Bimodal length of stay in the accommodation sharing economy

Emilio Gómez–Déniz
Department of Quantitative Methods in Economics and TiDES Institute, University of Las Palmas de Gran Canaria, Spain

Jorge V Pérez–Rodríguez
Department of Quantitative Methods in Economics, University of Las Palmas de Gran Canaria, Spain

Juan M Hernández
Department of Quantitative Methods in Economics and TiDES Institute, University of Las Palmas de Gran Canaria, Spain

Abstract
This paper analyses the determinants of tourist length of stay in sharing accommodation. Methodologically, the paper proposes a new approach to account for the observed bimodality of length of stay. More specifically, the determinants of length of stay are analyzed by modeling both the conditional mean and conditional modal frequencies, which represent shorter and longer tourist stays, unlike previous contributions on length of stay where modeling has been based exclusively on the conditional mean. The empirical analysis examines the length of stay in sharing accommodation lodgings (Airbnb and Homeaway) of tourists visiting the Canary Islands (Spain) before and after the outbreak of the COVID-19 pandemic. The results show a bimodal distribution of length of stay: 1 and 7 nights in the pre-pandemic period, reduced to 1 and 2 nights in the pandemic period. The results also indicate that the model approach followed in this paper is preferable to other previous bimodality models in terms of estimation simplicity and fit. In addition, the model allows an analysis of the determinants of length of stay and differentiation of the influence of each determinant on shorter and longer stays.

Keywords
shifted distribution, double-inflated model, length of stay, bimodality, COVID-19, sharing accommodation

Introduction
Length of stay or duration (we use both terms indistinctly in the paper) is the number of nights that tourists spend at a given destination. Length of stay constitutes a measure of tourism demand, as is

Corresponding author:
Jorge V Pérez–Rodríguez, Department of Quantitative Methods in Economics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, 35017, Spain.
Email: jv.perez-rodriguez@ulpgc.es
the case with number of tourist visits or tourist expenditure. The empirical literature on the tourism sector has extensively analyzed their determinants (see Boto-García et al., 2019, for an overview) and, in general, the most relevant results indicate that duration depends on several factors such as income, trip costs, and other variables related to the visit (type of accommodation, socio-economic and vacation characteristics, among others).

To analyze these factors, empirical studies have used heterogeneous econometric methods, such as survival analysis, discrete choice models and count data models, among others (Thrane, 2012; Alegre et al., 2011; Boto-García et al., 2019; Bavik et al., 2021). However, one of the most relevant statistical features of length of stay is its multimodality, a consequence of the heterogeneous preferences of tourists. Thus, shorter and longer stay can represent the two most frequent mode values of the length of stay distribution. The other specific feature of length of stay is excess dispersion, which indicates that the mean is lower than the variance. Both are important characteristics in the fit of the distribution of tourist length of stay and in the study of its determinants (Gómez-Déniz and Pérez-Rodríguez, 2019; Gómez-Déniz et al., 2020b).

In this context, the main contribution of this paper is twofold. First, it analyses length of stay in properties offered in peer-to-peer (P2P) accommodation using data from the real market. We study not only the length of stay distribution, but also factors characterizing it in the P2P market. Second, we propose a new statistical model for duration which accounts for the bimodality observed in tourism markets. In this paper, we characterize the bimodal features of duration using a model based on the general double-inflated model without and with covariates (i.e., including the analysis of length of stay determinants). The advantage of the double-inflated model is the simplicity of the estimation process, the higher log-likelihood values finally obtained, and the precision of the two modes, since we provide them in the model itself. Moreover, the model allows to estimate not only the conditional mean expectation but also the conditional frequency bimodal expectations, which could represent the most prominent shorter and longer tourist stays.

A further contribution of this paper is to estimate the effect of the COVID-19 pandemic on length of stay. This is done through selected sampling for the empirical analyses, which includes data of length of stay in P2P properties located in the Canary Islands (Spain) during two different months, December 2019 and December 2020. By considering these 2 months, we can compare the determinants of length of stay before and after the outbreak of the COVID-19 pandemic. We also include in the sample rental price and other characteristics of the properties that may determine length of stay.

The rest of this paper is organized as follows: The section An overview briefly describes the modeling and determinants of length of stay in the empirical literature, as well as some general characteristics of the P2P market in the Canary Islands. A general double-inflated model and the characteristics of the specific double-inflated model used here are described in the section The general double-inflated model. The empirical application is detailed in the section Empirical results. Finally, a discussion of the findings and the main conclusions can be found detailed in the section Discussion and conclusions.

**An overview**

**Length of stay models**

Length of stay has been mostly analyzed in the traditional accommodation sector (e.g. hotels, apartments, among others). Several econometric methods have been used to study the determinants
of tourist length of stay in this market, but no consensus has yet been reached on the most suitable model to deal with its characteristics (Almeida et al., 2021).

Previous studies have mainly focused on survival models to model visit duration (Gokovali et al., 2007; De Menezes et al., 2008; Thrane, 2012; Almeida et al., 2021; among others), although other contributions that have been proposed in the empirical literature on length of stay include count data models (see Alegre et al., 2011; Alén et al., 2014; Bavik et al., 2021; among others), or quantile regressions (e.g., Salmasi et al., 2012; Brida et al., 2014), as well as other specifications such as the Heckman model (Rodríguez et al., 2018) (for an overview see, for example, Almeida et al., 2021).

Some of the above models have accounted for special statistical characteristics of duration such as multimodality and excess dispersion, but separately. Multimodality has been treated using latent class count data models (Alegre et al., 2011; Yang and Zhang, 2015; among others) and quantile regressions (Salmasi et al., 2012). For example, Alegre et al. (2011) identified two market segments (classes) containing tourists with preferences for either short or longer stays, while Almeida et al. (2021) considered two segments for the Madeira Islands in a finite mixture model within the context of traditional survival models. For its part, excess dispersion has been considered in Alén et al. (2014) and Boto-García et al. (2019), among others. The latter authors analyzed tourist length of stay in a particular destination using a hurdle count data model that helps to discriminate determinants for same-day visitors and tourists staying at least one night.

However, authors like Gómez-Déniz and Pérez-Rodríguez (2019) and Gómez-Déniz et al. (2020b) have jointly modeled both characteristics to provide a more comprehensive point of view. More specifically, these authors considered bimodality in the length of stay, which allows to capture the patterns that the empirical data seems to follow. In general, their approach has shown a better statistical fit for the study of the determinants that affect the length of stay of tourists visiting a destination, although the two population modes obtained do not coincide exactly with the empirical ones, and, in both cases, the modal frequencies do not depend on covariates.

**Length of stay determinants**

In general, it can be seen in the empirical literature that the determinants selected and used to explain length of stay are commonly based on diverse vacation-related, individual, and sociodemographic factors.

Independently of the method used, the findings generally point to a positive dependence of duration of stay on income and time constraints (e.g. Salmasi et al., 2012) and a negative dependence on price (e.g. Hellstrom, 2006), while the effects of other socio-economic and demographic variables are varied. For example, Martínez–García and Raya (2008) found that a tourist’s country of origin, age, and choice of accommodation, as well as time of year and geographical area (urban vs rural), influence length of stay in low-cost tourism (defined as a particular type of tourism which uses low-cost airlines as a means of transportation). Brida et al. (2014) found that socio-economic, travel-related, psycho-graphic, and budgetary related characteristics were relevant in explaining tourist length of stay in Uruguay using quantile regression. Other more specific characteristics such as loyalty have also been analyzed. For example, repeat visits to a destination result in longer lengths of stay than first-time visitors (De Menezes et al., 2008; Gokovali et al., 2007; Santos et al., 2015; among others). Based on work status variables, for example, Alegre and Pou (2006) found that length of stay is longer for students and is negatively related to civil servants in the Balearic Islands. Finally, the intention to recommend a destination to others generally indicates a positive link between word-of-mouth and length of stay (Murphy et al., 2007; Bavik et al.,
2021; among others). These findings highlight the importance of customer satisfaction for increased length of stay.

