Applied Spatial Analysis and Polic (2021) 14:631–658
https://doi.org/10.1007/s12061-020-09370-3

Spending and length of stay by tourists flying to the Canary Islands (Spain) using low-cost carriers

Emilio Gómez-Déniz[1] · **Jorge Vicente Pérez-Rodríguez**[2]

Received: 13 May 2020 / Accepted: 23 November 2020 / Published online: 18 July 2021

Abstract
The AIM of this study is to examine and evaluate differences in expenditure and length of stay between tourists who use low-cost carriers and those who travel with full service providers. We consider the statistical dependence between these variables and propose a bivariate distribution that describes tourist expenditure (continuous variable) and length of stay (discrete variable) in terms of their conditional distributions. Covariates are included to reflect the factors that simultaneously affect both variables. In addition, an empirical analysis is made of data obtained by the Canary Islands Tourist Expenditure Survey. The results obtained show that our model achieves a reasonably good fit and that there are differences between LCC and FSC users regarding both expenditure and length of stay, in the use of nonhotel accommodation, as well as differences in expenditure in the case of repeated visits, and in the length of stay according to the visitors' age, nationality and travel party size.

Keywords Conditional distributions · Bivariate distributions · Tourism

Introduction

Tourist expenditure and length of stay are the main elements of travellers' tourism-related decisions. However, they are also of crucial importance to planners and managers and to the economies of tourist destination regions or countries. The economic impact of tourism is strongly dependent on these factors, and promotional campaigns need to be matched to tourists' decisions in this respect. An increase in

✉ Emilio Gómez-Déniz
emilio.gomez-deniz@ulpgc.es

1 Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017, Las Palmas de Gran Canaria, Spain

2 Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017, Las Palmas de Gran Canaria, Spain

the volume of travellers, for example, could require significant investment to increase capacity, an issue of major concern to policy makers.

A question of interest in this context is that of how the above variables are affected by the type of airline used to reach the holiday destination, i.e. whether low-cost carrier (LCC) passengers have different preferences and patterns of behaviour from traditional carrier or full-service carrier (FSC) passengers. Moreover, low-cost travellers are usually less flexible than traditional tourists in organising their vacation, because they suffer more severe budget restrictions and their decisions about travel dates, length of stay and trip expenditure are often more subject to external pressures than is the case of FSC users. Accordingly, LCC travellers are expected to present differences in their sociodemographic, travel and destination characteristics. Understanding these differences is important to tourist industry managers, and policymakers in the corresponding regional governments would also benefit from this information (Raya & Martínez-García, 2011).

Previous studies conducted in this area have investigated differences in behaviour patterns between LCC and FSC users using micro-data and studying factors affecting expenditure and length of stay independently. For example, Martínez-García and Raya (2008) and Raya and Martínez-García (2011) considered the impact on length of tourist stay, while Eugenio-Martin and Inchausti-Sintes (2016) and Ferrer-Rosell and Coenders (2017), among others, analysed the association between type of carrier and tourist expenditure. In the latter respect, empirical research into the relation between LCC travel and tourist expenditure has produced the following conclusions. Martínez-García and Raya (2008), Ferrer-Rosell and Seetaram (2014), and Eugenio-Martin and Inchausti-Sintes (2016) reported that LCC users spend more at their destination while Ferrer-Rosell and Coenders (2017) observed that users of LCC and FSC converge in their allocation of the vacation budget (between transportation and at-destination expenses, and within at-destination expenses), but diverge with regard to total trip expenditure. The latter findings provide information about how airline type convergence can be translated into tourist spending behaviour and about how the two types of airline users have converged. The authors observe, moreover, that business model convergence might reduce the differences between different types of airline users.

On the other hand, Ferrer-Rosell et al. (2014) and Ferrer-Rosell et al. (2015) recorded only small differences between LCC and FSC users in their respective analyses of the determinants of length of stay and expenditure allocation (among transportation, accommodation and other items).

However, in this paper we study both expenditure and length of stay from a different perspective, seeking to assess differences between LCC and FSC users. Accordingly, the present study makes two main contributions to the literature on LCC.

First, our paper focuses on the bidirectional causality between expenditure and length of stay, in order to analyse differences according to the type of airline (for example, the varying behaviour patterns of LCC and FSC users). To do so, we jointly model total tourist expenditure and length of stay, taking into account both dependence and simultaneity.

Empirical tourism research has established the existence of bidirectional causality between these variables, as a statistical finding. For example, duration is a function of expenditure (Hellström, 2006) and expenditure is determined by duration (Thrane, 2014; Pérez-Rodríguez & Ledesma-Rodríguez, 2019), among others.[1] In fact, it can be shown empirically that there is a weak positive correlation between aggregate expenditure and length of stay (Gómez-Déniz & Pérez-Rodríguez, 2019).

Our underlying theoretical model for total expenditure and length of stay has the antecedent of the recent work by Gómez-Déniz and Pérez-Rodríguez (2019) and Gómez-Déniz and Pérez-Rodríguez (2020). Gómez-Déniz and Pérez-Rodríguez (2019) proposed a univariate compound model, based on previous studies of mixed distributions, simultaneously implementing both length of stay and individual expenditure in order to determine aggregate tourist expenditure. In another approach, (Gómez-Déniz & Pérez-Rodríguez, 2020) proposed a bivariate model for the two variables which allowed correlation of any sign between them, but with given marginals whose means were not mutually dependent. Neither of these previous models, therefore, took account of dependency or of simultaneity, in contrast to the model we present, which allows the mean of each of the variables to be conditioned by the mean of the other.

Specifically, we model dependence by means of a bivariate structural model that describes tourist expenditure (continuous variable) and length of stay (discrete variable) in terms of their conditional distributions (Arnold & Strauss, 1991; Arnold et al., 1999). Simultaneity is introduced into the bivariate distribution, by making two assumptions: the conditional expectation of expenditure with respect to length of stay, and the conditional expectation of length of stay with respect to expenditure. By adopting a linear relationship for these conditional expectations, we assume that more expenditure implies a greater duration, and vice versa.

The idea underlying this approach is that we wish to obtain a bivariate model in which the mean of each variable depends on the value taken by the other. Although the statistical literature contains numerous methods for obtaining bivariate distributions, such as copulas, the conditional specification seems the most appropriate. In general, the models obtained in this way are based on formulations that incorporate a large number of parameters, but this requirement can be relaxed, as we show below, to obtain simple and practical formulations. The major advantage of these formulations, compared to those obtained by using copulas, is that they incorporate a parameter that controls the correlation, which is otherwise difficult to estimate because it must move within a range of allowed values.

The second major contribution of the present study is that the model proposed enables us to evaluate the effect of covariates on both expenditure and duration. Thus, we can distinguish between the effects of LCC and FSC and determine which variables simultaneously affect both expenditure and length of stay. Furthermore, the proposed model is well suited to marginally capture the skewness that may be present,

[1] It is noteworthy that the length of stay is considered as an additional argument in the underlying utility function see, for example, Hellström (2006), and is considered as an endogenous variable. Therefore, models which employ length of stay as a regressor and estimate the corresponding equation by OLS are open to question because duration is an 'endogenous' independent variable see also Thrane (2015).

as well as the long tail to the right that these two study variables tend to present in practice.

The model was evaluated empirically, based on data obtained from the 2017 Canary Islands Tourist Expenditure Survey, which provides information on tourists' sociodemographic, vacation and destination characteristics. Moreover, it enables us to distinguish LCC from FSC travellers.

Empirical findings show that our model provides a reasonable fit, and that there are differences between LCC and FSC users regarding both expenditure and length of stay in terms of non–hotel accommodation, between LCC and FSC users in terms of total expenditure during a repeated visit, and differences between LCC and FSC users in terms of length of stay according to the visitors' age, nationality and travel party size.

The rest of this paper is organised as follows. Section"Low-cost carriers operating in the Canary Islands" summarises the characteristics of LCCs flying to the Canary Islands. Section "Bivariate distribution based on conditional expectation" describes the proposed model for tourist expenditure and length of stay, together with the properties of the econometric model with dependence. A bivariate model incorporating the covariates derived is provided in "Bivariate regression model for (X, N) based on conditionals".Section "Empirical analysis" then presents the empirical analysis performed and the results obtained. Finally, "Conclusions" summarises the main conclusions drawn.

Low-cost carriers operating in the Canary Islands

Located over 1,100 kilometres south of mainland Spain, the Canary Islands have long been a popular winter sun destination for Northern European leisure travellers, especially those from the UK and Germany.

The growth in annual visitors to the Canary Islands, from 12 million in 2010 to 18 million in 2018, has been facilitated by emerging air service markets and by the use of a new route incentive fund (the Flight Development Fund) to support new air links in targeted markets, thus promoting the diversification of source markets and significant growth in arrivals from other origins, such as France, Italy and Poland.

The growing presence of low-cost carriers (LCCs) has given rise to a new business market model, with lower prices and greater competition among airlines (among full-service carriers, FSCs, too), facilitated by the advance of the Internet, which enables tourists to buy flights online, easily, quickly and more economically.

The expansion in LCC activity in the tourism market has had significant, positive effects, both direct and indirect, on the Spanish economy (Rey et al., 2011).[2]

The number of passengers at the five airports (Fuerteventura, Lanzarote, Gran Canaria, Tenerife Sur and Tenerife Norte) serving the seven main islands of the archipelago has grown considerably in recent years. An important element of this

[2]For an analysis of LCC effects on air transport demand for tourists visiting Spain, see Aguiló et al. (2007), and Rey et al. (2011).

expansion is the advance of specialist charter carriers, whose typically larger capacity aircraft have enabled them to counterbalance the islands' distance from the main international travel markets and thus to dominate the market, overtaking the major LCCs that operate within Europe. Among the 17 Spanish autonomous communities (i.e., self-governing regions), Catalonia, in mainland Spain, attracted most LCC passengers in the first two months of 2017, followed by the Canary Islands.

