

# The impact of review sentiments on occupancy: Evidence for signalling theory from peer-to-peer accommodation

Journal of Vacation Marketing

1–15

© The Author(s) 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/13567667231201406

[journals.sagepub.com/home/jvm](https://journals.sagepub.com/home/jvm)**Sunil Sahadev** 

University of Brighton, UK

**Andreas Seiler**

Aston University, UK

**Phil Scarf**

University of Cardiff, UK

## Abstract

Using data from 16,144 peer-to-peer properties in London, we study the impact of the two components of user-generated content – rating and sentiment – on occupancy rate. Our methodology is innovative because, firstly, we control for price variation when estimating these review-occupancy effects and secondly, we estimate interaction and curvilinear effects. We find that sentiment and rating have significant positive effects on occupancy rate; there is some evidence that sentiment and rating interact, one reinforcing the other; for a typical property among those analysed, an outstanding review increases occupancy by a fifth in relative terms. Thus, we interpret these associations as evidence that rating and sentiment signal value, and we estimate the strength of the signal in the peer-to-peer accommodation sector.

## Keywords

Accommodation, peer-to-peer, occupancy rate, rating, sentiment, signalling theory

## Introduction

The peer-to-peer accommodation sector has expanded rapidly over the past decade. According to Dolinica (2019, p. 248), ‘paid online peer-to-peer accommodation platforms facilitate the interaction of non-commercial providers of space(hosts) with end-users (guests)’. Impressive growth of platforms like Airbnb, Roomorama, HomeExchange, etc. have transformed patterns in accommodation booking, travel and hospitality consumption behaviour. According to the latest available World Bank report, in 2018, total peer-to-peer accommodation was approximately 7% of the total accommodation globally, that is about 8 million beds (World Bank, 2018). Over a very short time, Airbnb, the largest of the peer-to-peer accommodation sector operators,

has become the fourth-largest accommodation business (by rooms booked per night) with a market potential of \$3.4 trillion (Moore, 2020). According to Volgger et al. (2019), the consistent rise in the market share for Airbnb in the sector could be attributed to unique consumption dynamics at the macro-level like the ‘winner-takes-all’ or ‘superstar effects’.

There has been much research interest to explore the determinants of success in the peer-to-peer accommodation sector. For recent

## Corresponding author:

Sunil Sahadev, Sheffield Business School College of Business, Technology and Engineering, Sheffield Hallam University, UK.  
Email: [s.sahadev@shef.ac.uk](mailto:s.sahadev@shef.ac.uk)

reviews see Belarmino and Koh (2020) and Sainaghi et al. (2021). Conveying reliable signals about the quality and value of the accommodation is a vital component of business models in the sector (Xie and Mao, 2017; Fu et al., 2020). The influence of e-word-of-mouth on the performance of hotels is established in the existing literature (Yang et al., 2018). A hotel chain with an established brand name will have a variety of avenues to communicate its quality to a potential customer like mass media campaigns, third-party referrals, etc. rather than just through user-generated content. However, in the case of hosts in the peer-to-peer sector, the reliance on user-generated content is almost total as other brand-related communication platforms are unavailable. Thus, the importance of the signals provided by previous customers may be critical because a peer-to-peer accommodation provider (host) is unlikely to possess the credible brand image of a typical hotel chain. As Murillo et al. (2017) contend, the rapid expansion of peer-to-peer accommodation sector relies entirely on the success of the ‘commodification of trust’. Trust is built through online communities, where users rely on the content generated by other users to gauge the trustworthiness of hosts and properties. The role of user-generated content in the sector is therefore even more critical than in the case of branded hotel chains. It is thus very important to explore strength of one of the most important components of user-generated content – sentiment expressed through user-generated content on the occupancy rates of peer-to-peer accommodation providers.

Typically, user-generated content has two forms: (a) the review score – a number that quantifies user-opinions about a property and (b) textual reviews from users that provide rich, qualitative information about the attributes of a property (Li et al., 2019). One of the most important features of the textual review is the sentiment conveyed in these reviews (Park et al., 2020), and the hospitality sector is a rich field for conducting sentiment analysis (Mehraliyev et al., 2022).

Studies that have tried to evaluate the link between the characteristics of user-generated content and property outcomes like occupancy rate in the peer-to-peer accommodation sector are rare. While there are several studies on reviewer ratings and user-generated content itself (please see a recent review by Zheng et al. (2023), these articles do not empirically link reviewer rating or review sentiment with occupancy rates. A recent article published by

Zhang et al. (2021, p.1470) clearly explains this gap in literature through the following statement: ‘most existing work evaluate UGC effects on purchase intention/preference rather than more concrete outcomes such as actual arrivals’. In one of these rare studies, Van der Borg et al. (2017) showed that the rating (numerical value) of a property positively impacts its occupancy rate. However, this study does not consider the impact of qualitative sentiments on occupancy rates.

In the non-peer-to-peer sector, several studies have considered the impact of online reviews on factors such as revenue performance and occupancy rate (e.g., Duverger, 2013; Viglia et al. (2016); Yang et al., 2018). While many of these studies look at the impact of rating on property performance, very few studies have explored the impact of textual reviews. This is despite findings from studies in consumer psychology that show textual reviews are very helpful for consumer decision-making (e.g., Agnihotri and Bhattacharya, 2016; Chua and Banerjee, 2016; Lee et al., 2018).

In our study, we explore the effects of rating and a sentiment score derived from textual reviews upon occupancy rate in the peer-to-peer accommodation sector. Using signalling theory (Spence, 1973) as the theoretical lens, we suggest that both these forms of user-generated feedback signal the quality of a property, and thereby aid consumer decision-making and influence occupancy rate. We do this using a large sample of properties in London. We also investigate the interaction between rating and sentiment and curvilinear effects.

Our study contributes to the extant literature in three ways. Firstly, by exploring the impact of review sentiment on occupancy rate in peer-to-peer accommodation, we extend the extant knowledge on the determinants of occupancy rate in this sector. As mentioned in the previous section, since peer-to-peer accommodation sector have a higher reliance on user-generated content to convey quality to customers, our study is significant in understanding the factors that impact occupancy rates in this sector. Secondly, we contribute to the emerging literature on the nature of the impact of online user-generated content. That is, we are able to separate the effects of rating and sentiment on occupancy rate and thereby provide a more nuanced perspective on how user-generated content impacts occupancy rate in the accommodation sector. This is quite important since most of the extant studies

have used only one of the components of user-generated content to understand their impact on outcome variables. Also, through exploring the interaction effects of the two components of UGC, we develop our understanding of the mechanisms through which the two components impact outcomes. Few studies have explored the interactions between the two components of UGC. Further, most of the extant studies have used proxy variables to represent outcome variables like sales rank rather than a real outcome variable like occupancy – thereby limiting the validity of the results. By using a real outcome variable, this limitation is reduced in the current study.