The outbreak of the COVID-19 pandemic severely impacted travel intentions and tourist preferences. For example, Zheng et al. (2021) showed how travel fear due to the pandemic increases the adoption of certain coping strategies which contribute to the adoption of more cautious traveling. Some of these cautionary measures can lead to travelers changing their chosen destination. In this regard, some authors have found that post-pandemic preferred destinations are beach and open environments (Jeon and Yang, 2021). Other cautionary measures that can be adopted include reducing the length of stay and hence potential exposure to infection. Baños-Pino et al. (2021) estimated a 23.8% decline in length of stay after the outbreak of the COVID-19 pandemic in Asturias, a nature-based destination in Spain, while Santos and Moreira (2021) observed shorter lengths of stay in Portugal during the second half of 2020 compared to the same period in 2019.

In general, the above cited papers studied length of stay in the hotel industry. However, other studies have analyzed the relationship between selecting P2P accommodation as lodging in the destination and length of stay, analyzing information from surveys and statistically verifying and testing the relationships. For example, Tussyadiah and Pesonen (2016) showed that, when compared to hotels, the advantages of sharing accommodation for travelers include a lower cost and higher social interaction with local people, factors which favor increased length of stay in the destination. The preference of P2P users for longer stays has also been reported by visitors to Hong-Kong (Poon and Huang, 2017). To give some numbers, according to an online survey of Airbnb users (mostly hosted in Canadian cities) conducted by Guttentag and Smith (2017) the average length of stay in sharing accommodation was an estimated 4.54 nights. However, some studies have hypothesized that the observed steep growth in popularity (and consequently in the market share) of Airbnb accommodation with respect to traditional accommodation may be contributing to a decline in length of stay (Gosling et al., 2018).

However, the preference for longer stays when booking through sharing accommodation is not generalized, but depends on the type of destination. Specifically, in sun and sand destinations, such as those in the coastal areas of the South of Europe, a greater length of stay for visitors hosted in traditional accommodation has been observed. One reason is the prevalence of package tours, which are the most common form of travel for tourists staying in hotels in these destinations. These packages are associated with a weekly trip pattern, which tends to make the length of the stay significantly longer than when a package tour is not involved, as in the case of tourists staying in sharing accommodation. In fact, two modal values of around 7 and 14 nights were observed for visitors to the Balearic Islands in 2006, mostly hosted in hotels and tourist apartments, with an overall average of 8.59 days (Alegre et al., 2011). In the Azores, the length of stay for those arriving in charter flights was significantly higher than those arriving on non-charter flights (De Menezes et al., 2008). More recently, the rise in low-cost flights has been accompanied by increased expenditure in certain tourist segments but also by a generalized decrease in length of stay in the Canary Islands (Eugenio-Martin and Inchausti-Sintes, 2016). We would therefore expect the length of stay for Airbnb tourist hosting in the Canary Islands to be lower than those staying in hotels.

As far as we are aware, the question of multimodality and its determinants has been ignored until now in the analysis of length of stay in P2P accommodation. In this paper, we analyze this issue on the basis of observed length of stay data in sharing accommodation units. We also analyze data from the pre- and intra-COVID period in order to empirically estimate the effect of the pandemic on the length of stay in P2P accommodation.
The case study focuses on the P2P market in the Canary Islands, Spain. The archipelago has been a traditional ‘sun and beach’ destination since the 1960s, although in recent decades it has diversified its offer to include urban and rural tourism. The islands welcomed over 15 million tourists in 2019, with the major markets being, in order, the UK and Germany, and accounting for more than 50% of total tourist demand. Spanish tourism represents around 13% of tourist demand (FRONTUR, 2021).

As a result of the pandemic, tourist arrivals fell by 85% from March 2020 to February 2021, but have now recovered to the extent that tourist arrivals in April 2022 exceeded those of April 2019 (ISTAC, 2022).

The sharing accommodation sector first emerged in the archipelago in the last decade and has increased its market share each year since then. In fact, the share of P2P accommodation in the Canary Islands in 2016 was 8.6%, while around 11.9% of the total visitors in 2020 were hosted in such accommodation (ISTAC, 2021). Table 1 includes the total number of listings in the islands in 2019 and 2020 where length of stay in each listing was available, together with some other variables of interest. The data was obtained from AirDNA, a firm who provides metrics of Airbnb and Homeaway. For comparison, it also includes statistics for the hotel sector in the same period of time (ISTAC, 2022).

As can be observed, number of listings and average daily rate (ADR, in euros), defined as the quotient between total revenue and booked nights, remained stable in 2020, but number of users underwent a sharp decrease of 38%. Mean length of stay followed the decrease in number of users, in this case from 4.53 to 3.87 nights. These stays are shorter than in the hotel industry, generally conditioned as previously mentioned by tourist packages of 7 or 14 nights. In this sense, the average length of stay for tourists hosted in hotels is similar to other traditional sun and beach destinations, such as the Balearic Islands (Alegre et al., 2011). Moreover, the ADR is between 9.1% and 11.4% higher in hotels than in P2P accommodation. Interestingly, the effect of the COVID-19 pandemic on the two types of accommodation differed considerably. The hotel sector reduced its offer by 52.0%, and the number of users saw an even more drastic decrease. However, there was only a 5.4% reduction in the offer of P2P units and a smaller decrease in user numbers compared to the hotel sector.

The general double-inflated model

Since the works of Cohen (1966) and Lambert (1992), many papers have addressed in a theoretical way the subject of zero-inflated models. The essential idea of these models consists of starting from
a basic distribution (usually a discrete distribution belonging to the exponential family) and inflating the model by incorporating a parameter that tries to get the distribution to perfectly fit the frequency of the modal value of the empirical data, usually zero. The resulting distribution is a discrete mixture model from which the basic distribution is obtained for a particular value of the incorporated parameter. Some of the distributional assumptions for the parent distribution are the negative binomial (Cohen, 1966), the Poisson distribution (Bohning et al., 1999), the generalized Poisson distribution (Famoye and Singh, 2006) and the quasibinomial distribution (Gómez-Déniz et al., 2020a). In the tourist setting, inflated models are not very common. In this regard, the following are of particular interest: the Gurmu and Trivedi (1996) approach for a count dataset for recreational boating trips that shows zero counts, which is significantly higher than that expected for Poisson-distributed data, but without including covariates; in Boto-García et al. (2019) a hurdle count data model is used to identify the determinants to explain tourist length of stay; and finally in Van Truong et al. (2020) a study is made of the best fitting models for total population estimation concerning tourist accommodation data in which the outcome variable, number of guests, has an excess zero count. An extension of these models allowing the introduction of two values of inflation, usually zero and one, was provided in Zhang et al. (2016) and Liu et al. (2018). This idea is developed in our paper, introducing two modes that can be jointly estimated with the mean of the proposed distribution using a double-inflated model. More specifically, the model we propose characterizes the existence of two modes, \( r \) and \( s \), which are the most frequent values for the observed length of stay in the analyzed market.

If we consider the length of stay, \( X \), as a random variable censored at zero, we therefore assume a shifted (truncated or censored at zero) known distribution with probability mass function (pmf) \( f_\theta(x) \), \( x = 1, 2, \ldots \), depending on a parameter (or vector of parameters) \( \theta \). Henceforth, this distribution will be called the parent distribution.

Following the ideas in Liu et al. (2018), we propose the following double-inflated model for length of stay density, such that

\[
g(x) = \begin{cases} 
q_r, & x = r, \\
q_s, & x = s, \\
\gamma(\Theta)f_\theta(x), & x \neq r, x \neq s,
\end{cases}
\]

(1)

where \( q_r \geq 0, q_s \geq 0, q_r + q_s \leq 1, \)

\[
\gamma(\Theta) = \frac{1 - q_r - q_s}{1 - f_\theta(r) - f_\theta(s)}
\]

and \( \Theta = (q_r, q_s, \theta) \) is a vector of parameters.

**Stochastic representation of the probability mass function**

In this section, we provide a stochastic representation of the pmf given in (1). This representation facilitates interpretation of the proposed model.