Many carriers, including LCCs, operate more than one daily flight to the Canary Islands and large numbers of direct flights arrive from all over Europe. Las Palmas, on the island of Gran Canaria, is the busiest of the Canary Islands airports, and has handled in excess of ten million passengers in each of the last three years. The next busiest is Tenerife Sur, where eight to nine million passengers arrive every year.

One of the largest international LCCs operating in the Canary Islands is Ryanair, which has been among the main drivers of the archipelago's enhanced connectivity, together with Monarch, Norwegian, Vueling, easyJet and Germanwings.

Figure 1 shows the year-on-year variation in the volume of passengers flying to the Canary Islands in the period 1999-2019, including FSCs, LCCs and charter flights, and both international and mainland Spanish travellers. The figure highlights a significant feature of the patterns observed, namely that airline passenger numbers with FSCs and LCCs increased during the study period, while those of charter passengers fell. It is noteworthy that LCC passenger numbers have continued to rise and, in some cases, have overtaken those of the FSCs. In 2017, for example, according to the Tourist Expenditure Survey conducted by the Government of the Canary Islands, LCCs accounted for around 53% of all flights to the islands. In the first quarter of 2006, this share was only 19.88%, rising to 38.53% by the first quarter in 2014 see Eugenio-Martin and Inchausti-Sintes (2016). This pattern of growth clearly reflects the changing market structure and underlines the growing importance of LCCs in the Canary Islands destination market. The trend appears to be continuing, and estimates suggest that in 2017 as a whole, LCCs had 59% of the total market.

To our knowledge, only Eugenio-Martin and Inchausti-Sintes (2016) have investigated the factors relevant to tourist travel by LCC passengers to the Canary Islands. These authors studied tourist expenditure at origin and at destination, but did not consider the length of stay or the relation between length of stay and expenditure, taking into account differences between the types of airline used.

Bivariate distribution based on conditional expectation

Various econometric methods have been used to analyse tourist expenditure and length of stay, for example linear regression analysis and ordinary least squares (OLS)estimates. Length of stay has also been analysed using non-parametric and parametric duration (survival) analysis (Martínez-García & Raya 2008, 2011), among others. In the present paper, however, we focus on bivariate distributions based on conditional expectations.

In this section, we propose a non-linear bivariate model of tourist expenditure and length of stay which takes into account positive dependence between these variables. The model is built using a characterisation by conditional expectations, creating a

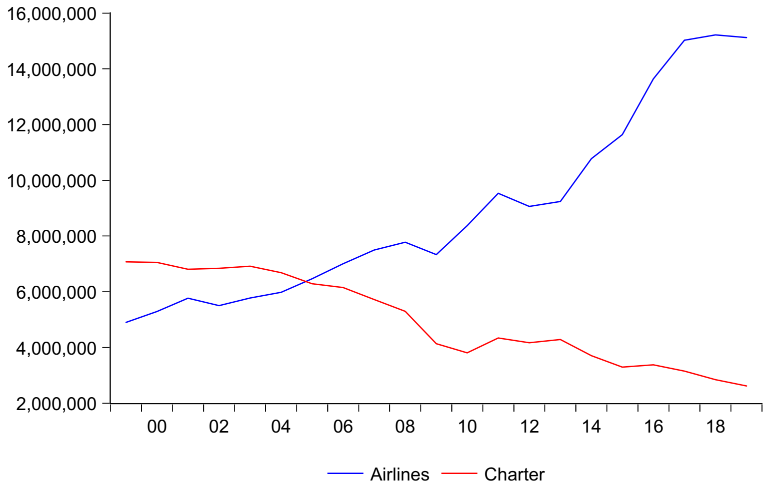

Fig. 1 Time path for airline (regular) and charter (non–regular) passengers to the Canary Islands (1999-2019). Source: Instituto Canario de Estadística, ISTAC)

class of bivariate distributions such that the conditional distributions belong to a specified exponential family.

Multivariate distributions can be specified through their conditional distributions rather than directly.[3] These conditional methodologies are comprehensively discussed in Arnold et al. (2001). If we assume that conditional distributions belong to certain parametric families of distributions, the joint distribution can be obtained as described in Arnold et al. (1999).[4] To obtain the joint distribution, it is first necessary to determine the resolution of certain functional equations, which facilitates highly flexible multiparametric distributions. As Arnold and Strauss (1991) point out, when we wish to specify a bivariate distribution it is sometimes convenient to visualise conditional distributions rather than marginal or joint distributions. In this context, it is useful in statistical modelling to have tractable multivariate distributions with given marginals in order to quantify the dependence effect of the variables in the model. Therefore we present this possibility, using given conditional distributions.

Accordingly, let us assume that tourist expenditure is represented as a random variable X and that the length of stay is also random, and denoted by N. We now wish to obtain the more general bivariate distribution (X, N) whose conditional distributions satisfy the following conditions based on expenditure conditional to the length of stay and also on the length of stay conditional to expenditure.

On the one hand, the expense incurred by an individual tourist at the destination depends on the length of stay see for example, Thrane (2014) and Pérez-Rodríguez and Ledesma-Rodríguez (2019). That is, if $X|N = n$ is the expenditure conditional

[3] To do this, the dependence structure can be modelled using bivariate copulas or other mathematical and statistical methods such as conditional distributions or mixing distributions.

[4] See also Arnold and Strauss (1991) for an introduction to this topic, and for applied works in this setting, see Sarabia et al. (2004), Sarabia et al. (2005), and Gómez-Déniz and Calderín (2014), among others.

to $N = n$ days spent at the destination, the expectation of the expenditure is linearly dependent on the length of stay. Hence,

$$E(X|N = n) = \alpha_1 + \alpha_2 n, \quad \alpha_1 > 0, \quad \alpha_2 > 0. \tag{1}$$

According to expression (1), assuming that n takes values within the set of integer numbers $\{1, 2, \ldots, n\}$ there is a minimum expense given by $\alpha_1 + \alpha_2$ in the case of a single-day stay. This expense increases as the length of stay increases.

On the other hand, it also seems logical to assume that the number of days spent at a tourist destination will depend on the amount spent by the tourist. For example, Hellström (2006) observed that the choice of the total number of nights spent at a destination is conditioned by the financial cost of travel and the financial cost incurred per day at the destination, among other factors. Hence,

$$E(N|X = x) = \alpha_3 x, \quad \alpha_3 > 0. \tag{2}$$

Note that Eq. 1 has an intercept but Eq. 2 does not. Apart from facilitating the construction of the mathematical model obtained from Eq. 2, it seems obvious that an initial expenditure of near-zero monetary units implies that the mean length of stay, and therefore the conditional mean, will also be close to zero.

In the present study, we seek to construct a bivariate distribution that satisfies the above conditions, taking into account that models constructed in this way are usually subject to the difficulty that both the marginal distributions and the joint distribution depend on a normalisation constant that is sometimes impossible to derive in a closed-form expression. In this respect, see for instance (Moschopoulos & Staniswalis, 1994). This difficulty is especially acute in the case we consider, in which one of the two variables is continuous and the other discrete.

A discrete random variable is said to follow a shifted Poisson distribution with parameter $\lambda > 0$ if its probability function (pf) is written as

$$f_N(n) = \Pr(N = n) = \frac{\lambda^{n-1}\exp(-\lambda)}{\Gamma(n)}, \quad n = 1, 2, \ldots \tag{3}$$

In the following, $N \sim \mathcal{SP}o(\lambda)$ denotes a random variable following a shifted Poisson distribution with a pf as given in Eq. 3. Moreover, a continuous random variable follows a gamma distribution with shape parameter $\zeta > 0$ and scale parameter $\beta > 0$ if its probability density function (pdf) is expressed as

$$f_X(x) = \frac{\beta^\zeta}{\Gamma(\zeta)} x^{\zeta-1}\exp(-\beta x). \tag{4}$$

In this case, $X \sim \mathcal{G}(\zeta, \beta)$. More flexible distributions than those given in Eqs. 3 and 4 can also be considered. For example, it can easily be confirmed that the relationship between the variance and the mean (dispersion index) for the shifted Poisson distribution is less than 1, which is undesirable for modelling the empirical distribution of the length of stay variable. In this case, the shifted negative binomial distribution could be used instead (3). For the distribution given in Eq. 4, the Weibull could also be used, although the method employed to obtain the bivariate distribution would have to be modified, since this distribution is not a member of the exponential family of distributions (see [Chapter 4, p. 88]arnoldetal1999 for details).

The following result provides the most general bivariate distribution with conditionals as given in Eqs. 5–6.

