationalized around three well-defined hypotheses.

RQ: Are the seasonal patterns of Airbnbs and hotels similar or different?

The higher the overlap (similarity), the higher the potential substitution threat, and vice versa. This paper explores this research question in an urban destination, represented by the city of Milan. In the methodology section, the reasons behind the choice of this context are presented.

**LITERATURE REVIEW**

This section is structured in three parts. The first section analyzes the results suggested in previous studies about Airbnb traveler profiles and their possible implications in term of seasonality. The second section explores the seasonal patterns of Milan – the case analyzed in this paper – based on published research. Based on the travelers’ characteristics and seasonal destination patterns, the hypotheses tested in the empirical findings are presented.

**Airbnb profile**
The literature on the peer-to-peer accommodation platforms is mainly centered around demand studies (Sainaghi et al., 2019). This is not new in the field of tourism and hospitality (Leiper, 1979; Smith, 1988). The paper’s exploration of the demand side can be classified around many topics, such as consumer behavior (Ert, Fleischer & Magen, 2016; Fagerstrøm et al., 2017; Goh, 2015; Tussyadiah & Pesonen, 2016, 2018), customer satisfaction (Mao & Lyu, 2017; Möhlmann, 2015; Priporas et al., 2017a, 2017b), and customer segmentation (Decrop et al., 2018; Del Chiappa, Lorenzo-Romero & Alarcón-del-Amo, 2018; Pera, Viglia & Furlan, 2016). For the purposes of this paper, some of the conclusions stated by some previous studies, which have compared peer-to-peer and hotel guests’ profiles, are discussed.

The work of Varma et al. (2016) has developed a comparative analysis and some recurrent differences between clients from hotels and Airbnbs are illustrated. Based on interviews with hotel executives, the authors state that the competition generated by Airbnb is very limited, considering that hotels and Airbnbs attract different segments. In particular, Airbnbs are used more by travelers with longer stays “but this does not hold true in the case of business travel” (Varma et al., 2016, p. 233). Airbnb clientele is more focused on the leisure side. The relevance of travel purposes (leisure versus business) as well as trip duration (stay) is confirmed by the work of Poon and Huang (2017). The authors compared Airbnb users and non-users and found many differences. In particular, Airbnbs are considered more suitable for leisure travel. Concerning trip duration, the authors analyzed this variable together with travel party size. “When traveling alone and traveling with a spouse/partner, respondents were more likely to stay at hotels if the trip was shorter, but would prefer Airbnb if the trip was longer. On the other hand, when traveling with friends, respondents were more likely to use Airbnb, regardless of trip duration” (Poon & Huang, 2017, p. 2435). Similarly, Young, Corsun, and Xie (2017) confirm the relevance of these three variables: motivation, party size, and trip duration. Peer-to-peer accommodation platforms are more associated with leisure travel. Party size and length of the trip are considered important determinants for choosing these accommodations. The relevance of the leisure segment is confirmed in some studies that explore the location patterns of Airbnb listings. In some cases, these studies develop comparisons with hotels. The work of Gutiérrez et al. (2017) found a positive correlation between Airbnbs “with the proximity of sightseeing spots, and land use associated with the leisure, hospitality and entertainment industries” (p. 288). Adamiak (2018) confirms that the number of Airbnb offers is positively related to the size of the city and its importance as a leisure tourism destination. The relevance of leisure clients is also confirmed by papers that explore the Airbnb pricing strategy. Gibbs et al. (2018a, 2018b) affirm the centrality of Airbnbs for leisure clients on one side, and the marginality of business travelers on the other. For this reason, they found a wider Airbnb price fluctuation within high-demand leisure destinations.

Based on these studies, some implications can be considered for the present paper. For a destination specialized in three different targets, represented by business, trade-fair, and leisure travelers (as discussed in the next section, dedicated to Milan), the hotels can attract all of these segments, while Airbnb listings should be more focused on leisure clients. Therefore, if these market segments show varied seasonal patterns, these two forms of supply should depict some different traits within those seasonal patterns.

