Modelling Heterogeneous Preferences for Nature-based Recreational Trips

David Boto-García* | Antonio Alvarez | José Baños Pino

botodavid@uniovi.es | alvarez@uniovi.es | jbanos@uniovi.es

Department of Economics

University of Oviedo

*Corresponding author.
Address: Avenida del Cristo s/n, Oviedo (Asturias). Faculty of Business and Economics, University of Oviedo.

Abstract:

This paper studies individual preferences for place-based attributes in the context of nature-based domestic tourism trips. We examine the regional characteristics that explain tourist destination choice focusing on taste heterogeneity for distance and temperature. We estimate a Random Parameter Multinomial Logit with Error Components that controls for: i) unobserved preference heterogeneity for the regional attributes, and ii) correlation in unobserved destination features. We examine the influence of a set of mean shifters in the marginal utilities for regional amenities, and how individuals are willing to trade distance in exchange for warmer (cooler) climates (i.e., marginal rates of substitution). Using a rich dataset of trips within Spain for nature-based purposes, we find large heterogeneity in preferences for temperature differentials and distance, with trip purposes acting as moderators.

Keywords: *nature-based recreational trips; Random Parameter Logit; temperature; distance; marginal rate of substitution*

JEL codes: C35, Q26, R21, Z30

Acknowledgements:

This paper benefited from suggestions by Juan Luís Eugenio, Jaume Rosselló-Nadal, Juan Prieto and Carlos Arias. We also want to thank seminar participants at the Universities of Oviedo and Las Palmas de Gran Canaria and two anonymous referees for their valuable comments. The first author acknowledges financial support from the Spanish Ministry of Education, Culture and Sport (FPU 16/00031).

1. INTRODUCTION

Tourism can be understood as a trade in services that involves the temporal displacement of consumers across regions, whose comparative advantage is determined by natural endowments. It is nowadays a fast-growing industry and a driver of economic growth, both in developed and developing countries (Paci and Marrocu, 2014; Faber and Gaubert, 2019). Tourism inflows have been shown to also produce important increases in inter-industry employment at local economies (Kadiyali and Kosová, 2013), especially for regions that are rich in natural amenities (Naranpanawa et al., 2019). The tourism sector is particularly relevant in Spain, a country that annually receives around 80 million international arrivals. Specifically, the tourism industry accounts for 11.7% of GDP and constitutes 12.8% of total employment (INE, 2019).

Despite the relevance of international tourism, the Spanish domestic travel market has increased its importance in the last decade. In the current pandemic context, domestic trips are expected to increase their contribution to regional GDP because under health risks people tend to travel within their country (Jeon and Yang, 2021). Among the different trip purposes, nature-based tourism is gaining increasing popularity and attention in the literature (Gosens and Rouwendal, 2018; Naranpanawa et al., 2019). It is estimated to represent 15% of total tourism in the world (UNWTO, 2018). In Spain, total spending from nature-based tourists is estimated to be about 9,000 million euros (SGAPC, 2017). Indeed, about 81% of nature-based travellers are residents. However, the economic modelling of domestic trips has been overlooked.

This paper examines how regional attributes affect individual destination choices for nature-based tourism within Spain. We devote special attention to the role of temperature relative to the place of residence and distance. Prior studies have shown that it is not only climate conditions at the destination that matter but climate differences between the origin and potential destinations (e.g. Rosselló-Nadal et al., 2011). Intuitively, individuals who are used to colder (warmer) climates may be looking to travel to warmer (colder) areas. Although there is a large body of literature on this for coastal destinations (Bujosa and Rosselló, 2013; Priego et al., 2015), the preferences for warmer or cooler destinations have been less studied in the context of nature-based trips, being Chan and Wichman (2020) and Dundas and von Haefen (2020) some exceptions but focused on daily recreational demand. We consider the ratio between temperature at each destination relative to that at the origin to assess how individual preferences for warmer (cooler) locations depend on climate conditions at the origin.

Similarly, there is inconclusive evidence on the effect of distance on tourist destination choice, since individuals exhibit heterogeneous preferences regarding travelling to nearby or distant destinations (Nicolau, 2008; 2010). Due to both travelling costs and the opportunity cost of time, *ceteris paribus*, distant locations are less preferred. However, in some settings, recreational travelling time might have a commodity value (e.g., Chavas et al., 1989). We explore whether the type of activities to engage in at the destination moderate or intensify the distance decay effect. This relates to previous evidence that links leisure trips to the satisfaction of needs (e.g., Dekker et al., 2014).

Unlike other studies that analyse aggregate flows between regions (de-la-Mata and Llano-Verduras, 2012; Massidda and Etzo, 2012; Cafiso et al., 2016; Pompili et al., 2019; Alvarez-Diaz et al., 2020; Panzera et al., 2021), we adopt a microeconomic perspective rooted in the product characteristics approach. We use microdata of monthly domestic trips by Spanish residents between February 2015 and September 2017. We focus on leisure trips for nature-based purposes to the 17 Spanish regions. We combine this dataset with: i) monthly regional data on tourism prices, temperature, rainfall and ski track kilometres available for practising winter sports; ii) weighted bilateral distances between the origin and potential destinations that take into account the probability of residence location within the region; and iii) time-invariant regional-specific characteristics such as tourism spots, natural parks, the size of protected natural areas and the presence of coast.

Our Random Parameter Multinomial Logit with Error Components model (Greene and Hensher, 2007) allows the random parameters to be a function of a mean parameter and several individual-specific characteristics such as age, income, party size and trip purpose, among others, plus a stochastic term varying across individuals. In this way, we explore the factors that shift the marginal utilities and explain the sources of preference heterogeneity. We allow the random parameters to be correlated to account for potential scale heterogeneity. Additionally, by including a set of error components, our model controls for similarity in preferences for destinations that share common unobserved features.

A particular feature of our analysis is that we do not only examine the effect of distance and relative temperatures on choice probabilities independently, but we also address the relationship between them. We derive the conditional means of the individual-specific marginal utilities for these attributes and compute the marginal rates of substitution at the individual level. In doing so, we follow Hess and Hensher (2010). Accordingly, we assess how much distance individuals are willing to cover in exchange for warmer (cooler) temperatures, *ceteris paribus*. Next, based on the model estimates, we derive own and cross regional elasticities for prices and relative temperature, showing how probabilistic demand changes under marginal variations in these two dimensions.

We contribute to the literature by examining how individual characteristics and travel motivations are related to marginal utilities for regional attributes. In particular, together with sociodemographics and time effects, we assess how trip purposes like mountaineering, practising winter or aquatic sports, or visiting natural areas moderate or intensify the disutility of distance and the preference for warmer destinations. Therefore, the paper extends the literature on regional tourism pull factors by adopting a microeconomic viewpoint to understand the determinants of domestic tourism consumption.

The paper is structured as follows. Section 2 presents the theoretical framework. Section 3 describes the database and the variables employed. Section 4 outlines the econometric modelling and the empirical strategy. The estimation results are presented and discussed in Section 5. Finally, Section 6 concludes with the main remarks.

2. THEORETICAL FRAMEWORK

Our analysis is based on the *Lancasterian* product characteristics approach (Lancaster, 1966), which assumes it is the characteristics of the goods from which utility is derived. Consider *J* alternative destinations characterized by *K* observable characteristics (X_{kj}). Each destination *j* can be regarded as a source of systematic utility (V_{ij}) in the form:

$$V_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} \quad (1)$$

where β_{ik} is the marginal utility of attribute *k* for individual *i*. Marginal utilities are allowed to vary in the population because consumers have different preferences over destination hedonic attributes. These preferences "are functions of their experience and personal characteristics, including both observed and unobserved components" (McFadden, 2001).

In their most general form, the marginal utilities are composed of two parts: i) a structural component that is the same for the entire population (b_k); and ii) an individual-specific component that adds stochastic variation to the marginal utilities. This latter component can be further partitioned into two elements: i) a vector of observable individual characteristics Z_i, and ii) a composite factor v_{ik} for random preference heterogeneity. Individual-specific characteristics (Z_i) might include sociodemographic features, the available time for travelling, disposable income and situational factors, such as trip purpose or party size composition. Therefore, the marginal utilities can be expressed as:

$$\beta_{ik} = g(b_k, Z_i, v_{ik}) \quad (2)$$

Since the characteristics of the goods are objectively given, consumers make choices among bundles of characteristics. Under this framework, preference rankings over goods can be derived conditional on the individual-specific marginal utilities. Because individuals are subject to a budget constraint, optimal choice implies choosing the alternative that maximizes utility while minimizing costs. The systematic utility can be expanded to include the associated price for each destination *j* (P_j) and the corresponding marginal (dis)utility γ:

$$V_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} - \gamma P_j \quad (3)$$

Conditional on the marginal utilities, Lancaster's approach is deterministic because all the sources of utility are known by the individual. At the empirical level, however, not all the relevant attributes of the destinations are observed. Since the model is sensitive to measurement error, an additive source of residual utility in the form of an independent and identically distributed disturbance term (ε_{ij}) is included. Therefore, the utility of destination *j* for individual *i* can be expressed as:

$$U_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} - \gamma P_j + \varepsilon_{ij} = \sum_{k=1}^{K} g(b_k, Z_i, v_{ik}) X_{kj} - \gamma P_j + \varepsilon_{ij} \quad (4)$$

The utility in (4) can be further expanded with another additive term ξ_j for unobserved (from the econometrician perspective) destination-specific characteristics (Berry and Pakes, 2007; Murdock, 2006). As a result, the utility is given by:

$$U_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} - \gamma P_j + \xi_j + \varepsilon_{ij} = \sum_{k=1}^{K} g(b_k, Z_i, v_{ik}) X_{kj} - \gamma P_j + \xi_j + \varepsilon_{ij} \quad (5)$$

The term ξ_j is generically defined to be of dimension *J*, but it can also be defined for a lower dimension. Destinations that have similar observed features might also share the same unobserved characteristics.

The *characteristics-based* demand model provides several advantages over the *taste for products* model, such as not imposing limits on the substitution patterns between alternative destinations or the setting of a utility bound as the number of alternatives increase (Berry and Pakes, 2007). It is important to highlight that the choice of destination is conditional on having decided to travel. We need to further impose that preferences for leisure activities are weakly separable (Deaton and Muellbauer, 1980) so that tourism demand is expressed independently of non-tour prices. As such, our model picks up at the second stage of budget decomposition, after income and time have been allocated for travelling. Furthermore, our analysis sets quantities (number of trips) to the unity. Therefore, we model probabilistic demand.

3. DATA

3.1. Database

Our database is drawn from the Spanish Domestic Tourist Survey conducted on a monthly basis by the Spanish National Statistics Institute to a representative sample of the Spanish population. The sample is obtained by multistage sampling, stratified by conglomerations with proportional section of primary (cities) and secondary units (census sections). This survey gathers information about all kind of trips conducted by Spanish residents, such as main destination, party size, length of stay, expenditure and sociodemographic characteristics, among others. Participants are interviewed at their homes by telephone about trips that took place two months before.

Our study covers the period February 2015-September 2017 (i.e. 32 months). The database does not have a panel structure but is a pool of monthly cross-sectional units. We only consider domestic trips whose main purpose is leisure, holidays or entertainment. International trips are not included. The sample is also restricted to those who declare nature and/or sport as their main travel purpose. After excluding some observations with missing values, we have valid data for 6,661 tourists that take a nature-based domestic trip to any of the 17 Spanish regions (NUTS 2). Trips to Ceuta and Melilla are excluded.

Summary statistics of a selection of sample characteristics are provided in Table 1. Our sample of domestic nature-based tourists comprises slightly less females (46 percent) than males, with a mean age of 43 years old. Respondents are relatively highly educated (60 percent) and mostly participating in the labour market (74 percent). About half of the sample is married, and the average number of travel companions is 2.2. Most respondents have middle incomes (56 percent), and the mean length of stay at the destination is 3.3 nights. Half of the sample travels during weekends (51 percent), being in the third quarter of the year when most trips take place (37 percent). Regarding trip activities, mountaineering and trekking is the most declared trip activity (72 percent), followed closely by visiting natural areas (56 percent). Interestingly, a non-negligible share of respondents travels to perform adventure and risky sports (22 percent), whereas 16 percent opts to visit rural areas and villages. Finally, only 7 percent of the sample practises winter sports.

Variable	**Description**	**Mean**	**SD**	**Min**	**Max**
age	Age (in years)	43.23	12.45	15	85
female	Female	0.46	0.49	0	1
primary	Level of education: primary studies	0.02	0.14	0	1
secondary	Level of education: secondary studies	0.36	0.48	0	1
tertiary	Level of education: university studies	0.60	0.48	0	1
employed	Labor status: employed	0.74	0.43	0	1
unemployed	Labor status: unemployed	0.08	0.28	0	1
retired	Labor status: retired	0.07	0.25	0	1
inactive	Labor status: inactive	0.09	0.28	0	1
married	Respondent is married	0.51	0.49	0	1
parsize	Travel party size (number)	2.21	1.22	1	7
inc1	Monthly income: < 1,500 euros	0.26	0.44	0	1
inc2	Monthly income: 1,500-3,500 euros	0.57	0.49	0	1
inc3	Monthly income: >3,500 euros	0.15	0.36	0	1
LOS	Length of the stay (days)	3.36	3.43	1	30
weekend	Travels during a weekend	0.51	0.49	0	1
q1	First quarter	0.20	0.40	0	1
q2	Second quarter	0.25	0.43	0	1
q3	Third quarter	0.37	0.48	0	1
q4	Fourth quarter	0.16	0.36	0	1
*winter_sports**	Trip activity: winter sports practice (i.e. skiing, snowboarding)	0.07	0.25	0	1
*mou_trek**	Trip activity: trekking and mountaineering	0.72	0.44	0	1
*rural**	Trip activity: visit to rural areas/villages	0.29	0.45	0	1
*nat_areas**	Trip activity: visit to natural areas (i.e. mountains, parks, forests)	0.56	0.49	0	1
*aquatic**	Trip activity: aquatic sports practice (i.e. surf, diving, sailing, windsurf, fishing)	0.16	0.36	0	1
*advent**	Trip activity: adventure/risky sports practice (climbing, canyoning, canoeing, kayaking, rafting, bungee jumping, skydiving, paintball)	0.22	0.41	0	1

Table 1.- Descriptive statistics (N=6,661)

*Note: these activities are not mutually exclusive

Aragon is the region with the largest number of tourists (15.3 percent), followed by Catalonia (14.3 percent). Conversely, Murcia and La Rioja are the regions with the lowest number of visitors (1.1 and 1.5 percent, respectively).