Theorem 1 *Assume that*

$$X|N = n \sim \mathcal{G}(\sigma(n), \eta(n)), \tag{5}$$

$$N|X = x \sim \mathcal{SP}o(\varphi(x)) \tag{6}$$

for given functions $\varphi(x) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, $\sigma(n) : \mathbb{N}^* \longrightarrow \mathbb{R}^+$ *and* $\eta(n) : \mathbb{N}^* \longrightarrow \mathbb{R}^+$, *being* $\mathbb{N}^* = \{1, 2, \dots\}$. *Then, the most general bivariate distribution with conditionals given in* Eqs. 5–6 *is given by*

$$\begin{aligned} f_{X,N}(x, n) = \frac{1}{x\Gamma(n)} \exp\Big\{& m_{00} + m_{10}n^* - (m_{01} + m_{11}n^*)x \\ & +(m_{02} + m_{12}n^*) \log x\Big\}, \end{aligned} \tag{7}$$

for $x > 0$, $n = 1, 2, \dots$, *where* $n^* = n - 1$, $m_{01} > 0$, $m_{02} \geq 0$, $m_{10} \geq 0$, $m_{11} \geq 0$, $m_{12} \geq 0$ *and the parameter* m_{00} *is the normalising constant, which is a function of the remaining parameters.*

Proof Observe that the conditional distributions given in Eqs. 5–6 can be rewritten as

$$f_{N|X}(N|X = x) = \frac{1}{\Gamma(n)} \exp\{-(n-1) \log \varphi(x) - \varphi(x)\},$$

$$f_{X|N}(X|N = n) = \frac{\eta(n)^{\sigma(n)}}{\Gamma(\sigma(n))} \exp\{-\eta(n)x + (\sigma(n) - 1) \log x\},$$

i.e. as members of the exponential family of distribution. Then the result is a simple particular case of Theorem 4.1 in Arnold et al. (1999). That is, the bivariate distribution has the following form,

$$f_{X,N}(x, n) = \frac{1}{x\Gamma(n)} \exp\left\{(1 \;\; n-1)\, \boldsymbol{M} \begin{pmatrix} 1 \\ -x \\ \log x \end{pmatrix}\right\}, \tag{8}$$

where $x > 0$, $n = 1, 2, \dots$ and $\boldsymbol{M} = (m_{ij})_{\substack{i=0,1 \\ j=0,1,2}}$ is a matrix with dimension 2×3 whose elements are constant parameters $m_{01} > 0$, $m_{02} \geq 0$, $m_{10} \geq 0$, $m_{11} \geq 0$, $m_{12} \geq 0$, m_{00} is the normalising constant given by

$$m_{00} = -\log \int_0^\infty \sum_{n=1}^{\infty} \frac{x^{m_{02}+m_{12}n^*-1}}{\Gamma(n)} \exp\left\{m_{10}n^* - (m_{01} + m_{11}n^*)x\right\} dx \tag{9}$$

and

$$\sigma(n) = m_{02} + m_{12}n^*, \tag{10}$$

$$\eta(n) = m_{01} + m_{11}n^*, \tag{11}$$

$$\varphi(x) = \exp\{m_{10} - m_{11}x + m_{12} \log x\}. \tag{12}$$

Hence the Theorem. □

The normalising constant given in Eq. 9 can also be obtained from one of the conditional distributions given in Eqs. 5–6.

Thus, the conditional distribution (5) of X given N is gamma with the parameters given in Eq. 10–11 and the conditional distribution (6) of N given X is shifted Poisson with the parameter (12). The distribution obtained when the conditional distributions are Poisson and gamma was initially considered by (Arnold et al., 1999, Chapter 4, p. 98) but the distribution described in Eq. 7, as far as we know, has not been studied in the statistical literature.

From Eq. 9, it is clear that computing the normalising constant may be difficult or even impossible. However, a univariate integration rule can sometimes be used to obtain Eq. 9. In this case, Gauss-Hermite rules see for instance (Davis & Rabinowitz, 1984) could be used to approximate the normalising constant. Nevertheless, an appropriate choice of the parameter could enable us to obtain a closed-form expression for the normalising constant. For example, the case $m_{11} = m_{12} = 0$ corresponds to that in which N and X are independent and the marginal distributions are shifted Poisson and gamma as in Eqs. 3 and 4, respectively. The case $m_{10} = m_{11} = 0$, $m_{12} = 1$ also provides a closed model with conditional mean $E(X|N = n)$ in the form of Eq. 1 but with the conditional mean of N given $X = x$, which results in x. The third submodel, which we consider in this analysis, is obtained for $m_{11} = 0$ and $m_{12} = 1$ from which the conditional means are of the type of Eq. 1 and 2 and thus suitable for our purpose.

In order to simplify the notation, we take $m_{02} = \alpha$, $m_{01} = \beta$ and $m_{10} = \log(\beta - \gamma)$, with $\beta > \gamma$. Then, the bivariate distribution given in Eq. 7 can be rewritten as

$$f_{X,N}(x, n) = \frac{(\gamma x)^{\alpha}[x(\beta - \gamma)]^{n-1} \exp(-\beta x)}{x\Gamma(n)\Gamma(\alpha)}, \quad x > 0, \ n = 1, 2, \ldots \tag{13}$$

Note that the bivariate distribution (13) has marginal distributions, one with continuous support and the other discrete. This type of bivariate distribution is uncommon in theoretical and applied statistical literature, allowing us to model phenomena that are not common in practice but which do occur, as in the present case. See for example, (Kotz et al., 2000), (Spanos, 1999) and (Gómez-Déniz & Calderín, 2014). Marginal, conditional distributions and marginal moments are detailed in the Appendix. The population correlation between the two variables is

$$\varrho(X, N) = \sqrt{1 - \frac{\gamma}{\beta}},$$

which is always positive and bounded between 0 ($\gamma \approx \beta$) and 1 ($\gamma \approx 0$). Thus, parameter γ controls the dependence or independence of the model.

Observe (see the Appendix) that with the assumption of these two conditional distributions expressions (1) and (2) are guaranteed.

Both the moment method and the maximum likelihood method appear to be feasible means of estimating the vector of parameters $\boldsymbol{\phi} = (\alpha, \beta, \gamma)$ of the distribution through sample observations, as shown in the Appendix.

Using a sample from the bivariate distribution, hypothesis testing can be performed with the parameters α, β and γ. We may also be interested in determining when the model might depend only on two parameters, i.e. for example when

$\beta = \gamma + 1$. In this case, (from the equations provided in the Appendix) the vector $\boldsymbol{\phi}$ can be estimated with the constraint that $\beta = \gamma + 1$. If we represent the new vector by $\boldsymbol{\phi}^*$, the critical region for the null hypothesis $H_0 : \beta = \gamma + 1$ is given by

$$2\left[\ell(\widehat{\boldsymbol{\phi}}; (\tilde{x}, \tilde{n})) - \ell(\widehat{\boldsymbol{\phi}^*}; (\tilde{x}, \tilde{n}))\right]$$

which asymptotically has a χ^2-squared distribution with one degree of freedom.

Bivariate regression model for (X, N) based on conditionals

In this section, we introduce a more realistic model in which covariates are included. The linear regression model, which makes no distributional assumptions, is likely to be unsatisfactory because certain combinations of parameters and regressors could violate the nonnegative restriction on the mean. To avoid this situation we propose a parametric model based on using the distributional assumptions presented in the previous section.

When a regression analysis is to be performed, it is often useful to model the mean of the response, which in the present case is the marginal mean. In Eq. 13 α is replaced by $\gamma\mu_1$ and β is replaced by $\gamma + (\mu_2 - 1)/\mu_1$, where $\mu_1 > 0$ and $\mu_2 \geq 1$. Then, the pdf (13) can be rewritten as

$$f(x, n) = \frac{(x\gamma)^{\gamma\mu_1}}{x\Gamma(n)\Gamma(\gamma\mu_1)} \left(\frac{x(\mu_2 - 1)}{\mu_1}\right)^{n-1} \exp\left[-\frac{x}{\mu_1}(\mu_2 + \gamma\mu_1 - 1)\right], \quad (14)$$

for which $E(X) = \mu_1$ and $E(N) = \mu_2$. After this reparameterisation we also obtain the cross moment, the covariance and the correlation, which are given by

$$\begin{aligned} E(XN) &= \mu_1\mu_2 + \frac{1}{\gamma}(\mu_2 - 1), \\ cov(X, N) &= \frac{\mu_2 - 1}{\gamma}, \\ \varrho(X, N) &= \sqrt{\frac{\mu_2 - 1}{\gamma\mu_1 + \mu_2 - 1}}, \end{aligned}$$

respectively, and therefore the pdf given in Eq. 14 is appropriate for including covariates. We write $(X, N) \sim \mathcal{BGSPoC}(\gamma, \mu_1, \mu_2)$ to denote a bivariate random variable (X, N) following the pdf given in Eq. 14. This bivariate distribution satisfies the condition that the conditional distributions are gamma and shifted Poisson. Graphs of the density function for different parameter values and their corresponding contour plots are shown in Fig. 2, revealing the presence of a wide range of densities.

