Journal of
Risk and Financial Management

Article

Skewed Binary Regression to Study Rental Cars by Tourists in the Canary Islands

Nancy Dávila-Cárdenes [1,*], José María Pérez-Sánchez [2], Emilio Gómez-Déniz [1] and José Boza-Chirino [1]

[1] Department of Quantitative Methods & TIDES Institute, Campus Universitario de Tafira, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain; emilio.gomez-deniz@ulpgc.es (E.G.-D.); jchirino@ulpgc.es (J.B.-C.)

[2] Department of Applied Economic Analysis & TIDES Institute, Campus Universitario de Tafira, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain; josemaria.perez@ulpgc.es

* Correspondence: nancy.davila@ulpgc.es

Abstract: Tourism is one of the economic sectors that contributes the most to the gross domestic product in many countries, moving, in turn, other economic sectors such as transport. In particular, the automotive industry constitutes an economic subsector that moves vast amounts of money. Concerning tourism and transport sectors, car rental is a crucial element contributing considerably to gross domestic product and job creation. Due to the effects that vehicle rental seems to have on various economic sectors, it seems interesting to know why a tourist chooses to rent a car during their vacation in a specific destination. This work aims to study those factors that can be considered relevant and affect the probability of renting a vehicle. The document addressed the following research topics: (a) identifying significant variables; and (b) can information on these factors help car rental firms? Empirically, it is shown that more tourists do not rent a car and this fact has to be considered. Thus, the classical logistic and Bayesian regression models do not seem adequate in this case, so that the authors will consider an asymmetric logistic regression model. This work analyzes 28,235 tourists who visited the Canary Islands during 2017. From a Bayesian point of view, asymmetric logistics regression is chosen as the best model because it detects relevant development factors not seen by standard logistic regressions. In light of the document's findings, various practice recommendations improve decision-making in this field. The asymmetric logit link is a helpful device that can help rental companies make decisions about their clients.

Keywords: asymmetric link; Bayesian estimation; car renting; logistic regression; models selection

Citation: Dávila-Cárdenes, Nancy, José María Pérez-Sánchez, Emilio Gómez-Déniz, and José Boza-Chirino. 2021. Skewed Binary Regression to Study Rental Cars by Tourists in the Canary Islands. *Journal of Risk and Financial Management* 14: 541. https://doi.org/10.3390/jrfm14110541

Academic Editor: Thanasis Stengos

Received: 24 September 2021
Accepted: 3 November 2021
Published: 10 November 2021

1. Introduction

Tourism has become one of the fastest-growing economic sectors globally, before the COVID-19 (affecting for the moment to the years 2020 and 2021) pandemic, contributing to the development of a growing number of new destinations. This dynamic has turned tourism into a fundamental driver for socio-economic progress. Consequently, the increasing flow of people moving around the world requires a well-developed system of mobility.

The World Travel and Tourism Council reported that, in 2018, the global travel and tourism industry grew by 3.9%. It created more than 319 million jobs worldwide and amassed 8.8 trillion in 2018. This massive growth in the industry also boosted other related industries' development, such as car rental. Nowadays, tourists prefer renting cars to explore a new place as it gives them the flexibility of planning their itinerary.

Recent trends in tourism point out an increasing emphasis on "self-service" tourism. Tourist habits worldwide seem to favour a higher number of shorter breaks to short-distance, leading to increased mobility. Furthermore, low-cost airlines and the increasing use of the Internet give holidaymakers direct access to tourism service providers and, by extension, an increase in tourist mobility in the host regions (Palmer-Tous et al. 2007).

As we can see, the tourism concept is linked with mobility because people go from their familiar environment to other places for recreation, leisure, business, etc. In their literature review, Currier and Falconer (2014) affirm that effective transport systems are key for developing the destination and generating sustainable visitor markets. However, although the relationship between transport and tourism has been considered significant in tourism research, it is interesting that concerning studies on tourist mobility in islands are scarce in the academic literature, let alone those around rental vehicles. It is not an objective in this work to analyze the impact that mass tourism undoubtedly has on the environment and the means of transport used by it, both to get to the place of enjoyment for the vacation and the one chosen to move to their own destination. In this sense, the reader can consult Martín et al. (2018), Martín et al. (2019) and the references provided therein.

The existing studies, including the car rental industry, focus on considering car rental as part of tourism's economic impact, as Martín et al. (2019) affirm. Other studies also analyze tourists' different services' satisfaction, including renting a car (Chadee and Mattson (1996)). Ekiz et al. (2010) point out that one of the critical determinants of customer satisfaction is service quality perception which has received considerable attention in marketing literature and consumer behavior. However, despite its undeniable contribution to the tourist experience and the economic impact, the car rental industry's quality issues have received less attention. Other papers that have faced the car rental analysis focus on designing systems to be used by a car rental company as in Patel et al. (2018), who developed a website for people to book their vehicles, including any requirements, from anywhere. Undoubtedly, vehicle rental services are relatively developed in industrialized economies and have been around for a long time. However, despite being one of the most structured self-drive tourism segments, the car rental industry is not well documented in the transport and tourism academic literature (Lohmann and Duval 2011).

In this paper, we focus on the rental car industry from the perspective of people who use the services to provide the industry with information for the purposes it considers. As Martín et al. (2019) obtain in their paper applied to Lanzarote, one of the eight Canary Islands, the island's tourism model is based on natural resources located all around the island, and visitors want to enjoy a tailor-made visit. However, the sun and the beach are an attraction for tourists visiting, in general, the islands where they want to move to different locations; the point is whether renting a car or not is the preferred option.

The Canary Islands have an economy based on tourism, and the car rental sector has been growing since 2011, driven mainly by the tourism sector. During the last few years, tourist arrivals have broken occupancy records in the islands, increasing the demand for car rental services. The Canarian Regional Federation of importers and car dealers collected in their annual report that the car rental sector turnover was around 520M euros in 2017, increasing 6% concerning 2016. The Finance Department of the Canary government received through the General Indirect Canary Tax (IGIC) around 49M euros (Gómez-Déniz et al. 2020).