Proposition 1. Let the random variable \( X \) be given by

\[
X = Z(1 - B_r) + rB_r(1 - B_s) + sB_rB_s
\]

(2)
where $B_r$ and $B_s$ follow a Bernoulli distribution with success probabilities $p_r$ and $p_s$, respectively, being $r \geq 1$, $s \geq 1$ and $r \neq s$. Let $Z$ now follow the pmf $f_\theta(x)$, $x = 1, 2, \ldots$, and $B_r$, $B_s$ and $Z$ be mutually independent. Then, it is verified that the pmf of $X$ is given by:

$$g(x) = \begin{cases}  
p_r(1 - p_s) + (1 - p_r)f_\theta(r), & x = r, \\
p_r p_s + (1 - p_r)f_\theta(s), & x = s, \\
(1 - p_r)f_\theta(x), & x \neq r, \quad x \neq s, 
\end{cases}$$

for $0 \leq p_r \leq 1$ and $0 \leq p_s \leq 1$.

**Proof:** Given the random variables $Z$, $B_r$, and $B_s$, we can establish the following relation between them:

$$Z = r \iff (V = r, B_r = 0) \cup (B_r = 1, B_s = 0),$$
$$Z = s \iff (V = s, B_r = 0) \cup (B_s = 1, B_r = 1),$$
$$Z = k \iff (V = k, B_r = 0), \quad k \neq r, \quad x \neq s$$

from which, taking into account the assumption of independence, expression (3) is easily obtained.

When $p_r = 0$, the distribution (3) reduces to $f_\theta(x)$ and other combinations of these parameters reduce the model to an $r$-inflated model and a hurdle model also widely used in the applied statistical literature.

Now, by denoting $q_r = p_r p_s + (1 - p_r)f_\theta(r)$ and $q_s = p_r(1 - p_s) + (1 - p_r)f_\theta(s)$, it is a simple exercise to see that the pmf of the double-inflated model can be rewritten in a more suitable form as the one provided in (1). Furthermore, other stochastic representations of the pmf given in (1) are possible using the ideas provided in Zhang et al. (2016).

The mean and the raw second order moment are given by:

$$E(X) = r q_r + s q_s + [\mu - r f_\theta(r) - s f_\theta(s)] \gamma(\Theta),$$

$$E(X^2) = r^2 q_r + s^2 q_s + [\sigma^2 - r^2 f_\theta(r) - s^2 f_\theta(s)] \gamma(\Theta),$$

where $\mu$ and $\sigma^2$ are the mean and variance of $f_\theta(x)$, respectively. From (4) and (5), the variance can be easily obtained.

**Specification of the parent distribution**

Recent papers on tourist length of stay, such as Gómez-Déniz and Pérez-Rodríguez (2019) and Boto-García et al. (2019), use two alternative models to identify the determinants (socio-demographic, distance, mode of transport, and specific attributes of the destination) of the tourist’s decision to visit for a single day or for a longer term. The particular model is selected according to the assumption made about the distribution of the unobserved heterogeneity for the length of stay (i.e., the zero-truncated negative binomial model and the zero-truncated Poisson log-normal model).

In this paper, we propose to use two variants of the two above cited models.

**The SP model.** We assume initially that the variable to be modeled, length of stay, $X$, is a random variable censored at zero, and therefore we assume a shifted (truncated or censored at zero) Poisson distribution (SP), with parameter $\lambda > 0$, and a pmf that is given by
This discrete distribution is different from the zero-truncated Poisson distribution (ZTP) described for instance by Palmer-Tous et al. (2007), who also considered the zero-truncated negative binomial distribution (ZTNB). The mean and variance of the distribution (6) are given by $E_\lambda(X) = 1 + \lambda$ and $\text{var}_\lambda(X) = \lambda$, respectively.

Given a sample $\tilde{x} = (x_1, \ldots, x_n)$, obtained from (6), the maximum likelihood (ML) estimator of the parameter is $\hat{\lambda} = \bar{x} - 1$, where $\bar{x} = (1/n)\sum_{i=1}^n x_i$ is the sample mean. The standard error is given by $(n/\hat{\lambda})^{-1/2}$, which is not significantly different from the classical (non-truncated) Poisson distribution.

Since $ID_{SP} = \text{var}_\lambda(X)/E_\lambda(X) = \lambda/(1 + \lambda) < 1$, the distribution shows infra-dispersion (i.e., the variance is lower than the mean), which would make it inappropriate for defining the length of stay, an event that empirical studies have shown presents over-dispersion (i.e., the variance is greater than the mean). Furthermore (see Cameron and Trivedi, 1986; Cameron and Trivedi, 1998; Palmer-Tous et al., 2007), the idea of using a distribution presenting infra-dispersion can produce biased estimators of the model parameters.

**The shifted negative binomial model.** In addition to the above, various factors (geographical, cultural, behavioral, etc.) influence an individual’s patterns of behavior. For instance, tourists act differently according to whether they are at home or their tourist destination. On the other hand, empirically, there is evidence that over-dispersion is related to the heterogeneity of the population. In this case, the parameter $\lambda$ is considered a random variable taking different values among different tourists and thus reflecting uncertainty about this parameter. Therefore, we believe that the pmf of $X$ in the tourist population is a mixed pmf given by:

$$p_x = E[f_\lambda(x)] = \int_0^{\infty} f_\lambda(x)\pi(\lambda)d\lambda,$$

where $\pi(\lambda)$ is the mixing density. In this analysis, we assume a gamma mixing distribution with shape parameter $\sigma = (\mu - 1)/\beta$, $\mu > 1$, and scale parameter $1/\beta > 0$. That is, the pdf of the mixing density is given by

$$\pi_{\mu,\beta}(\lambda) = \frac{\lambda^{(\mu-1)}\exp(-\lambda/\beta)}{\beta^{(\mu-1)}\Gamma((\mu-1)/\beta)}, \quad \lambda > 0, \quad \mu > 1, \quad \beta > 0. \quad (8)$$

By substituting (6) and (8) into (7), we get the unconditional pmf of the random variable length of stay

$$f_\theta(x) = \left(\frac{\mu - 1}{\beta + x - 2}\right)\left(\frac{1}{1+\beta}\right)^{\frac{(\mu-1)}{\beta}}\left(\frac{\beta}{1+\beta}\right)^{x-1},$$

for $x = 1, 2, \ldots$, where $\theta = (\beta, \mu)$. This is a shifted (displaced one unit) negative binomial (SNB) distribution with parameters $\beta(\mu - 1) > 0$ and $0 < \beta/(\beta + 1) < 1$. The mean and variance of this
distribution are given by \( E_\theta(X) = \mu > 1 \) and \( \text{var}_\theta(X) = (\mu - 1)(1 + \beta) \), respectively, being \( \beta \) a dispersion parameter. Thus, the index of dispersion of this distribution is:

\[
ID_{\text{SNB}} = \frac{\text{var}_\theta(X)}{E_\theta(X)} = \left(1 - \frac{1}{\mu}\right)(1 + \beta),
\]

which may exhibit infra or over-dispersion. When \( \beta > (\mu - 1)^{-1} \) the model shows over-dispersion, and when \( \beta < (\mu - 1)^{-1} \) the model shows infra-dispersion. This distribution, therefore, is appropriate for modeling length of stay. A detailed study of this model can be seen in Gómez-Déniz et al. (2021).

Nevertheless, empirically it is shown that the length of stay of tourists at their destination usually presents bimodality, and the distribution given in (9) does not seem to fit this type of data in a satisfactory way. That is, in practice there exist different probabilities for the length of stay of tourists. Assuming that the probability that the length of stay of \( r \) nights is \( p_r \), and that the probability that the length of stay of \( s \) nights is \( p_s \), then the variable total length of stay can be modeled like the one given in (2). Due to this, we will consider model (1) in which \( f_\theta(x) \) will be replaced by (9).

**Estimation**

Depending on the assumption that \( f_\theta(x) \) is a Poisson or negative binomial distribution, our model can be named double-inflated shifted Poisson (DISP) or double-inflated shifted negative binomial (DISNB), respectively.

