

Modelling Expenditure in Tourism using the Log-Skew Normal Distribution

E. GÓMEZ-DÉNIZ, N. DÁVILA-CÁRDENES, J. BOZA-CHIRINO

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,

emilio.gomez-deniz@ulpgc.es

<http://orcid.org/0000-0002-5072-7908> (EGD)

nancy.davila@ulpgc.es

<http://orcid.org/0000-0002-8886-8140> (NDC)

jchirino@ulpgc.es

<http://orcid.org/0000-0002-8494-8864> (JBC)

Abstract

Many factors are involved in tourist decision expenses. Such circumstances may give rise to some asymmetry in the distribution of tourism expenditure. We propose in this paper a reparameterization of the three-parameter log-skew normal distribution for modelling the expenditure at the country of origin, at destination, and total expenditure in a tourism setting. This distribution seems to fit the expenditure data satisfactorily in all the parts of the empirical distribution. In particular, the proposed model is well suited to capture the skewness and kurtosis that may be present and the long tail to the right that the three variables mentioned above tend to present in practice.

Keywords: Expenditure, Log-Skew Normal Distribution, Tourism

Mathematics Subject Classification (2020): 62J02, 62P25, 62P20.

Acknowledgements

EGD thanks to the Ministerio de Economía y Competitividad (project ECO2017-85577-P) for partial support of this work. (Ministerio de Economía y Competitividad, Spain).

Address for correspondence: E. Gómez-Déniz, Department of Quantitative Methods in Economics, University of Las Palmas de Gran Canaria, 35017-Las Palmas de G.C., Spain. E-mail: emilio.gomez-deniz@ulpgc.es

© [2022] This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

This document is the Accepted Manuscript version of a Published Work that appeared in final form in Current Issues in Tourism. To access the final edited and published work see <https://doi.org/10.1080/13683500.2021.1960282>

1 Introduction

Most empirical studies of tourism demand use the micro-level (microdata) information and consider tourism expenditure per person per day as the dependent variable (Wang and Davidson, 2010). The most widely used specifications of tourist spending models are those of dichotomous response, where the estimated coefficients indicate the probabilities that a tourist, in fact, makes the expenditure (Brida and Scuderi, 2012). Also, multinomial response models are used (Kim et al., 2010, Ferrer-Rosell et al., 2016, and Thrane, 2015), but they are not as usual as dichotomous responses. In general, the models that explain tourism demand and use spending as a dependent variable are supported by a set of covariates related to socioeconomic level, nationality, age, job, income, length of travel, type of travel, vacation accommodation, group travel, loyalty to the destination, among others, see Aguiló and Juaneda (2000), Fredman (2008), Craggs and Schofield (2009), Wang and Davidson (2010), Marcussen (2011), Thrane and Farstad (2011), García-Sánchez et al. (2013), Brida et al. (2013), Zheng and Zhang (2013), Thrane (2014), Marrocu et al. (2015), Disegna and Osti (2016), Aguiló et al. (2017), Gómez-Déniz and Pérez-Rodríguez (2019), Gómez-Déniz et al. (2020).

The demand models that use spending are based on tourists' preferences and budget constraints (Brida et al., 2013). As Lee and Choi (2019) explain, given that the tourist offer includes a series of goods and services that, due to their price, are exclusive for those tourists with sufficient resources and willing to make higher expenses, it is advisable to differentiate, for example, between luxury and normal items, not to do so it would produce a certain asymmetry in tourist spending. Different alternatives have been tried to solve the problem of asymmetry and obtain accurate estimates of tourism spending. The first, segmenting the market, Goryushkina et al. (2019) explain that a specifically identified group of clients with similar preferences and similar reactions is achieved by doing this. Craggs and Schofield (2009) segment tourism spending according to destination spending categories (shopping, coffee, bars) and identified statistically significant associations between different spending segments and a set of sociodemographic and behavioral variables. García-Sánchez et al. (2013) use different income segments and days of stay to estimate the determining factors of daily spending by foreign tourists. The second alternative is to disaggregate spending; Vinnicombe and Sou (2014) use tourism spending levels as a classification criterion. Alegre and Garau (2011) and Svensson et al. (2011) determine

the tourist demand based on different tourist spending levels. Pani et al. (2020) study spending patterns in tourist establishments, distinguishing three spending levels: those who spend a lot, those who spend medium, and those who spend little. Gómez-Déniz et al. (2020) divide tourists' total expenditure per person and day according to their location, differentiating between expenses at the country of origin and expenses at destination. Lee and Choi (2019) examine the effect of different attributes on tourism spending in the destination. Attributes considered frustrating have a negative asymmetric effect, and satisfactory attributes have a positive asymmetric impact.

The third alternative is by studying the empirical distribution of data; Wu et al. (2013) use the scobit model that includes an asymmetry parameter that improves the logit and corrects possible skewness biases. When the estimated value of the asymmetry parameter is different from unity, the scobit model captures tourist spending distribution's asymmetry. Cárdenas et al. (2015) study spending as a function of tourists' degree of satisfaction; they assume that the variable spending per tourist and day follows a gamma distribution. Gómez-Déniz et al. (2020) studied tourist spending per person and day, relate spending at destination and origin and obtain a skewed distribution to the right that fits a beta-prime distribution.

The mentioned works agree that data on tourist spending per person and day have asymmetric behavior, and the estimates of average expenditure per person and day are likely biased. Little attention has been given to the asymmetry and the long right tail of the empirical data on which we focus in this paper. For that, we propose a reparameterization of the three-parameter log-skew normal distribution for modeling the expenditure at the the country of origin, destination, and total expenditure in the tourism setting. This distribution has been studied by Lin and Stoyanov (2009) (see also Azzalini, 2013, Chap. 2 and Azzalini et al., 2003) and recently applied in the actuarial setting by Gómez-Déniz and Calderín-Ojeda (2020). This proposal seems to fit the expenditure data satisfactorily in all the part of the empirical distribution. In particular, the proposed model is well suited to capture the skewness and kurtosis that may be present and the long tail to the right that the three variables mentioned above tend to present in practice.

The organization of this paper is as follows. The model used to explain the expenditure and origin, destination, and both are shown in Section 2. Section 3 is devoted to an empirical application. Finally, some conclusions are drawn and promising fields for further research is proposed in the last Section.

2 Modeling the expenditure

In many situations, empirical data show slight or marked asymmetry and heavy tails, reflecting extreme values. These features imply that the Gaussian or normal distribution cannot model the data. Concerning tourist expenditure, "taking it as normally distributed along the whole real axis is unrealistic and may lead to inconsistent estimates. From a statistical point of view, spending is a zero-censored variable. Its distribution often presents a positive asymmetry for its density, which decreases with the amount", Brida and Scuderi (2012).

Let g and G be, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of symmetric distribution. A random variate Z is said to have a skew distribution if its pdf is given by

$$f_Z(z) = 2g(z)G(\lambda z), \quad -\infty < z < \infty, \quad \lambda \in \mathbb{R}. \quad (1)$$

This family of distributions has been widely studied as an extension of the normal distribution through a shape parameter, λ , which accounts for the skewness. In the case in which g and G are replaced in (1) by $\phi(z)$ and $\Phi(z)$, that is, the pdf and cdf of the standard normal distribution, respectively, the resulting distribution is called the skew-normal. It should be pointed out that the function g does not have to be precisely the derivative of the cdf G to ensure that the pdf given in (1) is a genuine pdf, although this case has not been studied in depth in the statistical literature. Following the notation provided in Henze (1986) we denote the family of distributions given by $g_Z(z) = 2\phi(z)\Phi(\lambda z)$ by $SN = \{SN(\lambda) : \lambda \in \mathbb{R}\}$. Nevertheless, attention here is shown to the generalized skew-normal density provided in Henze (1986) and also studied by Arnold and Beaver (2002). Its pdf is given by

$$f_Z(z) = \frac{\phi(z)\Phi(\lambda_0 + \lambda_1 z)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)}, \quad (2)$$

For multivariate extensions of this distribution, see Azzalini and Valle (1996); Azzalini and Capitanio (1999) and Arnold and Beaver (2002), among others. For an exhaustive and comprehensive study of the skew-normal distribution, see the recent book provided by Azzalini (2013).