**Milan seasonal patterns**
A relatively wide and recent amount of literature has explored the seasonal patterns of Milan (De Carlo et al., 2009; Guizzardi, Mariani, & Prayag, 2017; Baggio & Sainaghi, 2011; Sainaghi & Canali, 2009, 2011; Sainaghi & Mauri, 2018; Sainaghi, Mauri & d’Angella, 2018; Sainaghi et al., 2019). This destination is Italy’s economic capital (Sainaghi & Mauri, 2018). The city is located in Lombardy (one of the 20 Italian regions), which, alone, produces more than one fifth of the national gross domestic product (Sainaghi et al., 2019). Furthermore, Milan is a leading European destination for the trade-fair market (Sainaghi, Mauri & d’Agnella, 2018). Finally, the city is a world capital for fashion and design (with New York and Paris), hosted the Expo in 2015, and is known as the city of Leonardo da Vinci; it is also one of the top four Italian heritage cities (Sainaghi & Mauri, 2018).

This tourism positioning is well reflected within three main, relevant segments: business, trade-fair clients, and leisure travelers. These three market segments generate recurrent seasonal patterns. Generally speaking, the city shows a strong demand fluctuation between the weekend and weekdays. In particular, during the week, the business segment is prevalent and, considering the strong tie between Milan and the economy, the occupancy is higher than it is at weekends. The leisure segment, in fact, is estimated to be a less relevant market segment, representing approximately 20% of the tourist flows (Sainaghi & Mauri, 2018). However, since the Expo event, this percentage has increased due to the positive effect generated by the Expo during the weekend and, more generally during holiday periods (Sainaghi et al., 2019). A second seasonal pattern, always centered around business and leisure, varies in demand between working and non-working days. The first group includes all weekdays not affected by religious (such as Christmas and Easter) or civil holidays, such as 1st May (Labor Day), 25th April (Liberation Day), and 2nd June (Republic Day). Given the tradition in Italy to concentrate the summer closing of firms in August, the whole month is considered as a non-working period, as mentioned in previous researches (Sainaghi & Mauri, 2018; Sainaghi, Mauri & d’Angella, 2018). By contrast, non-working days include weekends and all holiday periods (as previously identified).

Finally, the presence of an important trade-fair center generates peaks in demand when relevant events are hosted. Although Fiera Milano organizes many events, a few business-to-business trade-fairs are considered to be relevant for the city and attract a large number of attendees. As analyzed in previous papers, these can considerably increase the average occupancy, average daily rate, and revenue per available room.

Hypotheses development

In this section some previous insights related to both the Airbnb tourist profile and Milan as a destination are considered to formulate four hypotheses, which guide the empirical analysis. In particular, the overlap between Airbnb’s strong identification with leisure on one side, and the prevalent business segment attracted by the city of Milan on the other, have guided the following first two hypotheses.

The first hypothesis is clearly focused on the fluctuation of weekly demand. As previously clarified, the city of Milan is mainly visited during the week by business clients, while during the weekend the city is more often visited by tourists. The hotels in this destination attract both market segments, but, considering the stronger focus of Milan on business clients, the hotel metrics (occupancy, average
daily rate, and revenue per available room) is surely higher during weekdays (Sainaghi & Mauri, 2018). By contrast, Airbnb listings should prevalently attract leisure market targets. Therefore, these seasonal traits should be very different during the weekdays and widely overlap during weekends, as stated in the following two hypotheses. In this paper, the focus is on seasonal patterns, which are expected to vary. Furthermore, a wide range of literature suggests that the services and attributes delivered by Airbnb listings and hotels are vastly different (Chen & Xie, 2017; Wang & Nicolau, 2017; Xie & Kwok, 2017; Young, Corsun & Xie, 2017). Therefore, if the seasonal patterns are desynchronized, there is stronger additional evidence of reduced substitution threat between the two lodging sectors (Airbnb and traditional hotels).

**Hypothesis 1.A** The *weekday* seasonality of Airbnb shows a *different* structure compared to hotels in Milan.

**Hypothesis 1.B** The *weekend* seasonality of Airbnb shows a *similar* structure compared to hotels in Milan.

The second hypothesis enlarges the time horizon moving from weekdays and weekends to working and non-working days, as previously defined. Also in this case, an asymmetric structure should characterize the lodging supplies of the two. In fact, hotels are expected to compete during both seasonal periods (but with higher metrics during working days), while Airbnb listings should concentrate more on holidays. In line with this first hypothesis (1.A and 1.B), the second should be able to catch the differences between these two lodging supplies (Airbnb versus hotels).