A classic logit model, based on the use of the logistic distribution, can analyze the factors that determine why tourists decide as a mobility alternative between renting a car or not. However, individual results are sometimes much more frequent with one category than another. That is, there is an apparent asymmetry between the two response variables. This fact can be decisive when the model specifies certain factors as significant and can therefore have unpredictable consequences in economic agents' decisions. Empirically, it has been shown that this situation is frequent in practice. In particular, for the database considered in this work, many more tourists decide not to rent than to rent a vehicle. Hence, a specification based on an asymmetric logit model seems better than the classical model, which assumes symmetry between the two response variables. Since the publication of Prentice (1976) pioneering work, numerous models that generalize discrete-choice models, mainly the logit and probit models, have developed other alternatives that in many cases require deep computational calculations that today's computers allow doing. In this context, Chen et al. (1999) applied a Bayesian procedure using a skewed link in their analysis of binary response data when one response is much more frequent than the other. Bazán et al. (2006) introduces a new skew-probit link for item response by considering an accumulated skew-normal distribution. Lemonte and Bazán (2018) consider a broad class of parametric link functions that contains as special cases both symmetric as well as asymmetric links when binary choices are considered. Caron et al. (2018) extends the asymmetry logit model to the multivariate one by using a link based on the Weibull distribution. Some applied papers related to this topic are the following: Bermúdez et al. (2008) used a skewed logit link for estimating the fraudulent conduct reflected in a Spanish database of insurance claims. Sáez-Castillo et al. (2010) applied the asymmetric logit model to analyze infection rates in a General and Digestive Surgery hospital department. Pérez-Sánchez et al. (2014) studied the risk variables underlying automobile insurance claims taking into account the asymmetry of the database. Alkhalaf and Zumbo (2017) studied logistic regression when some of the predictors have skewed cell probabilities and finally Mwenda et al. (2021) uses the logistic model proposed by Prentice (1976) to study correlated infant morbidity data.

The formal aspects of the different logistic regression models considered in this work are developed in Section 2. The description of the database is shown in Section 3. Section 4 discusses the results, and conclusions and future lines of research connected with this work are presented in the last section.

2. Methodology

In this section, classic logit, Bayesian, and asymmetric Bayesian logit models are described in detail. As it is well-known, logit and probit models are the highest popular models regarding binary outcomes. A binary response model is a regression model in which the dependent variable Y is a binary random variable that takes only the values zero and one. In our case, the variable $y = 1$ if a tourist rents a car and $y = 0$ otherwise. In this article, we use the logit model to estimate the probability of renting cars given a set of characteristics of the event; that is, given the predictor X, we estimate $\Pr(1|X = x)$, i.e., the conditional probability that $y = 1$ given the value of the predictor. As is known, the logit specification is a particular instance of a generalized linear model (see Weisberg 2005, chp. 12, for details). On the other hand, the logistic link function is a moderately not confusing alteration of the prediction curve and yields odds ratios. Both characteristics make it well-received among researchers in front of the probit regression. The standard logistic distribution has a closed-form expression and a shape notably similar to the normal distribution. Logit models have been used widely in several fields, including medicine, biology, psychology, economics, insurance, politics, etc. Recent applications of binary response specification in car renting are Gomes de Menezes and Uzagalieva (2012), Masiero and Zoltan (2013), Dimatulac et al. (2018) or Narsaria et al. (2020), among others. Gomes de Menezes and Uzagalieva (2012) analyze the demand function of car rentals in the Azores, taking into account the asymmetry by estimating a family of zero-inflated models.

2.1. Logistic Specification

To make the paper self-contained, we describe the logistic specification briefly. Let $\mathcal{Y}_i$ be a continuous and unobserved random variable associated with the event of renting a car for a person i which can be specified as $\mathcal{Y}_i = x_i'\beta + \varepsilon_i$, where $\beta = (\beta_1, \cdots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients, which represents the effect of each variable in the model, and it should be estimated and $x_i = (x_{i1}, ..., x_{ik})'$ is a vector (explanatory variables) of known constants, which can include an intercept, the vector of covariates for the tourist i in our case. The random variable ε is a disturbance term. We assume that

$$\begin{aligned} Y_i &= 1 \qquad \text{if } \mathcal{Y}_i > 0, \\ Y_i &= 0 \qquad \text{otherwise.} \end{aligned}$$

Thus, we have

$$p_i = \Pr(Y_i = 1) = \Pr(x_i'\beta + \varepsilon_i > 0) = 1 - F(-x_i'\beta),$$

where $F(\cdot)$ is the cumulative distribution function of the random variable ε. Furthermore, the marginal effect on p_i for a change in x_k results $f(-x_i\beta)\beta_k$, where $f(\cdot)$ is the probability density function of the random variable ε.

If we assume $F(\cdot)$ to be the standard normal cdf, $\Phi(\varepsilon)$, we get the probit model, and if we assume the logistic distribution, we have the logistic regression, which will be considered here. Then, for observation i in a sample of size n, we assume that

$$p_i = \Pr(Y_i = 1) = \frac{1}{1 + \exp(-x_i'\beta)} = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)},$$

and $\Pr(Y_i = 0) = 1 - p_i$. Recall that the probability density function of the standard logistic distribution is symmetric about 0. In summary, the logit specification adopts the following form:

$$\log\left(\frac{p_i}{1 - p_i}\right) = x_i'\beta, \quad i = 1, 2, \ldots, n.$$

Thus, the likelihood is given by

$$\ell(y|x, \beta) = \prod_{i=1}^{n} [F(x_i'\beta)]^{y_i} [1 - F(x_i'\beta)]^{1-y_i}, \tag{1}$$

where the β parameters are usually estimated by the maximum likelihood method. In this way, the model gives the probability of each tourist renting a car. The next step is to consider a cut-off for determining whether a tourist will rent or not. The classical logit (frequentist approach) model is implemented in most of the standard statistical packages as Mathematica (Champaign, IL, USA), STATA (Texas, TX, USA), and R (Vienna, Austria), among others. We have estimated the basic logit model using STATA 14.1. StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

2.2. Bayesian Symmetric Specification

In contrast to the frequentist approach, the Bayesian approach has gained popularity in the last few decades. In the past, the main motivation for using the standard logit regression model was basically by computational effort. However, software for implementing other methodologies became widely available in the last few decades due to the advances in computational sciences. From the pioneering work of (Zellner [1971] 1996), the applications of Bayesian methodology in econometrics theory have increased considerably. In the Bayesian approach, the β parameters are considered random variables assuming non-informative and centered normal prior distributions, making the comparisons with classical results easy. The Bayesian methods use the data and the prior knowledge to obtain the estimations, and these results usually are more accurate than those derived under classical methods.

Bayesian inference for logit studies satisfies the standard mechanism in Bayesian analysis consisting of the likelihood function of the data, the prior distribution over the unknown parameters, and the Bayes theorem to compute the posterior distribution of the parameters.