The basic log-skew normal distribution, obtained from the classical skew-normal distribution, has been studied by Lin and Stoyanov (2009) (see also

Azzalini, 2013, Chap. 2 and Azzalini et al., 2003). Nevertheless, we will pay attention to the distribution arising from exponentiation (4) by taking $\lambda_0 = \lambda_1 = \lambda$ and using a linear transformation to introduce a location and a scale parameter adding more flexibility to the model. It is easy to see that the resulting pdf has the expression given by

$$f(x) = \frac{\phi(\eta_{\mu,\sigma}(x)) \Phi((1 + \eta_{\mu,\sigma}(x))\lambda)}{\sigma x \Phi(\lambda_0)}, \quad x > 0, \quad (3)$$

where

$$\begin{aligned} \eta_{\mu,\sigma}(x) &= \frac{\log x - \mu}{\sigma}, \\ \lambda_0 &= \frac{\lambda}{\sqrt{1 + \lambda^2}}. \end{aligned}$$

Here $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Observe that when $\lambda = 0$ expression (3) reduces to the classical lognormal distribution. It can be seen that the λ parameter regulates the shape of the distribution. Furthermore, Lin and Stoyanov (2009) established that the distribution has heavy tails and therefore suitable for modelling expenditure in tourism data. The mean and second-order moment of the pdf given in (3) are given by,

$$E(X) = \frac{\Phi((1 + \sigma)\lambda_0)}{\Phi(\lambda_0)} \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad (4)$$

$$E(X^2) = \frac{\Phi((1 + 2\sigma)\lambda_0)}{\Phi(\lambda_0)} \exp[2(\mu + \sigma^2)]. \quad (5)$$

In advance when a random variable follows the pdf (3) we will write $X \sim \text{LSN}(\mu, \sigma, \lambda)$.

Let $F(z) = \Pr(Z \leq z)$ denote the cdf of $Z \sim \text{LSN}(\mu, \sigma, \lambda)$. Then, it is easy to see, by applying directly result B.21 in Azzalini (2013), that this cdf is given by

$$\begin{aligned} F(z) &= \Phi(\eta_{\mu,\sigma}(z)) + \frac{1}{\Phi(\lambda_0)} \left[T\left(\eta_{\mu,\sigma}(z), \frac{\lambda_0}{\eta_{\mu,\sigma}(z)}\right) + T\left(\lambda_0, \frac{\eta_{\mu,\sigma}(z)}{\lambda_0}\right) \right. \\ &\quad \left. - T\left(\eta_{\mu,\sigma}(z), \frac{(1 + \eta_{\mu,\sigma}(z))\lambda}{\eta_{\mu,\sigma}(z)}\right) - T\left(\lambda_0, \frac{\lambda^2 + (1 + \lambda^2)\eta_{\mu,\sigma}(z)}{\lambda}\right) \right], \end{aligned}$$

which is verified always that $\lambda \neq 0$. Here, $T(x, a)$ represents the Owen's function (see Owen, 1956) given by

$$T(x, a) = \frac{1}{2\pi} \int_0^a \frac{1}{1+t^2} \exp \left[-\frac{1}{2}x^2(1+t^2) \right] dt, \quad a \in \mathbb{R}.$$

Write now μ as

$$\mu = -\frac{\sigma^2}{2} + \log \left[\frac{\gamma \Phi(\lambda_0)}{\Phi((1+\sigma)\lambda_0)} \right],$$

thus the resulting pdf given in (3) has now mean given by $\gamma > 0$. Now, let $\mathbf{y}_i = (y_{1i}, \dots, y_{ki})'$ be a vector of k covariates associated with the i th observation, $i = 1, 2, \dots, n$, which is a vector of linearly independent regressors that are thought to determine the expenditure (at origin, destination or both) x . For the i th observation, the model takes the form

$$\begin{aligned} X_i &\sim \text{LSN}(\gamma_i, \sigma, \lambda), \\ \log(\gamma_i) &= \mathbf{y}'_i \boldsymbol{\beta}, \end{aligned}$$

for $i = 1, \dots, n$, where n denotes the number of observations and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ is the corresponding vector of regression coefficients. Observe that the logit link assumed ensures that γ_i falls within the interval $(0, \infty)$.

3 Empirical results

In this section, the versatility of the proposed log-skew normal model (LSN), as compared with the gamma (G), generalized gamma (GG) (see Stacy, 1962), inverse Gaussian (IG) and classical lognormal models (LN), is tested using the data obtained from the Canary Islands Tourist Expenditure Survey (Encuesta de Gasto Turístico, 2017). To do that, firstly will be described the variables to be used, these variables were taken from the survey carried out by the Canary Islands Institute of Statistics (ISTAC) from 39,000 personal interviews with tourists on the day of their departure, among the 16 million visitors to the Canary Islands in 2017. Then, a parameter estimation without using covariates will be done and a diagnosis of the used test will be carried out. Finally, estimation including covariates will be reported.

3.1 Variables

1. Dependent. In this study, the approach given to tourist spending aims to obtain a homogeneous group. We consider three categories of tourist spending per person per day are considered (total, origin, and destination). This classification is in line with the studies that segment different tourist expenditure types, Vinnicombe and Sou (2014); Disegna et al. (2017); Brida et al. (2018); Sangwon et al. (2020); and Pani et al. (2020), as explained by Svensson et al. (2011). Segmentation is a relevant instrument to understand visitor spending patterns.

Some descriptive statistics of the three variables of interest are shown in Table 1, R which shows that in mean the largest contribution to the total expenditure is provided by the spending at the country of origin. The maximum expenditure also occurs at the origin, mostly due to trip costs, such as flight and accommodation paid before travelling. This Table 1 also includes the bias and kurtosis values for the three variables. It is known (see, for instance, Groeneveld and Meeden, 1984) that the skewness's value using the standardized third central moment may become arbitrarily large and thus difficult to interpret. For this reason, we have also introduced the Bowley coefficient of skewness, $b_3 = (Q_3 + Q_1 - 2Q_2) / (Q_3 - Q_1)$, being Q_j the j th quartile of the data.¹ The moderate positive value of b_3 and the large value of the kurtosis for all variables in both cases suggests an empirical distribution that is moderately skewed to the right and with a long right tail. Thus, in all the cases, the expenditure is more concentrated below than above the mean, but there are enough large expenditures that can be considered. These properties should be considered in the empirical modeling.

Table 1: Descriptive statistics of the dependent variables

	Mean	Stand. Dev.	Minimum	Maximum	Skewness	b_3	Kurtosis
Expenditure at origin	93.0579	83.1315	1.00	2641.40	13.04	0.041	318.22
Expenditure at destination	43.5297	33.1604	1.36	357.14	2.38	0.127	12.180
Total expenditure	136.588	93.2523	9.33	2664.21	10.40	0.091	226.02

2. Independent. Three categories of spending per person per day are considered to estimate the factors that affect tourism spending: total

¹Recall that $-1 < b_3 < 1$, with 1 representing extreme right skewness and -1, extreme left skewness.

spending, spending at the country of origin, and destination spending. The factors are a set of socioeconomic and behavioral variables, for this study fourteen independent factors or variables have been selected on the basis that have already been used, contrasted and interpreted by other researchers (Laesser and Crouch, 2006; Wang and Davidson, 2010; Alegre et al., 2013; Brida et al., 2013; Zheng and Zhang, 2013; Marrocu et al., 2015; Aguiló et al., 2017; Sangwon et al., 2020; and Gómez-Déniz et al., 2020). Furthermore, these factors' effects may differ depending on the category of tourism expenditure considered (total, origin, and destination). The determinant factor may have varying impacts on a specific expenditure type, depending on its level (within expense-category analysis) Sangwon et al. (2020).

- (a) Income. This variable represents the different income levels of tourists. The literature regards income as a personal budget restriction that conditions people's purchasing capacity such that higher income levels bring about higher consumption levels, (see for instance Brščić and Šugar, 2020). Therefore, the literature reports a positive association between tourism income and spending (Thrane and Farstad, 2011). It is ordered taking the following values in euros: 1 = 12,000 – 24,000; 2 = 24,001 – 36,000; 3 = 36,001 – 48,000 ; 4 = 48,001 – 60,000; 5 = 60,001 – 72,000; 6 = 72,001 – 84,000; 7 = greater than 840,00 . The data reflect on Table 2 that on average, tourist's income is between 36000 and 48000 euros.
- (b) Job. The tourists with better jobs will probably be the ones with higher vacation budgets, so it is expected that travel expenditure increase with the occupational level, Wang et al. (2006). Following the work of Gómez-Déniz et al. (2020) that include the effect of the job category in the relationship between tourist spending per person per day at destination and origin, ten modalities are distinguished: 10 = entrepreneur, 9 = autonomous/liberal, profession, 8 = highly qualified employees, 7 = medium-level employees, 6 = low-level employees, 5 = other workers, 4 = students, 3 = retired, 2 = housewife 1 = unemployed.
- (c) Age. A variable grouped into intervals of years. Its purpose is to collect tourism spending behavior patterns attributable to

generational differences. Although, the age variable has been considered important to examine its relationship with tourism spending, Alegre and Pou (2006), Saayman and Saayman (2012), Brida et al. (2013), the relation between age and tourism expenditure is inconclusive. García-Sánchez et al. (2013) suggest that there is an inverted U-shape relationship between tourist age and their daily expenditure. It will be considered the logarithm of the age for the set of tourists. Tourist expenditure is lower among young and older tourists than among those of middle-aged.