**Hypothesis 2.A.** The *working* seasonality of Airbnb shows a *different* structure to that of hotels in Milan.

**Hypothesis 2.B.** The *holiday* seasonality of Airbnb shows a *similar* structure to that of hotels in Milan.

The third larger city market segment includes trade-fairs. In the study of Milan, some trade-fair events can considerable increase the occupancy and prices of hotels, as previously discussed. Trade-fair events attract mainly business attendants (employees of companies) and, therefore, firms should be more oriented toward hotels than Airbnb listings. The following hypothesis is stated.

**Hypothesis 3.** The trade-fair seasonality of Airbnb shows a *different* structure to that of hotels in Milan.

Finally, a last (fourth) hypothesis focuses on the general (total) seasonal structure of Airbnbs and hotels. As discussed in the previous section, Milan is a prevalent business and trade-fair destination. Therefore, these two market segments cover a large number of days. Considering the fact that during these seasonal periods, Airbnb listings should be considerably less competitive than hotels, while the overlapping occurs during weekend and holiday periods, the whole seasonal series of these two supplies (Airbnb and hotel) should differ. Therefore, the following last hypothesis is formulated.

**Hypothesis 4.** The whole seasonality of Airbnb shows a *different* structure to that of hotels in Milan.
**METHODOLOGY**

**The sample**

This study has chosen the city of Milan due to its prevalent focus on business and trade-fair clients on one side, but in association with its non-marginal presence of leisure travelers on the other. Previous papers that explore the effects generated by Airbnb in Europe are mainly focused on large leisure cities, such as Barcelona (Lambea Llop, 2017; Martín Martín, Guaità Martínez & Salinas Fernández, 2018; Nofre et al., 2018), Madrid (García-Ayllón, 2018), Paris (Heo, Blal & Choi, 2019), London (Ferreri & Sanyal, 2018), Venice (Oxoli, Prestifilippo & Bertocchi, 2017), and Berlin (Schäfer & Braun, 2016). Considering the explorative nature of the present study, the research team decided to focus only on the city of Milan. For this destination, the presence of a large business sector should be a useful context in which to test the previous hypotheses. Future studies can explore the generalizability of the current results (reported and discussed in the Findings section), enlarging the sample to include some strongly leisure-oriented destinations (such as Florence and Venice in Italy).

To compare the seasonal patterns of Airbnbs and hotels, two different series were used. For the hotel industry, Smith Travel Research (STR) data were collected. They include the daily data of demand (sold rooms), supply (available rooms) and revenue (Sainaghi, 2010; Sainaghi, Phillips & Corti, 2013; Sainaghi, Phillips & Zavarrone, 2017). This data has already been used in many tourism and hospitality studies (i.e., Enz, Peiró-Signes & Segarra-Oña, 2014; Makki, Singh & Ozturk, 2016; Pan and Yang, 2017) as well as in some recent articles exploring Milan’s seasonal patterns (Sainaghi & Mauri, 2018; Sainaghi et al., 2019; Sainaghi, Mauri & d’Angella, 2018). The STR series covers the period 2004–2018.

For Airbnb, AirDNA data were purchased by the research team. The only information available covers the period 2014 (from November) to June 2019. Therefore, the data include four completed years (2015–2018). To test the hypotheses, daily data was used. AirDNA considers the available and sold listings as well as the price for each day and for each listing. The sample includes all the Milan’s population represented by more than 55 thousand listings. A growing number of studies are using AirDNA data (Dalir, Mahamadaminov & Olya, 2020; Drogu et al., 2020; Gunter, Önder & Zekan, 2020). The authors decided to use these data because they are the only ones available for Milan on a daily basis.

Given the need to overlap the two series, the shorter data set defines the research period, spanning the four-year period 2015–2018.

**The method**

While several methods can be employed (Baggio & Sainaghi, 2016; Sainaghi & Baggio, 2017, 2019; Serra & Arcos, 2014) to estimate whether the patterns of two time series are synchronous, here we compare the occupancy series of Airbnb and hotels using the method proposed by Freeman et al. (2019) and Cazelles (2004). According to these two studies, the two time series are in synchrony; this is evident from the patterns of the two series following three main steps.