Now, let $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{ki})'$ and $\boldsymbol{z}_i = (z_{1i}, \ldots, z_{ki})'$ be two vectors of k covariates associated with the ith observation. These are two vectors of linearly independent regressors that are thought to determine (x, n). For the ith observation, the model takes the form

$$\begin{aligned} (X_i, N_i) &\sim \mathcal{BGSPoC}(\gamma, \mu_{1i}, \mu_{2i}), \\ \log(\mu_{1i}) &= \boldsymbol{y}_i'\boldsymbol{\delta}, \\ \log(\mu_{2i} - 1) &= \boldsymbol{z}_i'\boldsymbol{\eta}, \end{aligned}$$

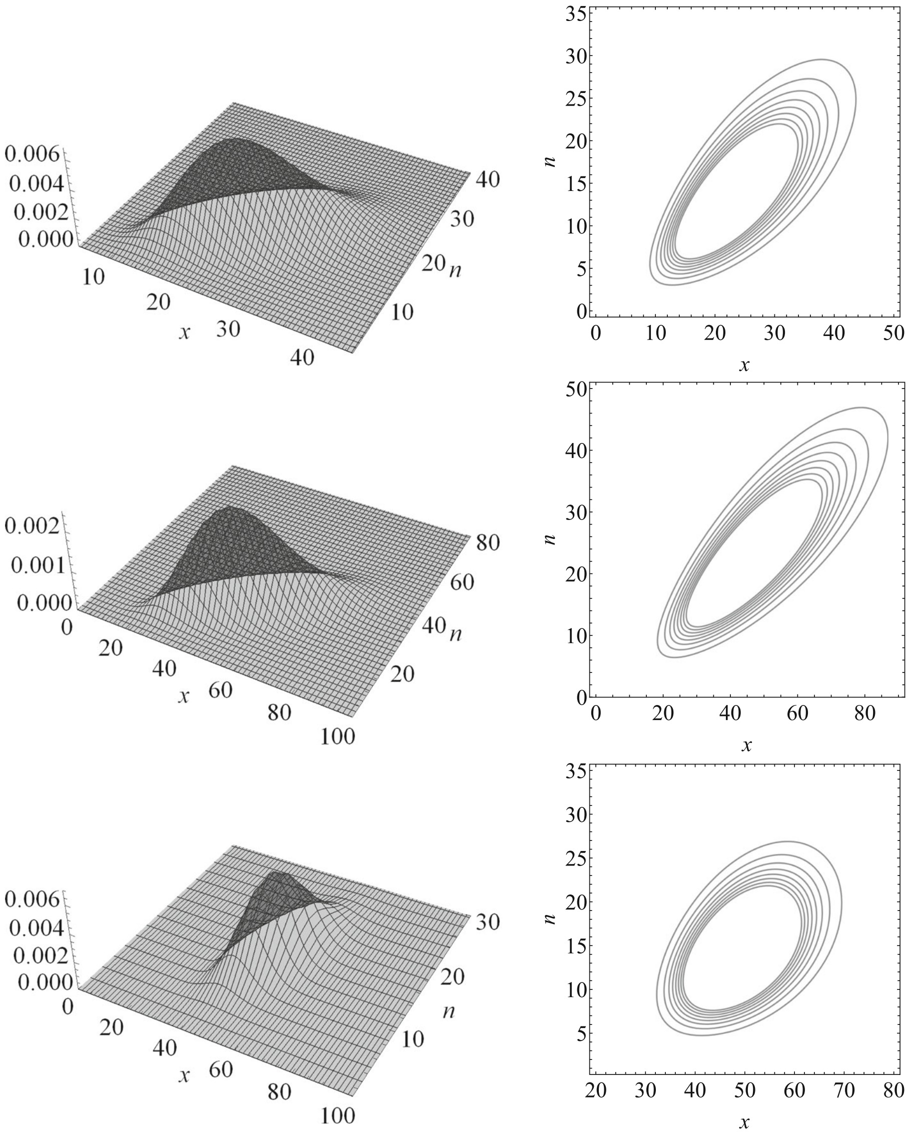

Fig. 2 Graphs of pdf for the bivariate distribution $f(x, n)$ in Eq. 14 and the corresponding contour plots. From top to bottom (μ_1, μ_2, γ) are given by (25, 15, 0.5), (50, 25, 0.25) and (50, 15, 1), respectively

for $i = 1, \ldots, t$ and where t denotes the number of observations and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)'$ the corresponding vectors of regression coefficients. In principle, each of the variables, X and N, can be influenced by different factors, hence the explanatory variables that are taken to explain $\mu_{\kappa i}$, $\kappa = 1, 2$, are not the same. Furthermore, observe that the logit link assumed ensures that μ_{1i} falls within the interval $(0, \infty)$ and μ_{2i} within the interval $(1, \infty)$.

Under this model the log-likelihood function takes the form given in the Appendix, in which the normal equations used to provide the estimators of the parameters are also shown. These are given in almost closed-form expression. The above model has the advantage of simplicity, in contrast to the normal equations, which require the use of the digamma function, $\psi(z) = \frac{d}{dz}\log(\Gamma(z)), \ z > 0$, to estimate the model parameters. Therefore, it is convenient to replace this derivative by using the approximation given by Eq. 22, shown in the Appendix.

Empirical analysis

Data

The database used was obtained from the 2017 Canary Island Tourist Expenditure Survey (Encuesta de Gasto Turístico). This survey is based on personal interviews with tourists on departure (i.e., domestic and foreign tourists and day-trippers who enter the Canaries at the airport) and is carried out by the Canary Islands Institute of Statistics (ISTAC). It provides information about tourists' total expenditure in the Canary Islands (composed of Gran Canaria, Fuerteventura, Lanzarote, Tenerife, La Palma, Gomera and Hierro). Many nationalities are represented, including visitors from mainland Spain, and from Germany, Austria, Belgium, Denmark, Finland, France, Netherlands, Ireland, Italy, Norway, Poland, Portugal, United Kingdom, Czech Republic, Russia, Sweden, Switzerland, Luxembourg and Others. In this study, we consider both package and non-package tourists, who stay for at least one night and for no more than 180 consecutive nights, and travel to the islands by LCC or FSC.[5] After filtering this database to exclude data with missing values and non-response, 22921 observations remained.

The following variables were included in the analysis:

1. Length of stay (number of days) in the Canary Islands.
2. Expenditure in the country of origin (i.e., flights and accommodation) and at the destination (€) on items such as restaurants, leisure and transport within the Canary Islands.
3. Low-cost carrier (LCC). A dummy variable which takes the value 1 if the tourist visit was made using a low-cost carrier, and 0 otherwise (traditional airlines or full-service carriers).
4. Household income. Two income variables are considered: High income, which takes the value 1 for incomes greater than 72001 euros, and medium income, which takes the value 1 for incomes between 24001 and 72000 euros, and 0 otherwise.
5. Travel party size. The number of persons composing the holiday package paid for in the country of origin.
6. Repetition. This dummy variable takes the value 1 if the respondent has previously visited the Canary Islands and 0 otherwise.
7. Age of the survey respondent.

[5] It is noteworthy that 99.28% of the tourists in the sample stayed for less than 30 nights.

8. Nationality. A dummy variable which takes the value 1 for the i-th country, and 0 otherwise. The individual countries considered are Germany, Austria, Belgium, Denmark, mainland Spain, Finland, France, Netherlands, Ireland, Italy, Norway, Poland, Portugal, United Kingdom, Czech Republic, Russia, Sweden, Switzerland and Luxembourg. The reference category is Other countries.
9. Non–hotel accommodation. A dummy variable which takes the value 1 if the tourist accommodation is other types of accommodation, such as the tourists' own property, or which belongs to friends or family, or campsites or apartments, and 0 otherwise. The reference category represents a 1 to 5-star hotel/aparthotel.

Table 1 shows descriptive statistics for tourist expenditure and length of stay, together with several explanatory variables associated with the filtered database, and distinguishing between FSC and LCC users. LCCs were used by 53% of all respondents. Among these LCC users, the average tourist spent around 1638 euros in the country of origin, and the travel party was composed of two persons. The respondents' average age was 43 years, and 72.30% had visited the Canary Islands at least once before. This Table 1 also includes the bias and kurtosis values for the variables length of stay and tourist expenditure (origin and destination). The positive value of the first and the large value of the second for all variables in both cases suggests an empirical distribution which is highly skewed and has a long right tail. These properties should be taken into account in the empirical modelling.

To obtain the bivariate models, total expenditure (X) was taken as the total expenditure in the country of origin plus total expenditure in the Canary Islands. The total expenditure is then expressed in natural logarithms. The results obtained indicate lower total expenditure and shorter stays for LCC passengers.

The equality of the two samples was analysed by non-parametric statistical tests, taking into account the LCC variable, thus distinguishing between FSCs and LCCs.

We tested the null hypothesis that the samples are equal for X and length of stay given LCC $=$ 0 and LCC $=$ 1, respectively. The non-parametric tests used (see Table 2) were the two-sample Kolmogorov-Smirnov (K-S) test for equality of distribution functions, the Kruskal-Wallis (K-W) equality-of-populations rank test and the two-sample Wilcoxon (W two sample) rank-sum test. In all cases, the results obtained lead us to reject the null hypothesis. For example, the W statistic is equal to 26.745 (p-value equal to 0.00). The same test was applied to the length of stay variable, producing a value of 18.420 (p-value equal to 0.00). Therefore, the LCC and FSC samples are not equal and represent differences between the tourists, according to all tests. This result is in line, at least for total expenditure, with Ferrer-Rosell and Coenders (2017), who reported that users of the two airline types diverge with regard to total trip expenditure. This conclusion was based on a statistical analysis method termed *compositional analysis with a total* to determine which variables affected the total expenditure of LCC users, among others.