3.2. Destination attributes

Distance

Although formally this is not a destination characteristic, from the viewpoint of a tourist any alternative destination is distant or nearby from her perspective. As such, distance can be considered as a destination feature that varies depending on the origin.

We use the Euclidean distance between the individual's place of residence and each possible destination, which is the most used measure (Marrocu and Paci, 2013; Cafiso et al., 2016; Gosens and Rouwendal, 2018; Pompili et al., 2019; Alvarez-Díaz et al., 2020; Panzera et al., 2021). In the survey, respondents report the regional area where they stay at the NUTS 3 regional disaggregation level (Spanish provinces, equal to 50). However, information on their place of residence is only provided at the NUTS 2 level (autonomous communities, equal to 17). The latter hinders the calculation of distances between origin and destination because they are not defined at the same regional level. If we computed the Euclidean distance in kilometres between NUTS2 regional centroids, that would set to zero the distances for all the trips that take place within autonomous communities. Given the heterogeneity in size between Spanish regions, that would equally assume zero distance both for *true* intra-regional trips and for *apparent* intra-regional trips.[1] In addition, that would reduce the variability in the distance variable.

To alleviate this, for each tourist in the sample we compute a Euclidean weighted measure of distance that considers tourists' place of origin probabilistically. We calculate the distance between the centroids of all Spanish provinces (NUTS 3) and we then compute bilateral distances between each province (p) and each autonomous community (c) in the following way:

$$d_{c,p} = \sum_{p \in c} \frac{pop_p}{pop_c} * d_{p,p'} \quad (6)$$

for $p = 1, \dots, 50$ and $c = 1, \dots, 17$.

where $d_{c,p}$ is the distance between each province destination and each autonomous community, $d_{p,p'}$ is the distance between pairs of provinces and pop_p and pop_c are the population in each province and autonomous community. In this way, distances between the origin and destinations consider the likelihood of the individuals living in each province based on population weights (adjusted biannually).

Finally, since our analysis is performed at the NUTS 2 level, we take the weighted distance mean within autonomous communities so that $d_{c,c'} = \sum_{\forall\, p \in c} \mathrm{d}_{c,p} * \frac{1}{n}$, where n indicates the number of provinces in each autonomous community. The resulting weighted distance ($d_{c,c'}$)

[1] In Spain, the regional NUTS 2 (autonomous communities) and NUTS 3 (provinces) definition is the same for Asturias, Cantabria, Navarre, La Rioja, the Balearic Islands, Murcia and Madrid. However, regions like Andalusia and Castile-and-Leon (NUTS 2) involve 8 and 9 provinces each, respectively.

is labelled *DIST* and is similar to the one implemented in Chandra et al. (2014) for modelling cross-border travelling. In spirit, it also shares some features with the circle-distance proposed by de-la-Mata and Llano-Verduras (2012). Although it cannot be considered the 'actual distance', it properly reflects the average distance connecting the actual spots of origin and destination.

Consistent with the well-documented distance decay effect by which tourism demand decreases as distance increases (de-la-Mata and Llano-Verduras, 2012; Gosens and Rouwendal, 2018; Pompili et al., 2019; Alvarez-Díaz et al., 2020), we expect distance to exert a negative effect on utility. Notwithstanding, our empirical modelling will allow for heterogeneity in this effect.

Climate

Climate is another important attribute for travel purposes that involve outdoor activities. Trips are normally planned in advance so that the actual weather is difficult to forecast. Therefore, we consider the expected (average) temperatures and rainfalls at each region in each month. This means we consider expected climate conditions rather than actual weather. Average temperature and rainfall per month during the 2010-2015 period were obtained from the Spanish National Meteorology Institute. We define two different variables.

First, we construct the variable *r_TEMP* as the ratio between the temperature at each possible destination and the temperature at the tourist's place of residence.

$$r\_TEMP_{imt} = \frac{T_{imt}}{T_{ijt}} \tag{7}$$

where T_{imt} is the temperature for individual *i* at region *m* during period *t* and T_{ijt} is the temperature at the origin.

The higher (lower) the ratio, the warmer (colder) the destination relative to the origin. Apart from being consistent with the empirical literature that shows that the preferences for warmer or colder destinations depend on climate conditions at the origin (e.g. Eugenio-Martín and Campos-Soria, 2010; Rosselló-Nadal et al., 2011), this ratio captures the non-linearity in relative temperature differences[2]. Unlike alternative approaches such as temperature differences (Agiomirgianakis et al., 2017), the ratio captures that an absolute difference in temperature between the origin and the destination does not have the same effect depending on the level. However, if we fix the temperature at the origin, then marginal increases in the ratio correspond with marginal increases in temperature at the alternative destinations[3]. In line with

[2] To understand this, consider the following two situations. A one-degree difference in favour of the destination gives a different value of the ratio depending on whether it is a destination with a temperature of 11 degrees Celsius (ºC) relative to an origin with 10ºC (11/10=1.1) or a destination with 21ºC relative to an origin with 20ºC (1.05).

[3] For example, for 10ºC at the origin, a marginal change of one-degree in destination from 11 ºC to 12 ºC leads to the same marginal change in the ratio (1.2-1.1=0.1) as a one-degree change from 21ºC to 22ºC (2.2-2.1=0.1).

prior findings (e.g. Helbich et al., 2014), we expect this variable to positively influence utilities (i.e. tourists are assumed to prefer warmer destinations).

Second, we define a dummy variable denoted by *RAIN* that takes value one if expected rainfalls at each possible destination at month *t* excess 60 litres per square meter (60 mm). This threshold was chosen for two reasons. First, it is the 75th percentile of the rainfall distribution during the study period, thereby reflecting destinations and periods with rainy climate. Second, Mieczkowski's subindex of 3 (Mieczkowski, 1985) precisely equals less than 60 mm rainfalls per month, and this threshold was also used by Eugenio-Martín and Campos-Soria (2010). We expect this variable to exert a negative effect on utility.

Prices

Prices constitute a third major determinant of tourist destination choice. According to microeconomic theory, prices must exhibit a negative relationship with demand. Therefore, the more expensive a destination is, the lower the probability of being selected, *ceteris paribus*. In line with related applications (Nicolau and Más, 2006; Massidda and Etzo, 2013), we use the regional consumer price indexes as a proxy of prices. Specifically, in line with Faber and Gaubert (2019), we employ the price index for accommodation and tourism-related services at each month for each destination (denoted as *TCPI*).[4] This data is drawn from the Spanish National Institute of Statistics. The year 2011 is the base period.[5]

A large body of research concerned about tourism flows argues that tourists compare prices across possible destination with the prevailing prices at their place of residence (Chandra et al., 2014; Alvarez-Díaz et al., 2020). Therefore, as an alternative measure, we define the ratio of *TCPI* between each possible destination *j* and the one at the place of origin *k* for individual *i* at month *t* so that:

$$r\_TCPI_{ijt} = \frac{TCPI_{ijt}}{TCPI_{ikt}} \tag{8}$$

We expect both *TCPI* and *r_TCPI* to be negatively related with choice probabilities, *ceteris paribus*.

Tourism spots

[4] We acknowledge that the use of price indexes exhibits some shortcomings. One of them is that CPI captures price variations over time relative to the base period, but it is not able to control for differences in price levels across regions (Marrocu and Paci, 2013). Nevertheless, the use of the CPI for tourism-related services is more specific and better captures the relevant prices for tourism than the generic CPI used in previous literature (Chandra et al., 2014; Alvarez-Diaz et al., 2020).

[5] Monthly regional prices cannot capture the within region and within month variability in prices. However, for modelling the choice of region *j* over any other, prospective tourists are assumed to compare mean price level differences across regions.

The use of aggregate zones (NUTS 2) instead of individual attractions requires controlling for the variability in utilities across individual alternatives that compose aggregate alternative *j* (Bekhor and Prashker, 2008). The number of tourism sightseeing spots per region (*TOU_SPOTS*) constitutes a relevant variable to measure the number of municipalities of interest in each region.[6] This variable is similar to the one used by Pompili et al. (2019). We expect it to have a positive effect on the probability of a destination being chosen.

Ski kilometers

We computed the sum of monthly available kilometres for alpine, Nordic and indoor skiing at each region during the ski season (November-April). This variable thus takes value zero for the rest of the year and is denoted by *SKI_KM*. This information has been gathered from the 2015, 2016 and 2017 Annual Reports of the Spanish Tourist Association for Ski and Mountain Resorts. This variable is expected to positively affect choice probabilities, especially for those who practice winter sports (Falk, 2010).

National parks

Because we are interested in the pull factors that attract nature-based tourists to destinations, we expect the number of national and natural parks to be positively valued by nature-based tourists. This information has been retrieved from the European Agency for the Environment. The variable is labelled as *NAT_PARKS*.

Size of protected natural areas

Together with the number of natural and national parks, we also consider the size of protected natural areas (in km^2). This variable is denoted as *SIZE_NAT* and gathers not only the surface of natural and national parks, but also other natural areas where people can recreate without holding such categorization. This data is drawn from the Ministry of Ecological Transition.

Coast

For those who seek to practise aquatic sports, coastal regions seem to be preferred. To control for this, we define a dummy variable denoted by *COAST* that takes value 1 if the region has a coastline.[7] Please note this dummy is complimentary to the use of *r_TEMP* because the temperatures of coastal north and south regions in Spain are quite different (see de-la-Mata and Llanos-Verduras, 2012).

Table 2 presents summary statistics of the attributes. Appendix A reports histograms for the time-varying attributes and kernel plots for the average temperatures per region and quarter.

---

[6] A municipality is considered by the Spanish National Institute of Statistics to be a tourism spot if it specifically concentrates tourism affluence. This definition is prior to our study period, thereby being an exogenous indicator.

[7] We favor the use of a binary indicator over the kilometres of coastline because Spanish regions are very heterogeneous in terms of the quality and recreational areas of their coastlines.

Attribute	**Description**	**Mean**	**Standard deviation**	**Min**	**Max**	**Source**	**Varies over time**
DIST	Euclidean distance between tourist's place of residence and each possible destination (in km)	532.4	458.8	0	2,182.8	Calculated using Google Maps	NO*
r_TEMP	Ratio of monthly average temperatures between each possible destination and the place of origin	1.02	0.26	0.28	3.46	Spanish National Statistics Institute (INE). Average values per month during the period 2010-2015.	YES
RAIN	Dummy variable that takes value one if monthly average rainfalls are higher than 60 liters per square meter	0.21	0.41	0	1	Spanish National Statistics Institute (INE). Average values per month during the period 2010-2015.	YES
TCPI	Monthly Tourism Consumer Price Index (accommodation and restaurant services, base 2010)	103.3	6.8	86.7	128.9	Spanish National Statistics Institute (INE)	YES
r_TCPI	Ratio of monthly TCPI at each possible destination relative to the corresponding one at origin	1.00	0.04	0.77	1.28		YES
TOU_SPOTS	Total number of tourism spots	6.1	5.6	1	24	Spanish National Statistics Institute (INE)	NO
NAT_PARKS	Number of natural and national parks	6.05	5.06	0	15	European Agency for the Environment	NO
SIZE_NAT	Size of protected natural areas (in km^2)	4,332	6,009.5	621	26,083	Ministry for Ecological Transition	NO
SKI_KM	Available kilometers for alpine ski, Nordic ski and indoor ski (in km)	25.12	88.42	0	484	Annual Reports from the Tourist Association for Ski and Mountain Resorts	YES
COAST	Dummy variable for whether the region has coast	0.5	0.44	0	1	Google Maps	NO

Table 2.- Summary statistics, notation, description and source for destination attributes

*Note: this variable varies slighting over time due to the biannual change in the population weights used in the calculation (see eq. 6)

4. EMPIRICAL MODEL

4.1. Econometric Modelling

Our destination choice model is based on the Random Utility Maximization Theory (RUM) developed by McFadden (1974). Under this framework, the latent utility of individual *i* for choosing alternative *j* (U_{ij}^*) is the sum of a systematic and a random component as follows:

$$U_{ij}^* = V_{ij} + \varepsilon_{ij} \tag{9}$$

where *i* indexes individuals, *j* indexes destinations, V_{ij} is a deterministic function of observable characteristics and ε_{ij} is a random error term which reflects unobserved factors. The systematic part of the utility function is an additively separable linear-in-parameters function of the *K* attributes of each alternative *j* (X_{kj}) so that:

$$V_{ij} = X_{kj}'\beta_k \tag{10}$$

If the random terms are IID type I extreme value (Gumbel) distributed, we obtain the standard Multinomial Logit Model (henceforth MNL):

$$P_{ij} = \frac{\exp(\lambda V_{ij})}{\sum_{j=1}^{J} \exp(\lambda V_{ij})} \tag{11}$$

where λ is a positive scale parameter that is inversely proportional to the standard deviation of the Gumbel error terms. Since λ and β are not separably identified, λ is normalized to 1 for identification.