The first defines a “rhythm” for the time series, transforming it into a series of symbols (discretization) by comparing each value with its neighbors’. We formally assess whether any given
point is a trough point, peak point, increase, decrease, or same (no variation): i) trough point (later identified with the letter “A” in the example reported in Figure 1): \( x(t+1)<x(t)\leq x(t+2) \) or \( x(t+1)<x(t+2)\leq x(t) \); ii) peak point (B): \( x(t)<x(t+2)\leq x(t+1) \) or \( x(t+2)<x(t)\leq x(t+1) \) or \( x(t+2)\leq x(t)<x(t+1) \); iii) increase (C): \( x(t)\leq x(t+1)<x(t+2) \); iv) decrease (D): \( x(t+2)\leq x(t+1)<x(t) \); and v) stability (E): \( x(t)=x(t+1)=x(t+2) \). Each point type is assigned to a specific symbol (letter) and the sequence of symbols contains only the information regarding the behavior of the series, automatically disregarding the information on trend and changes in amplitude (see Figure 1). In the example reported in Figure 1, all five cases are illustrated and identified by the corresponding letters.

![Figure 1. Transformation of a time series in symbols](image)

The second step comprises an evaluation of the mutual behaviors of the series using information theory. The metric used is the mutual information. This is, in short, a quantity that measures a relationship between two random variables that are sampled simultaneously and capture all dependencies between the variables (Latham & Roudi, 2009). In particular, it evaluates, on average, how much information is communicated by one time-series about the other and can be thought of as a measure of the extent to which the two series have synchronous behavior. Formally, given the series \( X \) and \( Y \), the mutual information \( I(X,Y) \) is calculated as: \( I(X,Y)=H(X)+H(Y)-H(X,Y) \), where \( H \) is the entropy of each series \( H(X) = -\sum p(x_i) \log_2(p(x_i)) \) and \( H(X,Y) \) the joint entropy of the series \( H(X,Y) = -\sum \sum p(x_i, y_j) \log_2(p(x_i, y_j)) \). We then normalize the mutual information using the formula: \( U(X,Y)= I(X,Y) / (H(x)+H(Y)) \). Thus \( U(X,Y) \) is in the interval \([0,1]\). It is easy to demonstrate that if \( X \) and \( Y \) are independent random variables, then \( H(X, Y) = H(X) + H(Y) \); therefore, the mutual information is zero. For our use, this means there is no synchronization of the two series (Latham and Roudi, 2009). All calculations were performed using an adapted version of the Python scripts available at https://github.com/people3k/pop-solar-sync.

During the last step of the procedure, we estimated the statistical significance of the values \( U(X,Y) \) obtained. To do so, we generated a number (500 in our case) of random surrogate pairs of series using a Markov process that preserves the short-term structure of our series (with a 1 time-step memory) and an average of the mutual information values were obtained. We assessed whether the computed mutual information for the Airbnb–hotel pair was statistically significant when compared to the mean mutual information of the surrogate time series via a one-sample t-test.

**Findings**

The findings are structured in five sub-sections. The first reports some descriptive statistics useful for understanding the information about the occupancy of Airbnb and hotels in Milan. The other four sections test the four hypotheses, focusing attention on the three different seasonal periods
(hypotheses 1, 2, 3) or on the whole time series, without considering any seasonal segmentation (hypothesis 4).

**Descriptive statistics**

To understand some distinctive characteristics of the Airbnb and hotel occupancy series, Table 1 articulates the values for each seasonal period considered in the hypotheses. A first remark concerns the different occupancy rates reported by the hotel and Airbnb listings. Generally speaking, the hotels considerably outperform the peer-to-peer lodgings. This is particularly evident focusing on the last row (hypothesis 4, “all”): hotels achieve 0.691 (69.1%) while Airbnb 0.223 (22.3%). The gap is roughly 0.45 (45%). Said differently, hotels, on average, account for occupancy that is 3.1 times higher than that of Airbnb listings. The total number of days was 1,461 (four years, one of which is a leap year). These wide differences suggest standardizing occupancy rates, as reported in Figure 2.

<table>
<thead>
<tr>
<th>Name</th>
<th>Series</th>
<th>Days</th>
<th>Occupancy</th>
<th>Std deviation</th>
<th>Mean std errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Airbnb</td>
<td>Hotels</td>
<td>Airbnb</td>
</tr>
<tr>
<td>1.A</td>
<td>midweek</td>
<td>1,043</td>
<td>0.211</td>
<td>0.703</td>
<td>0.086</td>
</tr>
<tr>
<td>1.B</td>
<td>weekend</td>
<td>418</td>
<td>0.253</td>
<td>0.661</td>
<td>0.086</td>
</tr>
<tr>
<td>2.A</td>
<td>workdays</td>
<td>877</td>
<td>0.214</td>
<td>0.744</td>
<td>0.089</td>
</tr>
<tr>
<td>2.B</td>
<td>holidays</td>
<td>584</td>
<td>0.236</td>
<td>0.612</td>
<td>0.095</td>
</tr>
<tr>
<td>3.</td>
<td>fair</td>
<td>122</td>
<td>0.317</td>
<td>0.836</td>
<td>0.136</td>
</tr>
<tr>
<td>4.</td>
<td>all</td>
<td>1,461</td>
<td>0.223</td>
<td>0.691</td>
<td>0.092</td>
</tr>
</tbody>
</table>