Model estimates

In this section, we show the results obtained for two bivariate regression models, one with covariates and one without. In this study, all estimation procedures were

Table 1 Descriptive statistics for all tourists distinguishing between LCC and FSC passengers

	LCC				FSC			
Variables	Mean	Standard deviation	Skewness	Kurtosis	Mean	Standard deviation	Skewness	Kurtosis
Length of stay	8.25 nights	6.09 nights	11.18	231.03	9.08 nights	6.40 nights	9.16	151.76
Expenditure: origin	1638.95€	1387.05€	2.70	18.70	2168.35€	1532.79€	2.29	14.69
Expenditure: Canary Islands	728.53€	736.24€	5.10	80.88	742.59€	3471.55€s	95.44	9609.46
High income	39.23%	–	–	–	42.35%	–	–	–
Medium income	20.39%	–	–	–	24.74%	–	–	–
Non–hotel accommodation	11.80%	–	–	–	7.02%	–	–	–
Repetition	72.30%	–	–	–	76.90%	–	–	–
Travel party size	2.40 persons	1.24 persons	–	–	2.40 persons	1.14 persons	–	–
Age of the respondent (years)	42.82	13.62	–	–	46.92	13.95	–	–
Observations	12203	–	–	–	10718			
Number of tourists after data cleaning	22921							

Filtered database

Table 2 Non-parametric test results for length of stay and expenditure at origin and destination

Variable	K-S for two-samples	p-value	K-W	p-value	W for two-samples	p-value
Length of stay (N)	0.1158	0.00	311.02	0.00	18.420	0.00
Expenditure at origin and destination (X)	0.1614	0.00	715.28	0.00	26.745	0.00

conducted using the Wolfram Mathematica (v. 12.0) and RATS (v. 7.00) packages. In the latter case, the approximation given in Eq. 22 was used, obtaining the same results as those obtained with Mathematica). For further information on these packages, see Ruskeepaa (2009) and Brooks (2009).

Distinguishing between LCC and FSC passenger behaviours, in a model without covariates

Table 3 shows estimates of $\widehat{\mu}_1$, $\widehat{\mu}_2$, $\widehat{\gamma}$, $\widehat{\varrho}(X, N)$ and p-values for the model estimated without covariates. The table also shows the maximum value for the logarithm of likelihood and the number of observations in each case. Total expenditure and length of stay results are shown for all tourists, and also for LCC and FSC users, separately.

Figures 3 and 4 show the sample (empirical) and estimated bivariate distributions, respectively. The estimated distribution was obtained from the marginal distributions with the parameters estimated by maximum likelihood, as shown in Table 3 for all tourists, and for LCC and FSC users. In general, as can be seen without performing any test, the curves obtained have similar patterns to those obtained empirically.

All parameters are statistically significant at 1% for all tourists and also for LSC and FSC users, considered separately.

The parameter $\widehat{\gamma}$ is statistically different from zero, indicating the existence of dependence between expenditure and length of stay. The estimated correlation

Table 3 Maximum likelihood estimates under the bivariate model without covariates

	All tourists		LCC		FSC	
Parameter	Estimate	p-value	Estimate	p-value	Estimate	p-value
$\widehat{\mu}_1$	7.642	0.00	7.5263	0.00	7.7742	0.00
$\widehat{\mu}_2$	8.634	0.00	8.2457	0.00	9.0765	0.00
$\widehat{\gamma}$	13.980	0.00	12.7329	0.00	16.9769	0.00
$\widehat{\varrho}(X, N)$	0.2582		0.2652		0.2401	
$\ell_{\max}$	-97096.30		-51485.56		-45013.99	
Sample correlation	0.2810		0.2714		0.2819	
Observations	22921		12203		10718	

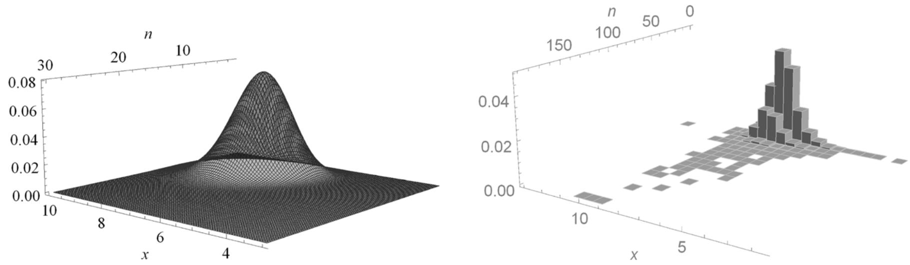

Fig. 3 Empirical smooth distribution (left) and fitted distribution (right) for N and X. All tourists

between the two variables, $\widehat{\varrho}(X, N)$, is low and very similar in all cases. This estimated correlation is close to the empirical Pearson correlation, found to be 0.2714 for LCC users and 0.2819 for FSC users.

The mean values for expenditure and length of stay, $\widehat{\mu}_1$ and $\widehat{\mu}_2$, are higher for FSC than for LCC users. This was expected because, as shown in Table 1, the estimated mean for these variables presents the same characteristic.

The main conclusion drawn from these data is that, on average, FSC users spend longer in the Canary Islands than LCC users, and that our model replicates this empirical finding reasonably well. We also recorded statistically significant differences between FSC and LSC users, in that the null hypothesis for the tests of equality was rejected in all three cases, indicating that both the mean and the dependence parameters differ according to the type of airline used. This result corroborates the divergence found by Ferrer-Rosell and Coenders (2017) in the case of expenditure in

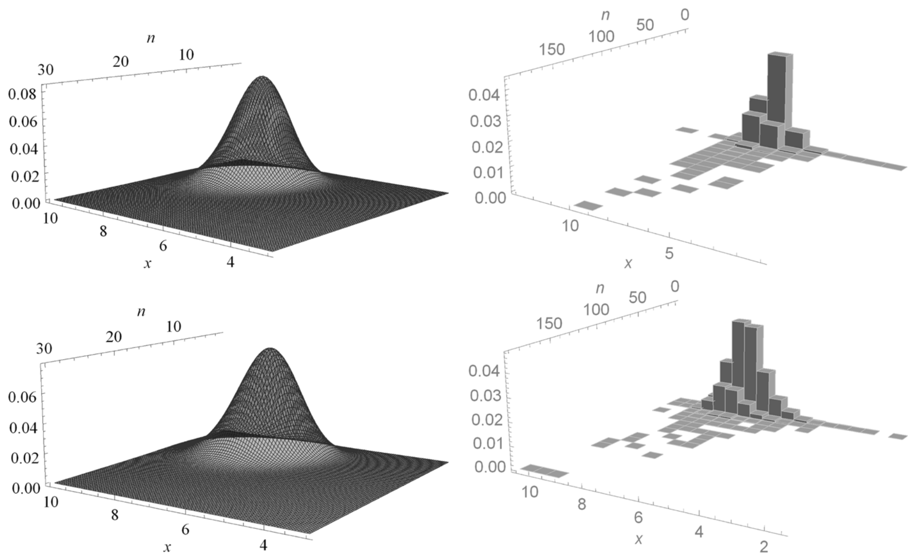

Fig. 4 Empirical smooth distribution (left) and fitted distribution (right) for N and X corresponding to LCC users (above) and FSC (below), respectively

Spain, extending it to the length of stay in the Canary Islands. Policy makers are able to implement policies focused on LCC users because FSC visitors spend more and make longer stays in the archipelago. Accordingly, if we wish to promote more visits by LCC tourists, perhaps a larger share of the marketing budget should be dedicated to this segment of the market.

Impact of the LCC variable on mean expenditure and duration

In this section we evaluate the validity of the bivariate regression model when covariates are included. In this case, instead of estimating a bivariate regression taking into account two models, one for LCC users and the other for FSC users, we estimate a single model including the LCC variable between regressors affecting the mean values of both variables.

Table 4 shows the results obtained for the model with covariates (105 parameters simultaneously estimated), including the LCC dummy variable, together with other determinants as controls which simultaneously affect both tourist expenditure and length of stay, such as certain individual characteristics (income, age and nationality) and certain vacation characteristics (repeated visit, travel party size, and type of accommodation). Most of these variables were also used by Ferrer-Rosell and Coenders (2017). The table also shows the multiplicative effects (interactions) of LCC with all other factors, in an approach previously taken by Eugenio-Martin and Inchausti-Sintes (2016), thus simultaneously analysing the expenditure at origin and at destination (i.e., total expenditure). This method enables us to distinguish the effects of LCC and FSC tourists, considered separately, on total expenditure and length of stay.

The results obtained show that many parameters are statistically significant at 5%.

In particular, the coefficient for the LCC variable is negative and not statistically significant for total expenditure but positive and statistically significant for length of stay. These results indicate that LCC users spend longer at their destination than FSC users, but that there are no differences in total expenditure between LCC and FSC users. In general, our results are in line with those of recent empirical literature, according to which there are differences between LCC and FSC users regarding the determinants of length of stay (Ferrer-Rosell et al., 2014), and Gómez-Déniz and Pérez-Rodríguez (2019), although the latter authors observed a negative effect of LCC use. However, our results contrast with those of recent literature on expenditure; for example, Ferrer-Rosell et al. (2015) reported finding differences in this respect between FSC and LCC passengers.

The following observations were made for the remaining variables representing controls such as individual and vacation characteristics that may influence tourists' total expenditure and length of stay. To our knowledge, this study is the first to make use of this methodology in the field considered, which means that our results cannot be compared directly with previous findings in this area.