The MNL presents certain shortcomings. First, it exhibits the well-known Independence of Irrelevant Alternatives (IIA) property (Debreu, 1960), by which the ratio of probabilities between two alternatives does not change if a third (irrelevant) one is included in the choice set. Second, it assumes taste homogeneity in respondents' preferences. These limitations have motivated researchers to develop alternative models. Among them, the Random Parameter Logit (hereafter RPL) has become the most accepted one. The RPL extends the MNL by allowing the parameters to vary randomly in the population according to a certain distribution (Revelt and Train, 1998; Train, 1998). Furthermore, it also allows the means of the parameter distributions to be heterogeneous. Therefore, the parameters can be expressed as follows:

$$\beta_{ki} = b_k + \delta' Z_i + \sigma_k \nu_{ik} \tag{12}$$

where b_k is the population mean, Z_i is a vector of choice invariant individual characteristics that shift the population mean parameter, δ is the associated vector of parameters to be estimated, ν_{ik} is the individual-specific heterogeneity that follows a probability distribution independent from ε_{ij}, and σ_k is the standard deviation of the distribution of β_{ki} around b_k.

The probability that individual *i* chooses destination j (P_{ij}) takes the form of a multidimensional integral over all possible values of β_i of the logit formula weighted by the density of β_i:

$$P_{ij} = \int P_{ij}|\beta_i \ f(\beta_i|\Omega)\, d\beta_i = \int \frac{\exp(V_{ij})}{\sum_{j=1}^{J}\exp(V_{ij})}\ f(\beta_i|\Omega)\, d\beta_i \tag{13}$$

where $P_{ij}|\beta_i = \frac{\exp(V_{ij})}{\sum_{j=1}^{J}\exp(V_{ij})}$ are the choice probabilities conditional on the vector of taste coefficients β_i, $f(.)$ is the density function of β_i, and Ω denotes the hyper-parameters of this distribution in the population (mean b and covariance matrix W) so that $\Omega = (b, W)$. The above integral does not have a closed solution, so choice probabilities are estimated by simulation techniques, taking random draws from the underlying distribution assumed for β_i.

RPL with correlated parameters

Most empirical applications that estimate a RPL model impose the random parameters to be uncorrelated (i.e. the random coefficients to have a diagonal covariance matrix). However, this has some limitations. First, this imposes constraints on the model estimation. Second, the uncorrelated RPL implies that the scale is constant. However, it is highly likely that the weight of the random component differs in the population (scale heterogeneity). This issue has received growing attention in recent years (e.g. Greene and Hensher, 2010; Keane and Wasi, 2013). Hess and Train (2017) argue that the best way to control for scale heterogeneity is to allow for correlation between the random parameters. The estimated correlation will capture common features in the magnitude of coefficient estimates across individuals.

Due to these reasons, we estimate a RPL with correlated parameters. The covariance matrix of the random coefficients is a lower triangular matrix with nonzero off diagonal elements (denoted by Γ) to be estimated:

$$\beta_i = b + \Delta\, \mathrm{Z_i} + \Sigma\, \nu_i \tag{14}$$
$$\text{where } Var\,(\beta_i |\, X_j, Z_i) = \Sigma = \Gamma\Gamma^t$$

RPL with error components (RPL-ECM)

To control for residual utility, empirical modelling usually includes a full set of Alternative-Specific Constants (hereafter ASCs). They capture the mean of the error term in the utility of each alternative and have been shown to improve overall fit (Klaiber and Von Haefen, 2019). The limitation of the ASCs is that they capture residual utility that is common to the entire sample. To relax this, one might consider allowing them to be randomly distributed in the same fashion as the attributes. However, in cross sectional data, specifying the ASCs to be random is not advisable (Greene, 2012). Alternatively, we could extend the RPL with a set of Error Components (Greene and Hensher, 2007) so that the utility function is given by:

$$U_{ij} = ASC_j + X_{kj}'\beta_k + \vartheta_n E_n + \varepsilon_{ij} \tag{15}$$

where ASC_j are a set of alternative-specific constants, E_n are random Error Components (henceforth ECs) that account for shared time-invariant correlation between choice alternatives not captured in the attributes contained in X_{kj}, and ε_{ij} are Type I Extreme Value distributed error terms. The ECs are standard normally distributed so that $E_n \sim N(0,1)$, with ϑ_n being the associated vector of parameters to be estimated (scale factors). Therefore, the random component of the utility is $\epsilon_{ij} = \vartheta_n E_n + \varepsilon_{ij}$.

The set of ASCs and the ECs are not separately identified when specified at the j-level. An interesting feature of the inclusion of ECs in the specification is the possibility of defining them at an upper level so that, conditional on the ASCs, they capture common unobserved heterogeneity to several destinations. Accordingly, the introduction of ECs causes utility to be correlated over alternatives because $Cov(E_n, E_l) = E[(\vartheta_n E_n + \varepsilon_{ij})(\vartheta_l E_l + \varepsilon_{ij})]) = \vartheta_n{}'W\,\vartheta_l$, where W is the covariance matrix of the ECs which is assumed to be diagonal.

4.2. Model specification

Our empirical model has the following form:

$$\begin{aligned} U_{ij} = {} & \alpha_1\, REG1_j + \alpha_2\, REG2_j + \alpha_3\, REG3_j + \alpha_4\, REG4_j + \alpha_5\, REG5_j + \alpha_6\, REG6_j \\ & + \beta_{1i}\, DIST_{ij} + \beta_{2i}\, r\_TEMP_{ij} + \beta_3\, RAIN_j + \beta_4\, TCPI_j \\ & + \beta_5\, TOU\_SPOTS_j + \beta_6\, NAT\_PARKS_j + \beta_7\, SIZE\_NAT_j + \beta_8\, SKI\_KM_j \\ & + \beta_9\, SKI\_KM_j * winter\_sports_i + \beta_{10}\, COAST_j + \beta_{11}\, COAST_j * aquatic_i \\ & + \vartheta_1 E1_{ij} + \vartheta_2\, E2_{ij} + \vartheta_3\, E3_{ij} + \vartheta_4\, E4_{ij} + \varepsilon_{ij} \end{aligned} \tag{16}$$

We specify the parameters for distance (β_{1i}) and the ratio of temperatures (β_{2i}) to be randomly distributed. We consider them to follow a normal distribution, which is the most used specification. The rest of the parameters (including *TCPI*) are treated as fixed. In this regard, one might wonder whether only considering two attributes as randomly distributed is contradictory with the issues raised about the necessity of allowing for free correlation between the attributes. We only allow *DIST* and *r_TEMP* to be random because the RPL model tends to be unstable when many coefficients are allowed to vary, especially when working with cross-sectional data (Revelt and Train, 1998)[8].

[8] Furthermore, Keane and Wasi (2013) tested several RPL specifications on ten datasets and indicate that the full covariance matrix is not needed to be estimated in all situations. Scale heterogeneity is addressed even when only a subset of the vector of parameters are allowed to be correlated.

We expect the number of kilometres for skiing (*SKI_KM*) and the presence of coast (*COAST*) to be more valued by those individuals whose main motivations is practising winter and aquatic sports, respectively. Therefore, the model includes two interaction terms between these two attributes and these two motivations (*SKI_KM*winter_sports* and *COAST*aquatic*). In the main analysis, we opt for using *TCPI* at each alternative destination *j* as our proxy for prices. Nonetheless, we repeat the baseline analysis replacing *TCPI* by *r_TCPI* (see subsection 5.1).

For parsimony, the ASCs are defined at the NUTS1 level (7 regions) because the inclusion of a full set of ASCs along with time-invariant destination attributes (in our case *TOU_SPOTS*, *NAT_PARKS*, *SIZE_NAT* and *COAST*) produces an identification problem. Note that size effects are captured in the ASCs. This practise of defining group-specific constants is common in the recreational demand literature (e.g. Parsons and Hauber, 1998). Concerning the ECs, we define four random components based on geographic location. Specifically, the first EC relates to regions located in the North of Spain. The second EC refers to regions in the centre without coastline. The third EC gathers preference for regions in the South-East part of the country (Mediterranean regions). The fourth EC is defined for the Canary Islands. Table 3 illustrates the composition of the ASCs and the ECs.

Regions	ASC	EC
Galicia, Asturias and Cantabria	*REG1*	*E1* (North)
Basque Country, Navarre, La Rioja and Aragon	*REG2*	
Community of Madrid	*REG3*	*E2* (Centre)
Castile and Leon, Castilla-La Mancha and Extremadura	*REG4*	
Catalonia, Valencian Community and the Balearic Islands	*REG5*	*E3* (South-East)
Andalusia and Region of Murcia	*REG6*	
The Canary Islands	Reference category	*E4* (Canary Islands)

Table 3.- ASC and error components (EC)

As discussed before, we allow preferences for distance and temperatures to depend on a set of individual characteristics. These mean shifters are grouped into six blocks: i) age, ii) income, iii) party size, iv) time effects, v) origin climate conditions and vi) goals (trip purposes).

- Age: the marginal utility for temperature and distance might vary according to the tourist's life stage, since it has been shown that destination choice preferences vary with age (e.g. Bernini et al., 2017). Accordingly, we include tourist's age in years (*age*).
- Income: assuming that tourism is a normal good, the higher the level of income the lower the dissuasive effect of distance. Hence, high-income tourists are expected to be willing to travel further away (Nicolau, 2010). We include two dummy variables for medium and high household income (denoted by *inc2* and *inc3*), leaving low income (*inc1*) as the reference category.

- Party size: trip party size is expected to exert a significant effect on the disutility of distance. Previous studies have shown that in deciding where to travel the composition of the travel group plays a relevant role (e.g. Morey and Krizberg, 2012). The variable *parsize* is defined as the number of individuals who participate in the trip.
- Time effects: preferences for climate and distance might depend on time constraints. This relates to McConnell's time allocation model for leisure demand (McConnell, 1999). First, we consider a dummy variable for weekend trips (Saturday and Sunday) denoted by *weekend*. Second, to capture seasonal effects, we include three dummy variables for whether the trip takes place in the first, second or fourth quarter (*q1*, *q2*, and *q4,* respectively). The third quarter acts as the reference category.
- Origin climate conditions: although in the utility function we consider the temperature at each possible destination relative to the origin, the effect of *level* conditions at the origin on the marginal utility for a gain in temperature may change depending on the season (quarter). To explore this, we define four dummy variables for above-mean temperatures per quarter as follows:

$$d\_warmorigin\_q_{jt} = \begin{cases} 1 & \text{if } \mathrm{T}_{jt} > \sum_{j=1}^{J}\sum_{t=1}^{T}\frac{T_{jt}}{J*T}\ \forall\ t\ \in q, \qquad for\ q = 1,2,3,4 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

  These four dummy variables capture the effect of origins that are warmer than the average per quarter.[9]
- Trip purposes: recent evidence has shown that goal pursuit determines site choice probabilities (Swait et al., 2020). To explore this, we consider the following trip purposes as taste shifters: the practice of winter sports (*winter_sports*), mountaineering, trekking or visiting natural areas (*mou_trek_nat*), visiting rural areas or villages (*rural*), the practice of aquatic sports (*aquatic*) and the practice of adventure/risk activities (*advent*).

Therefore, the random parameters for *DIST* (β_{1i}) and *r_TEMP* (β_{2i}) are specified as:

$$\beta_{1i} = b_1 + \delta_1\, Age_i + \delta_2\, Inc2_i + \delta_3\, Inc3_i + \delta_4\, Parsize_i + \delta_5\, weekend_i + \delta_6\, Q1_i + \delta_7\, Q2_i + \delta_8\, Q4_i + \delta_9\, d\_warmorigin\_1_i + \delta_{10}\, d\_warmorigin\_2_i + \delta_{11}\, d\_warmorigin\_3_i + \delta_{12}\, d\_warmorigin\_4_i + \delta_{13}\, winter\_sports_i + \delta_{14}\, mou\_trek\_nat_i + \delta_{15}\, rural_i + \delta_{16}\, aquatic_i + \delta_{17}\, advent_i + \sigma_1 v_{i1} \tag{18}$$

$$\beta_{2i} = b_2 + \theta_1\, Age_i + \theta_2\, Inc2_i + \theta_3\, Inc3_i + \theta_4\, Parsize_i + \theta_5\, weekend_i + \theta_6\, Q1_i + \theta_7\, Q2_i + \theta_8\, Q4_i + \theta_9\, d\_warmorigin\_1_i + \theta_{10}\, d\_warmorigin\_2_i +$$

[9] The cut-off points (mean temperatures per quarter) are 9.5ºC for the first quarter, 17.4ºC for the second, 23.7ºC for the third and 13.4ºC for the fourth.

$$\theta_{11}\, d\_warmorigin\_3_i + \theta_{12}\, d\_warmorigin\_4_i + \theta_{13}\, winter\_sports_i + \theta_{14}\, mou\_trek\_nat_i + \theta_{15}\, rural_i + \theta_{16}\, aquatic_i + \theta_{17}\, advent_i + \sigma_2 v_{i2} \tag{19}$$

One might consider some variables like income or party size to only shift the marginal utility for *DIST*, and others like $d\_warmorigin\_q$ to only affect the marginal utility for *r_TEMP*. However, to avoid imposing such restrictions, in our baseline analysis we use the same vector of moderators (Z_i) for the two random parameters.

5. RESULTS

Table 4 reports the estimates for the RPL model with correlated parameters (RPLc) and the correlated RPL with Error Components (RPLc-ECM). The two models have been estimated in NLOGIT 5 (ChoiceMetrics, 2012) using 1,000 Halton draws. To reduce the scale, the variables *DIST* and *SIZE_NAT* have been divided by 100 so that they refer to hundreds of kilometres and hundreds of square kilometres, respectively.

All parameter estimates have the expected signs and are statistically significant. Both models exhibit a good model fit according to McFadden pseudo ρ^2. The log-likelihood at convergence and the Akaike Information Criterion (AIC) indicate that the RPLc-ECM model provides a better fit. Accordingly, the discussion of results that follows is based on this model.