- (d) Gender. It is a sociodemographic dummy variable that takes the value 1 for males. This variable was also incorporated in the literature of tourism by Palmer-Tous et al. (2007), Kim et al. (2010), Saayman and Saayman (2012) and Gómez-Déniz and Pérez-Rodríguez (2019). The results are inconclusive. On the one hand, some of them explain that male tourists spend more than women while, in other studies, the results show that female tourists spend more, (Aguiló et al., 2017).
- (e) Nationality. This variable defines the tourist's country of residence to differentiate foreign from local travel market (Wang and Davidson, 2010, Alegre and Pou, 2006, Brida et al., 2013 and the recent work of Gómez-Déniz and Pérez-Rodríguez, 2019). Tourists are differentiated according to the following countries of residence: Germany, the United Kingdom, Spain, and the Nordic countries.
- (f) Length of stay. Duration of stay is regarded as one of the crucial elements determining travel expenditure (see in this respect, Palmer-Tous et al., 2007, García-Sánchez et al., 2013; Aguiló et al., 2017; Gómez-Déniz and Pérez-Rodríguez, 2019 and finally the work provided by Gómez-Déniz et al., 2020). It is assumed that the longer stays, the greater the budget to spend at the destination (Wang et al., 2017). However, the positive effect has been shown to decrease as days pass (see Thrane and Farstad, 2011 and García-Sánchez et al., 2013), even at a certain length, it becomes negative, Sangwon et al. (2020). The length of stay will be considered the logarithm of the length of stay for the set of tourists. It results in a minimum stay of 1 day and a maximum of 150 (Gómez-Déniz et al., 2020).
- (g) Repetition. It is a dichotomous variable that differentiates tourists

who have previous experience of the destination from those who visit the destination for the first time. Brida et al. (2013), in their literature review, they found abundant studies where there is no significant relationship between repeated destination and tourist spending. Almost a third of the reviewed works show a positive and significant relationship. Gómez-Déniz et al. (2020), also include the repeat variable to estimate the relationship between expenditures at the destination and origin per person per day. Kuo-Liang et al. (2013) found that previous travel experience does not generate significant differences in preference and consumption patterns.

- (h) Sun and beach. It is a dummy variable that takes the value 1 when the main reason for visiting the Canary Islands is to enjoy the sun and beach, and 0 otherwise. The good weather and the beaches are one of the main motivations of tourists who travel to certain destinations because they create the stage that allure tourists, Brščić and Šugar (2020). Sun and beach tourism is important because of the amount of income it generates. Sun, sand, and sea bring great economic benefits to destinations that develop their tourism sector based on that (Cabezas-Rabadán et al., 2020). However, the sun and beach profile tourist is included in the typology of those who spend less per person per day, Gómez-Déniz et al. (2020).
- (i) Accommodation. Most of the studies consulted explain that there is a positive effect between the category of accommodation and tourists' daily expenditure. Relatively higher room rates in hotels are linked to higher expenses García-Sánchez et al. (2013); Masiero and Nicolau (2012) and Aguiló et al. (2017). This positive effect between accommodation type and tourists' daily expenditure means that relatively higher room rates in hotels are linked with higher travel expenditure, Laesser and Crouch (2006)). The type of accommodation is divided into six categories: 6 = Five-star hotels, 5 = Four-star hotel/apart-hotel, 4 = One/two/three-star hotel/apart-hotel, 3 = Non-hotel accommodation, such as apartments or self-catering accommodation, 2 = Own home or that of friends or family, and 1 = Other types of accommodation. This variable was used by Gómez-Déniz et al. (2020) and

also by Alegre and Pou (2006), and Palmer-Tous et al. (2007) that included it as a dummy to distinguish between hotels and non-hotels accommodation.

- (j) Group. The number of persons travelling together using a holiday package. The literature has not presented definitive results of the effect on tourist spending, whether people travelling alone or in a group Wang et al. (2006); Wu et al. (2013); Serra, Correia y Rodrigues, 2015; Gómez-Déniz et al. (2020).
- (k) Booking in advance. This variable gives an idea of how far in advance holidays are planned. Booking holidays in advance usually allows getting lower prices. Travelers who organize their entire trip with tour operators tend to spend more than those who do not make any reservations in advance and only partial reserve elements of the trip, Sangwon et al. (2020). This variable is divided into different categories: 0 = the interviewee did not know when the reservation was made by someone else made it; 1 = tourist booked the holidays the same week of travelling; 2 = one week before travelling; 3 = two weeks in advance; 4 = three-four weeks in advance; 5 = one-two months in advance; 6 = the reservations were made two-three months in advance and 7 = more than three months before holidays. Brida et al. (2013) and Gómez-Déniz and Pérez-Rodríguez (2019) also consider this variable, including it as dichotomous, and they highlight that tourists who plan their vacations can often obtain a slight reduction in cost while maintaining their preferences.
- (l) Low cost. This is a dummy variable that takes the value one if the tourist has travelled in a low-cost airline company and 0 otherwise. Recently, Gómez-Déniz et al. (2020) interpreted that tourists who use low-cost carriers have less spending at the country of origin what favor the expenditure at the destination. Sánchez and Rubio (2008) explain that saving in air transport allows the tourist higher budget available
- (m) Trim 1. Dummy variable. Travelling in low season (spring-summer).
- (n) Trim 2. Dummy variable. Travelling in high season (autumn-winter).

Descriptive statistics of the variables explained above are shown in Table 2.

Table 2: Descriptive statistics of the independent variables

	Mean/Mode	Stand. Dev.	Minimum	Maximum	Relat. Frequency (dichotomous variables)
Length of stay	9.0501	6.2717	1.00	150	—
Repetition	—	—	0.00	1.00	0.768
Sun & Beach	—	—	0.00	1.00	0.90
Income	3.6154	2.0618	1.00	7.00	—
Age	43.582	13.281	16.00	86.00	—
Accommodation	3.8131	1.2193	1.00	6.00	—
Group	2.4443	1.1259	1.00	10.00	—
Bookadvance	6	—	0.00	7.00	—
Trim1	—	—	0.00	1.00	0.51
Trim2	—	—	0.00	1.00	0.25
Low Cost	—	—	0.00	1.00	0.519
Gender	—	—	0.00	1.00	0.495
Job	7	—	0.00	10.00	—
Germany	—	—	0.00	1.00	0.41
Britons	—	—	0.00	1.00	0.35
Spanish	—	—	0.00	1.00	0.42
Nordics	—	—	0.00	1.00	0.23
Observations	28754				

The expression of the pdf of the alternative distributions used here is provided in the Appendix of this work.

3.2 Parameters estimation without covariates

Parameter estimation for all the models considered in this paper has been completed using the maximum likelihood (ML) method. **Mathematica**® v.12.0 and corroborated with **WinRATS** v.7.0. Both the moment and the maximum likelihood method appear to be feasible means of estimating the vector of the distribution parameters through sample observations, as shown in the Appendix. Codes are available from the authors upon request. For details about these two software see Ruskeepaa (2009) and Brooks (2009), among others.

Since it can not obtain closed expressions for the maximum likelihood estimators and can not guarantee that the global maximum of the logarithm

function of the likelihood is reached, it is advisable to use several seed points as a starting method. It is also recommended to use different optimization methods (Newton-Raphson, Broyden-Fletcher-Goldfarb-Sanno, BGGS) that guarantee the same solution from any of these methods. The standard errors of the estimators can obtain by inverting the Hessian matrix. In this sense, both Mathematica and WinRats have at least two methods to reach it. The first is to retrieve them from the Cholesky factors (this package is available on the web upon request). The second, faster, is to obtain them by finite differentiation. Furthermore, the WinRats package also offers the possibility to get the maximum of the log-likelihood directly, giving us the Fisher information matrix elements. In fact, for the examples considered later, these two packages were used to get the maximum likelihood estimators quickly. Commands for fitting the skew-normal and log-skew normal distributions are also possible in Stata (see Marchenko and Genton, 2010).