As concerns the individual characteristics addressed, the following points are of interest. Firstly, income is a key variable in any model of tourist demand. Because our income variable is categorical, we include two categories: high and medium income, for which the reference category is low income. In our results, high income had a

Table 4 Maximum likelihood estimates and p-values under the bivariate model including covariates and interactions with LCC

	Without interactions				With interactions with LCC			
	Total expenditure		Length of stay		Total expenditure		Length of stay	
Parameter	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
LCC	−0.0181	0.151	0.3967	< 0.001				
Medium income	0.0093	< 0.001	−0.0716	< 0.001	0.0025	0.249	0.0144	0.209
High income	0.0210	< 0.001	−0.1417	< 0.001	0.0021	0.410	0.0283	0.043
Non–hotel accommodation	−0.0492	< 0.001	0.2382	< 0.001	−0.0301	< 0.001	−0.0308	0.061
Repetition	0.0023	0.168	0.1339	< 0.001	−0.0041	0.076	−0.0156	0.217
Travel party size	0.0400	< 0.001	0.0172	< 0.001	−0.0008	0.289	−0.0073	0.089
Age	0.0397	< 0.001	0.3906	< 0.001	0.0011	0.712	−0.0937	< 0.001
Country of residence								
Germany	−0.0128	0.012	−0.1204	< 0.001	0.0012	0.857	−0.1027	0.001
Austria	0.0076	0.310	−0.1109	0.001	0.0038	0.741	−0.1396	0.012
Belgium	−0.0152	0.012	−0.3466	< 0.001	0.0021	0.814	−0.0356	0.427
Denmark	−0.0346	< 0.001	−0.4297	< 0.001	0.0073	0.482	0.1771	0.001
Mainland Spain	-0.0685	< 0.001	−0.6079	< 0.001	0.0035	0.602	−0.0631	0.054
Finland	−0.0173	0.006	−0.2580	< 0.001	0.0081	0.437	0.0989	0.048
France	−0.0276	< 0.001	−0.3405	< 0.001	0.0006	0.935	−0.0419	0.270
Netherlands	−0.0304	< 0.001	−0.2766	< 0.001	0.0018	0.810	−0.0556	0.146
Ireland	−0.0339	< 0.001	−0.2896	< 0.001	0.0138	0.109	−0.0798	0.058

Table 4 (continued)

Parameter	Without interactions				With interactions with LCC			
	Total expenditure		Length of stay		Total expenditure		Length of stay	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Italy	−0.0147	0.038	−0.2462	< 0.001	−0.0098	0.258	−0.0275	0.516
Norway	−0.0164	0.008	−0.1757	< 0.001	0.0212	0.018	0.0948	0.023
Poland	−0.0384	< 0.001	−0.3909	< 0.001	0.0161	0.080	0.0574	0.223
Portugal	−0.0571	< 0.001	−0.6091	< 0.001	−0.0018	0.898	0.0854	0.304
United Kingdom	−0.0312	< 0.001	−0.3778	< 0.001	−0.0033	0.621	−0.0541	0.087
Czech Republic	−0.0400	0.001	−0.3750	< 0.001	0.0397	0.021	−0.0019	0.983
Russia	0.0446	< 0.001	0.1380	0.001	−0.0208	0.178	−0.1797	0.012
Sweden	−0.0292	< 0.001	−0.3120	< 0.001	0.0128	0.128	−0.0346	0.402
Switzerland	0.0047	0.498	−0.1365	< 0.001	0.001	0.8656	−0.025	0.608
Luxembourg	0.0013	0.927	−0.3229	< 0.001	−0.0001	0.99764	−0.002	0.990
constant	1.8230	< 0.001	0.7918	< 0.001				
γ	26.537	< 0.001						
$\ell_{\max} = -85637.70$								
CAIC = 172434								

stronger positive effect on total expenditure than medium income. For FSC users, this parameter is positive and statistically significant for total expenditure, and negative for length of stay. However, the parameters are not statistically significant for LCC users, except for high income, which is positive in the length of stay equation, indicating that these tourists stay for longer at their destination. In general, we conclude that significant differences are only observed between FSC and LCC users for high income and for length of stay, not for total expenditure.

Another significant factor is that of the tourists' age, which has a positive and statistically significant effect on both variables for FSC users, while for LCC users it is negative and significant for length of stay. Therefore, there are differences between FSC and LCC users for length of stay, with a greater impact being observed among FSC users than LCC users. In other words, LCC users stay for longer in the Canary Islands than do FSC users.

In general, the coefficients for the nationality dummy variables reflect a negative and statistically significant effect for FSC users both for total expenditure and for length of stay, with the exception of Russian tourists, for whom it is positive in both cases. However, in most cases the effect of nationality is not statistically significant at the 5% level for LCC users, with the exception of tourists from Germany, Austria, Denmark, Finland, Norway and Russia for length of stay, and from Norway and the Czech Republic for total expenditure. Overall, these results show that for most nationalities there are no differences between FSC and LCC in total expenditure, but there are some for length of stay.

As regards vacation characteristics, the following results were obtained.

The non–hotel accommodation variable represents the availability of second homes and/or the free accommodation provided by relatives and friends. This variable is included in our analysis in order to control for this type of accommodation, taking into account that 0.072% of the observations in our sample reflect stays of more than 30 nights and therefore are potentially distorting (see Figure 3). The motivations and characteristics of these visitors differ sharply from those of tourists in paid-for accommodation (the majority), who mainly stay for one or two weeks. Their inclusion, therefore, could lead to misleading results in the comparison of patterns of total expenditure and length of stay between LCC and FSC passengers. The parameters of this variable are negative for FSC users for total expenditure and negative for length of stay, and are statistically significant at 5%. However, for LCC users, the negative parameters for expenditure and length of stay are only statistically significant at 5% and 10%, respectively. For LCC users, this corresponds to a reduction in total expenditure and length of stay, in both cases. In summary, this type of accommodation decreases the total expenditure of both FSC and LCC users, but the length of stay is increased for FSC users and decreased for LCC users. Accordingly, differences exist between the two types of visitors with respect to this variable.

The 'repeated visit' variable is included in the view that it exerts a crucial influence on the visitors' image of the destination. Our study results indicate that repeated visits produce a positive and statistically significant impact on the length of stay of FSC users. However, the same variable has a negative and statistically significant effect (at 10%) on the expenditure of LCC users. This finding highlights the existence of a difference in this respect between FSC and LCC users.

Finally, the travel party size has a positive and statistically significant effect on the total expenditure and length of stay of FSC users. For LCC users, however, the same variable has a negative impact in both respects. This result is in line with our expectations, since LCC passengers normally face more severe budgetary restrictions. In summary, the size of the travel group produces differences between LCC and FSC users regarding length of stay (shorter by LCC users) but not for total expenditure.

Conclusions

In this study, a simultaneous analysis is performed of the differences between LCC and FSC users with respect to the impact produced on total expenditure and length of stay by visitors to the Canary Isles. To do so, we created a bivariate model based on the conditional distributions technique described by Arnold et al. (1999) to simultaneously model expenditure and length of stay and to account for dependence. In addition, we formulated a bivariate regression model based on conditionals, which allowed us to consider covariates.

These models were validated and the differences between LCC and FSC users analysed by reference to data obtained from the 2017 Canary Islands Tourist Expenditure Survey.

Our results for the model without covariates indicate that FSC users visiting the Canary Islands spend more and stay longer than LCC users. Our model replicates the empirical data reasonably well.

However, when covariates are included in the bivariate regression model, the results also reflect other interesting aspects. In general, our results do not reveal the type of convergence between FSC and LCC users that Ferrer-Rosell and Coenders (2017) observed using a different statistical methodology. In fact, we found tangible differences between the behaviour patterns of the two groups. In our view, these differences arise from the existence of interactions between the LCC variable and the explanatory factors of total expenditure and length of stay, for which many coefficients are statistically significant. For example, we recorded clear differences between LCC and FSC users when they stay in non–hotel accommodation, with the first group of visitors spending less in total total and staying less time at their destination. For the other factors considered, the results obtained are mixed. For example, in terms of personal income, only the visitors classed as ‘high income’ present significant differences between LCC and FSC users, with the former spending longer at their destination. LCC users who make repeated visits to the Canary Isles spend less in these repeated visits, in comparison with FSC users. This is so because the coefficient of the “repeated visit” variable is statistically significant at 10% in the expenditure equation for LCC users. However, there are no statistically significant effects for LCC users in terms of the length of stay equation (because the coefficient is not statistically significant). Moreover, there are differences (at 10% significance)between LCC and FSC users in terms of the relation between travel party size and length of stay. These differences arise from the fact that the larger the group size, the shorter the length of stay. However, no such difference is observed with respect to the relation between travel party size and expenditure. A similar pattern is apparent as concerns the impact of LCC vs. FSC use, in that this difference impacts on the relation between age and length of stay, but not on that between age and expenditure. Finally, regarding nationality, in general, there are no significant differences between LCC and FSC users in terms of the relation between this factor and total expenditure, although differences exist for some nationalities with respect to the length of stay.

The study results we present have certain policy implications, suggesting that hotel managers and policy makers should focus on the profile of airline users in order to target their marketing policies more effectively, and thus increase tourist spending and length of stay in the Canary Islands. For example, it should be taken into account that some high-income tourists travel by LCC and stay for longer than FSC users. In other words, LCC users are not necessarily low-income visitors.

Furthermore, when the age is increased by one year, LCC tourists are more likely to reduce the length of their stay than are FSC users. This finding suggests that older LCC tourists should not be a priority target for tourism marketing policies.

Finally, hotel managers and policy makers should also consider marketing policies focused on the visitors' country of origin, aimed at increasing the average length of stay.

In summary, although the model presented reflects the expenditure and length-of-stay patterns of low-cost tourists reasonably well, it does have an important limitation, concerning the linear formulations corresponding to the conditioned means, which should perhaps be relaxed to make the model more flexible. However, if this were done, it would not be possible to obtain the normalisation constant corresponding to the bivariate distribution in a closed form. To do so, the model would have to be estimated using methods other than those discussed in this paper, possibly based on the pseudolikelihood approach or on Bayesian methods using Monte Carlo and WinBugs techniques. For a more detailed analysis of these questions, see for instance (Arnold et al., 2001) and (Arnold & Strauss, 1988).