Let \( \bar{x} = (x_1, \ldots, x_n) \) be a random sample obtained from the basic model (1). Moment estimators of the model parameters can be easily obtained by equating \( q_r \) and \( q_s \) to the sample values \( \bar{f}(r) \) and \( \bar{f}(s) \), respectively. Now, by using (4) and (5) we get that the estimator of \( \mu \) is the sample mean \( \bar{x} \) and the estimator of \( \beta \) is equal to \( (s^2 - \bar{x} + 1)/(\bar{x} - 1) \), with sample mean, second order moment and variance given by \( \bar{x}, \bar{s}^2 \), and \( s^2 \), respectively. These estimators can be taken as initial values for estimating by the ML method. In this case, the log-likelihood function is given by

\[
\ell(\Theta, \bar{x}) = \sum_{i=1}^{n} I\{X_i = r\}\log q_r + \sum_{i=1}^{n} I\{X_i = s\}\log q_s + \sum_{i=1}^{n} I\{X_i \neq r, s\}\log f_\theta(x_i),
\]

from which the normal equations which provided the ML estimators are given by

\[
\frac{n_r}{q_r} + \frac{n - (n_r + n_s)}{1 - q_r - q_s} = 0, \quad (10)
\]

\[
\frac{n_s}{q_s} + \frac{n - (n_r + n_s)}{1 - q_r - q_s} = 0, \quad (11)
\]

\[
\frac{n - (n_r + n_s)}{1 - f_\theta(r) - f_\theta(s)} \left[ f'_\theta(r) + f'_\theta(s) \right] + \sum_{i=1}^{n} I\{X_i \neq r, s\} \frac{1}{f_\theta(x_i)} \frac{\partial f_\theta(x_i)}{\partial \theta} = 0, \quad (12)
\]

where \( n_r \) and \( n_s \) are the number of observations of \( x_r \) and \( x_s \) in the sample, respectively, and \( f'_\theta(\cdot) \) represents the derivative of \( f_\theta(\cdot) \) with respect to \( \theta \). From equations (10) and (11) it is obtained that the estimators of \( q_r \) and \( q_s \) are given by \( \hat{q}_r = n_r/n \) and \( \hat{q}_s = n_s/n \). The estimator of \( \theta \) is obtained from
(12) and is easily calculated depending on the assumption about \( f_\theta(x) \). Normal equations and Fisher’s information matrix of this basic model for the special choice of \( f_\theta(x) \) are not provided here, although two elements of this matrix have simple expressions given by

\[
E \left( \frac{\partial \ell(\Theta, \hat{x})}{\partial q_r^2} \right) = \frac{n_r}{q_r^2} \left[ 1 - q_r - q_s \right],
E \left( \frac{\partial \ell(\Theta, \hat{x})}{\partial q_s^2} \right) = \frac{n_s}{q_s^2} \left[ 1 - q_r - q_s \right].
\]

**Including covariates**

In this section, we model the conditional frequency of the two modes and the conditional mean of the parent distribution. To do so, we consider that these conditionals depend on several determinants or covariates.

The rationale for this approach is that a tourist’s preferences, which can be based on two or more segments for the length of stay (e.g. shorter and longer stays), can depend on several factors to explain the choice to belong to one or the other segment.

In this sense, this paper proposes a model which can explain the conditional mean of the duration of the visit but also both conditional frequency modes. A priori, there are no indications to consider the vector of covariates affecting them, but it is an interesting exercise that should be carried out.

Therefore, we link the model parameters \( q_r, q_s \) and \( \mu \) with the covariates as follows

\[
\mu_i = 1 + \exp \left( z_i^T \delta \right),
\]

\[
q_{ri} = \frac{\exp \left( \omega_{ri}^T \gamma_r \right)}{1 + \exp \left( \omega_{ri}^T \gamma_r \right) + \exp \left( \omega_{si}^T \gamma_s \right)},
\]

\[
q_{si} = \frac{\exp \left( \omega_{si}^T \gamma_s \right)}{1 + \exp \left( \omega_{ri}^T \gamma_r \right) + \exp \left( \omega_{si}^T \gamma_s \right)},
\]

where \( \delta, \gamma_r, \) and \( \gamma_s \) are vectors of regression coefficients, \( Z = (z_1, \ldots, z_n)^T \), \( \Omega_r = (\omega_{r1}, \ldots, \omega_{rn})^T \), and \( \Omega_s = (\omega_{s1}, \ldots, \omega_{sn})^T \) are the vectors of covariates for the \( r \) and \( s \) modes, respectively, and \( \mu = (\mu_1, \ldots, \mu_n) \), \( q_r = (q_{r1}, \ldots, q_{rn}) \), and \( q_s = (q_{s1}, \ldots, q_{sn}) \) are the vectors of parameters to be estimated together with the parameter \( \beta \). The expression given in (13) ensures that the mean of the distribution will be larger or equal to 1. Furthermore, expressions (14) and (15) are chosen as multinomial logit to ensure that \( q_{ri} + q_{si} \leq 1 \).

In practice, we can choose \( Z = \Omega_r = \Omega_s \).

**Empirical results**

**Data**

The data for the empirical section was provided by AirDNA and corresponds to daily reports for December 2019 (pre-COVID period) and December 2020 (intra-COVID period) of listings operating in the Canary Islands through Airbnb and HomeAway. The information covers those properties which were booked during these months and includes hosts with non–zero guests in the sample period. When the actual reservation was made in the previous month and/or was maintained during some days of the next month, the total length of stay in the accommodation unit was
considered. In addition to duration, other variables that were collected include daily rental price of the stay (including cleaning fees), type of lodging and the existence or not of different amenities in the property.

The motivation for the choice of these specific months was as follows. One the one hand, we were interested in examining the effect of the pandemic on length of stay. December 2020 can be considered an intra-COVID month since, although tourist arrivals to the islands had begun to recover, some restrictions on inbound tourism to the Canary Islands and return trips to the main countries of origin (England and Germany) were still in force. At that time, no vaccine was available and travelers needed to show a negative antigen or PCR test to be admitted into the destination, together with quarantine periods when returning to their home country. On the other hand, the selection of December guaranteed a high degree of data availability, since this is traditionally a month of major tourist flow to the Canary Islands.

The length of stay is a tourist’s decision that depends on several factors, including time and budget limitations, attributes of the destination and others described in the section Length of stay determinants. However, we cannot use the general determinants identified to explain length of stay in the tourism markets due to the lack of available data. Therefore, our model only includes supply-based factors, such as prices, and other covariates represented by the amenities (property-related characteristics).

| Variable          | Description                             | December 2019 | December 2020 |
|-------------------|-----------------------------------------|---------------|---------------|
| Listings          | Number of listings in sample            | 4640          | 9844          |
| Duration          | Mean duration in unit                   | 5.08          | 4.44          |
|                   | Standard deviation                      | (3.97)        | (4.17)        |
| Price             | Mean price of unit unit                 | 183.95        | 73.55         |
|                   | Standard deviation                      | (128.07)      | (55.17)       |
| Pets allowed      | Pets allowed in the property            | 0.17          | 0.22          |
| Accessibility     | Accessibility facilities                | 0.10          | 0.12          |
| Connectivity      | WiFi and/or internet connection         | 0.98          | 0.92          |
| Entertainment     | Entertainment elements                  | 0.94          | 0.85          |
| Beach environment | Location near beach                     | 0.15          | 0.16          |
| Basic facilities  | Basic facilities in the property        | 0.97          | 0.91          |
| Family            | Family friendly property                | 0.55          | 0.44          |
| Kitchen           | Kitchen available                       | 0.97          | 0.95          |
| Parking           | Parking available                       | 0.85          | 0.79          |
| Pool              | Pool available                          | 0.85          | 0.28          |
Table 2 shows the descriptive statistics of the variables and covariates of the model. Although there are many types of properties (cottages, bungalows, condominiums, etc.), apartments, houses and villas represent more than 80% of the total sample in both years and so only the statistics for these three types of lodging are included. As can be observed, Apartment is the most frequent type in the sample. The mean duration is higher for Villas than for the other types and was between 4 and 5 nights in 2019, but decreased in 2020. However, prices increased in 2020 for Villas and Houses, while they decreased for Apartments. Table 2 also includes the percentage of properties of each type that include a specific amenity. The amenities were classified into large groups, such as whether the property has accessibility facilities, connectivity (e.g. WiFi or internet access), etc. As can be observed in Table 2, the representation of amenities in each type of lodging varies little between the 2 years. There are no major differences between types of lodging with respect to the availability of amenities, although Apartments tend to be situated closer to the beach but with lower parking availability and pets are less frequently allowed.