Figure 2. Standardized occupancy
Focusing on the absolute rates (Table 1), the different seasonal periods show some important differences. For example, during the weekdays, they represent 1,043 days (71.4% of total days); the hotels exhibit a positive gap of Airbnb listings close to 0.50 (50%). Hotel occupancy is 3.3 times higher than those of Airbnb listings. The standardized values (Figure 2) confirm the gap. During weekends (418 days) the relationships among these two lodging supplies change significantly. In fact, the hotels outperform Airbnb listings; however, the ratio reduces to 2.6. The standardized values (Figure 2) show a lower hotel occupancy rate. Moving to the second hypothesis, the conclusions are in line with the first. Hotels consistently outperform peer-to-peer platforms during the 877 workdays (3.5 ratio between hotels and Airbnb occupancy) and the standardized values confirm this gap. However, during holidays (584 days) the ratio falls to 2.6; Figure 2 reveals a positive gap for Airbnb listings. Finally, during trade-fair days (122 days) both platforms achieve maximum values; however, the absolute gap is the highest (0.519). In fact, the hotel occupancy (0.836) is considerably higher than Airbnb (0.317). The standardized metrics illustrate a small positive gap exhibited by hotels.

Weekday and weekend periods

Table 2 reports the findings. The first column (name) identifies the hypothesis tested; the second (series) focuses on the seasonal period; and the third (mutual information) estimates the similarity or synchronization (dissimilarity or desynchronization) between the two series (Airbnb and hotels). The higher the value of the mutual information score, the higher the similarity and vice versa. A ratio of 0.20 identifies a good similarity (or synchronization), while a value lower than 0.10 depicts a strong dissimilarity or desynchronization (Latham & Roudi, 2009). The following columns are based on the comparison between the two lodgings series and the 500 random series. The fourth column (Rnd_mean) measures the random mean, which represents the average distance between the two series and the 500 randomized series, while the fifth (Rnd_sig) estimates whether the value is significant or not. The penultimate column depicts the t-value. The higher the absolute value, the more dissimilar the two lodgings series are compared to the 500 randomized series, while the p-value measures the probability of the t-value.

<table>
<thead>
<tr>
<th>Name</th>
<th>Series</th>
<th>Mutual information</th>
<th>Rnd_mean</th>
<th>Rnd_sig</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.A</td>
<td>midweek</td>
<td>0.0761</td>
<td>0.0034</td>
<td>0.0020</td>
<td>-809.95</td>
<td>≈0</td>
</tr>
<tr>
<td>1.B</td>
<td>weekend</td>
<td>0.2269</td>
<td>0.0088</td>
<td>0.0046</td>
<td>-1061.25</td>
<td>≈0</td>
</tr>
<tr>
<td>2.A</td>
<td>workdays</td>
<td>0.0896</td>
<td>0.0040</td>
<td>0.0023</td>
<td>-820.45</td>
<td>≈0</td>
</tr>
<tr>
<td>2.B</td>
<td>holidays</td>
<td>0.2182</td>
<td>0.0058</td>
<td>0.0029</td>
<td>-1624.81</td>
<td>≈0</td>
</tr>
<tr>
<td>3.</td>
<td>fair</td>
<td>0.0738</td>
<td>0.0313</td>
<td>0.0158</td>
<td>-59.95</td>
<td>≈0</td>
</tr>
<tr>
<td>4.</td>
<td>all</td>
<td>0.0845</td>
<td>0.0023</td>
<td>0.0013</td>
<td>-1428.31</td>
<td>≈0</td>
</tr>
</tbody>
</table>

Considering the similar results reported in all of the rows on Table 2, in relation to the last four columns on the right, a preliminary explanation is given that is valid for all of the cases. The t-value always accounts for a wide negative value (high in relation to the mutual information and the random
mean) and the p-value is always close to zero. Therefore, the null hypothesis (the two lodging series are similar to the 500 randomized ones) is rejected. Based on this result, it is possible to focus attention directly on the mutual information score and, according to its value, evaluate the synchronization (if the value is high) or desynchronization (if the value is small) between the two lodgings series.