Acknowledgments The authors are grateful to the Editor and to the two anonymous referees for their valuable comments and suggestions.

Funding Emilio Gómez-Déniz was partially funded by grant ECO2017–85577–P (Ministerio de Economía, Industria y Competitividad. Agencia Estatal de Investigación).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Appendix A

A.1 Marginal, conditional distributions and marginal moments

The marginal distributions of X and N are given by

$$X \sim \mathcal{G}(\alpha, \gamma), \quad \alpha > 0,\ \beta > \gamma,\ \gamma > 0, \tag{15}$$

$$N \sim \mathcal{SNB}\left(\alpha, p = \frac{\gamma}{\beta}\right), \quad \alpha > 0,\ \beta > \gamma,\ \gamma > 0, \tag{16}$$

where $\mathcal{SNB}$ refers to the shifted negative binomial distribution. That is,

$$f_X(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\gamma x), \quad x > 0, \tag{17}$$

$$f_N(n) = \binom{\alpha+n-2}{n-1} \left(\frac{\gamma}{\beta}\right)^\alpha \left(1-\frac{\gamma}{\beta}\right)^{n-1}, \; n = 1, 2, \ldots \tag{18}$$

The marginal distribution of N is obtained by integrating (13) with respect to x in the support $(0, +\infty)$ and the marginal distribution of X is calculated by summing (13) with respect to n in the support $\{1, 2, \ldots\}$.

Furthermore, it is readily apparent that the marginal means and variances are given by

$$\begin{aligned} E(X) &= \frac{\alpha}{\gamma}, \quad var(X) = \frac{\alpha}{\gamma^2}, \\ E(N) &= 1 + \frac{\alpha}{\gamma}(\beta-\gamma), \quad var(N) = \frac{\alpha\beta(\beta-\gamma)}{\gamma^2}, \end{aligned}$$

while the cross moment of X and N is

$$E(XN) = \frac{\alpha}{\gamma^2}[\beta + \alpha(\beta-\gamma)].$$

Simple calculations provide the covariance, given by

$$cov(X, N) = \frac{\alpha(\beta-\gamma)}{\gamma^2},$$

which is always positive.

The conditional distribution of X given $N = n$ is $\mathcal{G}(\sigma(n), \eta(n))$, where

$$\begin{aligned} \sigma(n) &= \alpha + n - 1, \\ \eta(n) &= \beta \end{aligned}$$

and the conditional distribution of N given $X = x$ is $\mathcal{SP}o(\varphi(x))$, with $\varphi(x) = (\beta-\gamma)x$. Observe that with the assumption of these two conditional distributions expressions (1) and (2) are guaranteed.

A.2 Estimation of the parameters

Here, we derive estimators based on the moments method and on maximum likelihood for the model with and without covariates, and also provide closed-form expressions for the Fisher information matrix.

A.3 Estimation of the model without covariates

Let us first consider the case of the model with no covariates. If

$$(\tilde{x}, \tilde{n}) = \{(x_1, n_1), (x_2, n_2), \ldots, (x_t, n_t)\}$$

is a sample obtained from the distribution (14) and $\bar{x} = (1/t)\sum_{i=1}^t x_i$, $\bar{n} = (1/t)\sum_{i=1}^t n_i$ and $\mu_{12} = (1/t)\sum_{i=1}^t x_i n_i$ are the corresponding sample moments, some computation provides the estimators based on these sample moments, which are given by

$$\widehat{\mu}_1 = \bar{x}, \quad \widehat{\mu}_2 = \bar{n}, \quad \widehat{\gamma} = \frac{\widehat{\mu}_2 - 1}{\widehat{\mu}_{12} - \widehat{\mu}_1 \widehat{\mu}_2}.$$

A.4 The score vector and Fisher information matrix

We now consider the maximum likelihood method. Let $\Theta = (\gamma, \mu_1, \mu_2)$ be the vector of parameters to be estimated. The log-likelihood function is proportional to

$$\begin{aligned}\ell((\tilde{x}, \tilde{n}); \Theta) \propto\ & t\left[\gamma\mu_1(\bar{x}^* + \log\gamma) - \bar{x}^* - \log\Gamma(\gamma\mu_1)\right] + \sum_{i=1}^{t} n_i \log x_i \\ & + t(\bar{n} - 1)\left[\log(\mu_2 - 1) - \log\mu_1\right] - \frac{t\bar{x}(\mu_2 + \gamma\mu_1 - 1)}{\mu_1},\end{aligned}$$

where $\bar{x}^* = (1/t)\sum_{i=1}^{t} \log x_i$.

Thus, the normal equations which provide the estimators of the parameters are given by

$$\bar{x}^* + \log\gamma + 1 - \psi(\gamma\mu_1) - \frac{\bar{x}}{\mu_1} = 0, \tag{19}$$

$$\mu_1^2\gamma\left[\bar{x}^* + \log\gamma - \psi(\gamma\mu_1)\right] - \mu_1(\bar{n} - 1) + \bar{x}(\mu_2 - 1) = 0, \tag{20}$$

$$\mu_1(\bar{n} - 1) - \bar{x}(\mu_2 - 1) = 0, \tag{21}$$

where $\psi(z)$ is the digamma function, the logarithmic derivative of the Euler gamma function. Some algebra manipulation provides the maximum likelihood estimators of μ_1 and μ_2, which are given by $\widehat{\mu}_1 = \bar{x}$ and $\widehat{\mu}_2 = \bar{n}$. Finally, the estimator of the parameter γ is the solution of the equation

$$\log\gamma - \psi(\gamma\mu_1) + \bar{x}^* = 0,$$

which can be solved numerically.

The second partial derivatives are as follows:

$$\begin{aligned}
\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial\gamma^2} &= \frac{t\mu_1}{\gamma} - t\mu_1^2\psi_1(\gamma\mu_1), \\
\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial\gamma\partial\mu_1} &= t\left[\bar{1} + x^* + \log\gamma - \psi(\gamma\mu_1) - \gamma\mu_1\psi_1(\gamma\mu_1)\right], \\
\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial\gamma\mu_2} &= 0, \\
\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial\mu_1^2} &= \frac{t(\bar{n} - 1)}{\mu_1^2} - t\gamma^2\psi_1(\gamma\mu_1) - \frac{2t\bar{x}}{\mu_1^3}(\mu_2 - 1), \\
\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial\mu_1\mu_2} &= \frac{t\bar{x}}{\mu_1^2}, \quad \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial\mu_2^2} = -\frac{t(\bar{n} - 1)}{(\mu_2 - 1)^2},
\end{aligned}$$

where $\psi_1(\cdot)$ is the first derivative of the digamma function.

The elements of the Fisher information matrix, $\mathcal{J}(\widehat{\Theta})$, are therefore

$$\mathcal{J}_{11}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\gamma^2}\right)\Bigg|_{\Theta=\widehat{\Theta}} = -\frac{t\widehat{\mu}_1}{\widehat{\gamma}} + t\widehat{\mu}_1^2\psi_1(\widehat{\gamma}\widehat{\mu}_1),$$

$$\mathcal{J}_{12}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\gamma}\partial\widehat{\mu}_1}\right)\Bigg|_{\Theta=\widehat{\Theta}} = -\sum_{i=1}^{t}\log x_i - t\left[(1+\log\widehat{\gamma} - \psi(\widehat{\gamma}\widehat{\mu}_1)\right.$$
$$\left. -\widehat{\gamma}\widehat{\mu}_1\psi_1(\widehat{\gamma}\widehat{\mu}_1)\right],$$

$$\mathcal{J}_{13}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\gamma}\widehat{\mu}_2}\right)\Bigg|_{\Theta=\widehat{\Theta}} = 0,$$

$$\mathcal{J}_{22}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\mu}_1^2}\right)\Bigg|_{\Theta=\widehat{\Theta}} = -\frac{t(\widehat{\mu}_2-1)}{\widehat{\mu}_1^2} + t\widehat{\gamma}^2\psi_1(\widehat{\gamma}\widehat{\mu}_1)$$
$$+\frac{2t}{\widehat{\mu}_1^2}(\widehat{\mu}_2-1),$$

$$\mathcal{J}_{23}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\mu}_1\widehat{\mu}_2}\right)\Bigg|_{\Theta=\widehat{\Theta}} = -\frac{t}{\widehat{\mu}_1},$$

$$\mathcal{J}_{33}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\mu}_2^2}\right)\Bigg|_{\Theta=\widehat{\Theta}} = \frac{t}{(\widehat{\mu}_2-1)}.$$

Here, $\widehat{\Theta}$ represents the maximum likelihood of Θ. Observe that the analytic expression for $E(\sum_{i=1}^{t}\log x_i)$ is not feasible. For large t, for computational purposes, this is evaluated by ignoring the expectation operator and replacing it by $\sum_{i=1}^{t}\log x_i$. The asymptotic variance-covariance matrix of $\widehat{\Theta}$ is obtained by inverting the information matrix.