As expected, *DIST* exerts (on average) a negative effect on tourists' utility, in line with Lyons et al. (2009), Chandra et al. (2014), De Valck et al. (2017) and Gosens and Rouwendal (2018), Pompili et al., (2019) and Alvarez-Díaz et al. (2020), among others. The spread parameter (standard deviation of the random component) is statistically significant. This implies that although for most individuals distance acts as a dissuasive factor, for some others it could be a desirable feature. The latter could be explained by travelling time being perceived as a commodity value that increases utility for some people (Chavas et al., 1989; Mokhtarian, 2005), *ceteris paribus*. Similarly, the positive and statistical significance of the *r_TEMP* mean coefficient indicates that, *on average*, utility increases as the temperature at the destination rises (relative to that at the origin). This is consistent with Bigano et al. (2006) and Lyons et al. (2009). However, both the significance and magnitude of the standard deviation of the random component suggest that the marginal utility of the ratio of temperatures is heterogeneous (see below).

Regarding the rest of the place-based attributes, the higher the prices at the destination, the less preferred the destination is. This is consistent with Massidda and Etzo (2012) and Alvarez-Díaz et al. (2020). The dummy variable for rainfalls (*RAIN*) is negative and significant, in line with prior expectations and the results of Lyons et al. (2009) and Alvarez-Díaz et al. (2020). Therefore, destinations with monthly average rainfalls over 60 litres per square metre are negatively valued. The number of tourism spots (*TOU_SPOTS*) is positively related with the likelihood of individuals travelling to that region. Additionally, both the number of national and natural parks (*NAT_PARKS*) and the surface size of natural areas (*SIZE_NAT*) positively affect a destination being chosen, in line with Alvarez-Díaz et al. (2020). Similarly, the positive coefficient of the *SKI_KM* variable indicates that nature-based travellers attach high importance to the availability of kilometres for skiing. When we look at the interaction term with *winter_sports*, we confirm that this positive utility is higher for these individuals. Interestingly, regions with coast (*COAST*) are negatively valued, *ceteris paribus*. However, the marginal utility for those who practise aquatic sports (*COAST***aquatic*) turns to be positive and significant. That is, the availability of the coast for nature-based tourism appears to be only relevant for this segment.

All ASCs are positive and statistically significant, except *REG3* (Madrid) and *REG6* (Andalusia and Murcia). This suggests that whereas there are no differences in utility between Madrid, Andalusia, Murcia and the Canary Islands conditional on the attributes, the rest of the regions have some residual features that increase their attractiveness. On the other hand, the standard deviations of the latent ECs are statistically significant for the Centre block of regions (*EC2*) and for the Canary Islands (*EC4*). Overall, it seems that unobserved features for the North and Mediterranean areas are better captured by the ASCs, while the corresponding ones for the Centre and the Canary Islands exhibit larger variation and are better accommodated through the ECs. Nonetheless, note that *REG4* exhibits its own mean effect apart from that from *EC2*.

Our estimates show that the negative marginal disutility of distance is moderated by age. This contradicts Lyons et al. (2009), who find that older people are averse to travelling farther away. Similarly, income also moderates the disutility of distance. This is in line with evidence that under economic constraints people travel to nearby regions (Cafiso et al., 2016). Conversely, travel party size increases the distaste for covering long distances. This might be explained by larger travel groups imposing higher transportation costs. Similarly, weekend trips are associated with tourists being more averse to long distances, possibly through time constraints. As for seasonal effects, it is only in the fourth quarter when the disutility of distance is higher. Interestingly, those from regions with above-mean temperatures exhibit higher disutilities for travelling to distant destinations in the second and third quarters. This suggests that when the place of origin is relatively warmer than the average, distance is a higher barrier.

Regarding the role of the trip purpose, we find that those who seek mountaineering, trekking and visiting natural areas are more discouraged to travel farther away. Nonetheless, the practice of aquatic or adventure sport moderates the disutility of distance, especially in the former case. As such, these two trip purposes alleviate the distance decay effect. This is consistent with previous research that finds differences in sensitivity to distance depending on travel purposes (Nicolau and Más, 2006; Swait et al., 2020; Panzera et al., 2020).

The marginal utility of *r_TEMP* is not related with age, income or party size. Conversely, those who travel during the weekend exhibit a reduced preference for warmer destinations (i.e. the positive effect of *r_TEMP* is moderated). Concerning seasonal differences, a climate gain is less valued in the first and fourth quarters (relative to the summer period). Accordingly, the pursuit of a warmer destination has less relevance in autumn and winter, although warm regions are still preferred. Interestingly, individuals from regions with above-mean temperatures in the summer season show higher preference for relatively cooler destinations. In this regard, there is consensus in the literature that higher temperatures are preferred up to a threshold (Bigano et al., 2006; Bujosa and Rosselló, 2013). By contrast, above-mean temperatures at the origin in the first quarter are associated with a higher preference for warmer regions. Our results are in line with Lyons et al. (2009), who document that i) on average tourists prefer destinations with high temperatures, with this preference being higher in the second and third quarters, and ii) mild climates are preferred in the first and fourth quarters.

	RPLc		RPLc-ECM	
Variable	Coef	Std.error	Coef	Std.error
$REG1$	3.019***	0.4003	3.431***	0.4167
$REG2$	1.912***	0.4189	2.118***	0.4340
$REG3$	1.046**	0.4211	–0.236	0.5237
$REG4$	1.535***	0.4141	1.773***	0.4294
$REG5$	1.177***	0.3835	1.480***	0.3913
$REG6$	0.319	0.4151	0.594	0.4195
$DIST$	–0.496***	0.0560	–0.519***	0.0598
r_TEMP	1.818***	0.7045	2.189***	0.7573
$RAIN$	–0.262***	0.0652	–0.328***	0.0725
$TCPI$	–0.012**	0.0058	–0.012**	0.0061
TOU_SPOTS	0.114***	0.0093	0.122***	0.0095
NAT_PARKS	0.052***	0.0085	0.054***	0.0089
$SIZENAT$	0.002***	0.0009	0.002**	0.0009
SKI_KM	0.001***	0.0002	0.001***	0.0002
$SKI\_KM * winter\_sports$	0.009***	0.0005	0.009***	0.0005
$COAST$	–1.144***	0.1648	–1.336***	0.1870
$COAST * aquatic$	2.736***	0.1249	2.799***	0.1274
SD $DIST$	0.413***	0.0154	0.440***	0.0181
SD r_TEMP	1.701***	0.3361	1.923***	0.3538
Cov($DIST$, r_TEMPT)	0.100	0.1253	0.080	0.1354
ϑ_1			0.196	0.5542
ϑ_2			1.024***	0.1342
ϑ_3			0.011	2.3180
ϑ_4			2.321***	0.2527
$DIST$ Mean shifters				
age	0.002***	0.0008	0.002***	0.0008
inc2	0.071***	0.0251	0.074***	0.0262
inc3	0.162***	0.0351	0.169***	0.0384
parsize	–0.088***	0.0095	–0.094***	0.0106
weekend	–0.543***	0.0238	–0.578***	0.0289
q1	–0.019	0.0491	–0.014	0.0518
q2	–0.011	0.0387	–0.008	0.0426
q4	–0.192***	0.0502	–0.207***	0.0546
d_warmorigin_1	–0.028	0.0510	–0.024	0.0538
d_warmorigin_2	–0.134***	0.0423	–0.142***	0.0455
d_warmorigin_3	–0.139***	0.0350	–0.132***	0.0382
d_warmorigin_4	0.099	0.0582	0.114*	0.0646

winter_sports	–0.089	0.0623	–0.081	0.0737
mou_trek_nat	–0.177***	0.0250	–0.188***	0.0285
rural	–0.035	0.0240	–0.041	0.0262
aquatic	0.280***	0.0262	0.295***	0.0282
advent	0.064**	0.0264	0.066**	0.0294
r_TEMP Mean shifters				
age	0.011	0.0091	0.012	0.0093
inc2	0.160	0.2673	0.205	0.2801
inc3	0.319	0.3503	0.384	0.3580
parsize	–0.158	0.986	–0.160	0.1040
weekend	–0.711***	0.2435	–0.700***	0.2626
q1	–0.960*	0.5015	–1.244**	0.5572
q2	–0.605	0.5670	–0.687	0.6202
q4	–1.161**	0.5093	–1.391**	0.5653
d_warmorigin_1	1.773***	0.4083	1.988***	0.4304
d_warmorigin_2	0.876	0.6010	1.022	0.6239
d_warmorigin_3	–3.168***	0.5677	–3.025***	0.6227
d_warmorigin_4	–0.410	0.5463	–0.320	0.5993
winter_sports	–0.570	0.4034	–0.616	0.4201
mou_trek_nat	–0.700***	0.2611	–0.724***	0.2785
rural	–0.791***	0.2640	–0.815***	0.2900
aquatic	0.213	0.3768	0.306	0.4068
advent	1.397***	0.2892	1.481***	0.3111
Log L	–11,194.7		–11,170.0	
AIC	22,497.6		22,456.0	
Pseudo-R2	0.406		0.408	
N	6,661		6,661	

Table 4.- Parameter estimates
*** p<0.01, ** p<0.05, * p<0.1

To facilitate interpretation, Table 5 presents how the marginal utility of *r_TEMP* differs by whether the origin exhibits below-mean or above-mean temperatures. Based on equations (16) and (19), the figures for the first quarter are obtained as follows:

$$\frac{\partial\, U_{ijt}}{\partial\, r\_TEMP_{ijt}} \frac{\partial\, r\_TEMP_{ijt}}{\partial\, q1_{it}} = \begin{cases} b_2 + \theta_6 & if\ d\_warmorigin\_1 = 0 \\ b_2 + \theta_6 + \theta_9 & if\ d\_warmorigin\_1 = 1 \end{cases} \tag{20}$$

The marginal utilities for the remaining quarters are computed in the same way. Non-significant coefficients are treated as zeroes.

	Below-mean temperature at origin	Above-mean temperature at origin
q1	0.945	2.933
q2	2.189	2.189
q3	2.189	–0.836
q4	0.798	0.798

Table 5.- Estimated marginal utilities for *r_TEMP* per quarter

As seen, in the summer season (third quarter), it is only those from warmer than average origins who prefer cooler destinations. However, the marginal utility does not vary between warmer and cooler origins either in the second and the fourth quarter, being the magnitude of the former notably larger. Strikingly, in the first quarter the preference for warmer regions is higher among those who live in relatively warmer regions.

Regarding trip purposes, trekking, mountaineering and visiting natural areas (*mou_trek_nat*) on the one hand, and visiting rural areas/villages (*rural*) on the other reduce the willingness to travel to warmer destinations. This suggests that for these purposes the value of a climate gain is of less relevance. Conversely, those who seek to practise adventure/risky sports (*advent*) show a higher preference for warmer locations. Finally, the marginal utility for *r_TEMP* is not related to *winter_sports* or to *aquatic*.

The estimated variance-covariance matrix of the random parameters in both the RPLc and the RPLc-ECM models along with the correlation between them is presented in Table 6. σ_{11} and σ_{22} denote the diagonal elements in the Cholesky matrix whereas σ_{12} refers to the below diagonal value.

	RPLc	RPLc-ECM
σ_{11}	0.413***	0.440***
σ_{12}	0.244	0.182
σ_{22}	1.683***	1.914***
Corr(β_{1i}, β_{2i})	0.143	0.094

Table 6.- Variance-covariance matrix estimates for RPLc and RPLc-ECM

As shown, the covariance between the random parameters is not significant (i.e. the two marginal utilities are independent). However, these estimates are conditional on the vector of taste shifters Z_i. We have re-estimated the RPLc-ECM model without them. Results are shown in Appendix B, Table A2. In this case, the covariance between the random parameters becomes significant (and positive), and the correlation amounts to 0.32. Accordingly, the marginal utilities for *DIST* and *r_TEMP* are *unconditionally* positively related. However, conditional on the taste shifters, the correlation vanishes. This clearly supports our modelling approach, which by means of introducing sources of preference heterogeneity captures the shared correlation between the marginal utilities of *DIST* and *r_TEMP*.

5.1 Robustness checks

Several alternative model specifications were examined. First, we tested whether results change depending on the distribution assumed for the random parameters. A triangular, a uniform, a truncated normal and a Weibull distribution were tested as alternatives. Our results are not driven by the distribution of unobserved heterogeneity (available upon request). Second, we replaced the *TCPI* at destination by the ratio of *TCPI* at destination relative to the one at the origin (*r_TCPI*). The estimates are reported in Appendix B, Table A3. The results are consistent with economic theory (utility is negatively related with destinations that are relatively more expensive) and remain largely unchanged. Third, both the RPL and the RPL-ECM were estimated imposing the restriction that the parameters for the mean shifters that were non-significant in Table 4 were zero. Results are displayed in Appendix B, Table A4. The magnitude and the direction of the effects are not affected. Fourth, to address concerns on the potential biases arising from the insularity of the Balearic and the Canary Islands, we estimated the model without considering these two regions. In doing so, trips to those regions and respondents travelling from there were excluded. Parameter estimates are reported in Appendix B, Table A5. Results remain consistent with the main analysis. Finally, we implemented the well-known travel cost method (Bujosa and Rosselló, 2013; Swait et al., 2020) to inspect the robustness of our modelling approach. The results obtained from this method are in line with our analysis. Details of its implementation and the parameter estimates are presented and discussed in Appendix C.

5.2 Marginal Rates of Substitution

There is a natural positive relationship between distance and the ratio of temperatures: regions with different climate conditions relative to the place of residence (either warmer or cooler) are located farther away. Since we have shown that, on average, individuals attach positive utility to warmer destinations and negative utility to distant regions, they seem to face a trade-off between these two features.