As can be observed in Figure 1, bimodality or multimodality of duration are present in the sample, with modes in 1 and 7 nights of stay for 2019, but 1, 2, 3, and 7 in 2020. That is, a weekly pattern is observed for P2P users, but with a clear preference for shorter-than-a week lengths of stay. Notably, bimodality is present in the pre-pandemic month and multimodality in the intra-COVID-19 month. More specifically, the second mode dropped from 7 to 2 nights.

**Estimation results**

This section summarizes the estimation results distinguishing the modelization with and without covariates.

**Comparing several unimodal and bimodal models without covariates.** In this section, we compare the estimation results with several unimodal and bimodal models recently used in the literature on length of stay without taking into account covariates. Specifically, we have used as unimodal models the SP distribution with pmf given in (6) and the SNB distribution with pmf given in (9). As bimodal
models, we have used the pmf provided in Gómez-Déniz and Pérez-Rodríguez (2019) (BD1) and the pmf more recently introduced in the literature by Gómez-Déniz et al. (2020b) (BD2). The pmfs of these last two distributions are provided in the Appendix of this paper. Finally, we have used the model proposed in this paper taking into account two alternatives for the parent distribution: the bimodal model assuming the SP distribution (DISP) and the bimodal model assuming the SNB distribution (DISNB).

Table 3 shows the ML results for the models above, including the number of listings, parameters estimated, p-values and the Aikake information criterion (AIC) value using distribution (1). It should be noted that DISP and DISNB bimodal models fix the modal frequency at the most frequent durations, 1 and 7 nights for the December 2019 database, and 1 and 2 nights for the December 2020 database, respectively. Therefore, we consider the parameters $\hat{q}_1$, $\hat{q}_2$, and $\hat{q}_7$ as the estimated frequency for each mode.

We also performed Vuong’s test (see Vuong, 1989; Denuit et al., 2009: p.43) to compare the estimates of the DISNB distributions with respect to the SP, BD1, and BD2 models, which are non-nested with the DISNB model. The results indicate that all the estimated coefficients are statistically significant at the 1% significance level. It can therefore be concluded that the DISNB model achieves the best fit.

It is noteworthy that the presence of two modes is captured by the estimated model, given that parameters $\hat{q}_1$, $\hat{q}_2$, and $\hat{q}_7$ are statistically significant. Henceforth, there are differences between the two samples. For example, the estimated frequency for the first mode ($\hat{q}_1$) increases with respect to 2019 and the frequency of the second mode ($\hat{q}_7$) decreases.

In order to check the goodness of fit between empirical and estimated data, Figures 2 and 3 show a smooth kernel distribution based on empirical data and the fitted pmf for the different models estimated in Table 3. As can be seen, the pattern followed by the empirical data in each sample is exclusively captured by the DISNB distribution. Obviously, the shape of the distribution can fit not only the modal values but also less frequent ones, both in the central part of the data and in the tail of the distribution.

**Factors explaining length of stay using the double-inflated model.** This section presents the estimation of length of stay with our model including covariates, allowing assessment of the factors that can explain both the conditional frequency modes and the conditional mean for visit duration.

The duration of a tourist’s stay is measured by the number of nights booked at the destination and analyzed with the DISNB model when covariates are included in all three conditional equations defined by (13)–(15), respectively. For simplicity, as previously explained, we restrict the analysis to the three major types of lodging in our sample of Airbnb listings: Apartment, House and Villa. These types of accommodation also show bimodality on 1 and 7 nights for December 2019, and on 1 and 2 nights for December 2020. The estimation allows differentiation of the impact of covariates on length of stay for different types of accommodation.

Table 4 shows the results obtained using the ML method when inflation is in 1 and 2 or 7 nights. For simplicity, we have assumed that all covariates have the same influence on the conditional length of stay and their conditional frequency modes (the two main segments of tourist preferences). Interpretation of the coefficients should be made with caution for two reasons. First, we have defined neither a general duration model explaining the conditional mean nor the frequency modes. And second, we have not estimated marginal effects either.

In general, the parameter $\beta$ in Table 4 is statistically significant at 5%, showing that the inflation parameters and the scale one of the parent SNB distribution, $1/\beta$, are relevant for all models. These results confirm the suitability of our estimated model.
Table 3. Estimation of the competitive models without covariates in both December 2019 and 2020, respectively.

| Coefficient | December 2019 | December 2020 |
|-------------|---------------|---------------|
|             | SP  | SNB | BD1 | BD2 | DISP | DISNB | SP  | SNB | BD1 | BD2 | DISP | DISNB |
| \( \hat{q}_1 \) | 0.2668*** | 0.2668*** |   |   |   |   | 0.3724*** | 0.3724*** |   |   |   |   |
| \( \hat{q}_2 \) | 0.0826*** | 0.1274*** |   |   |   |   | 0.0826*** | 0.1274*** |   |   |   |   |
| \( \hat{q}_7 \) | 0.1370*** | 0.1370*** |   |   |   |   | 2.9040*** | 2.9040*** |   |   |   |   |
| \( \lambda \) | 3.6796*** |   |   |   |   |   | 3.6796*** |   |   |   |   |
| \( \hat{\mu} \) | 4.2625*** | 5.5001*** | 3.5223*** | 4.9327*** | 4.9009*** |   | 3.9040*** | 5.4502*** | 3.0328*** | 5.4164*** | 4.1298*** |   |
| \( \beta \) | 4.6796*** | 2.6686*** | 3.7896*** |   |   |   | 4.6796*** | 2.6686*** | 3.7896*** |   |   |   |
| AIC         | 388,585 | 279,752 | 347,583 | 335,154 | 322,663 | 271,858 | 278,035 | 183,189 | 248,902 | 231,316 | 208,630 | 182,518 |
| Number of listings | 57,678 | 57,678 | 57,678 | 57,678 | 57,678 | 57,678 | 41,959 | 41,959 | 41,959 | 41,959 | 41,959 | 41,959 |

Vuong’s test

| DISNB vs DISP | z = 49.424[0.00] | z = 23.688[0.00] |
| DISNB vs BD1 | z = 66.551[0.00] | z = 64.507[0.00] |
| DISNB vs BD2 | z = 55.626[0.00] | z = 46.251[0.00] |
| DISNB vs SP  | z = 102.582[0.00] | z = 90.530[0.00] |

Notes. p-values appear in brackets. ***p < 0.01, **p < 0.05, *p < 0.1.
Figure 2. Observed and expected counts under the double-inflated model based on the use of the different distributions considered for December 2019.
Figure 3. Observed and expected counts under the double-inflated model based on the use of the different distributions considered for December 2020.
Table 4. Estimation of the double-inflated model (DISNB) with covariates. Periods corresponding to December 2019 and 2020, respectively.

| Variable           | December 2019 |          | December 2020 |          |
|--------------------|---------------|----------|---------------|----------|
|                    | Coefficient   | p-value  | Coefficient   | p-value  |
|                    |               |          |               |          |
| Panel A: Conditional mean |
| Constant           | 0.6336        | 0.01     | 0.0911        | 0.61     |
| log(Price)         | 0.1071        | 0.00     | 0.1499        | 0.00     |
| Pets allowed       | 0.0175        | 0.65     | 0.0944        | 0.04     |
| Accessibility      | 0.0400        | 0.43     | 0.0331        | 0.57     |
| Connectivity       | 0.1165        | 0.29     | 0.1602        | 0.02     |
| Entertainment      | 0.2899        | 0.00     | 0.1716        | 0.00     |
| Beach environment  | -0.0613       | 0.16     | -0.1563       | 0.01     |
| Basic facilities   | 0.0865        | 0.38     | 0.0687        | 0.34     |
| Family             | -0.1112       | 0.00     | -0.0672       | 0.06     |
| Kitchen            | 0.0994        | 0.08     | 0.2679        | 0.02     |
| Parking            | -0.0619       | 0.18     | 0.2040        | 0.00     |
| Pool               | -0.0531       | 0.27     | 0.1612        | 0.00     |
| Panel B: Conditional frequency mode, \( \hat{q}_i \) |
| Constant           | -0.0777       | 0.85     | 0.0713        | 0.69     |
| log(Price)         | -0.0418       | 0.40     | 0.0032        | 0.93     |
| Pets allowed       | -0.1230       | 0.15     | -0.1669       | 0.00     |
| Accessibility      | -0.3513       | 0.00     | -0.1177       | 0.06     |
| Connectivity       | -0.1465       | 0.54     | -0.0372       | 0.57     |
| Entertainment      | -0.6190       | 0.00     | -0.3377       | 0.00     |
| Beach environment  | -0.3102       | 0.00     | -0.0972       | 0.05     |
| Basic facilities   | -0.0285       | 0.87     | -0.0137       | 0.83     |
| Family             | 0.2380        | 0.00     | 0.0874        | 0.05     |
| Kitchen            | 0.3509        | 0.06     | -0.1257       | 0.15     |
| Parking            | -0.0217       | 0.82     | -0.4279       | 0.00     |
| Pool               | -0.0447       | 0.64     | -0.0379       | 0.44     |