Thirdly, recent studies have pointed to the positive bias inherent in user-generated content, especially provided by customers in the context of P2P accommodation sector. This positive bias has led the researchers to question the true impact of user-generated content on customer behaviour. In this study by using actual value of review ratings, consumer sentiment and occupancy rate at the property level, we are able to exhibit the extent of impact of user-generated content on customer decision-making. The results from this study therefore go towards clarifying the true extent of the impact of user-generated content in P2P settings.

In the remaining sections, we review the relevant literature, develop hypotheses, explain the methodology adopted in the study, and then present our results. We conclude by discussing the theoretical and practical implications that emerge from our study.

### **Online consumer reviews as signals of quality**

Signalling theory, developed by Spence (1973), explains how the problem of sub-optimal transactions due to information asymmetry is addressed through the exchange of a variety of signals. Signalling theory has been used in a variety of domains including human resource management, e-commerce, international trade, and entrepreneurship (Chen et al., 2020). The peer-to-peer accommodation sector is unique in terms of the effect of information asymmetry on consumption choice. This is because the sector is largely unregulated, with loosely enforced quality standards leading to higher levels of perceived risk for consumers compared to accommodation available through a hotel chain or a hotel brand (Huang et al., 2019). Thus, as Ye et al. (2019)

explain, peer-to-peer accommodation, unlike conventional hotels, cannot rely on traditional risk-mitigation measures like brand or reputation.

Customers typically perceive significant risks while booking accommodation through peer-to-peer portals. According to Huang et al. (2019) perceived risks have a high impact on the intention to book accommodation in the P2P sector. Potential customers face significant levels of psychological risks, physical risks, performance risks and social risks in the context of P2P accommodation. As information asymmetry is considered a significant contributor to perceived risk in e-commerce (Glover and Benbasat, 2010), signals that can reduce information asymmetry will greatly enhance trust and consumption in the peer-to-peer sector (Xie and Mao, 2017). Yao et al. (2019) in fact argue that information asymmetry between hosts and guests is higher in the case of peer-to-peer platforms like Airbnb than in the case of hotels. This is because guests have to go through a complicated process to sort out trustable signals from the information uploaded by hosts on Airbnb platform. Liang et al. (2018) have found that e-word-of-mouth or user-generated content can significantly reduce the perceived risk of consumers in peer-to-peer accommodation settings. Online reviews by consumers, including both the overall rating and the sentiment expressed through the qualitative opinions about the property, are expected to be very important and strong signals of the quality of a property.

Connelly et al. (2011) compared the utility of a signal across several dimensions. In the context of user-generated online reviews in peer-to-peer accommodation, based on the framework of Connelly et al. (2011), the dimensions that are applicable are (a) signal observability (related to the visibility and strength of the signal) and (b) signal reliability (related to the veracity and validity of the signal). User-generated online reviews are regarded as high in both observability and reliability. As online reviews are easily accessible to potential consumers and in fact peer-to-peer accommodation portals attempt to provide greater amplification for the display of user-generated online reviews, visibility is assured. Signal reliability is also assured because a potential consumer knows that the reviews are largely provided by users who have stayed in the property.

Cue-utilisation theory (Olson and Jacoby, 1972) is considered as a sub-theory of signalling

theory (Yang et al., 2020). According to this theory, customers tend to use cues with high confidence value more when they are faced with uncertainty. The confidence value of a cue is defined in terms of a consumer's perception about the confidence with which they can perceive and judge a cue. Rating and sentiment reflected in textual, user-generated reviews are easy for consumers to understand. It is therefore expected that consumers are confident to use these cues and act accordingly. As Dickinger (2011) found, online reviews posted by other travellers are perceived to be more up-to-date, more informative, more accessible and more reliable than information from travel service providers, and thus user-generated reviews are the more heavily utilised cues. Thus, in our study, we posit that user-generated rating and the sentiment derived from textual reviews are cues that consumers use to reduce risk when booking rooms in the peer-to-peer accommodation sector. Therefore, we would expect that the propensity to book a property is influenced by the signal provided by these two components of user-generated reviews.

### **Online consumer reviews and occupancy rate**

User-generated content in the accommodation sector comprises of two components: (a) an overall rating of the property or the review score – which is a quantitative score and (b) the qualitative feedback about the property. We develop separate hypotheses for the two components of user-generated content.

The impact of overall rating or the review score on sales has been established across several studies in the hospitality sector. Yang et al.'s (2018) meta-analysis lists 25 studies that looked at the impact of online review on hotel performance (not all of them were occupancy rates), of which 23 studies showed a significant positive relationship. However, two issues are noteworthy: (a) these studies were not in the peer-to-peer accommodation sector and (b) except for the studies by Xie et al. (2016) and Viglia et al. (2016), occupancy rate is not used the dependent variable. Most of these studies use a proxy variable – the revenue performance of the hotel – as the dependent variable, which is calculated as revenue per available room. While most of these studies show that the quantitative online review scores have a direct positive impact on hotel performance, some studies have

shown a weak relationship (e.g., Lu et al., 2014). To date only two studies (Van der Borg et al., 2017; Leoni et al., 2020) have considered the impact of overall rating scores of properties on occupancy rate in the peer-to-peer accommodation sector. The positive impact on occupancy rates shows how rating works as a credible signal for customers in their decision-making process. Based on the previous studies, we therefore hypothesise:

H1: the overall quantitative rating scores from user-generated reviews positively impacts the occupancy rate of a peer-to-peer property.