This first section focuses on the first hypothesis, which supposes strong differences during weekdays (1.A) and similarities during the weekend (1.B) days between hotels and Airbnb listings. During weekdays, the mutual information accounts for a value of 0.076, which clearly indicates a strong dissimilarity within the two seasonal series. The value confirms hypothesis 1.A: during weekdays, hotels and Airbnb listings show profound differences in their time series or, put differently, the two supply segments are desynchronized. A reasonable explanation is the different market segment focus of these two lodgings types. During weekdays, hotels attract more business and trade-fair guests, while Airbnb listings work prevalently with leisure clients. For this reason, despite having a quite similar standardized occupancy (see Figure 2, hypothesis 1), the two series illustrate a different structure (desynchronization).

The second hypothesis (1.B) moves from weekdays to weekend days and supposes more similar seasonal patterns. The mutual information accounts for a considerably higher value (0.226), which is supportive. Based on this finding, hypothesis 1.B is confirmed: during weekends, hotels and Airbnb listings show a partial synchronization in their time series. The explanation of this finding suggests that the two lodgings attract the same market segment (leisure clients). This mutual information is sufficiently high to indicate possible substitution threats. By contrast, the value is not large enough to show full competition. This is reasonable considering the profound differences between the two types of accommodation.

**Holidays and working days**

This section tests the second hypothesis articulated around holidays (2.A) and working (2.B) days. The holiday period accounts for 584 days (Table 1) and is considerably longer than the weekend period, which embraces 418 days. The holiday, in fact, includes the Easter and Christmas periods, as well as the entire month of August and other civil and religious holidays. Approximately 40% of all days included in the sample (1,461, Table 1) are classified as holidays. The remaining 60% are workdays (877).

The working period accounts for the highest occupancy for hotels (0.744) compared to both holidays (0.612) and midweek days (0.703). By contrast, for Airbnb listings, the holiday occupancy (0.236) is a little lower than at weekends (0.253). This lower value is mainly influenced by the fall in the occupancy rate during the month of August (evidence not reported). In fact, this period, despite being a non-working season, is a very hot month and, therefore, is not a very favorable period during which to visit Milan.

Moving from the occupancy rates reported in Table 1 to the mutual information score illustrated in Table 2, hypotheses 2.A (workdays) and 2.B (holidays) are tested. During workdays, the two series are strongly desynchronized (0.089). The result confirms hypothesis 2.A: during workdays, hotels and Airbnb listings show profound differences in their time series; the two supply segments are desynchronized. The explanation is in line with those previously reported for weekdays. In fact,
during workdays the hotels focus prevalently on business and trade-fair clients, while Airbnb listings attract leisure guests. The differences in the market segments generate a desynchronization in the two lodging series.

When focusing on holidays, the opposite occurs. In fact, the mutual information score peaks to 0.218 and suggests a fairly similar structure. This value confirms hypothesis 2.B: during holidays, hotels and Airbnb listings show a partial synchronization in their time series. As previously suggested, during holidays, the two lodging supplies attract the same market segment (leisure travelers) and, therefore, the two series are partially overlapped.

Trade-fair events

This section focuses on the third hypothesis and analyzes the effect generated by some trade-fair events. This analysis includes top trade-fair events that cover 122 days (Table 1); this is the shorter seasonal period. The mutual information score registers the lowest value reported in Table 2 (0.073), which clearly confirms the third hypothesis: during trade-fair events, hotels and Airbnb listings show profound differences in their time series; the two supply segments are desynchronized. Apparently, this result appears counter-intuitive, analyzing the standardized occupancy (Figure 2, hypothesis 3). In fact, the graph shows a very similar standardized rate for both hotels and Airbnb listings. However, the mutual information reveals a clear different structural pattern within the two series. This finding appears reasonable: during the top trade-fair events, the Milan hotels are fully occupied by trade-fair attendants, while Airbnb listings attract (as usual) leisure clients. Therefore, despite having a similar standardized performance value, the underlying patterns are widely different.