The assessment model presented uses the following information criteria: Negative loglikelihood (NLL), calculated by taking the negative of the value of the log-likelihood evaluated at the ML estimates; Akaike information criterion (AIC), calculated by twice the NLL, evaluated at the ML estimates, plus twice the number of estimated parameters; Consistent Akaike Information Criteria (CAIC), a corrected version of the AIC, proposed by Bozdogan (1987), to overcome the tendency of the AIC overestimating the complexity of the underlying model as it lacks the asymptotic property of consistency. To calculate the CAIC, a correction factor based on the sample size is used to compensate for the overestimating nature of AIC. The CAIC is defined as twice the NLL plus $k(1+\ln(n))$, where k is the number of free parameters, and n refers to the sample size. Furthermore, we also include the Kolmogorov-Smirnov test (KS), the Anderson Darling test (AD), and the Cramér-Von Mises test (CVM). Note that a model with lower statistics values is better than one with a higher value. All these results are shown in Table 3. The corresponding p -values are very low, as usual, for the expenditure in all their versions. Nevertheless, the incorporated tests KS, CVM and AD, are lower for the LSN proposed.

3.3 Diagnosis test

According to the maximum logarithm of likelihood value, AIC, and CAIC, the LSN produces a better fit than the other models. A Vuong's test (see Vuong, 1989) for non-nested model selection has also been computed and the

Table 3: Parameters estimates, their standard errors in parenthesis, maximum of the loglikelihood function and AIC for the data without including covariates

Expenditure at origin												
	$\hat{\lambda}$	p -value	$\hat{\mu}$	p -value	$\hat{\sigma}$	p -value	NLL	AIC	CAIC	KS	CVM	AD
G			2.561	< 0.01	0.027	< 0.01	33510.72	67025.50	67041.00	0.065	5.217	50.036
LN			4.325	< 0.01	0.669	< 0.01	33542.12	67088.20	67103.70	0.074	9.979	59.482
IG			93.058	< 0.01	141.745	< 0.01	34084.77	68173.50	68189.00	0.116	28.674	148.427
GG	0.539	< 0.01	8.344	< 0.01	0.761	< 0.01	33398.26	66802.50	66825.80	0.051	4.440	28.664
LSN	-2.438	< 0.01	5.897	< 0.01	1.164	< 0.01	33313.99	66634.00	66657.20	0.040	2.598	16.241
Expenditure at destination												
	$\hat{\lambda}$	p -value	$\hat{\mu}$	p -value	$\hat{\sigma}$	p -value	NLL	AIC	CAIC	KS	CVM	AD
G			2.267	< 0.01	0.052	< 0.01	28995.08	57994.20	58009.70	0.061	5.485	32.004
LN			3.537	< 0.01	0.695	< 0.01	28837.90	57679.80	57695.30	0.126	42.730	218.084
IG			121.843	< 0.01	59.493	< 0.01	30764.65	61533.30	61548.80	0.315	272.704	1298.180
GG	0.209	< 0.01	47.538	< 0.01	22.432	< 0.01	28825.96	57657.90	57681.20	0.027	0.572	3.213
LSN	-1.203	< 0.01	4.512	< 0.01	0.944	< 0.01	28821.94	57649.90	57673.10	0.022	0.365	2.263
Total expenditure												
	$\hat{\lambda}$	p -value	$\hat{\mu}$	p -value	$\hat{\sigma}$	p -value	NLL	AIC	CAIC	KS	CVM	AD
G			4.079	< 0.01	0.030	< 0.01	34812.48	69629.00	69644.40	0.053	5.229	58.185
LN			4.789	< 0.01	0.492	< 0.01	34524.17	69052.30	69067.80	0.035	2.311	14.313
IG			136.587	< 0.01	476.527	< 0.01	34647.86	69299.70	69315.20	0.043	4.263	26.388
GG	0.158	< 0.01	163.700	< 0.01	76.392	< 0.01	34531.88	69069.80	69093.00	0.043	2.273	17.290
LSN	-0.980	< 0.01	5.351	< 0.01	0.622	< 0.01	34518.82	69043.90	69066.90	0.031	1.910	12.065

results are provided in Table 4. Considering the values provided by this test and the p -values, we can conclude that the LSN is preferred to all others.

Table 4: Vuong's test (VT) and p -values comparing the LSN model in front of the others

	Expenditure					
	Origin		Destination		Total	
Distribution	VT	p -value	VT	p -value	VT	p -value
G	3.409	< 0.01	8.639	< 0.01	3.946	< 0.01
LN	6.383	< 0.01	2.339	< 0.01	> 100	< 0.01
IG	7.001	< 0.01	43.250	< 0.01	6.056	< 0.01
GG	9.026	< 0.01	2.826	< 0.01	4.361	< 0.01

Graphs of the histogram of the data and fitted densities are given in Figure 1. As can be seen, the proposed distribution seems to present a good fit for the empirical data.

Using expressions given in (4) and (5) and estimated values of parameters provided in Table 3 we have computed the population mean and standard

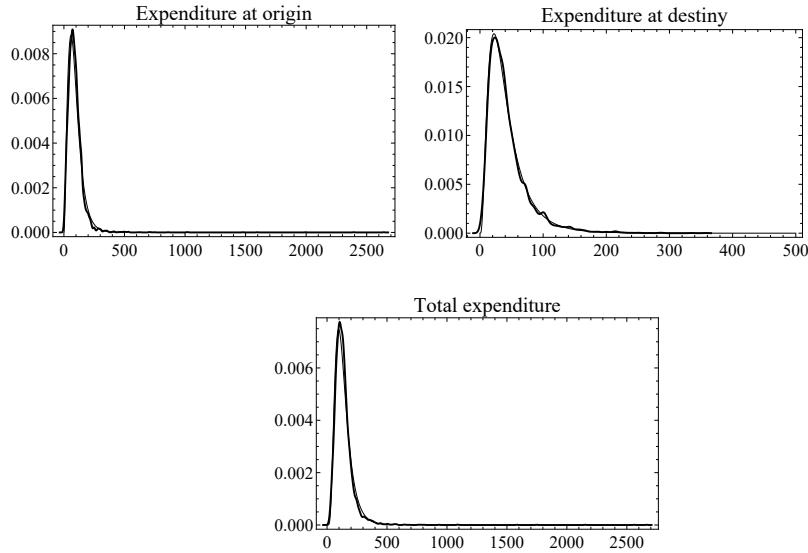


Figure 1: Smooth kernel density estimate of the empirical data (thick line) in front of the model fitted (thin line)

deviation of the three expenditures. The results corresponding to expenses at origin are 91.47 and 57.40, respectively; 43.43 and 32.49, respectively, for the spending at destination; and 135.43 and 69.23, respectively, for the total expenditure. Comparing with the empirical values provided in Table 1 we can see that the empirical means are significantly closer to the theoretical means for all types of expenditure. Nevertheless the variance is similar only for the expenditure at destination. This points out some sample bias estimates in these cases.

Table 3 shows that the values of the NLL, AIC, CAIC, and others are quite similar between the GG and LSN distributions. However, we must point out that all the statistics mentioned, in addition to Vuong's test, point to choosing between all of them, at least for the data studied here, the LSN distribution. Note that the only special function that appears in its formulation is the cdf of the normal distribution, which appears in all statistical packages. The generalized gamma distribution has the Euler gamma function in its formulation, also available in all statistical packages. However, it is much easier to work with the CDF of the normal distribution (also related to the error function) than with the Euler gamma function. To finish this

section, Figure 2 shows graphs of the different distributions used here, all of them with the same mean and variance, as well as their tail on the right highlighted in the graphs on the right. It has been assumed in the graph above, $\lambda = -0.5$ (LSN) and $\lambda = 1.5$ (GG), while in the graph below it was assumed $\lambda = 1.5$ (LSN) and $\lambda = 1$ (GG). As can be seen, the right tail of the LSN distribution is always greater than the others, thus being able to adjust better extreme values, few, but also common in tourist expenditure.