Another important signal derived from user-generated content is the sentiment conveyed in textual reviews. As Zhang et al. (2021) explain, online reviews written by the customers are considered to be more trustworthy than the inputs provided by the hosts. Further, as Zhu et al. (2020) explain sentiments expressed by customers through their qualitative review capture their consumption emotions and thus often complement the review ratings in P2P accommodation setting. Previous studies have shown that sentiments conveyed by user-generated content can act as a potent signal for consumers in their decision-making process (Li et al., 2019; Hu et al., 2014). When choosing accommodation, most of the attributes associated with a property are experience attributes and hence are better assessed through subjective, experience-based opinion. The qualitative comments provided by reviewers are often able to better capture such experience-based attributes than numerical scores (Hu et al., 2014). As Yin et al. (2014) explain, sentiments conveyed by the qualitative textual reviews serve as unique cognitive appraisals from previous customers that offer useful information cues for processing by potential customers. A positive sentiment score is therefore associated with the positive quality of the product/service thereby motivating potential customers to choose the particular product/service.

Several studies in the past have found positive relationships between positive sentiments from user-generated reviews and product sales. For instance, Hu et al. (2014) found a direct relationship between sentiment score and sales of books in Amazon.com, Li et al. (2019) with the sales for tablet computers, Chong et al. (2016) in the context of sales for electronic goods; Eslami and Ghasemaghaei (2018) in the sales of

musical instruments; Wu et al. (2018) in the sales of cameras, etc. It should also be noted that none of these studies used actual sales data. Instead, a proxy – sales rank – was used. In fact, in one of the rare studies which used actual sales data – in the context of movie ticket sales – Lee et al. (2017) could not find a direct relationship between reviewer sentiment score and sales. Interestingly, previous literature on peer-to-peer accommodation has not considered the impact of sentiments conveyed by textual reviews on sales-related variables like occupancy rates of properties, though Lawani et al. (2019) found, in a study conducted in Boston, that the strength of the positive sentiment score expressed by Airbnb consumers impacts the price of properties. Based on the above arguments, it is therefore hypothesised that:

H2: Strength of the sentiment expressed through user-generated qualitative reviews positively impact occupancy rates in peer-to-peer accommodation.

While it is argued that the numerical rating and sentiment score of the qualitative reviews both separately influence consumer decision-making, it is also possible that they show interaction effects in influencing sales of the products. This argument can be developed based on the perspective of consistency of signals (Erdem and Swait, 1998; Connelly et al., 2011). For instance, when the rating is high, but the sentiment conveyed by the qualitative review is not positive, the consumer may get a confusing message, thus reducing the impact of the rating on purchase intentions. Instead, when the sentiment derived from the qualitative review is strongly positive, it will reinforce the positive impact of the rating on consumer's decision to select a property. Few studies have explored the interaction effects of these two variables in the context of sales of products or services, let alone in the case of peer-to-peer accommodation. Thus, our third hypothesis is:

H3: Rating and sentiment expressed through user-generated reviews interact to impact positively on occupancy rate in peer-to-peer accommodation.

Previous studies have reported curvilinear effects of online consumer reviews (Hernandez-Ortega, 2020). Based on cue-utilisation theory (Easterbrook, 1959), consumers develop their views about the quality of an offer through the

analysis of the extrinsic and intrinsic cues available to them. Extrinsic cues are product or offer related but are not part of the physical product (Shirai, 2020). User-generated content about a property, both the quantitative review rating and the sentiment conveyed through the qualitative reviews can be considered as extrinsic cues in this context. It is generally believed that consumers will rely significantly on extrinsic cues to infer product quality when intrinsic cues are not available (Sabri et al., 2020) as is the case of P2P properties. Easterbrook (1959, p.193), explained the idea of 'point of optimum proficiency' on the continuum of cue utilisation. According to this concept, there is an optimum point at which cues achieve maximum efficiency after which increase in the intensity of cues actually reduces its performance impact. Based on this idea it can be argued that, as attention to online review-based cues increase, initially, its efficacy also increases as other peripheral cues will be ignored, but as the attention level increases beyond a point, even important elements of review-based cues will start getting excluded as all the peripheral cues have already been excluded. The salience of review-based cues will thus start getting effected. Thus, we hypothesise:

H4: Rating has a curvilinear effect on peer-to-peer accommodation occupancy rate such that once rating has increased beyond a point its impact on the occupancy rate diminishes.

H5: Sentiment has a curvilinear effect on peer-to-peer accommodation occupancy rate such that once sentiment has increased beyond a point its impact on the occupancy rate diminishes.

## Quantitative methodology

### *Price-variance as a control variable*

Price is considered as one of the most important antecedents of occupancy rate in the hospitality sector. In the extant literature, several studies have used mean price as a determinant of occupancy rate (e.g., Gunter and Onder, 2018; Leoni et al., 2020; Van der Borg et al., 2017). However, previous studies have found that the rental price charged by a property is to a very large extent determined by the price charged by other hosts of similar properties (by size and type) in the same location. For a brief review, see Chica-Olmo et al. (2020). Landlords therefore are expected to have very little latitude in pricing

their properties. Thus, to meaningfully trace the impact of price on occupancy rate, it is important to see how the price of the property relative to its neighbouring properties influences its occupancy rate. For instance, Leoni et al. (2020) in their model use the difference between the mean price charged by nearby properties and the price of the particular property as an independent variable in their model.

We take a similar but more nuanced approach to control for the impact of price in the relationship between online reviews and occupancy rate. We first determine, using a regression model, the price a property would expect to charge based on its neighbourhood, size, and type (whether an entire home or private room in a home). We then calculate the difference between the actual price and this model-derived expected price. We call this difference the 'price variance', and use *price variance* as a control variable in our model of occupancy rate. In this way, the price variance signals the monetary value of a property. A negative *price-variance* indicates that a property is cheaper than one would expect given its size and neighbourhood, etc. Using this value-signal in our occupancy-rate model we can investigate the effect of review attributes on occupancy rate while controlling for monetary value (*price variance*).

### Other control variables

As is typical of the previous studies we have cited, we further use the following control variables: (a) the number of online reviews for the property (*n reviews*), (b) the number of photos of the property in the online interface (*n photos*), (c) facility available (or not) for instant booking in the online interface (*instant book*), (d) the minimum number of nights of stay (*min stay*), (e) the mean response time of the host (*response time*), and (f) the type of the property (*property type*).

### Regression modelling

As indicated above, to control for the effect of property price on occupancy rate, two linear models were fitted. In the first, the *landlord* model, *price* (mean-daily-rate), the response variable or independent variable, is explained by the explanatory variables (a) *neighbourhood mean price*, (b) *property type*, (c) *max guests*, (d) *cancellation* and (e) the interaction of *property type* and *max guests*.