(continued)
Table 4. (continued)

| Variable                  | December 2019 | December 2020 |
|---------------------------|---------------|---------------|
|                           | Villa         | House         | Apartment     | Villa         | House         | Apartment     |
|                           | Coefficient   | p-value       | Coefficient   | p-value       | Coefficient   | p-value       |
|                           |              |               |              |               |              |               |
| **Panel C: Conditional frequency mode, \( \tilde{q}_2 \) and \( \tilde{q}_2 \)** |              |               |              |               |              |               |
| Constant                  | -2.5151       | 0.00          | -4.0385      | 0.00          | -3.4750       | 0.00          | -1.2194       | 0.04          | 0.0033        | 0.99          | -0.2288       | 0.24          |
| log(Price)                | 0.2165        | 0.00          | 0.4958       | 0.00          | 0.3546        | 0.00          | -0.0818       | 0.31          | -0.2090       | 0.00          | -0.1131       | 0.01          |
| Pets allowed              | -0.1829       | 0.04          | -0.2734      | 0.00          | -0.0406       | 0.39          | 0.0293        | 0.80          | 0.0755        | 0.27          | -0.0374       | 0.52          |
| Accessibility             | -0.0048       | 0.97          | -0.0368      | 0.68          | 0.0501        | 0.35          | -0.2531       | 0.13          | 0.1362        | 0.15          | -0.0079       | 0.90          |
| Connectivity              | -0.1076       | 0.66          | 0.1148       | 0.23          | 0.1031        | 0.09          | 0.2656        | 0.46          | -0.2333       | 0.02          | -0.1441       | 0.03          |
| Entertainment             | 0.2220        | 0.14          | -0.0549      | 0.55          | 0.0545        | 0.49          | 0.0230        | 0.92          | -0.0580       | 0.47          | 0.1418        | 0.07          |
| Beach environment         | -0.1569       | 0.09          | -0.0361      | 0.63          | 0.0756        | 0.03          | -0.0733       | 0.58          | -0.0690       | 0.42          | -0.0970       | 0.02          |
| Basic facilities          | 0.2488        | 0.19          | -0.0846      | 0.44          | 0.0070        | 0.91          | 0.0914        | 0.63          | 0.0089        | 0.94          | -0.2980       | 0.00          |
| Family                    | -0.0339       | 0.61          | 0.0580       | 0.32          | -0.0580       | 0.08          | 0.1320        | 0.17          | -0.1543       | 0.01          | 0.1548        | 0.00          |
| Kitchen                   | 0.1866        | 0.27          | 0.2211       | 0.12          | 0.2144        | 0.04          | -0.2044       | 0.23          | -0.0514       | 0.74          | -0.2642       | 0.00          |
| Parking                   | -0.0828       | 0.40          | 0.1898       | 0.01          | 0.0655        | 0.06          | -0.0254       | 0.85          | -0.0681       | 0.41          | -0.0800       | 0.06          |
| Pool                      | 0.2582        | 0.02          | 0.1733       | 0.01          | 0.2301        | 0.00          | -0.2968       | 0.03          | -0.1484       | 0.06          | -0.1542       | 0.00          |

| \( \beta \)   | 2.4120       | 0.00          | 3.8189       | 0.00          | 3.8136       | 0.00          | 2.5284       | 0.00          | 4.1199       | 0.00          | 4.5041       | 0.00          |
| Log L         | -10.619.17   |              | -22.575.4438 |              | -73.849.1072 |              | -10.657.84   |              | -16.874.01   |              | -47.969.89   |              |
| AIC           | 10.545.17    |              | 22.501.44    |              | 73.775.11    |              | 10.583.84    |              | 16.800.01    |              | 47.895.89    |              |
| Number of listings | 4640       |              | 9844         |              | 31.538       |              | 4784         |              | 7823         |              | 22.496       |              |
We expect a negative sign for coefficients of prices, as economic theory suggests. The amenities are dummy variables and the coefficients should be interpreted as an increase of duration with respect to the base category. We expect a positive sign for coefficients of amenities because we assume that the supply-related characteristics of the property are positively observed by the guests.

However, the ML results for coefficients of covariates are mixed in all three panels for each type of accommodation. We describe results dividing price and amenity effects, providing explanations for the specific outcome.

**Price effects.** Focusing on the ML estimates for the conditional expectation of duration in the three types of lodging for both years, results indicate that coefficients for price per stay (in logs) are statistically significant and have a positive effect on duration. The positive influence of price is higher in the pre-COVID period in all types of lodging. To check the robustness of these results, we conduct a linear regression estimation for duration against price (in logs) using an ordinary least squares (OLS) estimator, obtaining again a positive relationship between the two variables in most cases.

Focusing on the conditional expectation of modal frequencies, which can represent the most frequent stays, we observe that, in December 2020, price shows negative and statistically significant effects at 5% significance level for Houses in both modes ($\hat{q}_1$ and $\hat{q}_2$) and for Apartments in the second mode, respectively. In this sense, these results agree with other previous studies analyzing bimodal length of stay. For example, Alegre et al. (2011) found that tourist expenditure on hotels and apartments, as a proxy of price per stay, had a negative effect on duration for both segments considered (7 and 14 nights). However, in December 2019, price has a positive impact on duration in both modal frequencies for Apartments ($\hat{q}_1$ and $\hat{q}_7$), and Villas and Houses in the second mode.

The positive relationship between price and length of stay does not match the standard economic theory. An explanation for this outcome can be found in the specific characteristics of the P2P accommodation market. For example, P2P units include exclusive factors influencing the market value, which are not shared with the traditional accommodation sector (hotels) and are not included in the list of amenities considered in this analysis. Some of these have been identified as price determinants, such as personal attributes of the host (Ert et al., 2016), being a superhost (Wang and Nicolau, 2017), single or multi-unit host (Chen and Xie, 2017), and reputation (Moreno-Izquierdo et al., 2019). Spatial effects, such as competition among properties, have also been found to be price determinants (Chen and Xie, 2017; Tang et al., 2019). All these factors may also influence length of stay and are not included in the model, leading to a possible bias in results.

**Amenity effects.** Many of the amenities used in the conditional expectation of duration are statistically significant at level 1% or 5%. Moreover, many of them have positive coefficients for all types of accommodation and periods, as we expected, but there are also other coefficients which present negative signs.

We initially focus on results for the conditional mean. Regarding the positive coefficients, entertainment and kitchen amenities appear as common factors to explain duration in all three types of accommodation in 2019. However, in 2020, the results changed, with basic facilities a relevant amenity only for Villas and Apartments. Other variables are also relevant. For example, the coefficient for pool availability in Houses and Apartments is, in general, positive in both periods, showing that the existence of a pool in a property increases length of stay compared to listings without a pool. Some of the amenities present negative coefficients, indicating a negative influence on duration with respect to the base category. For example, this is the case of a family friendly lodging in the conditional mean for all three types of lodging. The negative effect is more
pronounced, in general, in the intra-pandemic period (December 2020). This result points to a lower length of stay in family friendly lodgings.