To examine this, we compute the marginal rate of substitution of distance for a warmer temperature (MRS). That is, how individuals are willing to trade distance in exchange for a temperature gain. Under the linear specification, the MRS equals the ratio of the partial derivatives of the latent utility with respect to *r_TEMP* and *DIST*. It can be understood as a sort of 'willingness to pay' if we consider distance as a payment vehicle. This is similar to De Valck et al. (2017). Since individuals have different sensitivities, this estimate is individual-specific and obtained as follows:

$$MRS_{r\_TEMP,DIST_i} = \frac{MU_{r\_TEMP}}{MU_{DIST}} = \frac{\frac{\partial U_{ij}^*}{\partial r\_TEMP_i}}{\frac{\partial U_{ij}^*}{\partial DIST_i}} = \frac{\beta_{2,i}}{\beta_{1,i}} \tag{21}$$

In equation (21) we use a conditional estimator of the marginal utilities rather than the structural parameters derived from the model estimates. As highlighted in Hensher et al. (2006), the derivation of the MRS as the ratio of individual-level parameters reduces the incidence of extreme values in comparison to drawing them from the unconditional population distributions. By applying Bayes' theorem, we compute the conditional expectation of the individual-specific marginal utilities by conditioning on all available information about each individual. Technical details together with kernel density plots are shown in Appendix D.

As it is widely known, the ratio of two normal distributions has a discontinuous distribution with a singularity problem when the denominator takes value zero. To avoid this, for the calculation of the MRS we have omitted those individuals who have a (statistically) zero marginal utility for either of the two attributes.[10] For this purpose, we follow the procedure outlined by Hess and Hensher (2010).[11] This leaves the sample with 4,368 valid observations (65.5% of the original sample). Although only 0.7% display values of $CV_{DIST,i}$ larger than 2 (in absolute terms), this percentage rises to 33.6% for $CV_{r\_TEMP,i}$. This suggests that the fraction of individuals that do not consider the ratio of temperatures is much larger than for the case of distance.

Figure 1 plots the histogram of the MRS, which is distributed to both sides of zero.[12] The discontinuity in the values around zero is a direct consequence of having dropped observations with estimated conditional marginal utilities in the neighbourhood of zero. Without altering the sign, negative values refer to the number of kilometres (in hundreds) that individuals are willing to travel to obtain a marginal *increase* in temperature relative to the origin ($r\_TEMP > 1$) while positive values indicate the reversal: the number of kilometres (in hundreds) individuals are willing to travel to obtain a marginal *decrease* in temperature relative to the origin ($r\_TEMP < 1$). The mean of the distribution of the MRS is –1.59. This means that, *on average*, tourists who pay attention to both attributes are willing to cover 159 kilometres to obtain a marginal increase in temperatures relative to the place of origin. About 70% of the sample exhibits a negative MRS (i.e. they trade distance for climate gains). By contrast, the remaining 30% is willing to cover distance in exchange for relatively cooler climate conditions. Overall, the MRS shows how the willingness to cover longer distances increases as individuals attach higher importance to a different climate at the destination (either warmer or cooler).

[10] This way to proceed is consistent with Daly et al. (2012). These authors provide mathematical proof showing that when the domain of the distribution of the denominator is restricted not to have support in an arbitrarily interval close to zero, inverse moments exist. Their theorem applies to independent parameters. For jointly normal variables, independence directly follows from a lack of correlation (i.e. $Cov(\beta_{1i}, \beta_{2i}) = 0$). In our data, conditional on Z, the MUs for distance and relative temperatures have been shown to be uncorrelated.

[11] These authors compute the coefficient of variation (CV) of the individual-specific conditional mean and standard deviation estimates for each attribute ($i.e. CV_{ki} = \frac{\sigma_{ki}}{\beta_{ki}}$). When this noise-to-signal ratio is higher than a given threshold, the individual-specific preference distribution is said to be *over-dispersed*. Hess and Hensher arbitrarily chose 2 as the threshold value because normal distributions tend to be over-dispersed when $CV_{ki} > 2$. This practise was followed by Scarpa et al. (2013).

[12] Since the obtained distribution has long tails, Figure 1 restricts the estimated MRS to lie on the interval (– 12.87, 9.69), which gathers 99.7 percent of the data ($\mu \pm 3\sigma$).

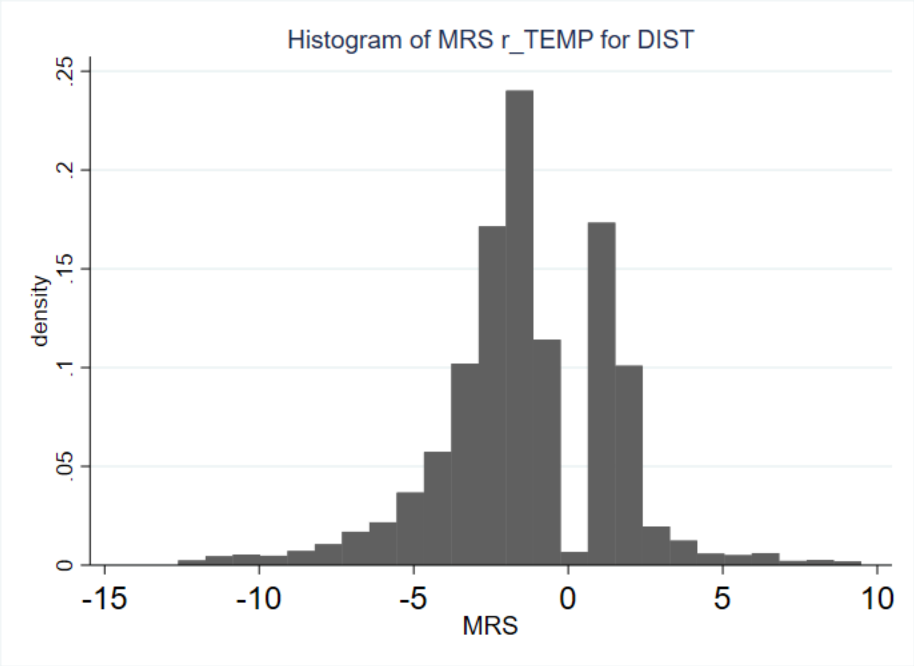

Figure 1.- MRS r_TEMP for DIST

5.3 Price and temperature elasticities

Apart from considering taste heterogeneity, scale heterogeneity and cross-correlation between alternatives, another advantage of the RPLc-ECM model is that it does not exhibit the IIA property. This means that any marginal change in an attribute does not only affect own choice probabilities but also impacts the choice probabilities of all the remaining regions. From a policy perspective, it seems valuable to evaluate tourists' reassignments under a shock in attribute *k* in a region *j*, *ceteris paribus*. Specifically, we aim to evaluate how changes in the price index for tourism services or the ratio of temperatures would impact the destination choice of nature-based tourists.

The marginal effect of a change in attribute *k* in region *j* on any generic destination *m* is given by:

$$\theta_{im(k|j)} = P_{im}(1(j = m) - P_{ij})\, \beta_{ik} \tag{22}$$

where P_{ij} and P_{im} are the choice probabilities as defined in Equation (13), β_{ik} are the MU of attribute *k*, and $1(j = m)$ is an indicator function of whether destination *j* equals destination *m*. Based on this formula, the own marginal effect when $j = m$ is:

$$\theta_{ij(k|j)} = P_{ij}\, (1 - P_{ij})\beta_{ik} \tag{23}$$

and the cross marginal effect is:

$$\theta_{im(k|j)} = P_{im}(-P_{ij})\, \beta_{ik} \tag{24}$$

To facilitate the interpretation, we compute the elasticities (i.e. the percentage change in the probability that individual *i* chooses destination *m* if there is a one-percent increase in the value of attribute *k* in alternative *j*) as follows:

$$\eta_{im\,(k|j)} \frac{\partial \ln P_{im}}{\partial \ln X\,(k|j)} = \frac{X_{k|j}}{P_{im}} \vartheta_{im(k|j)} = X_{k|j}(1(j = m) - P_{ij})\,\beta_{ik} \tag{25}$$

Tables 7 and 8 report the elasticities of *r_TEMP* and *TCPI*, respectively. They reflect the percentage change in choice probabilities of destination *m* (in columns) if there is a one-percent increase in either the relative temperature or the price index in destination j (in rows). The values on the diagonal refer to own elasticities ($j = m$), while the rest are cross elasticities ($j \neq m$). The matrix is not symmetric since the elasticity of a change in *j* on *m* is different from the reversal.

Figures 2 and 3 depict the own elasticities with respect to *r_TEMP* and *TCPI*, respectively. Darker colours refer to higher values. All the regions exhibit positive own *r_TEMP* elasticities (i.e. a marginal increase in temperature relative to the origin increases the likelihood of that region being chosen). Interestingly, regions in the North-West area (Cantabria, Galicia and Asturias), North-West (La Rioja and Aragon), Andalusia and Madrid have elasticities higher than the unity. However, there is no clear spatial pattern in the own *r_TEMP* elasticities.

Turning to the cross *r_TEMP* elasticities, we document large cross elasticities between the South area (Andalusia and Murcia) and the North-West one. Specifically, rises in temperatures in the South relative to the origin reduce North-Western regions' choice probabilities in greater magnitude than the opposite. This implies that when the South becomes warmer in comparison to the origin, *ceteris paribus*, the percentage reduction in visitors to the North-West is larger than the reversal. Something similar applies to Aragon, whose choice probabilities are notably reduced when there is a marginal increase in temperatures in Andalusia and Murcia relative to tourists' place of origin. This implies that the pursuit of different climate conditions takes place in favour of Southern regions.

This finding is contrary to the ones by Bujosa and Rosselló (2013) and Priego et al. (2015), who document that under a climate change scenario with a rise in temperatures, Northern regions would increase their visitors while Eastern regions will reduce their market shares. Nevertheless, our results cannot be directly compared with theirs since they analysed coastal tourism. Moreover, they only consider the summer season and ten regions (those with coast). In our case, all the regions are analysed, and elasticities are average values over the whole year. In this regard, since we are analysing choice elasticities with reference to a ratio, the estimated values are affected by the level of the denominator (i.e. the temperature at the origin). Regions have different temperatures (see Table A1 and Figure A6 in Appendix A) and different shares of outbound tourists. Since the elasticities are average values over the sample, they are affected by i) how many tourists share the same origin, and ii) how much different regions are among them in terms of temperature.

	CAN	AST	GAL	ARA	BQC	LRJ	NAV	MAD	CMA	CLE	EXT	BIS	CAT	VAL	AND	MUR	CIS
CAN	**2.389**	–0.012	–0.014	–0.015	–0.006	–0.011	–0.003	–0.008	–0.004	–0.004	–0.005	–0.006	–0.005	–0.006	–0.016	–0.007	–0.006
AST	–0.083	**1.452**	–0.066	–0.068	–0.039	–0.054	–0.014	–0.055	–0.022	–0.022	–0.024	–0.036	–0.028	–0.034	–0.081	–0.033	–0.028
GAL	–0.028	–0.014	**1.796**	–0.026	–0.012	–0.022	–0.006	–0.012	–0.009	–0.009	–0.008	–0.008	–0.010	–0.013	–0.019	–0.011	–0.011
ARA	–0.223	–0.135	–0.182	**1.871**	–0.084	–0.130	–0.038	–0.088	–0.061	–0.053	–0.067	–0.080	–0.069	–0.074	–0.182	–0.096	–0.075
BQC	–0.095	–0.073	–0.066	–0.077	**0.770**	–0.056	–0.001	–0.033	–0.016	–0.016	–0.017	–0.028	–0.026	–0.033	–0.083	–0.029	–0.028
LRJ	–0.266	–0.153	–0.237	–0.244	–0.092	**1.161**	–0.022	–0.135	–0.065	–0.047	–0.080	–0.097	–0.072	–0.076	–0.206	–0.115	–0.089
NAV	–0.092	–0.071	–0.063	–0.078	–0.033	–0.056	**0.220**	–0.015	0.002	0.003	0.006	–0.004	–0.013	–0.023	–0.076	–0.016	–0.026
MAD	–0.056	–0.054	–0.058	–0.038	–0.023	–0.032	–0.015	**1.202**	–0.021	–0.019	–0.032	–0.042	–0.015	–0.017	–0.079	–0.022	–0.025
CMA	–0.033	–0.025	–0.028	–0.034	–0.020	–0.028	–0.006	–0.012	**0.520**	–0.027	–0.025	–0.013	–0.021	–0.021	–0.027	–0.024	–0.020
CLE	–0.042	–0.031	–0.034	–0.041	–0.025	–0.036	–0.002	–0.015	–0.029	**0.431**	–0.022	–0.014	–0.023	–0.025	–0.033	–0.024	–0.023
EXT	–0.077	–0.070	–0.061	–0.072	–0.049	–0.057	–0.022	–0.039	–0.066	–0.063	**0.534**	–0.043	–0.044	–0.045	–0.072	–0.054	–0.047
BIS	–0.109	–0.112	–0.088	–0.098	–0.072	–0.079	–0.038	–0.058	–0.45	–0.043	–0.051	**0.673**	–0.062	–0.064	–0.116	–0.072	–0.066
CAT	–0.062	–0.049	–0.050	–0.053	–0.043	–0.055	–0.018	–0.014	–0.030	–0.030	–0.024	–0.029	**0.600**	–0.047	–0.050	–0.048	–0.038
VAL	0.163	–0.109	0.128	–0.149	–0.051	–0.104	–0.028	–0.063	–0.033	–0.012	–0.043	–0.058	–0.039	**0.678**	–0.141	–0.076	–0.051
AND	–0.453	–0.320	–0.393	–0.371	–0.165	–0.274	–0.075	–0.415	–0.131	–0.115	–0.156	–0.229	–0.143	–0.162	**1.598**	–0.205	–0.141
MUR	–0.107	–0.085	–0.092	–0.113	–0.068	–0.087	–0.043	–0.029	–0.062	–0.054	–0.058	–0.064	–0.079	–0.072	–0.092	**0.871**	–0.066
CIS	–0.046	–0.036	–0.032	–0.040	–0.025	–0.029	–0.006	–0.018	–0.013	–0.013	–0.013	–0.018	–0.018	–0.021	–0.041	–0.020	**0.734**

Table 7.- *r_TEMP* elasticities

*The values indicate the percentage change in choice probabilities for regions in columns if there is a one-percent increase in *r_TEMP* in the regions in rows.