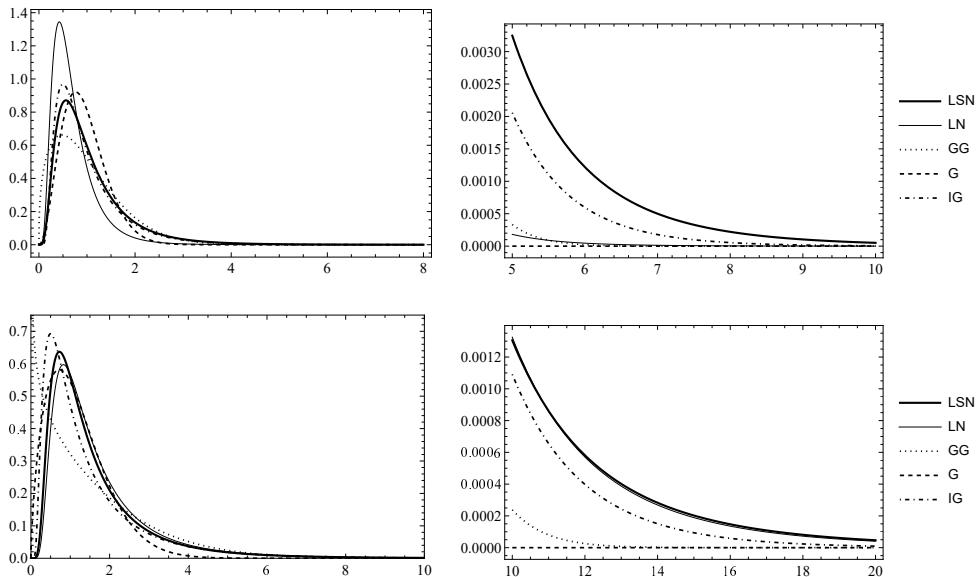


Figure 2: Pdfs of the different distributions considered (left) and right tail of the same (right) with the same mean and variance

3.4 Parameters estimation including covariates

3.4.1 Primary specifications

To reinforce the justification for the contribution of this study, when using covariates, we start by estimating the basic ordinary least squares (OLS) model and a generalization of the latter, the generalized additive model (GAM). Given the results obtained in Table 3 concerning the NLL, AIC, and CAIC, we cannot specify an alternative model (G, IG, GG, LN) that behaves in the three types of expenditure sufficiently better than the others.

We have chosen to fit a basic GAM model, given by $Y_l = \beta_0 + \sum_j f_j(x_l) + u_l$, where $f_j(\cdot)$ is the smoothing spline for the independent variables x_l and u_l , independent normal random variates. To read about these models, see, for example, Hastie and Tibshirani (1986). Many statistical packages, especially in R, allow the incorporation of a parametric model for the response variable through the appropriate link (see Wood, 2017), whose work constitutes an irreplaceable reference in the matter. We have used the mgcv package with cubic regression splines. The results for the OLS and GAM models obtained are shown in Table 5. The GAM model includes two terms to be smoothed, length of stay and age. Observe that the rest of the variables are dichotomous or categorical.

Regarding the results obtained from the estimation of these two models, see Table 5, both provide the same results on significance and sign for the explanatory variables in origin, destination, and total expenditure.

Table 5: Parameters estimates, p -values, maximum of the loglikelihood function, AIC and CAIC for the data after including covariates by OLS and GAM specifications

Variable	Expenditure at origin			Expenditure at destination			Total expenditure			
	OLS	GAM	OLS	GAM	OLS	GAM	OLS	GAM	OLS	
log(Length of stay)	-37.192	< 0.001	—	—	-12.623	< 0.001	—	—	-49.815	< 0.001
Repetition	2.943	0.199	2.867	0.211	-2.140	0.018	-2.150	0.017	0.802	0.751
Sun & Beach	-10.412	< 0.001	-9.880	0.003	-2.554	0.051	-2.056	0.117	-12.967	< 0.001
Income	4.405	< 0.001	4.328	< 0.001	1.540	< 0.001	1.486	< 0.001	5.946	< 0.001
log(Age)	9.155	0.005	—	—	2.013	0.121	—	—	11.168	0.002
Accommodation	15.373	< 0.001	15.604	< 0.001	-1.417	< 0.001	-1.166	< 0.001	13.955	< 0.001
Group	-8.234	< 0.001	-8.810	< 0.001	9.850	< 0.001	9.722	< 0.001	-18.084	< 0.001
Bookadv	0.295	0.676	0.299	0.674	0.044	0.872	0.058	0.835	0.340	0.663
Trim1	-6.821	0.004	-6.676	0.006	0.139	0.884	0.165	0.863	-6.682	0.012
Trim2	-3.011	0.222	-2.908	0.238	0.673	0.488	0.680	0.483	-2.337	0.390
Low Cost	-12.537	< 0.001	-12.420	< 0.001	1.825	0.027	1.809	0.028	-10.711	< 0.001
Gender	-0.740	0.710	-0.300	0.880	0.823	0.296	0.749	0.343	0.082	0.970
Job	0.561	0.202	0.348	0.461	0.839	< 0.001	0.836	< 0.001	1.401	0.004
Germany	2.543	0.381	2.602	0.371	-7.945	< 0.001	-8.476	< 0.001	-5.402	0.092
Britons	-2.406	0.446	-2.305	0.465	1.573	0.206	1.306	0.293	-0.832	0.811
Spanish	-13.665	< 0.001	-14.603	< 0.001	1.006	0.361	0.343	0.759	-12.659	< 0.001
Nordics	13.567	0.002	13.080	0.003	-0.466	0.790	-0.623	0.723	13.100	0.007
constant	97.581	< 0.001	4.007	< 0.001	83.836	< 0.001	4.219	< 0.001	181.418	< 0.001
Observations	6277			6277			6277			
NLL	36190.45	36176.00		30350.44	30323.95		36815.64	30.323		
AIC	72416.90	72401.93		60736.90	60703.00		73667.30	67236.07		
CAIC	72556.30	72528.30		60876.30	60823.30		73806.70	73765.20		

The results for the smooth terms are summarized in Table 6 by the effective degrees of freedom (EDF), which measures the complexity of a penalized smooth term. As it is well-known, EDF can be interpreted as an estimate of how many parameters are needed to represent the smooth. If the EDF is

equal to 1, a linear relationship cannot be rejected. In our study, the EDFs estimated show clearly non-linearity. We have chosen length of stay and age as smoothed variables. Besides being continuous variables, concerning age García-Sánchez et al. (2013) ensure that spending may decrease for the age among the oldest and the youngest tourists. There is a U-inverted-shaped relationship between tourist age and daily expenditure. Regarding the length of stay, average expenditure per person per day diminishes for longer trips, due to the economies of scale, in this line coincide Thrane and Farstad (2011), and Thrane (2014). As shown in Table 6, the smooth terms for length of stay and age are far from being linear as reflected the effective degree of freedom for the three modalities of expenditure. The better fit between these two models is for the GAM model that provides a lower value for the NLL, AIC, and CAIC.

Table 6: Results of the GAM for smooth terms

Smooth terms	Expenditure at origin		Expenditure at destination		Total expenditure	
	EDF	<i>p</i> -value	EDF	<i>p</i> -value	EDF	<i>p</i> -value
s(log(length of stay))	7.730	< 0.001	3.677	< 0.001	6.463	< 0.001
s(log(age))	6.024	< 0.001	7.254	0.011	6.572	< 0.001

3.4.2 The log-skew normal specification

We proceed to estimate the proposed model including the variables described at the beginning of the section and the results are shown in Table 7.

The estimates confirm that for the studied tourist destination, most of the estimated coefficients are statistically significant. Only booking in advance was not significant in any of the expenditures.

The following explanatory variables are significant for all the expenditure categories. Looking at the coefficients in Table 7 we obtain the length of stay (in log), sun and beach, income, age (in log), accommodation, group, low cost, and Spanish nationality.