Focusing on the conditional expectation of modal frequencies, results again are mixed. Regarding positive results, family friendly properties extended length of stay for tourists in the first mode \( \hat{q}_1 \) for Villas and Apartments in 2019 and 2020. Interestingly, pool availability positively influenced duration in the tourist segment with the largest mode \( \hat{q}_7 \) in 2019, but exerted a negative effect for the tourist segment with mode \( \hat{q}_1 \) in Apartments in 2019 and with mode \( \hat{q}_2 \) in any type of property. That is, the availability of a pool helps increase length of stay for users who stay around 1 week but exerts a negative influence for those users staying just one or 2 days. Other amenities present a negative influence on duration for users with shorter lengths of stay, such as being located near the beach. The nature of these amenities (pool and beach) may explain this outcome as they are amenities related to being on vacation and having time to enjoy them. As revealed by the estimations, users who stay one or 2 days are not attracted by this kind of vacation amenity and are unlikely to extend their stay because of their availability or proximity.

**Discussion and conclusions**

This paper presents a bimodal model which allows modeling of both the mean and the modes of the length of stay distribution. Moreover, it also allows an assessment of the determinants of length of stay using conditional expressions depending on covariates in both the mean and the modal frequencies. The latter aspect is a new approach in this setting.

The empirical analysis was conducted using information for the P2P market in the Canary Islands, Spain. More specifically, we used a database of listings from AirBnb and HomeAway online platforms, differentiating two sample periods, namely, pre-COVID (December 2019) and intra-COVID (December 2020).

The results show that the proposed model, without covariates, fits the data better than other models considered in the applied statistical literature. Therefore, the model with covariates will also better represent the empirical data. Compared to the bimodal models proposed in Gómez-Déniz and Pérez-Rodríguez (2019) and Gómez-Déniz et al. (2020b), we can highlight the following: (i) estimation of our model is much simpler since for the previously proposed bimodal models the estimators of the parameters are highly sensitive to the seed point from which to start to obtain the ML estimation; (ii) the model proposed in this paper has the advantage of explaining the factors that affect the conditioned mean of the dependent variable but also the two inflation parameters, and therefore there is a benchmark for the affectation of variables in both modes considered which allows the economic agents involved to have information sources that are not available to other models; and (iii) the simplicity of the model, based on the use of the shifted Poisson and negative binomial distribution, allows the development of an EM-type algorithm that would allow more robust estimation procedures to be obtained, if necessary; and finally, (iv) the stochastic representation helps (although it has not been carried out here) the simulation study to assess the performance of the estimation procedure. The main disadvantage is that the researcher needs to previously set the value of the modes, which is impossible to estimate together with the rest of the parameters. However, the estimation results show that the model reasonably captures bimodality in the sample. Therefore, this approach can be used as an alternative model for length of stay.

The results can be used to compare the characteristics of the length of stay of P2P users and those tourists staying in traditional accommodation. For example, previous empirical studies found bimodality of 7 and 15 nights in the hotel sector (see Gómez-Déniz and Pérez-Rodríguez, 2019 and Gómez-Déniz et al., 2020b for the Canary Islands, and Alegre et al., 2011 for the Balearic Islands).
It is interesting to note that the modes do not coincide in the two markets in the same destination (Canary Islands). The bimodality of length of stay in sharing accommodation in the Canary Islands appears on 1 and 7 nights in the pre-pandemic period (before March 2020), while the second mode fell from 7 to 2 nights in the intra-pandemic period. That is, the findings indicate that tourists in the Canary Islands staying in P2P accommodation units prefer shorter lengths of stay than tourists staying in hotels at the same destination.

The preference of P2P users for shorter lengths of stay than hotel users can be explained by the characteristics of tourism in traditional ‘sun and beach’ destinations in Europe. Commonly, hotels have agreements with tour operators who sell holiday packages in the country of origin of tourists. These packages tend to be weekly based, normally of 7 or 14 nights duration. The observed duration modes in hotels match the duration of the holiday packages sold by tour operators. However, P2P users do not arrange trips through tour operators but directly with the hosts. Therefore, the average length of stay is more similar to that observed for P2P users in other regions (around 4 or 5 days).

The strong reduction in length of stay in the intra-pandemic period can be attributed to a change in tourists’ preferences after the lockdown period (March–June 2020). In this regard, our results agree with previous studies which detected a reduction of length of stay in nature-based destinations (Baños-Pino et al., 2021), complementing them with new findings in the same direction for ‘sun and beach’ destinations.

The model including covariates allowed estimation of some factors influencing length of stay, such as price and property amenities. Regarding price, results are mixed in the sense that its effect is negative for some types of accommodation, and positive for others. The positive effect of price on length of stay does not agree with predictions of the standard economic theory and suggests the existence of some factors influencing length of stay that are not included in the empirical estimation in this paper. Some may be specific to the P2P accommodation market, such as host attributes (photos, being a superhost, reputation). Other possible factors include spatial effects (competition). All of them have been detected as price determinants in other previous studies.

The estimation of bimodality in the length of stay of P2P users enabled new findings about the influence of property-related factors. More specifically, the direction of the influence differs between the two market segments determined by the modes. For example, amenities such as pool and location near the beach positively affect duration for those tourists who stay around 7 days, but negatively for those tourists whose length of stay is just 1 or 2 days. These findings reveal different consumption preferences of these amenities between market segments characterized by shorter and longer lengths of stay. We also detected some differences in length of stay between the pre-COVID and intra-COVID period. This is the case of family friendly lodgings, where there was a decrease in length of stay after the lockdown period. However, we could not include other supply-based factors (such as destination attributes) or sociodemographic and holiday characteristics due to the lack of information on these variables, which is the major limitation of our study.

Some practical implications can be derived from the empirical results. It is evident that the pandemic caused a sharp decrease in length of stay in P2P lodgings in the destinations, and managers and policy makers need to implement measures to increase this variable. Moreover, the results reveal two market segments corresponding to the two modes detected by the application of the model. That is, property managers can adopt different strategies directed to each market segment aimed at promoting amenities that increase their length of stay. For example, one such amenity is the pool, which increases length of stay for clients staying around 7 days, but not for those staying 1 or 2 days. Therefore, managers can obtain better returns by tailoring its promotion to one or the other market segment.
Certain limitations of this study should also be acknowledged. First, the empirical model ignores some factors influencing length of stay. Among them, we consider especially relevant the spatial effects. More specifically, relative prices among properties located nearby can better fit the length of stay since clients usually compare lodgings (prices and amenities) in a neighborhood before making a choice. Future research can follow this up. Second, the second period analyzed (December 2020) was an intra-COVID period, and the effect of the pandemic on length of stay could be transitory. It would be useful to conduct similar studies on length of stay on the basis of new and updated data in order to analyze the post-COVID period.

Author’s Note

The views expressed here are those of the authors and not necessarily those of the institution with which they are affiliated.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Emilio Gómez–Déniz was partially funded by grant ECO2017-85577-P (Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación, Spain); Juan María Hernández and Jorge V. Pérez–Rodríguez were funded by Universidad de Las Palmas de Gran Canaria (ULPGC, Spain) Program together with the Ministerio de Ciencia, Innovación y Universidades under grant GOB-ESP2019-07 and the ULPGC Program through grant COVID 19-04.

ORCID iDs

Emilio Gómez–Déniz https://orcid.org/0000-0002-5072-7908
Jorge V Pérez–Rodríguez https://orcid.org/0000-0002-6738-9191
Juan M Hernández https://orcid.org/0000-0001-6897-5179

Supplemental Material

Supplemental material for this article is available online.

Note

1. However, the researcher can restrict the specification considering several possibilities. For example, we can define a model where the conditional mean exclusively depends on explanatory variables, or where the conditional frequency modes only depend on exogenous factors. These restrictions will simplify the model to be estimated and focus on the determinants which really affect the mean and/or modes.

References

Alegre J and Pou L (2006) The length of stay in the demand for tourism. Tourism Management 27(6): 1343–1355.

Alegre J, Mateo S and Pou L (2011) A latent class approach to tourists’ length of stay. Tourism Management 32: 555–563.
Alén E, Nicolau J, Losada N, et al. (2014) Determinant factors of senior tourist’s length of stay. *Annals of Tourism Research* 49: 19–32.

Almeida A, Machado LP and Xu C (2021) Factors explaining length of stay: lessons to be learnt from Madeira Island. *Annals of Tourism Research Empirical Insights* 2(1): 100014. DOI: 10.1016/j.annale.2021.100014.

Baños-Pino JF, Boto-García D, Del Valle E, et al. (2021) The impact of COVID-19 on tourists’ length of stay and daily expenditures. *Tourism Economics*. Epub ahead of print 09 December 2021. DOI: 10.1177/13548166211053419.