CAN: Cantabria; AST: Principality of Asturias; GAL: Galicia; ARA: Aragon; BQC: The Basque Country; LRJ: La Rioja; NAV: Navarre; MAD: Community of Madrid; CMA: Castilla-LaMancha; CLE: Castile and Leon; EXT: Extremadura; BIS: The Balearic Islands; CAT: Catalonia; VAL: Valencian Community; AND: Andalusia; MUR: region of Murcia; CIS: The Canary Islands.

	CAN	AST	GAL	ARA	BQC	LRJ	NAV	MAD	CMA	CLE	EXT	BIS	CAT	VAL	AND	MUR	CIS
CAN	**–1.227**	0.008	0.010	0.013	0.010	0.008	0.006	0.003	0.005	0.006	0.005	0.004	0.006	0.008	0.010	0.006	0.007
AST	0.042	**–1.183**	0.037	0.038	0.039	0.034	0.038	0.035	0.029	0.028	0.034	0.043	0.032	0.032	0.049	0.034	0.030
GAL	0.014	0.008	**–1.254**	0.016	0.011	0.016	0.009	0.005	0.010	0.010	0.007	0.007	0.010	0.013	0.010	0.001	0.009
ARA	0.129	0.083	0.114	**–1.121**	0.100	0.099	0.076	0.036	0.078	0.074	0.061	0.062	0.081	0.095	0.097	0.076	0.075
BQC	0.084	0.087	0.072	0.081	**–1.138**	0.074	0.089	0.036	0.069	0.070	0.067	0.071	0.081	0.081	0.081	0.073	0.061
LRJ	0.154	0.111	0.195	0.170	0.135	**–1.044**	0.119	0.070	0.130	0.138	0.104	0.101	0.141	0.161	0.118	0.131	0.116
NAV	0.120	0.152	0.119	0.123	0.160	0.131	**–1.043**	0.082	0.132	0.134	0.140	0.163	0.155	0.147	0.127	0.148	0.109
MAD	0.020	0.026	0.024	0.015	0.013	0.015	0.012	**–0.805**	0.013	0.012	0.021	0.027	0.009	0.010	0.032	0.012	0.014
CMA	0.026	0.028	0.028	0.029	0.030	0.031	0.036	0.024	**–1.166**	0.069	0.067	0.034	0.040	0.034	0.026	0.042	0.027
CLE	0.037	0.038	0.039	0.040	0.045	0.046	0.051	0.030	0.099	**–1.137**	0.087	0.043	0.057	0.052	0.035	0.053	0.039
EXT	0.057	0.071	0.060	0.058	0.063	0.062	0.075	0.079	0.129	0.121	**–1.105**	0.091	0.071	0.064	0.064	0.083	0.056
BIS	0.071	0.095	0.077	0.070	0.073	0.076	0.086	0.100	0.070	0.065	0.089	**–1.117**	0.079	0.074	0.088	0.094	0.064
CAT	0.041	0.043	0.040	0.045	0.050	0.049	0.056	0.017	0.051	0.051	0.044	0.048	**–1.169**	0.057	0.039	0.061	0.042
VAL	0.133	0.130	0.139	0.139	0.153	0.151	0.154	0.064	0.138	0.146	0.127	0.134	0.161	**–1.069**	0.128	0.147	0.125
AND	0.206	0.209	0.182	0.180	0.163	0.153	0.142	0.182	0.126	0.122	0.140	0.177	0.137	0.143	**–0.992**	0.148	0.130
MUR	0.054	0.059	0.056	0.059	0.059	0.062	0.069	0.030	0.065	0.059	0.064	0.073	0.077	0.066	0.055	**–1.173**	0.053
CIS	0.030	0.032	0.025	0.029	0.035	0.026	0.033	0.015	0.027	0.027	0.027	0.028	0.032	0.030	0.030	0.029	**–0.955**

Table 8.- TCPI elasticities

*The values indicate the percentage change in choice probabilities for regions in columns if there is a one-percent increase in TCPI in the regions in rows.

CAN: Cantabria; AST: Principality of Asturias; GAL: Galicia; ARA: Aragon; BQC: The Basque Country; LRJ: La Rioja; NAV: Navarre; MAD: Community of Madrid; CMA: Castilla-LaMancha; CLE: Castile and Leon; EXT: Extremadura; BIS: The Balearic Islands; CAT: Catalonia; VAL: Valencian Community; AND: Andalusia; MUR: region of Murcia; CIS: The Canary Islands.

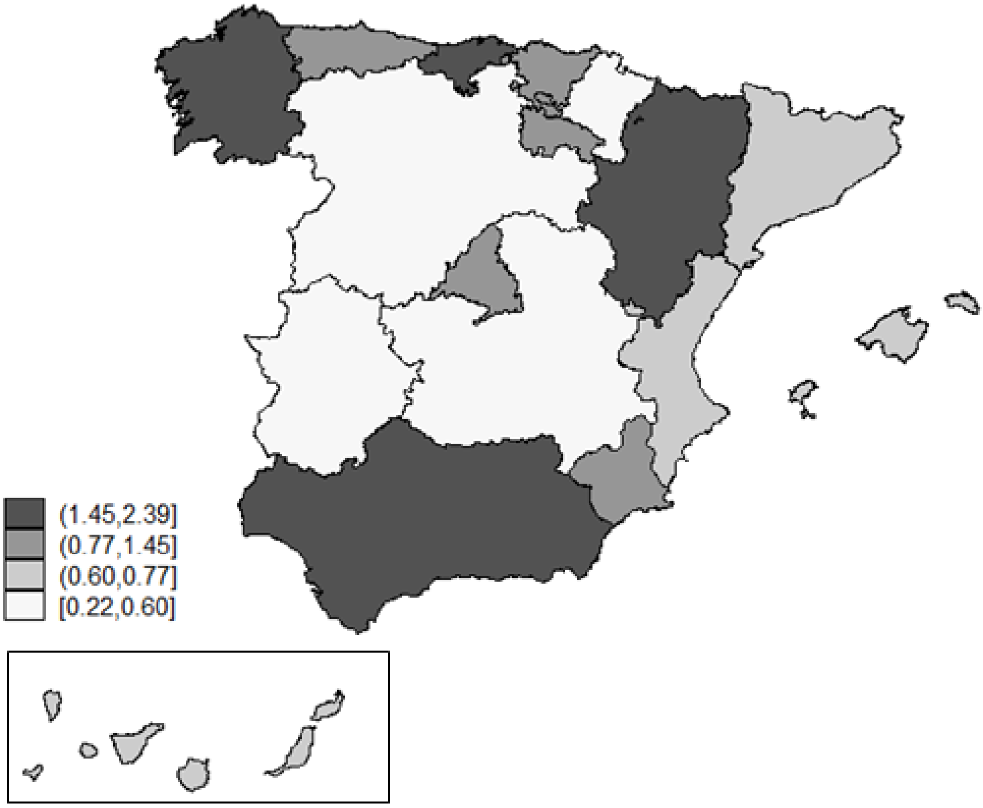

Figure 2.- Own choice-r_TEMP elasticities

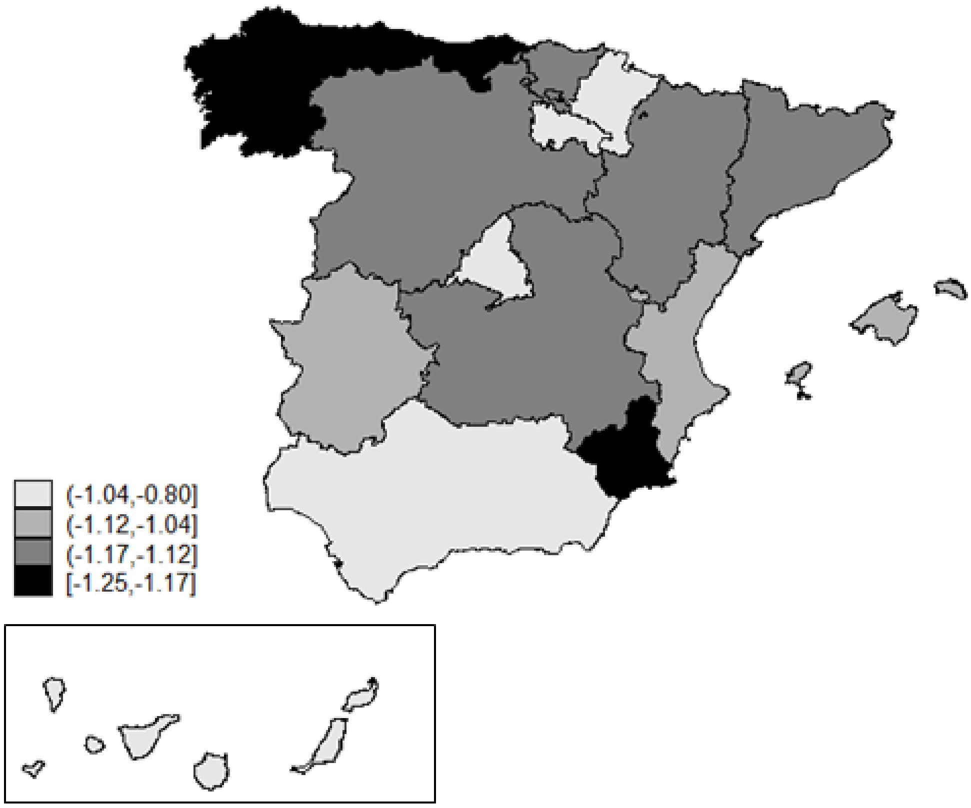

Figure 3.- Own choice-TCPI elasticities

Concerning the own *TCPI* elasticities (on-diagonal values in Table 8, depicted in Figure 3), all of them are negative, in line with economic theory. Except for Andalusia, the Canary Islands and Madrid, they are elastic (larger than one in absolute value). Since *TCPI* is a price index, the elasticities are interpreted as the percentage change in own choice probabilities if there is an inflation rate of one percent. The North-West regions are the most sensitive to price increases, closely followed by the Valencian Community, Murcia and Castilla-La Mancha.

As for the cross *TCPI* elasticities, the percentage change in choice probabilities under a one-percent inflation rate is larger for Andalusia, Navarre, La Rioja and the Valencian Community. In other words, tourism price inflation in these regions produces significant positive shifts in choice probabilities in the rest of regions. Among them, Andalusia and the Valencian Community are the two regions in which price increases lead to the largest reassignment of tourists. Most notably, cross elasticities are asymmetric. This is in line with Chandra et al. (2014), who find that Canadians are more sensitive to price changes in the USA than vice versa. Another interesting result is that the magnitude of the cross choice-price elasticities is larger for neighbouring regions, as it happens for Navarre-the Basque Country (0.160), Valencian Community-Murcia (0.147), Castilla-LaMancha-Castile and Leon (0.07), and Extremadura-Castilla-LaMancha (0.129). This implies that these regions are close substitutes in their characteristics, so a one percent increase in their prices makes other regions' choice probabilities to rise significantly. Additionally, we document a Northwest-South substitution pattern by which the cross choice-price elasticities between these areas are the largest in size.

Before ending this subsection, some limitations need to be acknowledged. First, for the *r_TEMP* case, the elasticities assume percentage increases in temperature in the region being analysed keeping everything else constant. That is, both the temperature of the remaining regions and all the other attributes, including the dummy for high rainfall (*RAIN*), are assumed not to change. This *ceteris paribus* condition might be a strong assumption. Second, since the model only considers domestic nature-based tourism, no transfers between domestic and international tourism are allowed.

6. CONCLUSIONS

This paper analyses the destination attributes that drive domestic trips for nature-based purposes in Spain. We match monthly microdata with both time-varying and time-invariant regional characteristics. We pay attention to the effect of distance and relative temperatures in site choice probabilities, while controlling for other destination-specific amenities. In doing so, we provide new evidence on the drivers of heterogeneity in marginal utilities.

We estimate a correlated Random Parameter Logit with Error Components model that controls for unobserved preference heterogeneity for the attributes and the destinations. Our results point to substantial preference heterogeneity for distance and temperature differentials between origin and destination. Whereas on average distance is a dissuasive factor, this distaste for distant regions is moderated by age and income. Conversely, travel party size, travelling on weekends and in the fourth quarter reinforce the negative effect of distance. Interestingly, trips with the aim of practising aquatic or adventure sports make individuals more prone to travelling farther away, whereas trekking and mountaineering are associated with a higher distaste for distance.

As for the temperature at the destination relative to the origin, tourists prefer warmer regions, especially in the third quarter. Remarkably, we show that individuals from origins with above mean temperatures are deterred from travelling to warmer locations in the summer period. This highlights the existence of relevant non-linearities in the preference for higher temperature. Concerning trip purposes, the preference for warmer destinations is larger for those who seek to practise adventure sports, but moderated by the purpose of trekking and mountaineering and visiting rural areas. Additionally, site choice probabilities are negatively related to high rainfall.

Our results also indicate that the number of available kilometres for skiing (especially for winter sport tourists), the number of tourism spots, the number of national parks and the size of protected natural areas are positively valued by individuals when choosing where to travel. As predicted by economic theory, choice probabilities are negatively influenced by tourism prices. Interestingly, whereas inland regions are on average preferred for nature-based trips, coastal destinations are best suited for those who like to practise aquatic sports.