The set of unknown parameters is represented by the vector $\beta = (\beta_1, \ldots, \beta_k)$. Thus, the logit Bayesian model has the following stochastic representation:

$$\log\left(\frac{p_i}{1 - p_i}\right) = x_i'\beta, \tag{2}$$

$$\beta \sim \pi(\beta), \tag{3}$$

where $\pi(\cdot)$ is the prior distribution of β. The selection of the prior distribution can involve informative prior distributions if the researcher knows something about the parameters or non-informative prior if there is little information about these coefficients. A problem arises when informative prior distributions are chosen: the information must be given on the logit scale, i.e., on the β parameters directly.

We suppose, as it is usual that the parameters of the logit models follow a normal distribution, $\beta_j \sim N(\mu_j, \sigma_j^2)$, $j = 1, \ldots, k$, where μ is zero, and σ is usually chosen to be large enough to be considered as non-informative.

By combining the prior assumption with the likelihood in (1), we obtain the posterior distribution for the parameters β, which is proportional to

$$\begin{aligned}\pi(\beta|y,x) &\propto \ell(y|x,\beta)\pi(\beta)\\ &= \left\{\prod_{i=1}^{n}\int_0^{\infty}[F(x_i'\beta)]^{y_i}[1-F(x_i'\beta)]^{1-y_i}\right\}\prod_{j=1}^{k}\frac{1}{\sigma_j\sqrt{2\pi}}\exp\left(-\frac{\beta_j^2}{2\sigma_j^2}\right).\end{aligned}$$

Multiple integrations for calculating the marginal distribution are required because it does not have a closed-form expression. The literature in this respect uses a Gibbs sampler (see Carlin and Polson 1992 and Gilks et al. 1995, for further details) as implemented by WinBUGS to obtain approximately the properties of the marginal posterior distributions for each parameter. WinBUGS (1.4, Cambridge, UK), the MS Windows operating system version of BUGS: Bayesian Analysis Using Gibbs Sampling, is a flexible software program that carries out Markov chain Monte Carlo (MCMC) simulations for a broad diversity of Bayesian models (WinBUGS was developed jointly by the Medical Research Council Biostatistics Unit (University of Cambridge, UK) and the Imperial College School of Medicine at St. Mary's, London; see Lunn et al. (2000)).

2.3. Bayesian Asymmetric Specification

The regression logit model outlined above is too simple to be used for any serious empirical work when the sample data present asymmetry between the two values of the binary response variable, as occurs in our database. In this context, a Bayesian approach is a powerful tool providing more flexible models in regression analysis.

The main idea of the Bayesian regression model (Zellner [1971] 1996 and Koop 2003) is to consider that the regression coefficients are random and fit a distribution function (the prior distribution). We propose two alternative Bayesian estimations of the logit model. The first model appears as a special case using a symmetric link function, and secondly, an asymmetric link function.

From the asymmetric standpoint, an approach based on data augmentation (see Albert and Chib 1995) can be used. In this case, it is easily shown that the asymmetric logit link is equivalent to considering the following:

$$Y_i = \begin{cases} 1, & w_i \geq 0, \\ 0, & w_i < 0, \end{cases}$$

where $w_i = x_i'\beta + \delta z_i + \varepsilon_i, z_i \sim G, \varepsilon_i \sim F$ and $i = 1, 2, \ldots, n$. We assume that z_i and ε_i are independent, and that F is the standard logistic cumulative distribution function. Moreover, G is the cumulative distribution function of a random variable with positive support. Thus, candidates for this distribution are the exponential and the half-normal distribution, among others. In this paper, it will be assumed for G the half-standard normal distribution with probability density function given by

$$g(z_i) = \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_i^2\right), \quad z_i > 0.$$

In this model, $\delta \in (-\infty, \infty)$ is the skewness parameter and so the skewness of the regression model is measured by δz_i. If $\delta > 0$, the probability of $p_i = 1$, i.e., the probability that the ith tourist will rent a car increases. On the other hand, if $\delta < 0$, the probability of not renting a car increases. Obviously, the symmetric logit model is a special case of the previous model obtained for $\delta = 0$.

Figure 1 shows the cdf given by $\int_0^\infty F(x_i\beta + \delta z_i)g(z_i)\,dz_i$ for special values of parameter δ. Since the latter integral has not a closed-form expression, we have proceeded by numerical integration, obtaining a cloud of points used to plot a line through these points after assuming $\beta = 1$. It is observed that, as the delta parameter takes values different from zero (corresponds to the symmetric case), the shape of the curve varies.

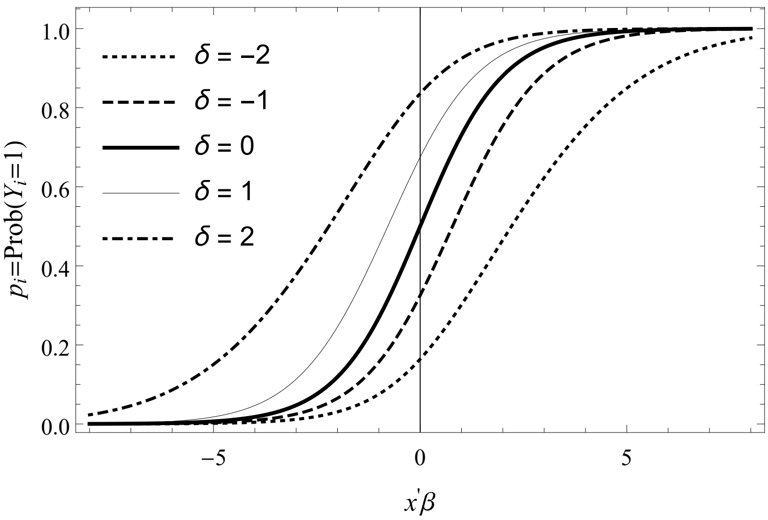

Figure 1. Cumulative density function (logistic kernel mean function) of the skewed logit model for special values of skewness parameter δ. The case $\delta = 0$ corresponds to the classical logistic distribution.

Variation of the marginal effect in front of p_i and for special values of the parameter δ can be seen in Figure 2. This figure clearly shows the relationship between p_i and the marginal effect ($\partial p_i/\partial x'\beta$). Its maximum value goes from $p_i = 0.5$ (case of symmetry with $\delta = 0$) to the left or right as the parameter δ decreases or increases. As can be seen, the marginal effects takes on its maximum values at different probability levels depending upon the value of δ.