In the second model, the *visitor* model, *occupancy rate* (demand) is the dependent variable and control variables are: (a) *n reviews*, (b) *n photos*, (c) *instant book*, (d) *min stay*, (e) *response time*, (f) *price variance* and (g) *property type*, and explanatory variables are (a) *sentiment* and (b) *rating* with linear terms, quadratic terms, and interaction depending on the model.

*Price variance* is defined as  $price\ variance = price_i - E(price_i)$ , where  $price_i = K + \beta_1 * (mean\ price\ of\ neighbourhood\ properties)_i + \beta_2 * (listing\ type)_i + \beta_3 * (maximum\ number\ of\ guests)_i + \beta_4 * (cancellation\ policy)_i + \beta_5 * (listing\ type)_i * (maximum\ number\ of\ guests)_i + \epsilon_i$ .

In the dataset, it was seen that many variables are highly skewed, so transformation towards multivariate normality is desirable. We use the Yeo–Johnson transformation (Yeo and Johnson, 2000). This is a generalisation of the log transformation in which input variables are not required to be strictly positive. The transformation is optimised using a function in R, part of the *recipes* package (<https://cran.r-project.org/web/packages/recipes/recipes.pdf>). The log transformation is applied to *price* (since *price* is strictly positive). Occupancy rate is not transformed because it is not skewed.

Further, to control for multicollinearity, instead of an OLS methodology, we used the lasso (least absolute shrinkage and selection operator) regression methodology. Lasso regression produces simple, sparse models and significantly reduces multicollinearity (Tibshirani, 1996).

### Data

We used data provided by the Consumer Data Research Centre on Airbnb rentals in London for the period 1 June 2017 to 31 May 2018. These data are a subset of the data collected by AirDNA LLC, which tracks the daily performance of over 4.5 million properties in 60,000 markets worldwide. The Consumer Data Research Centre data are held in two separate databases. One holds textual review data, the other holds all other data (price, occupancy, rating, property description, etc.). The two databases can be linked using the unique identity of a property. Notionally, the dataset lists 207,117 London properties. However, occupancy data is missing for approximately half of these (100,135). Furthermore, very many properties with occupancy data have no review data (60,068). This is because at least one of the following is true: a property has zero occupancy; a property has no reviews; review

data for a property is missing. A property may have no reviews because either it has not been rented (zero occupancy) or no visitor (renter) has submitted a review. A property has missing review data because there is no matching property (with the same identity) on the review database.

The databases were linked using the R package dplyr 0.8.3. The function ‘inner\_join()’ returns all rows from database 1 where there are matching values in database 2. In the case of multiple matches, inner\_join() returns all combination of the matches, so that our data preprocessing handled the circumstance in which review data are aggregated for each property.

We also excluded exotic property types (e.g., shared rooms), of which there are few instances (115 in total), and focused on entire homes and private rooms in a home. Thus, a property was included in our analysis only if its occupancy rate was not missing, it had at least one review, and it was either an *entire home* or a *private room in a home*. This resulted in a dataset of  $n = 16,114$  properties (8874 entire homes and 7240 private rooms) across 33 London neighbourhoods (Table 1) with occupancy-and-review data for the *period* 1 June 2017 to 31 May 2018.

### Definition of variables

The variables considered in this study are shown in Table 2. Thus, the dataset we used contained

**Table 1.** Number of properties in each neighbourhood.

Neighbourhood	Count	Neighbourhood	Count
Westminster	1581	Greenwich	244
Hackney	1548	Barnet	233
Tower Hamlets	1429	Merton	211
Camden	1214	Hounslow	186
Islington	1131	Croydon	106
Lambeth	1112	Redbridge	100
Kensington and Chelsea	1109	Bromley	98
Southwark	1102	City of London	88
Hammersmith and Fulham	880	Kingston upon Thames	88
Wandsworth	742	Enfield	81
Brent	515	Harrow	63
Haringey	467	Hillingdon	52
Lewisham	420	Sutton	39
Newham	319	Bexley	29
Richmond upon Thames	309	Barking and Dagenham	20
Ealing	301	Havering	18
Waltham Forest	279		

the value of each of these variables for each property (the unit of analysis). The descriptions of the variables are provided in the Consumer Data Research Centre dataset, except for those that we derived from the data: *neighbourhood mean price* and *sentiment*.

We measured sentiment by analysing the text of reviews using the ‘bing’ lexicon (<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>). This lexicon was first used in Hu and Liu (2004). It partitions the set of all words into three subsets: positive words; negative words, and uninformative words. For each property, we counted the number of positive words and the number of negative words in the total review text over the period, and calculated three sentiment measures: the total number of positive words divided by the total number of reviews of the property over the period (positive sentiment); the same but for negative words (negative sentiment); their difference, which we call *sentiment*. Thus, *sentiment* is the mean sentiment score per review. Properties with *sentiment* above 40 were classed as outliers and excluded. There were only three excluded properties. The next highest *sentiment* was 30.

### Results

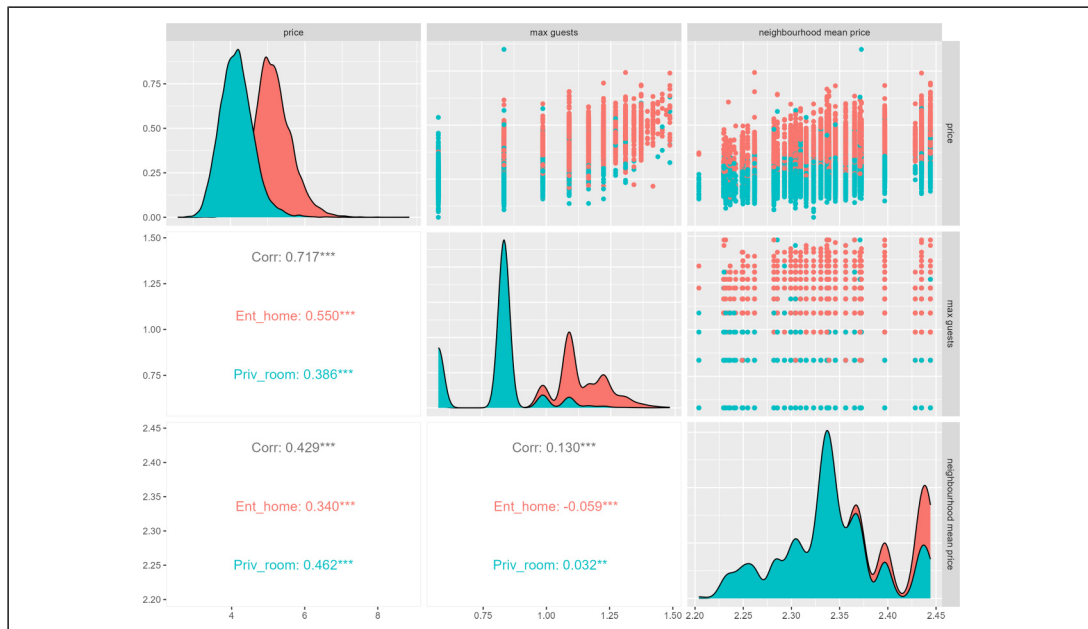
Matrix scatterplots of all variables in the *landlord* model (Figure 1) and all variables in the *visitor* model (Figure 2) are shown. These plots provide visual checks on collinearity, effectiveness of transformation, and outliers and influential observations. These are important to ensure validity of the regression models. The plots also provide confirmation (or otherwise) of associations between variables.