Based on the model estimates, we document that distance and the relative temperature of the destination with respect to the origin are not statistically significant for 0.7% and 34% of the sample, respectively. For those for whom both attributes are relevant, we find they are willing to travel about 160 kilometres for a marginal increase in relative temperatures. Interestingly, about 70% of the sample trade distance for temperature gains. However, the remaining 30% are found to cover long distances to reach cooler locations. Furthermore, the own price elasticities indicate that all destinations are highly price-elastic. The largest cross price elasticities are found among cross-bordering regions. We also document a Northwest-South substitution pattern by which price increases in one area positively impact choice probabilities in the other. Most importantly, cross-price elasticities are asymmetric.

This paper contributes to the literature on regional and tourism economics in different ways. First, given the mixed evidence on the (dis)taste for distance and temperatures encountered in the literature, the proposed model specification deepens into this by allowing the marginal utilities for these two dimensions to be heterogeneous. We link the distribution of preferences in the sample with a set of sociodemographic, temporal and trip-related variables. Therefore, we do not only model taste heterogeneity but also identify the factors that shift sensitivities.

Second, unlike most applications that only use temperatures at the destination, our model specification acknowledges that temperature rises may be differently valued depending on the level at the origin. To the authors' knowledge, this is the first empirical application that considers relative temperatures for modelling individual choices. Moreover, among the set of taste shifters, we consider indicators of quarterly above mean temperatures at the origin. Therefore, we provide new evidence on the non-linearities of the preference for temperature.

Third, our analysis deals with unobserved heterogeneity, allowing for shared correlation between the taste for temperatures and distance. We show that the marginal utilities for *DIST* and *r_TEMP* are unconditionally positively related. However, conditional on the taste shifters, the correlation vanishes. This suggests that modelling the sources of preference heterogeneity allows us to capture the shared correlation between the marginal utilities of *DIST* and *r_TEMP*. From the model estimates, we compute the marginal rates of substitution of distance for temperature gains and provide some intuition about the trade-offs between these two dimensions.

Our results have some implications. Given the growing importance of the tourism domestic market for the economic development of some regions, our findings could be valuable for regional authorities in charge of tourism planning and destination management. Our results might enhance their understanding of the factors that attract prospective tourists to their regions, highlighting the pull effect of the regional natural resources for outdoor recreation. In the pandemic context, nature-based trips are expected to become even more popular. Prompting nature-based tourism could also alleviate the intrinsic seasonality of tourism revenues. All Spanish regions choice probabilities are highly price-elastic, with Murcia, Asturias, Cantabria and Galicia exhibiting the largest choice sensitivities. The high cross-price elasticities between Northwest and South regions point to a reallocation of tourists from the North to the South and vice versa after an increase in tourism prices. Policymakers need to be aware of this in the development of public policy interventions that involve tourism taxes. From the viewpoint of practitioners, our modelling approach offers an improved theoretically consistent way of analysing domestic leisure trips that can be extended to other types of tourism trips. Nonetheless, the generalization of our findings to other tourist segments or countries should be made with caution, since our results could be specific of the Spanish domestic travel market.

REFERENCES

Agiomirgianakis, G., Serenis, D. and Tsounis, N. (2017). Effective timing of tourism policy: The case of Singapore. *Economic Modelling* 60: 29-38.

Alvarez-Diaz, M., D'Hombres, B., Ghisetti, C. and Pontarollo, N. (2020). Analysing domestic tourism flows at the provincial level in Spain by using spatial gravity models. *International Journal of Tourism Research*, 22: 403-415.

Bekhor, S. and Prashker, J.N. (2008). GEV-based destination choice models that account for unobserved similarities among alternatives. *Transportation Research Part B* 42: 243-262.

Bernini, C., Cracolici, M.F. and Nijkamp, P. (2017). Place-based attributes and spatial expenditures behavior in tourism. *Journal of Regional Science* 57(2): 218-244.

Berry, S. and Pakes, A. (2007). The pure characteristics demand model. *International Economic Review* 48(4): 1193-1225.

Bigano, A., Hamilton, J.M., and Tol, R.J. (2006). The impact of climate on holiday destination choice. *Climatic Change* 76(3–4): 389–406.

Bujosa, A. and Rosselló, J. (2013). Climate change and summer mass tourism: the case of Spanish domestic tourism. *Climatic Change* 117: 363-375.
Cafiso, G., Cellini, R. and Cuccia, T. (2016). Do economic crises lead tourists to closer destinations? Italy at the time of the Great Depression. *Papers in Regional Science* 97(2): 369-386.
Chan, N.W. and Wichman, C.J. (2020). Climate change and recreation: evidence from North American Cycling. *Environmental and Resource Economics*, 76: 119-151.
Chandra, A., Head, K. and Tappa, M. (2014). The economics of cross-border travel. *The Review of Economics and Statistics* 96(4): 648-661.
Chavas, J.P., Stoll, J. and Sellar, C. (1989). On the commodity value of travel time in recreational activities. *Applied Economics* 21(6): 711-722.
Daly, A.J., Hess, S. and Train, K.E. (2012). Assuring finite moments for willingness to pay in random coefficients models. *Transportation* 39(1): 19–31.
De-la-Mata, T. and Llano-Verduras, C. (2012). Spatial patters and domestic tourism: an econometric analysis using inter-regional monetary flows by type of journey. *Papers in Regional Science*, 91(29): 437-470.
De Valck, J., Landuyt, D., Broekx, S., Liekens, I., De Nocker, L. and Vranken, L. (2017). Outdoor recreation in various landscapes: which site characteristics really matter? *Land Use Policy* 65: 186-197.
Deaton, A.S. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review* 70(3): 312-326.
Debreu, G. (1960). Review of Individual Choice Behavior by R.D. Luce. *American Economic Review* 50(1): 186-188.
Dekker, T., Hess, S., Arentze, T. and Chorus, G. (2014). Incorporating needs-satisfaction in a discrete choice model of leisure activities. *Journal of Transport Geography* 38: 66-74.
Dundas, S.J. and von Haefen, R.H. (2020). The effects of weather on recreational fishing demand and adaptation: implications for a changing climate. *Journal of the Association of Environmental and Resource Economists* 7(2): 209–242.
Eugenio-Martín, J.L. and Campos-Soria, J.A. (2010). Climate in the region of origin and destination choice in outbound tourism demand. *Tourism Management* 31: 744-753.
Faber, B. and Gaubert, C. (2019). Tourism and economic development: evidence from Mexico's coastline. *American Economic Review*, 109(6): 2245-2293.
Falk, M. (2010). A dynamic panel data analysis of snow depth and winter tourism. *Tourism Management* 31: 912-924.
Gosens, T. and Rouwendal, J. (2018). Nature-based outdoor recreation trips: Duration, travel mode and location. *Transportation Research Part A* 116: 513-530.
Greene, W.H. (2012). *NLOGIT5 Reference Guide*. Chapter 29, 556-566.
Greene, W.H. and Hensher, D.A. (2007). Heteroscedastic control for random coefficients and error components in mixed logit. *Transportation Research Part E* 43: 610-623.
Greene, W.H. and Hensher, D.A. (2010). Do scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation* 37: 413-428.
Helbich, M., Böcker, L. and Dijst, M. (2014). Geographic heterogeneity in cycling under various weather conditions: evidence from Greater Rotterdam. *Journal of Transport Geography* 38: 38-47.
Hensher, D.A., Greene, W.H. and Rose, J. (2006). Deriving Willingness-to-Pay estimates of Travel-Time Savings from Individual-based Parameters. *Environment and Planning A: Economy and Space* 38(12): 2365-2376.
Hess, S. and Hensher, D.A. (2010). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B* 44: 781-790.
Hess, S. and Train, K. (2017). Correlation and scale in mixed logit models. *Journal of Choice Modelling* 23: 1-8.
Jeon, C.Y. and Yang, H.W. (2021). The structural changes of a local tourism network: comparison of before and after COVID-19. *Current Issues in Tourism*, forthcoming.
Kadiyali, V. and Kosová, R. (2013). Inter-industry employment spillovers from tourism inflows. *Regional Science and Urban Economics*, 43: 272-281.
Keane, M. and Wasi, N. (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics* 28: 1018-1045.

Klaiber, H.A. and Von Haefen, R.H. (2019). Do Random Coefficients and Alternative Specific Constants improve policy analysis? An empirical investigation of model fit and prediction. *Environmental and Resource Economics* 73(1): 75-91.

Lancaster, J. K. (1966). A new approach to consumer theory. *Journal of Political Economy* 74(2): 132–157.

Lise, W. and Tol, R.S.J. (2002). Impact of climate on tourism demand. *Climatic Change* 55: 429-449.

Lyons, S., Mayor, K. and Tol, R.S.J. (2009). Holiday destinations: understanding the travel choices of Irish tourists. *Tourism Management* 30: 683-692.

Marrocu, E. and Paci, R. (2013). Different tourists to different destinations. Evidence from spatial interaction models. *Tourism Management* 39: 71-83.

Massidda, C. and Etzo, I. (2012). The determinants of Italian domestic tourism: a panel data analysis. *Tourism Management*, 33: 603-610.

McConnell, K.E. (1999). Household Labor Market Choices and the Demand for Recreation. *Land Economics* 75(3): 466-477.

McFadden, D. (1974). Conditional logit analysis of qualitative choice Behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

McFadden, D. (2001). Economic choices. *American Economic Review* 91(3): 351-378.

Mieczkowski, Z. (1985). The Tourism Climatic Index: A Method of Evaluating World Climates for Tourism. *Canadian Geographer* 29(3): 220–233.

Mokhtarian, P.L. (2005). Travel as a desired end, not just a means. *Transportation Research Part A* 39(2-3): 93-96.

Morey, E.R. and Krizberg, D. (2012). It's not where you do it, it's who you do it with? *Journal of Choice Modelling* 5: 176-191.

Murdock, J. (2006). Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management* 51: 1-25.

Naranpanawa, N., Rambaldi, A.M. and Sipe, N. (2019). Natural amenities and regional tourism employment: a spatial analysis. *Papers in Regional Science*, 98: 1731-1757.

Nicolau, J.L. (2008). Characterizing tourist sensitivity to distance. *Journal of Travel Research* 47: 43-52.

Nicolau, J.L. (2010). Variety-seeking and inertial behavior: the disutility of distance. *Tourism Economics* 19(1): 251-264.

Nicolau, J.L. and Más, F.J. (2006). The influence of distance and prices on the choice of tourist destinations: the moderating role of motivations. *Tourism Management* 27: 982-996.

Paci, R. and Marrocu, E. (2014). Tourism and regional growth in Europe. *Papers in Regional Science* 93: S25–S60.

Panzera, E., de Graaff, T. and de Groot, H.L.F. (2021). European cultural heritage and tourism flows: the magnetic role of superstar World Heritage Sites. *Papers in Regional Science*, 100, 101-122.

Parsons, G.R. and Hauber, A.B. (1998). Spatial boundaries and choice set definition in a random utility model of recreation demand. *Land Economics* 74(1): 32-48.

Pompili, T., Pisati, M. and Lorenzini, E. (2019). Determinants of international tourist choices in Italian provinces: a joint demand-supply approach with spatial effects. *Papers in Regional Science*, 98: 2251-2273.

Priego, F.J., Rosselló, J. and Santana-Gallego, M. (2015). The impact of climate change on domestic tourism: a gravity model for Spain. *Regional Environmental Change*, 15: 291-300.

Revelt, D. and Train, K. (1998). Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level. *The Review of Economics and Statistics* 80(4): 647-657.

Rosselló-Nadal, J., Riera, A., and Cárdenas, V. (2011). The impact of weather variability on British outbound flows. *Climatic Change* 105: 281-292.

Scarpa, R., Zanoli, R., Bruschi, V. and Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics* 95(1): 165-180.

Swait, J., Franceschinis, C. and Thiene, M. (2020). Antecedent Volition and Spatial Effects: can multiple goal pursuit mitigate distance decay? *Environmental and Resource Economics* 75(2): 243-270.

Train, K. (1998). Recreation demand models with taste variation. *Land Economics* 74: 230–239.

ONLINE SUPPLEMENTARY MATERIAL FOR

Modelling Heterogeneous Preferences for Nature-based Tourism Trips

APPENDIX A

Here we provide further information on some of the attributes considered in the analysis. Figures A1-A5 show the histograms of the variables *DIST*, *TCPI*, *RAINFALL* (continuous), T*EMPERATURE* and *r_TEMP*, respectively. Table A1 presents the yearly mean temperature per region. Figure A6 depicts the monthly average temperature by region. Figure A7 shows a smooth kernel density plot for temperature per quarter.