Bavik A, Correia A and Kozak M (2021) What makes our stay longer or shorter? A study on Macau. *Journal of China Tourism Research* 17(2): 192–209.

Bohning D, Dietz E, Schlattmann P, et al. (1999) The zero–inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A* 162(2): 195–209.

Boto-García D, Baños-Pino JF and Álvarez A (2019) Determinants of tourists’ length of stay: a hurdle count data approach. *Journal of Travel Research* 58(6): 977–994.

Brida J, Pereyra J and Scuderi R (2014) Repeat tourism in Uruguay: modelling truncated distributions of count data. *Quality and Quantity* 48: 475–491.

Cameron A and Trivedi P (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1: 29–54.

Cameron C and Trivedi P (1998) *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Chen Y and Xie KL (2017) Consumer valuation of airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management* 29(9): 2405–2424.

Cohen AC (1966) A note on certain discrete mixed distributions. *Biometrics* 22: 566–572.

de Menezes AG, Moniz A and Vieira JC (2008) The determinants of length of stay of tourists in the Azores. *Tourism Economics* 14(1): 205–222.

Denuit M, Maréchal X, Pitrebois S, et al. (2009) *Actuarial Modelling of Claim Counts Risk Classification, Credibility and Bonus-Malus Systems*. Germany: John Wiley & Sons.

Ert E, Fleischer A and Magen N (2016) Trust and reputation in the sharing economy: the role of personal photos in Airbnb. *Tourism Management* 55: 62–73.

Eugenio-Martin JL and Inchausti-Sintes F (2016) Low-cost travel and tourism expenditures. *Annals of Tourism Research* 57: 140–159.

Famoye F and Singh K (2006) Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science* 4: 117–130.

FRONTUR (2021) Series mensuales de entradas de turistas y excursionistas. Islas de Canarias. 2012-2021 (Metodología 2016). Retrieved from: http://www.gobiernodecanarias.org/istac/jaxi-istac/menu.do?uripub=urn:uuid:ccdf465c-2230-421d-99f6-d6a1669d6032 (accessed 1 September 2021).

Gokovali U, Bahar O and Kozak M (2007) Determinants of length of stay: a practical use of survival analysis. *Tourism Management* 28: 736–746.

Gómez-Déniz E and Pérez-Rodríguez J (2019) Modelling bimodality of length of tourist stay. *Annals of Tourism Research* 75: 131–151.

Gómez-Déniz E, Gallardo D and Gómez H (2020a) Quasi-binomial zero-inflated regression model suitable for variables with bounded support. *Journal of Applied Statistics* 47(12): 2208–2229. DOI: 10.1080/02664763.2019.1707517.

Gómez-Déniz E, Pérez-Rodríguez J, Reyes J, et al. (2020b) A bimodal discrete shifted Poisson distribution. a case study of tourists’ length of stay. *Symmetry* 12(3): 1–15.

Gómez-Déniz E, Boza-Chirino J and Dávila-Cárdenes N (2021) Tourist tax to promote rentals of low-emission vehicles. *Tourism Economics* 27(7): 1461–1481.
Gossling S, Scott S and Hall CM (2018) Global trends in length of stay: implications for destination management and climate change. *Journal of Sustainable Tourism* 26(12): 2087–2101.

Gurmu S and Trivedi P (1996) Excess zeros in count model for recreational trips. *Journal of Business & Economic Statistics* 14(4): 469–477.

Guttentag DA and Smith SL (2017) Assessing Airbnb as a disruptive innovation relative to hotels substitution and comparative performance expectations. *International Journal of Hospitality Management* 64: 1–10.

Hellstrom J (2006) A bivariate count data model for household tourism demand. *Journal of Applied Econometrics* 21: 213–226.

ISTAC (2021) Encuestas de alojamiento turístico [tourist accommodation survey]. Retrieved from: http://www.gobiernodecanarias.org/istac (accessed 1 September 2021).

ISTAC (2022) Encuestas de alojamiento turístico [tourist accommodation survey]. Retrieved from: http://www.gobiernodecanarias.org/istac (accessed 1 June 2022).

Jeon CY and Yang HW (2021) The structural changes of a local tourism network: comparison of before and after COVID-19. *Current Issues in Tourism* 24(23): 3324–3338.

Lambert D (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34(1): 1–14.

Liu W, Tang Y and Xu A (2018) A zero-and-one inflated Poisson model and its application. *Statistics and Its Interface* 11: 339–351.

Martínez-García M and Raya JM (2008) Length of stay for low cost tourism. *Tourism Management* 29: 1064–1075.

Moreno-Izquierdo L, Ramón-Rodríguez AB, Such-Devesa MJ, et al. (2019) Tourist environment and online reputation as a generator of added value in the sharing economy: the case of Airbnb in urban and sun-and-beach holiday destinations. *Journal of Destination Marketing & Management* 11: 53–66.

Murphy L, Mascardo G and Benckendorff P (2007) Exploring word-of-mouth influences on travel decisions: friends and relatives vs. other travellers. *International Journal of Consumer Studies* 31(5): 517–527.

Palmer-Tous T, Riera-Font A and Roselló-Nadal J (2007) Taxing tourism: the case of rental cars in Mallorca. *Tourism Management* 28: 271–279.

Poon KY and Huang WJ (2017) Past experience, traveler personality and tripographics on intention to use Airbnb. *International Journal of Contemporary Hospitality Management* 29(9): 2425–2443.

Rodríguez X, Martínez-Roget F and González-Murias P (2018) Length of stay: evidence from Santiago de Compostela. *Annals of Tourism Research* 68: 9–19.

Salmasi L, Celidoni M and Procidano I (2012) Length of stay: price and income semi-elasticities at different destinations in Italy. *International Journal of Tourism Research* 14: 515–530.

Santos N and Moreira C (2021) Uncertainty and expectations in Portugal’s tourism activities. Impacts of COVID-19. *Research in Globalization* 3: 100071.

Santos GEDO, Ramos V and Rey-Maquieira J (2015) Length of stay at multiple destinations of tourism trips in Brazil. *Journal of Travel Research* 54(6): 788–800.

Tang LR, Kim J and Wang X (2019) Estimating spatial effects on peer-to-peer accommodation prices: towards an innovative hedonic model approach. *International Journal of Hospitality Management* 81: 43–53.

Thrane C (2012) Analyzing tourists’ length of stay at destinations with survival models: a constructive critique based on a case study. *Tourism Management* 33: 126–132.

Tussyadiah IP and Pesonen J (2016) Impacts of peer-to-peer accommodation use on travel patterns. *Journal of Travel Research* 55(8): 1022–1040.

Van Truong N, Shimizu T, Kurihara T, et al. (2020) Generating reliable tourist accommodation statistics: bootstrapping regression model for overdispersed long-tailed data. *Journal of Tourism, Heritage & Services Marketing* 6(2): 30–37.
Vuong Q (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.

Wang D and Nicolau JL (2017) Price determinants of sharing economy based accommodation rental: a study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management* 62: 120–131.

Yang Y and Zhang HL (2015) Modeling tourists’ length of stay: does one model fit all? *Tourism Analysis* 20: 13–23.

Zhang C, Tian G-L and Ng K-W (2016) Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods. *Statistics and Its Interface* 9: 11–32.

Zheng D, Luo Q and Ritchie BW (2021) Afraid to travel after COVID-19? Self-protection, coping and resilience against pandemic ’travel fear’. *Tourism Management* 83: 104261.

**Author biographies**

**Emilio Gómez-Déniz** is professor of Mathematics at University of Las Palmas de Gran Canaria (Spain). His research focuses on the distributions theory, Bayesian statistics, robustness, Bayesian applications in economics with emphasis in actuarial statistics (credibility, ruin theory, etc.).

**Jorge V Pérez–Rodríguez** is professor of Econometrics at Department of Quantitative Methods (University of Las Palmas de Gran Canaria, Spain). His research focuses on the tourism economics and financial statistics and econometrics.

**Juan M Hernández** is professor at the University of Las Palmas de Gran Canaria, Spain. Currently, his main line of research is the use of the complex systems perspective for the analysis of tourism and fishery economics.