Regarding the length of stay variable, when the number of days at destination increases, the expenditure per person's probability decreases. These results are in line with those obtained by García-Sánchez et al. (2013) and Sangwon et al. (2020). The duration of the trip negatively affects tourists' daily expenditure, although this negative effect decreases as the stay lengthens. According to the authors' results, increasing the trip's duration from one

Table 7: Parameters estimates, p -values, maximum of the loglikelihood function, AIC and CAIC for the data after including covariates

Variable	Expenditure					
	Origin		Destination		Total	
	Estimate	p -value	Estimate	p -value	Estimate	p -value
log(Length of stay)	-0.495	< 0.01	-0.271	< 0.01	-0.413	< 0.01
Repetition	0.015	0.301	-0.046	0.016	7.1E-3	0.951
Sun & Beach	-0.054	0.017	-0.059	0.019	-0.037	0.031
Income	0.034	< 0.01	0.034	< 0.01	0.032	< 0.01
log(Age)	0.103	< 0.01	0.057	0.021	0.096	< 0.01
Accommodation	0.200	< 0.01	-0.057	< 0.01	0.124	< 0.01
Group	-0.060	< 0.01	-0.240	< 0.01	-0.114	< 0.01
Bookadv	0.007	0.095	0.006	0.252	0.006	0.065
Trim1	-0.082	< 0.01	0.013	0.497	-0.057	< 0.01
Trim2	-0.058	< 0.01	0.027	0.184	-0.037	< 0.01
Low Cost	-0.132	< 0.01	0.048	< 0.01	-0.074	< 0.01
Gender	0.005	0.702	0.032	0.057	0.016	0.125
Job	0.003	0.162	0.017	< 0.01	0.009	< 0.01
Germany	0.081	< 0.01	-0.174	< 0.01	0.002	0.890
Britons	-0.042	0.068	0.050	0.056	-0.030	0.079
Spanish	-0.202	< 0.01	0.051	0.023	-0.122	< 0.01
Nordics	0.112	< 0.01	0.004	0.995	0.066	< 0.01
constant	6.713	< 0.01	6.427	< 0.01	5.057	< 0.01
λ	-1.718	< 0.01	-1.091	< 0.01	0.910	< 0.01
σ	0.830	< 0.01	0.810	< 0.01	0.449	< 0.01
Observations	6277		6277		6277	
NLL	32079.192		28086.987		33202.772	
AIC	64198.400		56214.000		66445.500	
CAIC	64353.300		56368.900		66600.400	

to two days reduces the daily expenditure per person by 6.33%, while increasing the time from 14 to 15 days reduces it only by 2.95%. Thrane and Farstad (2011) and Thrane (2014) also find a relationship between the length of stay and total tourism expenditure, the effect of length of stay on expense diminishes for longer trips due to a fixed budget and economies of scale. Also, for Svensson et al. (2011), the size of stay reduces spending per person and day of tourists (1.6% less spending per additional day).

Concerning sun and beach, those tourists whose purpose is enjoying the sun and beach decrease their probability of expenditure at the country of origin, destination, and total. Alegre et al. (2011) found that when motivation is considered, the characteristics of a sun-and-beach holiday have a negative effect on total expenditure and are not significant for either expenditure at origin or destination.

Regarding income, it has a positive effect on tourism expenditure as expected. This result coincides with those obtained by Thrane and Farstad (2011) and Brščić and Šugar (2020). Income positively affects total, at the country of origin, and destination expenditure. Gómez-Déniz et al. (2020) get positive and significant effects of income when explaining the relationship between expenditure at destination and origin. The higher the tourist income, the higher the proportion of spending at the destination.

There is a positive relationship between age and all the expenditure modalities. The older the tourist are, the higher the expenditure at the country of origin. Tourist of this profile spend more on accommodation and transportation and both expenses are made at the origin in line with (see, for instance, Van Loon and Rouwendal, 2017). In Gómez-Déniz et al. (2020) the coefficient of age (in logs) is negative and statistically significant for German and British tourists. Also Brida and Scuderi (2012), in a review of studies of determinants of tourist expenditure did not find conclusive results between age and tourist expenditure, finding as results that age and expenditure are directly, inversely related, or no significant.

The accommodation has a positive and significant impact on total and origin expenditure; the effect is negative at the destination. Results indicate that as much vacation budget tourists spend on accommodation, fewer resources are spent on destination. Gómez-Déniz et al. (2020) also obtained a negative effect on the category of housing and spending at the destination. The type of accommodation is a variable that presents significant relationships with tourist spending. Svensson et al. (2011) conclude that the heavy spender's segment uses hotels as alternative lodging. Other accommodation

categories imply significantly less spending. D'Urso et al. (2020) explain the positive relationship between spending and accommodation as a direct effect on tourist's satisfaction.

Concerning the variable group, those tourists that travel in a group often sharing costs reducing expenses. The estimated effects on expenditures per person per day are negative and significant. In the same line, Brida et al. (2018) shows that in general, those who spend more are those who travel alone. Gómez-Déniz et al. (2020) also find that travel party size negatively affects spending per person per day. Also, Aguiló et al. (2017) find that groups of people sharing expenditures during the trip, using a budget, and deciding ordinary expenditure, reduces expenses. According to Laesser and Crouch (2006) group travel is associated with 10% lower expenditure than average. Thrane and Farstad (2011) obtained results that explain that there is a proportional relationship between the size of the group and the decrease in spending per person per day.

Travel with a low-cost carrier negatively affects the origin and total expenditure per person per day; however, its effect on the destination is positive. Tourists travelling in low-cost companies save money compared to those who travel in regular companies. Still, as Gómez-Déniz et al. (2020) explain, the reduction in transport costs could interpret as an unexpected increase in available income and the investment of such resources at the destination. Nevertheless, tourists do not always transfer saved money to expenditure at the destination as found in Dobruszkes and Monodou (2013), and Eugenio-Martín and Inchausti-Sintes (2016).

The nationality has different effects in origin and at the destination depending on the country. The expenditure results negative at origin and positive at the destination for Spanish tourists. However, being German increases the probability of spending increases on the origin and decreases at the destination. These results are in line with Sangwon et al. (2020), and Gómez-Déniz et al. (2020), depending on the country of residence, tourists have different propensities to spend both at the origin and in the tourist destination.

The explanatory covariate, repetition, is significant and negative for expenditure at the destination. The probability of expenditure decreases for those tourists who know better the holiday destination. As Alegre and Cladera (2011) affirm, a high repetition rate is one of the most relevant characteristics of sun and sand destinations, but if it is expensive, tourists are less likely to repeat the visit. Tourists that spend less have a high propensity to repeat

destination.

Job has a significant and positive impact on total and at the destination expenditure. The occupation or employment status of tourists was also used by Gómez-Déniz and Pérez-Rodríguez (2019). These authors estimate the average tourist expenditure per person and day among Germans, British, and Spanish who travelled to the Canary Islands. Only for Spanish tourists, they find that professional level has a positive and significant effect on spending. Aguiló et al. (2017) find that students and unemployed are occupational modalities with a negative impact on tourist spending. The rest of the labor modalities have a positive effect. Given these results, they recommend interpreting the professional occupation concerning personal income and the possibility of spending more at the destination.

Regarding seasonality constitutes a variable of considerable importance in this scenario, (see, for instance, the recent work of Perles-Ribes et al., 2021). This variable is reflected in this work as TRIM1 (low season, spring-summer) and TRIM2 (high season, autumn-winter), with a negative marginal effect on origin and total expenditure and not significant at destination. Considering that in the Canary Islands, nowadays, the entire year is considered as high season since tourists visit the islands throughout the year. Hence, there do not appear to be differences between the two seasons in spending behavior.

The results obtained are very similar for the estimators, at least in the signs calculated and p -values to the previous OLS and GAM. The LSN offers the lowest value for the NLL, AIC, and CAIC, in all cases, with and without covariates. The LSN also provides a lower value of the KS statistics, AD CVM tests. Furthermore, with the OLS and the GAM, it is not ensured that the response variable falls into its support for any of the last two models. However, the parametric model based on the use of LSN does ensure it, in our case, $(0, +\infty)$. We have done, for comparative purposes, the adjustment for OLS and GAM as a first approximation in the case of using covariates. Based on the values obtained for the NLL, AIC, and CAIC, the LSN model shows a better fit and therefore preferable to the others.

4 Conclusions

In this study, three types of tourism spending, expenditure at the country of origin, destination, and total expenditure, are modeled via a three-parameter log-skew-normal distribution. According to the results, the proposed model

fits the empirical data in the range of the empirical distribution and captures the skewness, kurtosis, and the long tail to the right the three variables mentioned above present in practice. To the best of our knowledge, this distribution has not been used in the tourism context. It has the advantage that it adapts itself to asymmetry during modeling when data possess this characteristic. On the other hand, the regression coefficients have the same interpretation as those in the lognormal model, Chai and Bailey (2008). We compared the LSN against the OLS and GAM, the advantage of the LSN is that it ensures the variable falls into its support, $(0, +\infty)$ when prediction is desired. The results obtained are very similar in terms of the estimators, at least in the signs obtained and p -values. However, based on the values obtained for the NLL, AIC, and CAIC, the LSN model shows a better fit and therefore preferable to the others.