The following likelihood function is thus obtained:

$$l(y|x,\beta,\delta) = \prod_{i=1}^{n} \int_0^\infty [F(x_i'\beta + \delta z_i)]^{y_i}[1 - F(x_i'\beta + \delta z_i)]^{1-y_i} g(z_i)dz_i. \tag{4}$$

In the context of Bayesian analysis, a prior distribution must be specified for β and δ, say, $\pi(\beta,\delta)$. We assume non-informative and centred normal prior distributions for both parameters in order to facilitate comparison with frequentist estimations, i.e., $\beta_j \sim N(0,\sigma_j^2), \forall j = 1,\ldots,k$, and $\delta \sim N(0,\sigma_\delta^2)$, considering $\sigma_j > 0, \forall j = 1,\ldots,k$, and σ_δ sufficiently large, noting the absence of prior knowledge about the parameters of interest, which facilitates comparison with the frequentist model.

By combining these prior assumptions with the likelihood shown in (4), we obtain the posterior distribution for the parameters β and δ, which is proportional to the prior times the likelihood,

$$\begin{aligned} \pi(\beta,\delta|y,x) &\propto l(y|x,\beta,\delta)\pi(\beta,\delta) \\ &= \left\{ \prod_{i=1}^{n} \int_0^\infty [F(x_i'\beta + \delta z_i)]^{y_i}[1 - F(x_i'\beta + \delta z_i)]^{1-y_i} g(z_i)dz_i \right\} \pi(\beta,\delta). \end{aligned}$$

This posterior distribution summarizes all the prior and data-based information about the unknown parameters, β and δ.

Again, we need to factor the posterior distribution, simulate the marginal posterior distribution of the parameters (or hyperparameters), and then simulate the other parameters conditional on the data and the simulated parameters. Thus, we can sample (β, δ) from this posterior distribution using the WinBUGS package.

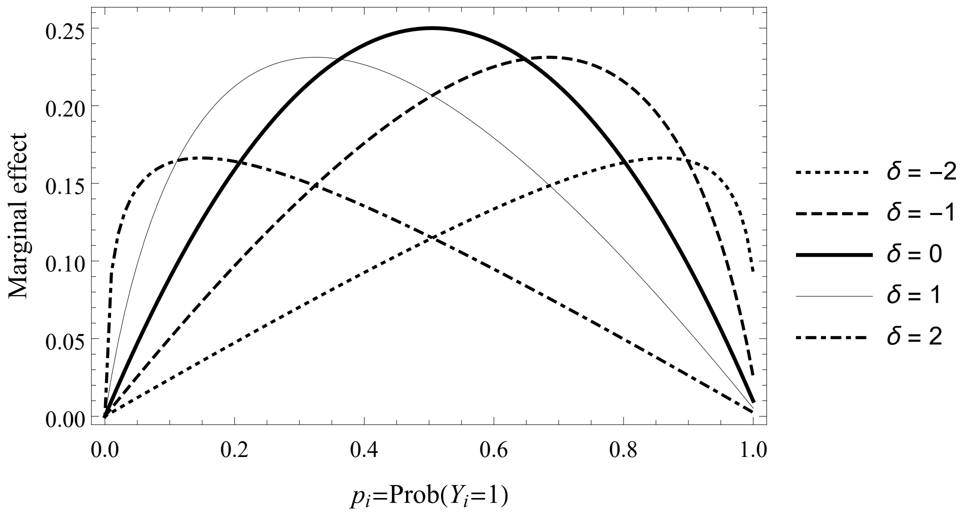

Figure 2. Marginal effect of the skewed logit model with different values of skewness parameter δ. The case $\delta = 0$ corresponds to the classical logistic distribution.

3. Description of Database

A database of a tourist survey provided by the Canarian Islands Statistical Institute (ISTAC) was used. The original database gathered approximately 39,000 personal interviews on tourists at their departure time, among about 16 million people who visited the Canary Islands in 2017. Specifically, the current analysis includes those tourists who rented (or did not) a car for at least one day. This information is essential since it would allow for knowing the profile of tourists who rent a car and plan effective measures to improve the industry. After data cleansing, to analyze the factors that might affect the probability of renting a vehicle, 28,235 pooled observations were considered. Of them, 21,933 did not rent a car, and only 6302 did, showing an apparent asymmetry in the database. To estimate the probability of renting a car, we divided the variables included in our analysis into three categories: variables associated with the trip, variables related to trip motivation, and those related to socio-economic characteristics. The main descriptive statistics of these variables are shown in Table 1.

Explanatory variables associated with the trip (General variables)

1. Origin spent. A quantitative variable defining expenses at origin per person and day. Expenditure of tourists is approximately 99.92 euros on average.
2. Destination spent. A quantitative variable defining expenses at destination per person and day. Expenditure of tourists is approximately 40.68 euros on average.
3. Nights. A quantitative variable representing the length of stay. It results in approximately nine days on average, with a minimum stay of one day and a maximum of 180.
4. Previous visits. A dummy variable takes one whether the tourist has visited the Canary Islands before the current trip and 0 otherwise. Approximately 77% of visitors repeat visits.
5. Accommodation. A dummy variable takes one if the tourist has been accommodated at a hotel and 0 otherwise.
6. Party. A dummy variable takes 1 if the tourist has travelled with someone else and 0 otherwise.
7. Booking. A dummy variable takes one if the tourist has booked the holidays at home and 0 otherwise.

8. Low cost. A dummy variable, which takes one if the tourist has travelled in a low-cost carrier.
9. Season. A categorical variable expressing the time of the year the tourist traveled: January–May, dummy variable which represents traveling from January to May; June–September, a dummy variable for traveling from June to September; and October–December, the reference dummy variable representing trips from October to December.

It was more often that visitors to the Canary Islands stayed at hotels (54.8%). On average, tourists traveled in groups (81.8%); most tourists booked their holidays before traveling (98.5%); 51.8% used low-cost carriers, and 37.1% of visitors traveled to the Canary Islands from January to May.

Explanatory variables associated with trip motivation:

10. SunBeach. A dummy variable takes one whether the main reason for visiting the Islands is enjoying sun and beach, and 0, otherwise.
11. Holiday. A dummy variable takes 1 when the reason for traveling is holidays, and 0 otherwise.

According to these two variables, results on Table 1 show that most visitors travel for enjoying sun and beach (90%) and holidays (94%).

Explanatory variables associated with socio-economic characteristics:

12. Age in years. It can be seen in Table 1 that, on average, tourists are in their forties. The minimum age of those interviewed is 16 years old, and 9 are the oldest ones.
13. Gender. A dummy variable takes 1 for males. 49.5% of visitors are men.
14. Income. An ordered categorical variable which takes the following values: (1): from 12,000 to 24,000 euros; (2): from 24,001 to 36,000 euros; (3): from 36,001 to 48,000 euros; (4): from 48,001 to 60,000 euros; (5): from 60,001 to 72,000 euros; (6): from 72,001 to 84,000 euros; and (7): greater than 84,000 euros. The data reflect on Table 1 that on average, tourist's income is between 36,000 euros and 48,000 euros.
15. Job. A dummy variable takes one if the tourist is employed and 0, otherwise. Approximately 82% of visitors are employed.
16. Nationality. Tourists are separated according to the following countries of residence: Germany, The United Kingdom, Spain, Nordic countries, and others. Mostly, incoming tourists are from the United Kingdom, followed by other countries, Spain, Germany, and Nordic countries. The dummy reference variable is 'Other'.