Annual values

This last section compares the two series without considering any temporal segmentation. The number of days (four years) is equal to 1,461 (Table 1). The occupancy gaps between the two supplies (Airbnb and hotels) are wide. In fact, in the case of Airbnb listings, the mean is 0.22, while hotels more than triple this percentage (0.69). The mutual information accounts for a lower value (0.084), which is in line with the amount achieved by hypothesis 1.A (0.076), 2.A (0.089), and 3 (0.073). Based on this result, the fourth hypothesis is confirmed: considering all of the days (without any temporal segmentation), hotels and Airbnb listings show profound differences in their time series. This last result is in line with the Milan context, prevalently focused on business and trade-fair clients. The workdays (during which business are the main target for hotels) accounts for 877 days (60%), while there are 122 trade-fair days (8%). During these two periods, the two lodging supplies show profound differences (as previously explained). Not surprisingly, the annual mutual information reveals a general desynchronization between the two series.

CONCLUSIONS

The conclusions are articulated in threefold: some theoretical, as well as practical, implications are traced and some study limitations are identified.
Theoretical implications

As discussed in the introduction and in the literature review, the debate about the substitution threat generated by Airbnb listings on traditional hotels is an open question, and the conclusions are contradictory, moving from a strong substitution to the absence of any overlapping (Blal, Singal & Templin, 2018; Guttentag, 2015; Heo, Blal & Choi, 2019; Oskam & Boswijk, 2016; Xie & Kwok, 2017; Zervas, Proserpio & Byers, 2017). Other studies limit the competition to budget and low-end hotels and motels (Guttentag & Smith, 2017; Hajibaba & Dolnicar, 2017).

This study sheds new light on this debate. The substitution threat is strongly influenced by the market segments attracted by the destination under investigation or, put differently, is seasonally bound. The evidence from Milan suggests that Airbnb listings are more competitive during weekends and holiday periods, when the attracted market segment clearly focuses on leisure clients. In particular, during these two periods, the synchronization between the two lodgings series is relatively high and can generate some competitive threats.

By contrast, during weekdays and working days, Airbnb listings show desynchronization, suggesting, in the case of Milan, the absence of any reasonable substitution threat between the two offers. This result can be extended to the year time series (hypothesis four).

These findings confirm some qualitative assumptions that were stated in previous studies: Airbnb is strongly specialized in the leisure segment, while peer-to-peer listings are less attractive for business market segments. Based on this evidence, the higher the destination specialization on leisure segments, the higher the potential substitution threat. The opposite occurs when the destination is prevalently specialized (as in the case of Milan) on business and trade-fair market segments. Therefore, future studies that explore the substitution threat, must take in account the destination specialization in terms of market segments and underlying seasonal patterns.

This paper introduces important innovations to analyze the substitution threat and, more generally, the overlap between hotels and Airbnb listings. The first innovation is to clearly identify the main destination market segments (in the case of Milan, leisure, business, and trade-fair guests) and the corresponding seasonal patterns. The second methodological innovation is the use of mutual information to perceive and measure the degree of synchronization between the series. This approach can open new research opportunities in other destination contexts. We note here that the methodology, relatively simple from a conceptual and computational point of view, is of general utility and can thus be used in all situations in which a possible correlation between two phenomena needs to be investigated.

Practical implications

The practical implications of this paper can be articulated at two levels: on a managerial level (for owners, managers and entrepreneurs of hotels and Airbnb listings) and on a destination level (for destination marketing and management organizations). At a managerial level, the findings clearly suggest the absence of substitution threats between the two lodging supplies in the case of business and related segments (such as meetings, incentives, congress, and exhibitions). Therefore, hotels can
manage their performance metrics accordingly (occupancy and rates). By contrast, the peer-to-peer listings show a good synchronization with hotels in the case of leisure segments, creating a possible competitive threat. This result can have some important consequences on leisure-oriented destinations. However, as explained in the study limitations, to measure the substitution threat, the two lodgings types must be varied in their segmentation (considering location, category, size, rating scores, experience, etc.). This study focuses only on the general level and clearly suggest the presence of synchronization between the two options in the case of leisure guests. The potential substitution threat is limited to weekend days and holidays; however, these two seasonal periods cover a considerable number of days (respectively 29% and 40%). Therefore, the potential overlap is not marginal in a business context, such as in the city of Milan. The hotel industry can react to this potential substitution threat generated by Airbnb listings using a wide array of tactics and strategic tools. Some examples of such tools, based on previously published papers, are the topic of the following discussion. First, the two lodging suppliers (Airbnb listings and hotels) are considerably different in terms of services and attributes delivered (Chen & Xie, 2017; Wang & Nicolau, 2017; Xie & Kwok, 2017; Young, Corsun & Xie, 2017). Therefore, the hotels should segment the market, focusing on the service-sensitive targets on one side, and the less price-sensitive travelers on the other side. Second, many Airbnb listings are offered by micro-entrepreneurs less skilled in using sophisticated pricing strategies, such as revenue management, dynamic prices, and minimum length of stay (Gibbs et al., 2018a, 2018b). Therefore, hotels are often more experienced at extracting value (revenue) during the high season than Airbnb hosts – with the partial exception of professional hosts (Kwok & Xie, 2019; Xie, Kwok & Heo, 2019; Wegmann & Jiao, 2017). Finally, the security level assured by hotels is usually described as higher than that offered by Airbnb listings. This gap can be another powerful segmentation criterion, able to attract more risk-averse guests to traditional hotels. These examples clearly demonstrate the wide range of strategic and tactical policies that hotels can use to differentiate themselves from Airbnb listings.