The results highlight that the effects of the covariates on tourism expenditure are different depending on the respective spending modalities. Regarding the length of stay, this is negative for all types of expenditure. In other words, as the number of days of the stay increases, the expenses per person and day decrease. Although we must consider the results of Sangwon et al. (2020) that state that the expenditure negative effect is reduced as shorter is the stay. Usually, the fixed expenses of the trip, which are accommodation and travel, are higher the fewer the days of holiday. The expenditure will reduce proportionally to more days of the duration of the trip. The economies of scale help to reduce all forms of spending when stays are longer. Concerning the income variable, the probability of increasing expenditure at country of origin, destination, and total increases as higher income. This variable is key to explaining expenditure. The higher the tourist income, the higher the proportion of spending at the destination. With regard to accommodation, this is usually an expense incurred by tourists at origin; for this reason, the more invested in housing at origin, the fewer expenditures are made for this concept at the destination. These results coincide with those of Gómez-Déniz et al. (2020). The category of accommodation also indicates that those who invest the most in accommodation use hotels, and any other category of accommodation implies lower spending according to Svensson et al. (2011).

Another significant variable for the three modalities of tourist expenditure is travelling in a group. Travelling in a group reduces travel expenses both at origin and destination and in total spending. Sharing a budget reduces expenditures per person. Authors such as Thrane and Farstad (2011) point

out that a proportional relationship exists between the size of the group and the decrease in spending per person per day. It is also significant the way of transport since travelling with a low-cost carrier, it positively affects expenditures at the destination, this may be a consequence of a transferring part of the savings produced by using low-cost companies in spending at the destination.

Furthermore, there are some significant covariates for one type of expenditure such as job, seasonality, repetition, and some nationalities. The job only has a significant and positive effect on spending at the destination. A higher professional category of the tourist increases the probability of a higher level of expenditure at the destination. It is expected that those tourists with a higher level of employment are willing to spend more on activities at the destination, increasing their total spending. Gómez-Déniz et al. (2020) link the professional profile of tourists by nationality to expenditures at the destination having similar results.

The seasonality in both seasons, high and low, reduces expenditures at origin. Enjoying the sun and beach is the primary purpose of visiting the Canary Islands; consequently, travelling in the high season reduces expense at origin, but it also does in the low season. It may be related that the islands are well known for having steady weather conditions all over the year. These results are interesting for policymakers and tourism managers to know tourist spending patterns in different segments markets.

Given the long right tail of the distribution, perhaps it can serve as a prior distribution for the parameter of interest of a given likelihood, discrete or continuous. For example, if it is a question of studying the length of stay distribution (usual in tourism scenarios), it usually starts from a Poisson distribution. Assumption of heterogeneity among the group of tourists we can assume the parameter of the random Poisson distribution and following the LSN distribution studied here. The resulting unconditional distribution is sure to have very heavy tails. Given that it will not have a closed expression, exhaustive numerical calculations will be required to work with it and obtain the posterior distribution. Using techniques based on WinBugs and Markov Chain Monte Carlo techniques will prove to be useful tools in this setting. These types of problems are undoubtedly interesting and worthy of future research.

This study contributes at the academic level by incorporating a model that assumes the asymmetry of tourist spending. The results are also crucial for policymakers and tourism managers because it allows to know them as

tourist spending behaves in three market segments: origin, destination, and total. By understanding the visitor's expenditure patterns, those responsible for tourism destinations can focus their policies to design the better development of the areas. As Aguiló et al. (2017) explain, tourist spending is the key economic driver, and knowing these patterns is helpful when trying to expand the market share in terms of spending and consolidated destinations.

Future research may point to slightly modifying the proposed model, including the possibility of using panel data and studying whether the model also fits well to not usual scenarios that may arise in the tourism environment. For example, in the last two years, due to the pandemic situation that has considerably affected the sector, spending has significantly decreased. We wonder if the model can also capture this scenario, which could be prolonged over time.

References

Aguiló, E. and Juaneda, C. (2000). Tourist expenditure for mass tourism markets. *Annals of Tourism Research*, 27(3):624–637.

Aguiló, E., Rosselló, J., and Vila, M. (2017). Length of stay and daily tourist expenditure: A joint analysis. *Tourism Management Perspectives*, 21:10–17.

Alegre, J. and Cladera, M. (2006). Repeat vvisitation in mature sun and sand holiday destinations. *Journal of Travel Research* 44:288–297.

Alegre, J., Cladera, M. and Sard, M. (2011). Analysing the influence of tourist motivations on tourist expenditure at a sun-and-sand destination. *Tourism Economics*, 17(4): 813–832.

Alegre, J. and Garaus, J. (2011). The factor structure of tourist satisfaction at sun and sand destinations. *Journal of Travel Research*, 50(1):78–86.

Alegre, J., Mateo, S., and Pou, L. (2013). Tourism participation and expenditure by Spanish households: The effects of the economic crisis and unemployment. *Tourism Management*, 39:37–49.

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

Arnold, B. and Beaver, R. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11(1):1–54.

Azzalini, A. (2013). *The skew-normal and related families*. Cambridge University Press, Cambridge.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B*, 61(3):579–602.

Azzalini, A., Cappello, D., and Kotz, S. (2003). Log-skew-normal and log-skew-t distributions as model for family income data. *Journal of Income Distribution*, 11:12–20.

Azzalini, A. and Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726.

Bozdogan, H. (1987). The general theory and its analytical extension. *Psychometrika*, 52(3):345–370.

Brida, J., Disegna, M., and Osti, L. (2013). Visitors’ expenditure behaviour at cultural events: The case of Christmas markets. *Tourism Economics*, 19(5):1173–1196.

Brida, J., Lanzilotta, B., Moreno, L., and Santinaque, F. (2018). A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in Uruguay. *Tourism Management*, 69:62–68.

Brida, J. and Scuderi, R. (2012). Determinants of tourist expenditure: a review of micro econometric models. *Munich Personal RePEc Archive (MPRA)*, (Paper 38468).

Brooks, C. (2009). *RATS Handbook to Accompany Introductory Econometrics for Finance*. Cambridge University Press.

Brščić, K. and Šugar, T. (2020). Users’ perceptions and satisfaction as indicators for sustainable beach management. *Tourism and Hospitality Management*, 26(1):33–48.

Cabezas-Rabadán, C., Rodilla, M., Pardo-Pascual, J., and Herrera-Racionero, P. (2020). Assessing users' expectations and perceptions on different beach types and the need for diverse management frameworks along the Western Mediterranean. *Land Use Policy*, 81:219–231.

Cárdenas, P., Pulido, J., and Pulido, M. (2015). The influence of tourist satisfaction on tourism expenditure in emerging urban cultural destinations. *Journal of Travel & Tourism Marketing*, 33(497):497–512.

Chai, H. and Bailey, K. (2008). Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Statistics in Medicine*, 27(18):3643–3655.

Craggs, R. and Schofield, P. (2009). Expenditure-based segmentation and visitor profiling at the Quays in Salford, uk. *Tourism Economics*, 15(1):243–260.

Disegna, M., D'Urso, P., and Massari, R. (2017). Analysing cluster evolution using repeated cross-sectional ordinal data. *Tourism Management*, 69:524–536.

Disegna, M. and Osti, L. (2016). Tourists' expenditure behaviour: The influence of satisfaction and the dependence of spending categories. *Tourism Economics*, 22(1):5–30.

Dobruszkes, F. and Monodou, V. (2013). Aviation liberalization as a means to promote international tourism: The EU-Morocco case. *Journal of Air Transport*, 29:23–34.

D'Urso, P., Disegna, M. and Massari, R. (2020). Satisfaction and tourism expenditure behaviour. *Social Indicators Research*, 149: 1081–1106.

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

Ferrer-Rosell, B., Coenders, G., and Martínez-García, E. (2016). Segmentation by tourist expenditure composition: An approach with compositional data analysis and latent classes. *Tourism Analysis*, 21(6):589–602.

Fredman, P. (2008). Determinants of visitor expenditures in mountain tourism. *Tourism Economics*, 14(2):297–311.

García-Sánchez, A., Fernández-Rubio, E., and Collado, M. (2013). Daily expenses of foreign tourists, length of stay and activities: Evidence from Spain. *Tourism Economics*, 19(3):613–630.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2020). On the usefulness of the logarithmic skew normal distribution for describing claims size data. *Mathematical Problems in Engineering*, 2020:1–9.