Table 1. Descriptive statistics of variables.

Variable	Minimum	Maximum	Mean/Mode	Standard Deviation
Renting	0	1	0.223	—
(number of observations)	(21,933)	(6302)		
General variables				
Origin spent	0.52	1988.76	99.919	64.413
Destination spent	0	500	40.679	37.105
Nights	1	180	8.917	7.238
Previous visits	0	1	0.767	—
Accommodation	0	1	0.548	—
Party	0	1	0.818	—
Booking	0	1	0.985	—
Low cost	0	1	0.518	—
Season:				
Jan-May	0	1	0.371	—
Jun-Sep	0	1	0.295	—
Oct-Dec	0	1	0.333	—
Trip motivation variables				
SunBeach	0	1	0.903	—
Holiday	0	1	0.939	—
Socio-economic variables				
Age	16	92	44.827	14.275
Gender	0	1	0.495	—
Income	1	7	3.540	2.038
Job	0	1	0.818	—
Nationality:				
German	0	1	0.173	—
British	0	1	0.276	—
Spanish	0	1	0.185	—
Nordic	0	1	0.099	—
Other	0	1	0.267	—
Observations	28,235			

Figure 3 shows a histogram of the dependent variable which reflects a significant imbalance in the two categories of outcome considered.

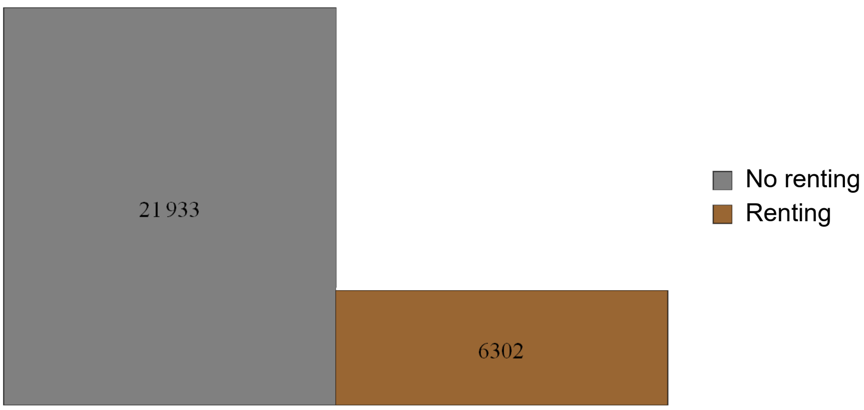

Figure 3. Histogram of the dependent variable.

4. Empirical Results and Discussion

In this section, the classic logit model and asymmetric Bayesian logit model explained in Section 2 were carried out to study the factors that determine why people who make tourism decide as a mobility alternative between renting a car or not.

4.1. Brief Explanation of the Computations

A total of 500,000 iterations were carried out (after a burn-in period of 100,000 simulations) for simulating the posterior distributions for the asymmetric Bayesian model in WinBUGS. Three different chains were carried out and the convergence was evaluated for all parameters using tests provided within the WinBUGS Convergence Diagnostics and Output Analysis (CODA) software. As it is known, this method uses the complete conditional distributions of the parameters, thus the conditional distributions of each parameter given the other parameters and the data, and requires that random numbers from these distributions be generated. The posterior marginal densities are approximated by using a random sample from the complete conditional distributions. The source codes of Bayesian estimations are available upon request from the authors.

4.2. Interpretation of the Results

The results under frequentist and non-informative asymmetric Bayesian estimations are shown in Table 2 which contains the estimated coefficients ($\hat{\beta}$), standard deviations (sd), p-values (frequentist), and MC errors (Bayesian). As it can be analyzed, both estimations are very similar in terms of signs and significance. The table also shows the marginal effects for both frequentist and asymmetric Bayesian estimations. The marginal effect on p_i on a change on x_k, for a continuous variable, can be computed as

$$\frac{\partial p_i}{\partial x_k} = \int_0^\infty \frac{\partial}{\partial x_k} F(x_i'\beta + \delta z_i g(z_i)\, dz_i = \beta_k \int_0^\infty f(x_i'\beta + \delta z_i) g(z_i)\, dz_i,$$

where $f(\cdot)$ is the pdf of the logistic distribution. As in the classical logistic models, the impact of changes in a variable x_k depends not only on β_k, but also on the value of $x_i'\beta$. Thus, a lot of caution should be needed here. Observe that, for $\delta = 0$, the marginal effect coincides with the one obtained from the classical logit link, since the integral reduces to 1. For dichotomous variables, taking values 0 and 1, the marginal effect for the variable x_k is given by

$$\int_0^\infty F(x_i'^*\beta + \delta z_i) g(z_i)\, dz_i - \int_0^\infty F(x_i'^*\beta + \delta z_i) g(z_i)\, dz_i,$$

where $x_i'^*$ represents the set of variables in which the k variable changes and the rest of the variables remain constant. Since there is a marginal effect for each individual in the sample and some variables are continuous and others dichotomous, we computed the marginal effect for all the individuals and took their mean value.

In the light of these results, the following significant variables regarding the general variables were obtained: origin and destination spending, number of nights, accommodation, party, booking, low cost, and season. Concerning expenditures, the vehicle rental is usually an expense made at a tourist destination, so the expectation of renting a car increases when the spending at the destination increases and when the expenditure at the origin country decreases. In this line, Aguiló et al. (2017) also finds a positive and significant relationship between destination expenditure and transport expenses at the destination. On the other hand, the higher the number of nights, the higher probability of renting a car. This result is similar to those of Palmer-Tous et al. (2007) and Aguiló et al. (2017) who explain that the accommodation days increases the daily expenditure transport at the destination. In addition, long vacation periods also increase the probability of renting a car for more days (Thrane and Farstad 2011 and Gómez-Déniz et al. 2020). According to the results, tourists with hotel accommodation are unwilling to move around the island, decreasing the likelihood of renting a car. Palmer-Tous et al. (2007) and Gómez-Déniz et al. (2020) show that tourists who frequently rent cars are those who do not stay in hotels. Furthermore, tourists who lodge in hotels spend fewer nights at their destination than those who choose other types of accommodation (Alegre and Pou 2006). In addition, traveling with others increases the likelihood of renting cars since rental expenditures are distributed among the group. However, these economies of scale tend to disappear when the group exceeds nine members (Marrocu et al. 2015). Regarding booking at home, the model's estimates suggest that tourists who make reservations in the countries of origin are the most likely to rent cars. It suggests that they rent a car at origin, but they pay it at the destination. According to our results, the likelihood of renting cars increases when using low-cost carriers. It can be explained because, when traveling by low-cost airlines, expenses at origin decrease, and part of these savings can be spent at the destination, including renting cars, (Gómez-Déniz and Pérez-Rodríguez 2019). On the other hand, traditional airline users, with higher transport expenditures, tend to decrease their destination spending (Ferrer-Rossell and Coenders 2017). Finally, visitors who travel from January to May, the high season in the Canary Islands, may be looking for more pleasant temperatures, decreasing the probability of renting cars.