Results from the lasso regression models are presented in Tables 3 and 4. Therein, we apply a tune-grid to determine the amount of regularisation. By convention, tuning parameters are optimised by fitting and evaluating a sequence of models using cross-validation. The bounds of 95% confidence intervals (CI) are presented in lieu of the p-values. A confidence interval that does not contain the value zero is equivalent to  $p < 0.05$ . In Table 4 (*visitor* model), model 1 is a regression model with just the control variables, and model 2 includes just linear components of rating and sentiment, both of which are positive and significant:  $\beta_{\text{sentiment}} = 0.012$  ( $p < 0.05$ );  $\beta_{\text{rating}} = 0.083$  ( $p < 0.05$ ). Hence H1 and H2 are supported. In model 3, the interaction is introduced and is significant ( $\beta = 0.003$ ,  $p < 0.05$ ).

**Table 2.** List of all the variables included in the models with description.

Variable name	Variable type	Description
Property type	Categorical	Entire home or private room in a home
Price	Continuous	Mean of the daily-rate (price in US\$) over the <i>period</i> , including cleaning fees distributed across the length of each reservation
Occupancy rate	Continuous	(Days with a reservation)/(total number of days with a reservation and available excluding blocked days)
N reviews	Discrete	Total number of reviews of the property
Max guests	Discrete	Maximum number of guests the property can accommodate
Response time	Continuous	Mean time it takes a host to respond to new booking inquiries (minutes)
Cancellation	Categorical	Strict or nonstrict cancellation policy
Minimum stay	Discrete	Minimum night stay required by host
N photos	Discrete	Number of photographs of the property
Instantbook	Categorical	Property can be booked without any host-guest communication (yes, no)
Rating	Continuous	Rating of the property on a scale of 1–5
Neighbourhood	Categorical	The London district (Table 1) numbered 1–33
Neighbourhood mean price*	Continuous	Mean daily-rate, over the <i>period</i> , of all properties in the same neighbourhood as the property
Sentiment*	Continuous	Mean sentiment score per review (see Section 4.6)

Prices and occupancy rate were available for the period 1 June 2017 to 31 May 2018. Variables indicated \* were calculated.



**Figure 1.** Variables in the landlord model (transformed data) showing: a smoothed empirical density plot of every variable (diagonal entries); every variable plotted against every other (upper right off diagonal); and the respective Pearson correlations (lower left). Response variable (occupancy rate) is on the top row. Entire homes (pink); room in a home (turquoise).

Thus, it is inferred that sentiment and rating interact to influence occupancy rate, so that H3 is supported.

In model 4, we introduce quadratic terms of both sentiment and rating. The analysis shows that both the quadratic terms are significant ( $\beta$

$= 0.878$ ,  $p < 0.05$ ) for sentiment and ( $\beta = 0.996$ ,  $p < 0.05$ ) for rating. To understand the nature of the quadratic effects, occupancy rate was plotted against rating and against sentiment (Figure 3). They indicate that as rating and sentiment each increases beyond a point, the rate of





**Figure 2.** Variables in the visitor model (transformed data) showing: a smoothed empirical density plot of every variable (diagonal entries); every variable plotted against every other (upper right off diagonal); and the respective Pearson correlations (lower left). Response variable (occupancy rate) is on the top row. Entire homes (pink); room in a home (turquoise).

**Table 3.** Regression coefficients and their upper and lower 95% confidence bounds for the ‘landlord’ model.

	Coefficient	Lower	Upper
Intercept	-5.045		
Neighbourhood mean price	3.572	3.464	3.685
Property type	-0.045	-0.115	0.000
Max guests	1.689	1.641	1.735
Cancellation	-0.043	-0.054	-0.032
Property type × max guests	-0.525	-0.583	-0.441
Model summary $R^2$	0.718		

increase in the occupancy rate diminishes. Thus, hypotheses H4 and H5 are supported.

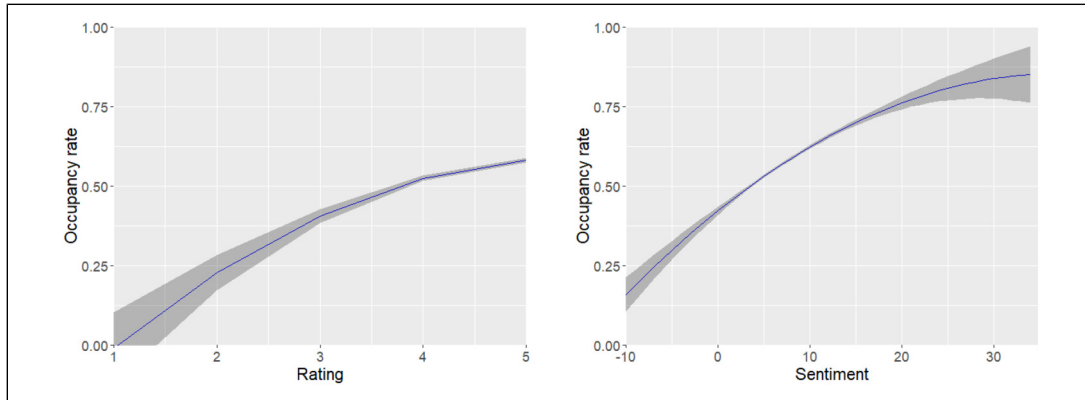
Interestingly, in model 3 (Table 4), with linear terms for *rating* and *sentiment*, the interaction is statistically significant but *sentiment* is not – the coefficient for *sentiment* is near-zero. This points to the complex relationship between sentiment and occupancy rate in peer-to-peer properties. It suggests that *sentiment* is only important when *rating* is high. Thus, good sentiment alone is not sufficient to influence occupancy, while on the contrary rating alone is important. In this way, rating might be regarded as a

primary signal of quality and then if rating is good, sentiment acts as secondary signal.