At a destination level, the findings can orient destination marketing and management organizations (Bornhorst, Ritchie & Sheehan, 2010; Sainaghi, 2006). In fact, as in the case analyzed, the development of Airbnb listings can support a partial re-equilibration in the destination mix. The rise of Airbnb listings can support the increase in leisure clients, as per the Milan destination following the Expo events (Guizzardi, Mariani & Prayag, 2017; Sainaghi & Mauri, 2018; Sainaghi, Mauri & d’Angella, 2018; Sainaghi et al., 2019). However, this increase can also have negative aspects, especially in mass destination, which accounts for a critical carrying capacity score. In these contexts, the overlap between the two lodging segments increases (both supplies attract leisure guests) and can trigger a price reduction or favor a cheap price specialization of a relevant percentage of Airbnb listings, attracting new price-sensitive clients and reducing the destination’s sustainability.

Limitations

This is an explorative paper and in line with similar previous studies focused on competitive threats (Heo, Blal & Choi, 2019; Zervas, Proserpio & and Byers, 2017), it is centered around a single case study. The findings’ generalization is partially limited. However, the paper adopts a longitudinal approach and considers 1,461 observations (days), creating a consistent temporal pattern. A future research agenda is proposed to verify whether, within a multi-destination study, the evidence reported
can be confirmed. In particular, the new sample could include some strongly leisure-oriented destinations, such as the cities of Florence and Venice. A second limitation is related to the difficulties in separating business and similar market segments (meeting, incentive, congress, and exhibition) on one side, and, on the other, leisure clients in other destination contexts. In fact, in the case of Milan, leisure travelers focus prevalently on weekends and holidays, as confirmed in many previous studies (De Carlo et al., 2009; Guizzardi, Mariani & Prayag, 2017; Sainaghi & Baggio, 2011; Sainaghi & Canali, 2009, 2011; Sainaghi & Mauri, 2018; Sainaghi, Mauri & d’Angella, 2018; Sainaghi et al., 2019). In relation to other destinations, the leisure segment could be more mixed for business guests, resulting in a more difficult separation of the two. However, as previously explained, the mutual information score is very useful for solving this problem. In fact, as shown in the case of Milan, despite showing a similar standardized occupancy to hotels, during weekdays and workdays, Airbnb listings depict a clear desynchronization, suggesting that the two lodging supplies attract different market segments. This study has analyzed the two lodgings types (without segmenting) based on relevant variables, such as location, category, size, rating scores, and experience. Another limitation concerns the two data sources (STR and AirDNA) used in this study. As analytically explained by Agarwal, Koch and McNab (2019), the metrics calculated by AirDNA – including occupancy, average daily rate and revenue per available room (or night) – do not perfectly adhere to STR definitions for these variables. Furthermore, AirDNA metrics tend to exhibit a notable upward. Finally, in this study we have only considered occupancy, ignoring possible seasonal patterns (such as weekends, holidays, and trade-fair events) where the number of available listings could be expected to rise, and prices to fluctuate. These additional metrics could improve the analysis. In fact, the higher occupancy during peak times can be reinforced by higher rates and a higher number of available listings. These limitations can open up a future research agenda, to show whether these variables can illustrate more precise substitution threats.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their valuable comments and STR and AirDNA for the helpful collaboration. R.B. acknowledges the financial support of the Ministry of Education and Science of the Russian Federation in the framework of the Competitiveness Enhancement Program of the Tomsk Polytechnic University.

REFERENCES