The socio-economic characteristics of tourists (age, gender, job, income, and nationality) are significant variables in explaining the likelihood of renting cars. Concerning age and gender, results reflect that young people and men have a higher tendency to rent cars. Moreover, the willingness to rent cars increases for those employed and with higher income. Finally, those from Germany and the mainland in Spain are more likely to rent a car regarding the tourist's nationality. On the contrary, British and Nordic tourists are less likely to rent a car.

Concerning the marginal effects, holidays are the variable with a higher positive value and British with a higher negative value for frequentist and asymmetric Bayesian estimations methods. In addition, as we observed in Figure 2 in Section 2, most of the marginal effects from the frequentist estimation are greater than those from the asymmetric Bayesian results. We believe this is because part of the information provided by the marginal effects in the frequentist model is now included in the δ parameter for the asymmetric Bayesian model.

As we can also see in Table 2, the Bayesian estimated coefficients differ considerably from those of the frequentist model, although the signs remain the same. The estimation of the intercept represents the most remarkable difference. Maybe this is because the estimated constant may include part of the asymmetry in the frequentist model. Note that δ is positive, confirming the applicability of the asymmetric model to our database and providing evidence that this model increases the probability of detecting the rentals.

4.3. Checking the Models

In order to evaluate the quality of fitting of the two models, we propose four different measures: (i) the percentage of correct fittings calculated by considering the estimates probabilities; (ii) the deviance information criterion (DIC), given by $\text{DIC} = -2\ln\left(l(y|x,\hat{\beta})\right)$; (iii) the Akaike information criterion (AIC) defined as $\text{AIC} = 2(k - \ln\left(l(y|x,\hat{\beta})\right))$ (k represents the number of explanatory variables); and (iv) the Bayesian information criterion (BIC), given by $\text{BIC} = k\ln n - 2\ln\left(l(y|x,\hat{\beta})\right)$. DIC, AIC, and BIC statistics measure the relative quality of statistical models for a given set of data and models with smaller values should be preferred to models with larger ones. See Akaike (1974) and Spiegelhalter et al. (2002) for details.

The percentage of correct fittings and the results of the AIC and DIC criteria appear at the bottom of Table 2. For our database, we obtained a DIC of 27,862.584, an AIC of 27,904.584, a BIC of 28,077.798 for the frequentist logit model; and the asymmetric Bayesian logit model provided a DIC of 4647.38, an AIC of 2369 and a BIC of 2550. This table also shows that the accuracy, i.e., the proportions of rentals and non-rentals that the models correctly classified, is around 77.65% for the frequentist model (corresponding only to 124 rentals and 21,801 non-rentals) and 99.99% for the asymmetric Bayesian model (corresponding to 6302 rentals and 21,933 non-rentals). The threshold probability used to fit a rental was the sampling frequency of rentals, 0.223. As we can observe, the asymmetric Bayesian model fits the rentals and non-rentals better. Obviously, these results are explained by the increase in the probability of fitting the $y_i = 1$ cases induced by the asymmetric model, since the parameter δ is positive and highly significant, pointing out the asymmetric character of the response variable and the need of taking this into account.

Table 2. Frequentist and non-informative asymmetric Bayesian estimations.

	Frequentist				Asymmetric Bayesian			
Variables	$\hat{\beta}$	**Robust sd**	***p*-Value**	**ME**	$\hat{\beta}$	**sd**	**MC Error**	**ME**
Origin spending	-0.004 ***	3×10^{-4}	0.000	-6.4×10^{-4}	-3.246 ***	0.312	0.022	-0.002
Destination spending	0.004 ***	4×10^{-4}	0.000	6.4×10^{-4}	1.791 ***	0.187	0.013	9.9×10^{-4}
Nights	0.008 ***	0.002	0.000	1.3×10^{-3}	0.698 ***	0.184	0.010	3.5×10^{-4}
Repeat	-0.002	0.035	0.958	-3.2×10^{-4}	-0.121	0.449	0.034	-6.9×10^{-5}
Accommodation	-0.100 ***	0.033	0.001	-0.016	-1.422 ***	0.434	0.029	-8.03×10^{-4}
Party	0.591 ***	0.045	0.000	0.087	7.383 ***	0.727	0.066	0.004
Booking	0.470 ***	0.143	0.001	0.067	4.734 ***	1.462	0.144	0.002
Low cost	0.217 ***	0.031	0.000	0.035	2.775 ***	0.414	0.030	0.001
Jan-May	-0.098 ***	0.036	0.007	-0.016	-1.285 ***	0.456	0.029	-7.3×10^{-4}
Jun-Sep	-0.039	0.037	0.289	-0.006	-0.507	0.472	0.031	-2.9×10^{-4}
SunBeach	-0.069	0.054	0.198	-0.011	-0.968 *	0.635	0.057	-5.6×10^{-4}
Holiday	0.977 ***	0.083	0.000	0.125	12.33 ***	1.119	0.108	0.006
Age	-0.004 ***	0.001	0.000	-6.4×10^{-4}	-0.823 ***	0.226	0.013	-4.5×10^{-4}
Gender	0.141 ***	0.030	0.000	4.7×10^{-4}	1.760 ***	0.387	0.024	0.001
Income	0.072 ***	0.008	0.000	0.012	1.865 ***	0.241	0.016	9.4×10^{-4}
Job	0.217 ***	0.044	0.000	0.034	2.791 ***	0.601	0.052	0.0015
German	0.142 ***	0.044	0.001	0.023	1.806 ***	0.565	0.038	0.001
British	-1.053 ***	0.044	0.000	-0.150	-13.770 ***	0.977	0.087	-0.007
Spanish	0.469 ***	0.044	0.000	0.081	5.881 ***	0.688	0.056	0.003
Nordic	-0.767 ***	0.629	0.000	-0.106	-9.944 ***	1.001	0.074	-0.005
Intercept	-3.079 ***	0.183	0.000		-58.330 ***	3.765	3.765	
δ					29.090 ***	1.767	0.176	
Observations		28,235				28,235		
% Correct fit		77.61				99.99		
DIC		27,862.584				4647.380		
AIC		27,904.584				2369.000		
BIC		28,077.798				2550.000		

*** indicates 1% significance level. * indicates 10% significance level.

5. Conclusions

This paper introduced a simulation-based approach by applying a Monte Carlo Bayesian Gibbs sampling for fitting a tourism rental database using a dichotomous model. Our approach identifies the likelihood of the data by using an asymmetric logit model and then assuming a proper prior distribution of the model's parameters. Combined with the Gibbs sampler, these considerations allow us to simulate based on the posterior distribution of these parameters.