Of the control variables, *price variance* has a negative coefficient that is significant. As expected, this shows that lower-priced properties tend to have higher occupancy. Unsurprisingly, the coefficients of *n reviews* and *minimum stay* are positive and significant. Further, the instant booking feature tends to increase occupancy. On the other hand, the number of photos displayed appears to have no effect, while response time has a negative impact. The negative coefficient on property type implies that properties of

**Table 4.** Regression coefficients and their upper and lower 95% confidence bounds for the 'visitor' model.

Response: occupancy rate	Model 1 initial model			Model 2 + rating + sentiment			Model 3 + rating + sentiment + interaction			Model 4 + rating + sentiment + interaction + quadratic			Model 5 + rating + sentiment + quadratic		
	Coeff.	Lower	Upper	Coeff.	Lower	Upper	Coeff.	Lower	Upper	Coeff.	Lower	Upper	Coeff.	Lower	Upper
Intercept	0.280			-0.101			-0.055			0.330			0.330		
Residual price	-0.119	-0.131	-0.108	-0.140	-0.151	-0.128	-0.139	-0.153	-0.129	-0.143	-0.155	-0.133	-0.143	-0.155	-0.133
N reviews	0.061	0.058	0.064	0.047	0.043	0.051	0.046	0.042	0.051	0.049	0.045	0.053	0.049	0.045	0.053
N photos $\times 10^3$	2.293	-2.186	7.724	0.741	-3.137	5.418	0.032	-3.041	5.826	0.718	-3.904	5.896	0.718	-3.564	5.681
Instantbook	0.023	0.014	0.034	0.043	0.033	0.053	0.042	0.033	0.053	0.043	0.034	0.054	0.043	0.033	0.054
Minimum stay	0.225	0.194	0.253	0.215	0.186	0.239	0.210	0.184	0.243	0.214	0.189	0.243	0.214	0.189	0.240
Response time	-0.010	-0.012	-0.009	-0.010	-0.011	-0.009	-0.010	-0.011	-0.009	-0.010	-0.011	-0.009	-0.010	-0.011	-0.009
Property type	-0.053	-0.060	-0.043	-0.064	-0.072	-0.056	-0.063	-0.073	-0.057	-0.065	-0.074	-0.056	-0.065	-0.073	-0.057
Sentiment				0.012	0.009	0.014	0.000	-0.013	0.000	2.938	1.016	3.879	2.938	2.256	3.710
Sentiment quad										0.878	0.182	1.563	0.878	0.122	1.600
Rating				0.083	0.071	0.095	0.074	0.063	0.088	4.345	3.715	4.964	4.344	3.705	4.934
Rating quadratic										0.996	0.406	1.542	0.996	0.409	1.549
Rating $\times$ sentiment							0.003	0.002	0.005	0.000	-0.001	0.001	0.000		
Model summary $R^2$	0.178			0.203			0.204			0.205			0.205		



**Figure 3.** Occupancy rate as a function of *rating* and *sentiment* (model 4); fitted effects (solid line) shown for the quadratic model with confidence envelope for the effect (grey ribbon).

type private room have slightly lower occupancy rates than entire homes.

Finally, to indicate the practical relevance of these results, the sizes of the sentiment effect and rating effect were calculated. The following methodology was used: In model 2, the *occupancy rate* was determined for the ‘average’ property, which was the *occupancy rate* when all variables in the visitor model are held at their mean value. Then, this was repeated with the variable of interest (*sentiment* or *rating*) held at a high value, that is, at its mean plus two standard deviations. The difference in the two values of *occupancy rate* is then the benefit of being exceptional as opposed to average, sentiment or rating-wise. Thus, exceptional is defined as being in the top 2.5% on the measure. A two standard deviation increase (above the mean) in *rating* increases the expected occupancy from 56% to 63% (an 12.5% increase in relative terms). This increase of two standard deviations is approximately equivalent to a one-point increase in the rating. A two standard deviation increase (above the mean) in *sentiment* increases the expected occupancy rate from 56% to 61% (a 9% increase in relative terms). When both *sentiment* and *overall rating* are set simultaneously to two standard deviations above the mean, the expected *occupancy rate* increases from 56% to 68% (21% increase in relative terms). Alternatively, we can hold *occupancy rate* constant by increasing the price (because *occupancy rate* depends on *price variance* in the *visitor model*). In this way, simultaneous two standard deviation increases in *rating* and *sentiment* increase the expected price (holding *occupancy rate* constant) by \$16 per night. Thus, it is clear that rating and sentiment have a

real positive impact on the occupancy and equivalently the price of peer-to-peer accommodation.

## Conclusions

Using a large dataset of properties in London, this study measures the effect of user-generated content on occupancy rate in the peer-to-peer accommodation sector. Results indicate that the sentiment expressed in textual reviews and the numerical rating of a property have a significant impact on its occupancy rate. Such user-generated content is understood to signal value and thereby influence consumer choice. This study contributes to signalling theory in general, and the emerging discussion on the impact of user-generated content on demand in particular.

In this study, user-generated content, both quantitative reviewer rating and the qualitative sentiments are considered as cues that drive consumer decision-making in P2P settings. Using cue-utilisation theory – a sub-theory of signalling theory, we demonstrate how the cues impact consumer decision-making, both independently as well as through an interaction between the two cues. This result contributes to the extant theory on cue utilisation, especially in the context of consumer decision-making. The curvilinear effect demonstrates the validity of Easterbrook’s (1959) concept of ‘point of optimum proficiency’ of cues. The curvilinear effect of both the qualitative sentiment scores and reviewer rating goes to show the possibility of an optimum value in cue utilisation.