A.5 Estimation of the model with covariates

For the sake of simplicity, we assume $\boldsymbol{\eta} = \boldsymbol{\delta}$ and write $\mu_{1i} = \mu_{1i}(\boldsymbol{\delta})$ and $\mu_{2i} = \mu_{2i}(\boldsymbol{\delta})$. Let $\Theta = (\gamma, \boldsymbol{\delta})$. The log-likelihood is then proportional to

$$\ell((\tilde{x},\tilde{n});\Theta) \propto \gamma\sum_{i=1}^{t}\mu_{1i}\log(\gamma x_i) - \sum_{i=1}^{t}\log\Gamma(\gamma\mu_{1i})$$
$$+\sum_{i=1}^{t}(n_i-1)\left[\log x_i + \log(\mu_{2i}-1) - \log\mu_{1i}\right]$$
$$-\sum_{i=1}^{t}\frac{x_i}{\mu_{1i}}(\mu_{2i}+\gamma\mu_{1i}-1).$$

Thus, the normal equations, for $i = 1, \ldots, t$, are given by

$$\begin{aligned}
\frac{\partial \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma} &= \sum_{i=1}^{t} \mu_{1i} \left[1 + \log(\gamma x_i) - \psi(\gamma \mu_{1i})\right] - t\bar{x} = 0, \\
\frac{\partial \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \delta_j} &= \gamma \sum_{i=1}^{t} y_{ij} \mu_{1i} \left[\log(\gamma x_i) - \psi(\gamma \mu_{1i})\right] \\
&\quad + \sum_{i=1}^{t} (n_i - 1) \left[\frac{1}{\mu_{2i} - 1} \frac{\partial \mu_{2i}}{\partial \delta_j} - \frac{1}{\mu_{1i}} \frac{\partial \mu_{1i}}{\partial \delta_j}\right] \\
&\quad - \sum_{i=1}^{t} \frac{x_i}{\mu_{1i}^2} \left[\mu_{1i} \frac{\partial \mu_{2i}}{\partial \delta_j} + (1 - \mu_{2i}) \frac{\partial \mu_{1i}}{\partial \delta_j}\right] = 0,
\end{aligned}$$

where $j = 1, \ldots, k$, $\frac{\partial \mu_{1i}}{\partial \delta_j} = y_{ij} \mu_{1i}$ and $\frac{\partial \mu_{2i}}{\partial \delta_j} = z_{ij}(\mu_{2i} - 1)$. Finally, after computing the second partial derivatives we obtain the elements of the Fisher information matrix, as follows:

$$\begin{aligned}
\mathcal{J}_{11}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma^2}\right)\Bigg|_{\Theta = \widehat{\Theta}} = -\sum_{i=1}^{t} \widehat{\mu}_{1i} \left[\frac{1}{\widehat{\gamma}} - \widehat{\mu}_{1i} \psi_1(\widehat{\gamma} \widehat{\mu}_{1i})\right], \\
\mathcal{J}_{12}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma \partial \delta_j}\right)\Bigg|_{\Theta = \widehat{\Theta}} = -\sum_{i=1}^{t} y_{ij} \widehat{\mu}_{1i} \left[1 + \log(\widehat{\gamma} x_i)\right. \\
&\quad \left. - \psi(\widehat{\gamma} \widehat{\mu}_{1i}) - \widehat{\gamma} \widehat{\mu}_{1i} \psi_1(\widehat{\gamma} \widehat{\mu}_{1i})\right], \\
\mathcal{J}_{22}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \delta_j^2}\right)\Bigg|_{\Theta = \widehat{\Theta}} = -\sum_{i=1}^{t} y_{ij}^2 \mu_{1i} \widehat{\gamma} \left[\log(\widehat{\gamma} x_i)\right. \\
&\quad \left. - \psi(\widehat{\gamma} \widehat{\mu}_{1i})\right] + y_{ij}^2 \widehat{\mu}_{1i}^2 \left[-\widehat{\gamma}^2 \psi_1(\widehat{\gamma} \widehat{\mu}_{1i}) + \frac{1 - \mu_{2i}}{\mu_{1i}}\right] + y_{ij}^2 (\mu_{2i} - 1), \\
\mathcal{J}_{23}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \delta_j \partial \delta_l}\right)\Bigg|_{\Theta = \widehat{\Theta}} = -\sum_{i=1}^{t} y_{ij} z_{il} \widehat{\gamma} \widehat{\mu}_{1i} \left[\log(\widehat{\gamma} x_i) - \psi(\widehat{\gamma} \widehat{\mu}_{1i})\right] \\
&\quad + y_{ij} z_{il} \widehat{\mu}_{1i}^2 \left[-\widehat{\gamma}^2 \psi_1(\widehat{\gamma} \widehat{\mu}_{1i}) + \frac{1 - \widehat{\mu}_{2i}}{\widehat{\mu}_{1i}^2}\right] + z_{il}^2 (\widehat{\mu}_{2i} - 1), \quad j \neq l.
\end{aligned}$$

Standard errors can be obtained conventionally from the inverse of the matrix.

A.6 Approximation of the digamma function

A practical approach to the digamma function is given by the following expression, which is well known in the statistical literature,

$$\log\left(\Gamma(z)\right) \approx \frac{1}{2}\log(2\pi) + \left(z - \frac{1}{2}\right)\log(z) - z + \frac{z}{2}\log\left(z \sinh\left(\frac{1}{z}\right)\right). \tag{22}$$

References

Aguiló, E., Rey, B., Roselló, J., & Torres, C. (2007). The impact of the post liberalization growth of lccs on the tourism trends in spain. In *Rivista di Politica Economica* (pp. 39–59).

Arnold, B., Castillo, E., & Sarabia, J. (1999). *Conditional specification of statistical models*. New York: Springer.

Arnold, B., Castillo, E., & Sarabia, J. (2001). Conditionally specified distributions: an introduction (with comments and a rejoinder by the authors). *Statistical Science*, 16(3), 249–274.

Arnold, B., & Strauss, D. (1988). Pseudolikelihood estimation. *Sankhya Series B*, 53(21-24), 233–243.

Arnold, B., & Strauss, D. (1991). Bivariate distributions with conditionals in prescribed exponential families. *Journal of the Royal Statistical Society*, 53, 365–375.

Brooks, C. (2009). *RATS Handbook to accompany introductory econometrics for finance*. Cambridge: Cambridge University Press.

Davis, P., & Rabinowitz, P. (1984). *Methods of numerical integration*. San Diego: Academic Press.

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

Ferrer-Rosell, B., & Coenders, G. (2017). Airline type and tourist expenditure: are full service and low cost carriers converging or diverging? *Journal of Air Transport Management*, 63, 119–125.

Ferrer-Rosell, B., Coenders, G., & Martínez-García, E. (2015). Determinants in tourist expenditure composition: the role of airline type. *Tourism Economics*, 21, 9–32.

Ferrer-Rosell, B., Martínez-García, E., & Coenders, G. (2014). Package and no-frills air carriers as moderators of length of stay. *Tourism Management*, 42, 114–122.

Ferrer-Rosell, B., & Seetaram, N. (2014). Determinants of per diem tourist expenditure - the role of self-reported income. In *Paper presented at the 4th Advances in Hospitality and Tourism Marketing and Management Conference* (pp. 25–27).

Gómez-Déniz, E., & Calderín, E. (2014). Unconditional distributions obtained from conditional specifications models with applications in risk theory. *Scandinavian Actuarial Journal*, 7, 602–619.

Gómez-Déniz, E., & Pérez-Rodríguez, J. (2019). Modelling distribution of aggregate expenditure on tourism. *Economic Modelling*, 78, 293–308.

Gómez-Déniz, E., & Pérez-Rodríguez, J. (2020). Modelling dependence between daily tourist expenditure and length of stay. Tourism Economics (in press).

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

Kotz, S., Balakrishnan, N., & Johnson, N. (2000). *Continuous multivariate distributions. In: Vol 1: Models and Applications*. New York: Wiley.

Martínez-García, E., & Raya, J. (2008). Length of stay for low-cost tourism. *Tourism Management*, 29(6), 1064–1075.

Moschopoulos, P., & Staniswalis, J. G. (1994). Estimation given conditionals from an exponential family. *American Statistician*, 48(4), 271–275.

Pérez-Rodríguez, J., & Ledesma-Rodríguez, F. (2019). Unconditional quantile regression and tourism expenditure: The case of the Canary Islands. Tourism Economics (in press).

Raya, M., & Martínez-García, E. (2011). Nationality and low-cost trip duration. a microeconometric analysis. *Journal of Air Transpor Management*, 17, 168–174.

Rey, B., Myro, R., & Galera, A. (2011). Effect of low-cost airlines on tourism in spain. a dynamic panel data model. *Journal of Air Transport Management*, 17(3), 163–167.

Ruskeepaa, H. (2009). *Mathematica navigator. Mathematics, Statistics, and Graphics*, 3rd edn. Cambridge: Academic Press.

Sarabia, J., Castillo, E., Pascual, M., & Sarabia, M. (2005). Mixture Lorenzv curves. *Economics letters*, 89, 89–94.

Sarabia, J., Gómez-Déniz, E., & Vázquez-Polo, F. (2004). On the use of conditional specification models in claim count distributions: an application to bonus-malus systems. *ASTIN Bulletin*, 34, 85–89.

Spanos, A. (1999). *Probability theory and statistical inference*. Cambridge: Cambridge University Press.

Thrane, C. (2014). Modelling micro-level tourism expenditure: recommendations on the choice of independent variables, functional form and estimation technique. *Tourism Economics*, 20(1), 51–60.

Thrane, C. (2015). On the relationship between length of stay and total trip expenditures: a case study of instrumental variable (iv) regression analysis. *Tourism Economics*, 21(2), 357–367.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.