Comparing the frequentist and the asymmetric Bayesian logit estimation results, we see that the Bayes logistic model gives posterior estimations for the parameters quite different from the classical ones. Any model with classical inference should give almost the same estimations as a Bayesian inference with non-informative normal priors. However, the asymmetric consideration of the Bayesian model presents, in absolute values, estimations strongly higher than those obtained with the frequentist logit. Moreover, the frequentist model shows a lack of fit due to the incorrect classification of one case (rentals). The asymmetric Bayesian model is more suitable for fitting data when one response appears more often than the other. Due to data distribution, the asymmetry must be included in the logistic model to represent reality in a better way. Since the Bayesian asymmetric logit model presented here is only used for fitting purposes, it is necessary to search for an asymmetric link function to model the rental car database to obtain the best predictive model. Therefore, a natural extension of this paper is looking for asymmetric link functions which help us to get better predictions.

We can observe that the results are robust using both estimation methods to analyze the determinants that explain the probability of renting cars. However, it is essential to remark that there is one crucial factor for the Bayesian estimation: the frequentist one does not detect the sun and beach as the main purpose for traveling. The probability of renting cars decreases for those visitors traveling to the Canary Islands and searching for a sun and beach destiny (90.3% of the sample). We believe these tourists arrive at the beaches using another way of transport: hotel buses, on foot, etc. or stay at their hotels enjoying its facilities. Considering this result, stakeholders could consider promoting a complementary purpose for visiting the islands to attract some part of this group of tourists to the rent a car sector, informing about alternative beaches that have more difficult access than the most popular beaches. Some of the determinants found in this paper are consistent with the literature (see, for example, Gomes de Menezes and Uzagalieva 2012). In this sense, variables such as destination spending, nationality, hotel accommodation, and traveling with someone else are crucial factors in renting cars. Among them, only hotel accommodation has a negative impact. In addition, this paper detects new factors in explaining the probability of renting cars from both frequentist and Bayesian estimation methods. On the one hand, the length of stay, booking in advance, traveling with low-cost carriers, gender (men), income, and having a job are positively correlated with the probability of renting cars. On the other hand, in the season January to May and June to September, British and Nordic tourists and the age decrease the likelihood of renting cars.

The consideration of socio-economic aspects and the geographical characterization of the studied space have not been the object of this work. Regarding the latter, it is interesting to note that the Canary Islands are made up of eight islands, each with its economic, environmental, and population peculiarities that could make the results obtained in this work differ if they were analyzed separately. However, with the assignment of an a priori distribution for the parameters of interest, Bayesian consideration implicitly assumes some population heterogeneity. Therefore, from this point of view, we do not believe that the results obtained, studying the effect on the dependent variable of interest separately, if the islands are considered independently, will differ significantly from the results obtained in this analysis.

Regarding tourist mobility, there are different standpoints to approach it. However, this paper has focused only on the probability of renting cars; no aspects concerning sustainability have been considered. In this sense, in line with the environmental efficiency, Gómez-Déniz et al. (2020) propose to promote low emission car rental in tourist destinations.

In another vein, the COVID pandemic's impact on the tourism sector, particularly in the vehicle rental sector, is an element that should be addressed in the future. This sector has been forced to dispose of part of the vehicle fleet to survive during the crisis caused by the pandemic mentioned, causing direct and indirect effects on the economy that undoubtedly deserve to be studied.

Author Contributions: Conceptualization, E.G.-D., N.D.-C., J.M.P.-S. and J.B.-C.; methodology, J.M.P.-S., E.G.-D., N.D.-C. and J.B.-C.; software, J.M.P.-S. and E.G.-D.; validation, N.D.-C., J.M.P.-S., E.G.-D. and J.B.-C.; formal analysis, N.D.-C., J.M.P.-S., E.G.-D. and J.B.-C.; investigation, E.G.-D., N.D.-C., J.M.P.-S. and J.B.-C.; resources, N.D.-C., J.M.P.-S., E.G.-D. and J.B.-C.; data curation, J.M.P.-S., E.G.-D., N.D.-C. and J.B.-C.; writing—original draft preparation, N.D.-C., J.M.P.-S., E.G.-D. and J.B.-C.; writing—review and editing, N.D.-C. and J.M.P.-S.; visualization, N.D.-C., J.M.P.-S., E.G.-D. and J.B.-C.; supervision, N.D.-C., J.M.P.-S., E.G.-D. and J.B.-C.; funding acquisition, E.G.-D. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank the Ministerio de Economía, Industria y Competitividad, Spain (project partially funded by grant ECO2017-85577-P), for the partial support of this work for Emilio Gómez-Déniz.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors agree and appreciate all the comments made by the reviewers and the academic Editor, which have improved the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

Aguiló, Eugeni, Jaume Rosselló, and Mar Vila. 2017. Length of stay and daily tourist expenditure: A joint analysis. *Tourism Management Perspectives* 21: 10–17. [CrossRef]

Albert, Jim, and Siddhartha Chib. 1995. Bayesian residual analysis for binary response regression models. *Biometrika* 82: 747–69. [CrossRef]

Alegre, Joaquín, and Llorenç Pou. 2006. The length of stay in the demand for tourism. *Tourism Management* 27: 1343–55. [CrossRef]

Alkhalaf, Arwa, and Bruno D. Zumbo. 2017. The impact of predictor variable(s) with skewed cell probabilities on Wald tests in binary logistic regression. *Journal of Modern Applied Statistical Methods* 16: 40–80. [CrossRef]

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–23. [CrossRef]

Bazán, Jorge L., Márcia D. Branco, and Heleno Bolfarine. 2006. A Skew Item Response Model. *Bayesian Analysis* 4: 861–92. [CrossRef]

Bermúdez, Lluis, José María Pérez-Sánchez, Mercedes Ayuso, Emilio Gómez-Déniz, and Francisco José Vázquez-Polo. 2008. A bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance, Mathematics and Economics* 42: 779–86. [CrossRef]

Carlin, Bradley, and Nicholas Polson. 1992. Monte Carlo Bayesian methods for discrete regression models and categorical time series. *Bayesian Statistics* 13: 577–86.