Our study makes a novel contribution as previous studies that have explored similar relationships either did not use real occupancy rate (e.g.,

Duverger, 2013; Park et al., 2020) of properties or used the sentiments from review text (e.g., Viglia et al., 2016; Van Der Borg et al., 2017). This was an important limitation of studies in the extant literature especially considering the critical importance of qualitative user-generated content in assessing the quality of peer-to-peer accommodation.

Thus, using occupancy rate, rather than other proxy variables like revenue performance, and using a measure of sentiment calculated from review text, we are able to provide greater depth and validity to the current understanding of the influence of user-generated content on outcome variables. Our study also explores how rating and sentiment scores interact to provide a more nuanced explanation to occupancy rate in peer-to-peer accommodation. To our knowledge, no previous study has explored the interaction effect even though such interactions have been suggested in extant theory. Interaction effects are quite important since, based on the theory of consistency of signals, the qualitative reviews provided by users and overall rating scores have a joint effect on outcome variables like occupancy rates or sales of products. Our models also include curvilinear effects of rating and sentiment scores on occupancy rate and found that beyond a point, the impact of rating and sentiment scores on occupancy rate decreases. The curvilinear effects can be interpreted as a necessary consequence of the fact that occupancy rate is bounded, that is, it cannot exceed 100%.

Further, recent studies in P2P sector (eg. Bridges and Vasques, 2019), have specifically pointed out the positive bias of user-generated content and thus raised questions about its real impact in driving outcomes like occupancy rates. This study uses real data and shows that the two dimensions of user-generated content namely review rate and sentiments, positively impact occupancy rates, hence we show that despite the positive bias, user-generated content has a real impact on occupancy rate. This result is also significant since we have also control for other potential variables that could impact occupancy rates.

### **Managerial implications**

This study therefore provides important insights into the relationship between user-generated content and occupancy rate in the context of peer-to-peer accommodation. While consumer choice in the peer-to-peer accommodation

sector is largely guided by user-generated content, the role of sentiment scores in customer choice had still not been conclusively established. Results from the study help to establish the strength of sentiment scores in impacting consumer choice by specifying its role in improving occupancy rates. The interaction and curvilinear effects provide important pointers to decision-makers regarding the optimum utilisation of user-generated content in promoting their properties. The impact of online reviews is enhanced by positive sentiments revealed through textual content highlighting the need to have both positive sentiment scores and a high rating. Hence, satisfied customers who may provide positive numerical reviews but very little or almost no textual reviews hardly help to generate a positive attitude towards the property among potential customers. Similarly, customers who may provide a near negative numerical review but negative textual reviews can harm the attitudes of potential customers to the property. These insights can be very useful for practitioners in devising digital marketing strategies. The curvilinear effects highlight the limited impact of both rating and sentiment scores above a certain level. Once high levels of rating and sentiment scores are achieved, their impact in driving occupancy rate is limited and hence other factors should be considered to increase occupancy rates.

While our results support all our hypotheses, there are limitations. Firstly, our findings apply to peer-to-peer properties in London, a city with unique characteristics. Nonetheless, occupancy rates are generally uniformly high in London, and the effects of interest may be easier to estimate in regions with more variable occupancy rates. Secondly, we have used linear models in our analysis when there may exist critical thresholds for sentiments and ratings and these thresholds may vary with price. Nonetheless, such thresholds may vary between consumers, so that effects are blurred and the review-occupancy relationship reverts to a smooth one. Thirdly, although the dataset is from a reliable source, there is limited detail about how rating is calculated, and sentiment can be measured in many ways. In this context, it would be interesting in further study to measure sentiment such that more weight is placed on more recent reviews.

### **Author's note**

Prof. Sunil Sahadev Professor of Marketing and Responsible Enterprise Sheffield Business

School College of Business, Technology and Engineering, Sheffield Hallam University, UK.

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Sunil Sahadev  <https://orcid.org/0000-0001-9648-8079>

### References

- Agnihotri A and Bhattacharya S (2016) Online review helpfulness: role of qualitative factors. *Psychology & Marketing* 33: 1006–1017.
- Belarmino A and Koh Y (2020) A critical review of research regarding peer-to-peer accommodations. *International Journal of Hospitality Management* 84: 102315.
- Bridges J and Vásquez C (2018) If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism* 21(18): 2057–2075.
- Chen L, Baird A and Straub D (2020) A linguistic signaling model of social support exchange in online health communities. *Decision Support Systems* 130: 113233.
- Chica-Olmo J, González-Morales JG and Zafra-Gómez JL (2020) Effects of location on Airbnb apartment pricing in Málaga. *Tourism Management* 77: 103981.
- Chong AYL, Li B, Ngai EW, et al. (2016) Predicting online product sales via online reviews, sentiments, and promotion strategies. *International Journal of Operations & Production Management* 4: 358–383.
- Chua AY and Banerjee S (2016) Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior* 54: 547–554.
- Connelly BL, Certo ST, Ireland RD, et al. (2011) Signaling theory: a review and assessment. *Journal of Management* 37: 39–67.
- Dickinger A (2011) The trustworthiness of online channels for experience-and goal-directed search tasks. *Journal of Travel Research* 50: 378–391.
- Dolnicar S (2019) A review of research into paid online peer-to-peer accommodation: launching the annals of tourism research curated collection on peer-to-peer accommodation. *Annals of Tourism Research* 75: 248–264.
- Duverger P (2013) Curvilinear effects of user-generated content on hotels' market share: a dynamic panel-data analysis. *Journal of Travel Research* 52(4): 465–478.
- Easterbrook JA (1959) The effect of emotion on cue utilization and the organization of behavior. *Psychological Review* 66(3): 183–201.
- Erdem T and Swait J (1998) Brand equity as a signaling phenomenon. *Journal of Consumer Psychology* 7: 131–157.
- Eslami SP and Ghasemaghaei M (2018) Effects of online review positiveness and review score inconsistency on sales: a comparison by product involvement. *Journal of Retailing and Consumer Services* 45: 74–80.
- Fu S, Cheng X, Bao Y, et al. (2020) Staying in a hotel or peer-to-peer accommodation sharing? A discrete choice experiment with online reviews and discount strategies. *Internet Research* 31: 654–676.
- Glover S and Benbasat I (2010) A comprehensive model of perceived risk of e-commerce transactions. *International Journal Of Electronic Commerce* 15: 47–78.
- Gunter U and Önder I (2018) Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics* 24(3): 270–293.
- Hernandez-Ortega B (2020) What about “U”? The influence of positive online consumer reviews on the individual's perception of consumption benefits. *Online Information Review* 44: 863–885.
- Hu M and Liu B (2004) Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '04)*, Seattle, USA, 168–177.
- Hu N, Koh NS and Reddy SK (2014) Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems* 57: 42–53.
- Huang D, Liu X, Lai D, et al. (2019) Users and non-users of P2P accommodation: differences in perceived risks and behavioral intentions. *Journal of Hospitality and Tourism Technology* 10: 369–382.
- Lawani A, Reed MR, Mark T, et al. (2019) Reviews and price on online platforms: evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics* 75: 22–34.
- Lee JH, Jung SH and Park J (2017) The role of entropy of review text sentiments on online WOM and movie box office sales. *Electronic Commerce Research and Applications* 22: 42–52.