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

Gómez-Déniz, E., Pérez-Rodríguez, J., and Boza-Chirino, J. (2020). Modelling tourist expenditure at origin and destination. *Tourism Economics*, 26(3):437–460.

Goryushkina, N., Gaifutdinova, T., Logvina, E., Redkin, A., Kudryavtsev, V., and Shol, Y. (2019). Basic principles of tourist services market segmentation. *International Journal of Economics and Business Administration*, VII(2):139–150.

Groeneveld, R. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, 33:391–399.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):297–318.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13(4):271–275.

Kim, S., Prideaux, B., and Chon, K. (2010). A comparison of results of three statistical methods to understand the determinants of festival participants' expenditures. *International Journal of Hospitality Management*, 29:297–307.

Kuo-Liang, C., C, C. C.-M., and Meyer, T. (2013). A comparison study of travel expenditure and consumption choices between first-time and repeat visitors. *Tourism Management*, 35:275–277.

Laesser, C. and Crouch, G. (2006). Segmenting markets by travel expenditure patterns: The case of international visitors to australia. *Journal of Travel Research*, 44:397–406.

Lee, J. and Choi, M. (2019). Examining the asymmetric effect of multi-shopping tourism attributes on overall shopping destination satisfaction. *Journal of Travel Research*, 59(2):295–314.

Lin, G. and Stoyanov, J. (2009). The logarithmic skew-normal distribution are moment-indeterminate. *Journal of Applied Probability*, 46:909–916.

Marchenko, Y. and Genton, M. (2010). A suite of commands for fitting the skew-normal and skew-t models. *The Stata Journal*, 10(4):507–539.

Marcussen, C. (2011). Determinants of tourist spending in cross-sectional studies and at Danish destinations. *Tourism Economics*, 17:833–855.

Marrocu, E., Paci, R., and Zara, A. (2015). Micro-economic determinants of tourist expenditure: A quantile regression approach. *Tourism Management*, 50:13–30.

Masiero, L. and Nicolau, J. (2012). Tourism market segmentation based on price sensitivity: Finding similar price preferences on tourism activities. *Journal of Travel Research*, 51(4):426–435.

Owen, D. (1956). Tables for computing bivariate normal probabilities. *Tables for computing bivariate normal probabilities*, 27:1075–1090.

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.

Pani, A., Sahu, P., and Majumdar, B. (2020). Expenditure-based segmentation of freight travel markets: Identifying the determinants of freight transport expenditure for developing marketing strategies. *Research in Transportation Business & Management*. In press <https://doi.org/10.1016/j.rtbm.2020.100437>.

Perles-Ribes, J.F.; Moreno-Izquierdo, L. and Torregrosa, T. (2021). Satisfaction, seasonality and tourist expenditure in consolidated tourist destinations. *Anatolia*. In press <https://doi.org/10.1080/13032917.2021.1901752>.

Ruskeepaa, H. (2009). *Mathematica Navigator. Mathematics, Statistics, and Graphics. Third Edition*. Academic Press. USA.

Saayman, A. and Saayman, M. (2012). Determinants of spending at three major sporting events in South Africa. *International Journal of Tourism Research*, 14:124–138.

Sánchez, A. and Rubio, E. (2008). Some impacts of the low-cost air carriers over the tourist activity. *Amfiteatru Economic*, 10:154–166.

Sangwon, P., Woo, M., and Nicolau, J. (2020). Determinant factors of tourist expenses. *Journal of Travel Research*, 59(2):267–280.

Stacy, E. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 33(3):1187–1192.

Svensson, B., Moreno, P., and Martín, D. (2011). Understanding travel expenditure by means of market segmentation. *The Service Industries Journal*, 31(10):1683–1698.

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). Examining tourists' long-distance transportation mode choices using a multinomial logit regression model. *Tourism Management Perspectives*, 15:115–121.

Thrane, C. and Farstad, E. (2011). Domestic tourism expenditures: the non-linear effects of length of stay and travel party size. *Tourism Management*, 32:46–52.

Van Loon, R. and Rouwendal, J. (2017). Travel purpose and expenditure patterns in city tourism: evidence from the Amsterdam metropolitan area. *Journal Cultural Economic*, 41:109–127.

Vinnicombe, T. and Sou, P. (2014). Market segmentation by expenditure: an underutilized methodology in tourism research. *Tourism Review*, 69(2):122–136.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307–333.

Wang, D. and Davidson, M. (2010). A review of micro-analyses of tourist expenditure. *Current Issues in Tourism*, 8(5):333–346.

Wang, Y., Rompf, P., Severt, D., and Peerapatdit, N. (2006). Examining and identifying the determinants of travel expenditure patterns. *International Journal of Tourism Research*, 13(5):507–524.

Wang, Y., So, K. K. F., and Sparks, B. (2017). Technology readiness and customer satisfaction with travel technologies: A cross-country investigation. *Journal of Travel Research*, 56(5):563–577.

Wood, S. (2017). *Generalized Additive Models. An Introduction with R. Second Edition*. CRC Press. x

Wu, L., Zhang, J., and Fujiwara, A. (2013). Tourism participation and expenditure behaviour: analysis using a scobit based discrete-continuous choice model. *Annals of Tourism Research*, 40:1–17.

Zheng, B. and Zhang, Y. (2013). Household expenditures for leisure tourism in the USA, 1996 and 2006. *International Journal of Tourism Research*, 15(2):197–208.

Appendix

PDF of the alternative distributions

The pdf of the gamma, generalized gamma and inverse Gaussian distributions used in this work are given by,

$$\begin{aligned}
 f(x) &= \frac{\sigma^\mu}{\Gamma(\mu)} x^{\mu-1} \exp(-\sigma x), \quad x > 0, \mu > 0, \sigma > 0, \\
 f(x) &= \frac{\lambda \sigma^\mu}{\Gamma(\mu)} x^{\lambda \mu - 1} \exp(-\sigma x^\lambda), \quad x > 0, \mu > 0, \sigma > 0, \lambda > 0, \\
 f(x) &= \sqrt{\frac{\sigma}{2\pi x^3}} \exp\left[-\frac{\sigma(x-\mu)^2}{2x\mu^2}\right], \quad x > 0, \mu > 0, \sigma > 0,
 \end{aligned}$$

respectively.

Normal equations

Let us first consider the case of the model with no covariates and let $\tilde{x} = \{x_1, x_2, \dots, x_n\}$ a sample obtained from the distribution (3). It is easy to see that if $\Theta = (\mu, \sigma, \lambda)$ is the vector of parameters to be estimated, then the log-likelihood is proportional to

$$\ell(\tilde{x}; \Theta) = \sum_{i=1}^n \left[\log \Phi((1 + \eta_{\mu, \sigma}(x_i))\lambda) - \frac{1}{2}\eta_{\mu, \sigma}(x_i)^2 \right] - n [\log \sigma - \log \Phi(\lambda_0)].$$

The normal equations which provide the maximum likelihood estimators result

$$\begin{aligned} \frac{\partial \ell(\tilde{x}; \Theta)}{\partial \mu} &= \sum_{i=1}^n \eta_{\mu, \sigma}(x_i) - \lambda \sum_{i=1}^n \frac{\phi((1 + \eta_{\mu, \sigma}(x_i))\lambda)}{\Phi((1 + \eta_{\mu, \sigma}(x_i))\lambda)} = 0, \\ \frac{\partial \ell(\tilde{x}; \Theta)}{\partial \sigma} &= -\lambda \sum_{i=1}^n \frac{\eta_{\mu, \sigma}(x_i) \phi((1 + \eta_{\mu, \sigma}(x_i))\lambda)}{\Phi((1 + \eta_{\mu, \sigma}(x_i))\lambda)} + \sum_{i=1}^n \eta_{\mu, \sigma}(x_i)^2 - n = 0, \\ \frac{\partial \ell(\tilde{x}; \Theta)}{\partial \lambda} &= \frac{n\lambda_0 \phi(\lambda_0)}{\Phi(\lambda_0)} - \lambda(1 + \lambda^2) \sum_{i=1}^n \frac{(1 + \eta_{\mu, \sigma}(x_i))\phi((1 + \eta_{\mu, \sigma}(x_i))\lambda)}{\Phi((1 + \eta_{\mu, \sigma}(x_i))\lambda)} = 0. \end{aligned}$$