Caron, Renault, Debajyoti Sinha, Dipak K. Dey, and Adriano Polpo. 2018. Categorical data analysis using a skewed Weibull regression model. *Entropy* 20: 176. [CrossRef]

Chadee, Doren, and Jan Mattson. 1996. An empirical assessment of customer satisfaction in tourism. *The Service Industries Journal* 16: 305–20. [CrossRef]

Chen, Ming-Hui, Dipak K. Dey, and Qi-Man Shao. 1999. A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association* 94: 1172–86. [CrossRef]

Currier, Christine, and Peter Falconer. 2014. Maintaining sustainable island destinations in Scotland: The role of the transport-tourism relationship. *Journal of Destination Marketing & Management* 3: 162–72.

Dimatulac, Terence, Hanna Maoh, Shakil Khan, and Mark Ferguson. 2018. Modeling the demand for electric mobility in the Canadian rental vehicle market. *Transportation Research Part D* 65: 138–50. [CrossRef]

Ekiz, Erdogan, Pradeep Kumar Nair, and Kashif Hussain. 2010. Measuring The Service Quality in Car Rental Services: Purifying RENTQUAL Instrument with Asian Tourists. Paper presented at the 4th Tourism Outlook & 3rd ITSA Conference, Kuala Lumpur, Malaysia, 30 November–3 December. Edited by L. L. Kian and M. H. M. Hanafiah.

Ferrer-Rossell, Berta, and Germa Coenders. 2017. Airline type and tourist expenditure: Are full service and low cost carriers converging or diverging? *Journal of Air Transport Management* 63: 119–25. [CrossRef]

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1995. Introducing Markov Chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*. Edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter. London: Chapman and Hall.

Gomes de Menezes, Antonio, and Ainura Uzagalieva. 2012. The Demand of Car Rentals: A Microeconometric Approach with Count Models and Survey Data. *Centro de Estudos de Economia Aplicada do Atlántico, CEEApla Working Paper* 12: 1–24.

Gómez-Déniz, Emilio, and Jorge Vicente Pérez-Rodríguez. 2019. Modelling distribution of aggregate expenditure on tourism. *Economic Modelling* 78: 293–308. [CrossRef]

Gómez-Déniz, Emilio, José Boza-Chirino, and Nancy Dávila-Cárdenes. 2020. Tourist tax to promote rentals of low-emission vehicles. *Tourism Economics* 6: 1354816620946508. [CrossRef]

Koop, Gary. 2003. *Bayesian Econometrics*. Hoboken: John Wiley & Sons, Inc.

Lemonte, Artur J., and Jorge L. Bazán. 2018. New links for binary regression: An application to coca cultivation in Peru. *Test* 27: 597–617. [CrossRef]

Lohmann, Guy, and David Duval. 2011. *Critical Aspects of the Tourism-Transport Relationship. Contemporary Tourism Reviews*. Edited by C. Cooper. Oxford: Goodfellow Publishers.

Lunn, David J., Andrew Thomas, Nicky Best, and David Spiegelhalter. 2000. Winbugs: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 10: 325–37. [CrossRef]

Marrocu, Emanuela, Raffaele Paci, and Andrea Zara. 2015. Micro-economic determinants of tourist expenditure: A quantile regression approach. *Tourism Management* 50: 13–30. [CrossRef]

Martín, José M., José A. Rodríguez, Karla A. Zermeño, and José A. Salinas. 2018. Effects of Vacation Rental Websites on the Concentration of Tourists—Potential Environmental Impacts. An Application to the Balearic Islands in Spain. *International Journal of Environmental Research and Public Health* 15: 347. [CrossRef] [PubMed]

Martín, José M., José M. Guaita, Valentín Molina, and Antonio Sartal. 2019. An Analysis of the Tourist Mobility in the Island of Lanzarote: Car Rental Versus More Sustainable Transportation Alternatives. *Sustainability* 11: 739. [CrossRef]

Masiero, Lorenzo, and Judit Zoltan. 2013. Tourists intra-destination visits and transport mode: A bivariate probit model. *Annals of Tourism Research* 43: 529–46. [CrossRef]

Mwenda, Ngugi, Ruth N. Nduati, Mathew Kosgei, and Gregory Kerich. 2021. Skewed logit model for analyzing correlated infant morbidity. *PloS ONE* 16: e0246269. [CrossRef] [PubMed]

Narsaria, Isha, Meghna Verma, and Ashish Verma. 2020. Measuring Satisfaction of Rental Car Services in India for Policy Lessons *Case Studies on Transport Policy* 8: 832–838. [CrossRef]

Palmer-Tous, Teresa, Antoni Riera-Font, and Jaume Rosselló-Nadal. 2007. Taxing tourism: The case of rental cars in Mallorca. *Tourism Management* 28: 271–79. [CrossRef]

Patel, G., A. Koli, R. Kadam, R. Bhat, and P. Kshirsagar. 2018. On Hire: Car Rental System. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)* 5: 214–15.

Pérez-Sánchez, José María, Miguel Ángel Negrín-Hernández, Cataline García-García, and Emilio Gómez-Déniz. 2014. Bayesian asymmetric logit model for detecting risk factors in motor ratemaking. *Astin Bulletin* 44: 445–57. [CrossRef]

Prentice, Ross L. 1976. A generalization of the probit and logit methods for dose-response curves. *Biometrika* 32: 761–68. [CrossRef]

Sáez-Catillo, Antonio J., María J. Olmo-Jiménez, Joé M. Pérez, Miguel A. Negrín, Ángel Arcos-Navarro, and Juan Díaz-Oller. 2010. Bayesian analysis of nosocomial infection risk and length of stay in a department of general and digestive surgery. *Value in Health* 13: 431–39. [CrossRef]

Spiegelhalter, David, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B* 64: 583–639. [CrossRef]

Thrane, Christer, and Eivind Farstad. 2011. Domestic tourism expenditures: The nonlinear effects of length of stay and travel party size. *Tourism Management* 32: 46–52. [CrossRef]

Weisberg, Sandford. 2005. *Applied Linear Regression*. Hoboken: John Wiley & Sons, Inc.

Zellner, Arnold. 1996. *An Introduction to Bayesian Inference in Econometrics*. Hoboken: John Wiley & Sons, Inc. First published 1971.