- Lee PJ, Hu YH and Lu KT (2018) Assessing the helpfulness of online hotel reviews: a classification-based approach. *Telematics and Informatics* 35: 436–445.
- Leoni V, Figini P and Nilsson JOW (2020) It's not all about price. The determinants of occupancy rates in peer-to-peer accommodation: a methodological contribution. *International Journal of Contemporary Hospitality Management* 32: 1693–1711.
- Li X, Wu C and Mai F (2019) The effect of online reviews on product sales: a joint sentiment-topic analysis. *Information & Management* 56: 172–184.
- Liang LJ, Choi HC and Joppe M (2018) Understanding repurchase intention of Airbnb consumers: perceived authenticity, electronic word-of-mouth, and price sensitivity. *Journal of Travel & Tourism Marketing* 35: 73–89.
- Lu Q, Ye Q and Law R (2014) Moderating effects of product heterogeneity between online word-of-mouth and hotel sales. *Journal of Electronic Commerce Research* 15: 1–12.
- Mehraliyev F, Chan ICC and Kirilenko AP (2022) Sentiment analysis in hospitality and tourism: a thematic and methodological review. *International Journal of Contemporary Hospitality Management* 34: 46–77.
- Moore E (2020) Airbnb's red-hot IPO will bring problems too. *Financial Times*. Available at: <https://www.ft.com/content/757409f2-9747-4a2f-ae3-ad143df6cd6b>.
- Murillo D, Buckland H and Val E (2017) When the sharing economy becomes neoliberalism on steroids: unravelling the controversies. *Technological Forecasting and Social Change* 125: 66–76.
- Olson JC and Jacoby J (1972) Cue utilization in the quality perception process. *ACR Special Volumes*.
- Park E, Kang J, Choi D, et al. (2020) Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. *Current Issues in Tourism* 23: 605–611.
- Sabri O, Van Doan H, Malek F, et al. (2020) When is transparent packaging beneficial? *International Journal of Retail and Distribution Management* 48(8): 781–801.
- Sainaghi R, Abrate G and Mauri A (2021) Price and RevPAR determinants of airbnb listings: convergent and divergent evidence. *International Journal of Hospitality Management* 92: 102709.
- Shirai M (2020) Understanding consumer perceptions of quality based on package size: the moderating role of the evaluation context. *International Journal of Retail and Distribution Management* 48(7): 749–747.
- Spence M (1973) Job market signaling. *Quarterly Journal of Economics* 87(3): 355–374.
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58: 267–288.
- Van der Borg J, Camatti N, Bertocchi D, et al. (2017) The rise of the sharing economy in tourism: Exploring Airbnb attributes for the Veneto Region. University Ca'Foscari of Venice, Department of Economics Research Paper Series No. 5.
- Viglia G, Minazzi R and Buhalis D (2016) The influence of e-word-of-mouth on hotel occupancy rate. *International Journal of Contemporary Hospitality Management* 28(9): 2035–2051.
- Volgger M, Taplin R and Pforr C (2019) The evolution of 'Airbnb-tourism': demand-side dynamics around international use of peer-to-peer accommodation in Australia. *Annals of Tourism Research* 75: 322–337.
- World Bank (2018) Tourism and sharing economy: Policy and potential for sustainable peer2peer accommodation. Working paper. Available at: <http://documents1.worldbank.org/curated/en/161471537537641836/pdf/130054-REVISED-Tourism-and-the-Sharing-Economy-PDF.pdf>.
- Wu J, Du L and Dang Y (2018, July) Research on the impact of consumer review sentiments from different websites on product sales. In 2018 *IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)* (pp. 332–338). IEEE.
- Xie K and Mao Z (2017) The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *International Journal of Contemporary Hospitality Management* 29: 2240–2260.
- Xie KL, Chen C and Wu S (2016) Online consumer review factors affecting offline hotel popularity: evidence from tripadvisor. *Journal of Travel & Tourism Marketing* 33(2): 211–223.
- Yang K, Kim HM and Tanoff L (2020) Signalling trust: cues from Instagram posts. *Electronic Commerce Research and Applications* 43: 100998.
- Yang Y, Park S and Hu X (2018) Electronic word of mouth and hotel performance: a meta-analysis. *Tourism Management* 67: 248–260.
- Yao B, Qiu RT, Fan DX, et al. (2019) Standing out from the crowd—an exploration of signal attributes of Airbnb listings. *International Journal of Contemporary Hospitality Management* 31(12): 4520–4542.
- Ye S, Ying T, Zhou L, et al. (2019) Enhancing customer trust in peer-to-peer accommodation: a “soft” strategy via social presence. *International Journal of Hospitality Management* 79: 1–10.
- Yeo I-K and Johnson RA (2000) A new family of power transformations to improve normality or symmetry. *Biometrika* 87: 954–959.

- Yin D, Bond SD and Zhang H (2014) Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly* 38(2): 539–560.
- Zhang L, Wei W, Line ND, et al. (2021) When positive reviews backfire: the effect of review dispersion and expectation disconfirmation on Airbnb guests' experiences. *International Journal of Hospitality Management* 96: 102979.
- Zhang Y, Gao J, Cole S, et al. (2021) How the spread of user-generated contents (UGC) shapes international tourism distribution: using agent-based modeling to inform strategic UGC marketing. *Journal of Travel Research* 60(7): 1469–1491.
- Zhu L, Lin Y and Cheng M (2020). Sentiment and guest satisfaction with peer-to-peer accommodation: when are online ratings more trustworthy? *International Journal of Hospitality Management* 86: 102369.