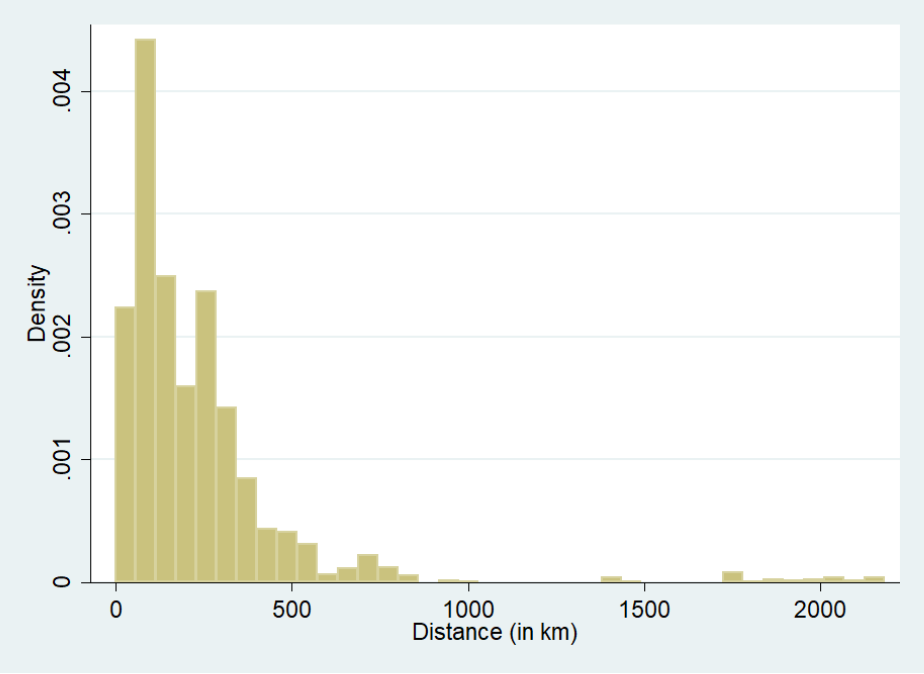

Figure A1.- Histogram of *DIST* (in kilometres)

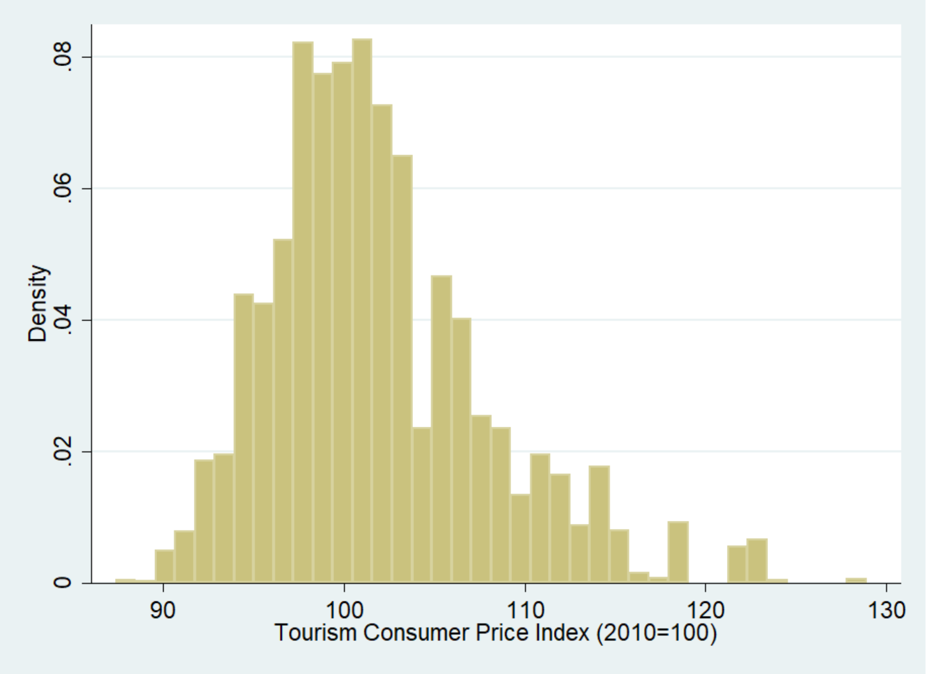

Figure A2.- Histogram of *TCPI*

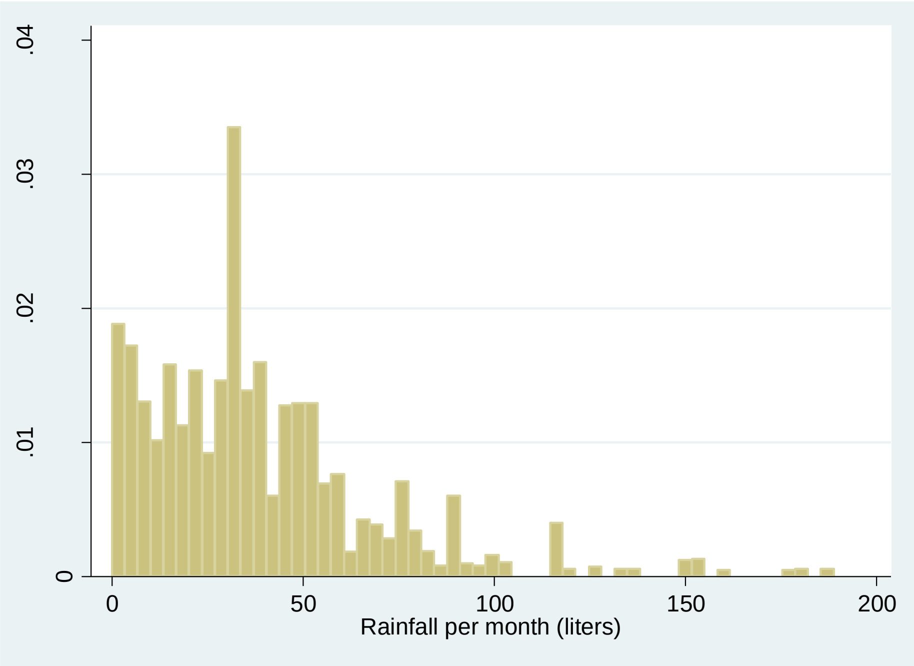

Figure A3.- Histogram of *RAINFALL* (liters per month)

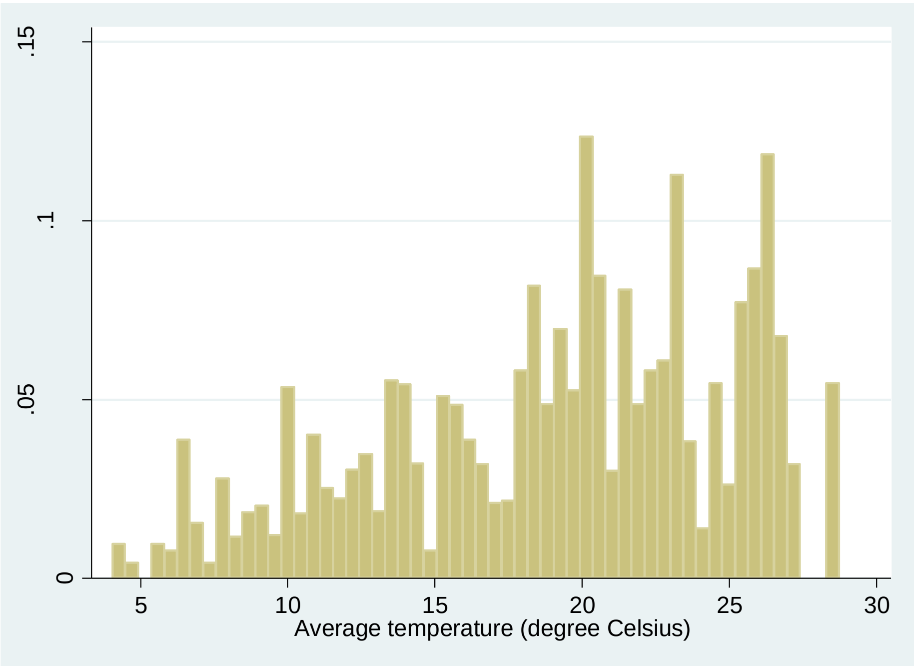

Figure A4.- Histogram of *TEMPERATURE* (degree Celsius)

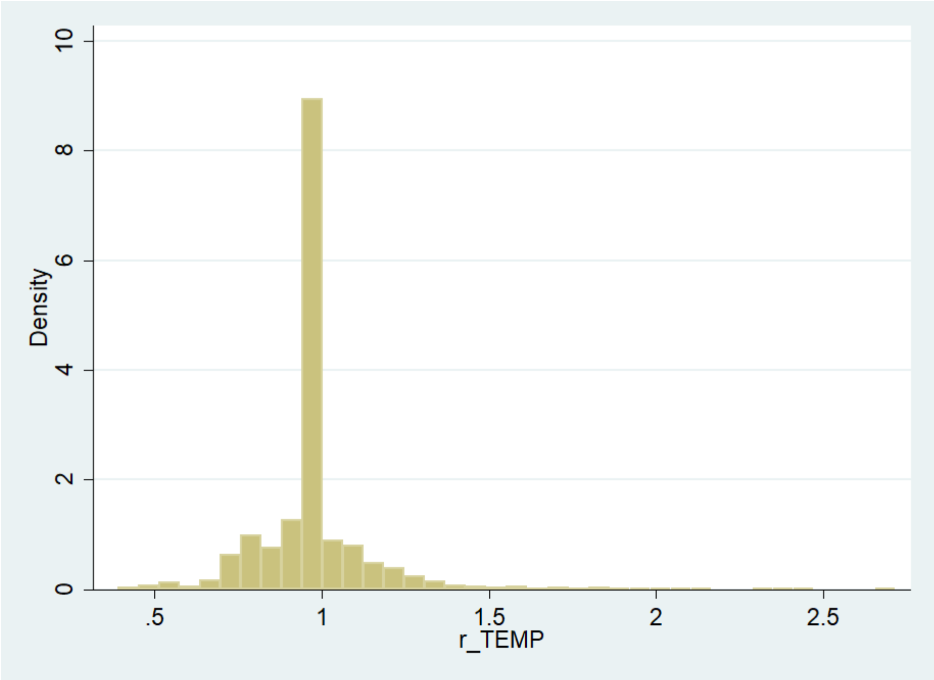

Figure A5.- Histogram of *r_TEMP*

Region	Yearly Average Temperature (ºC)
AND	18.47
ARA	14.66
AST	13.51
BIS	17.94
CIS	18.08
CAN	14.99
CLE	12.06
CMA	15.03
CAT	16.60
VAL	18.51
EXT	17.04
GAL	14.04
MAD	15.66
MUR	19.44
NAV	13.44
BQC	13.67
LRJ	14.31

Table A1- Yearly average temperature per region

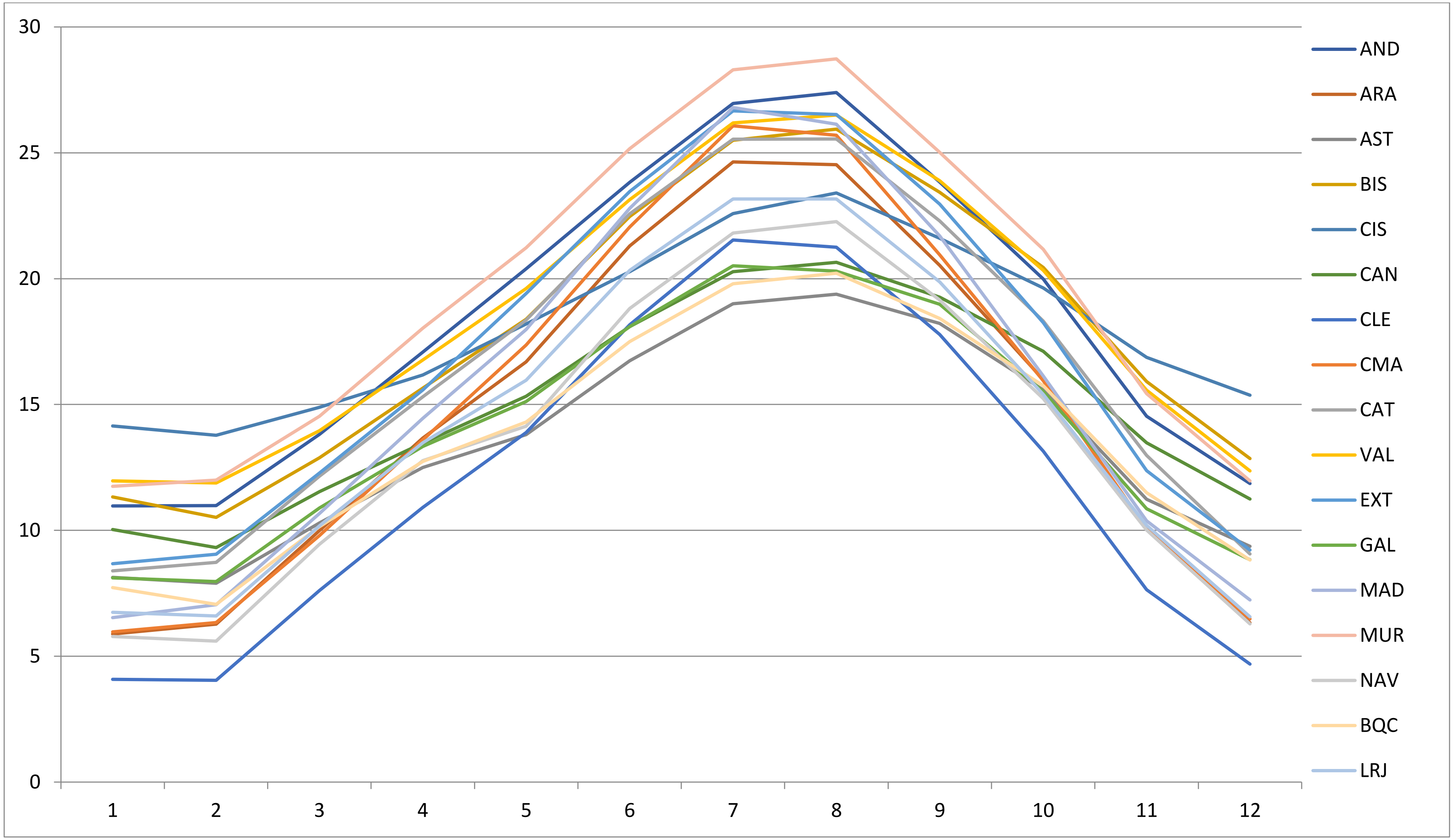

Figure A6.- Monthly average temperature by region

CAN: Cantabria; AST: Principality of Asturias; GAL: Galicia; ARA: Aragon; BQC: The Basque Country; LRJ: La Rioja; NAV: Navarre; MAD: Community of Madrid; CMA: Castilla-LaMancha; CLE: Castille and Leon; EXT: Extremadura; BIS: The Balearic Islands; CAT: Catalonia; VAL: Valencian Community; AND: Andalusia; MUR: region of Murcia; CIS: The Canary Islands.

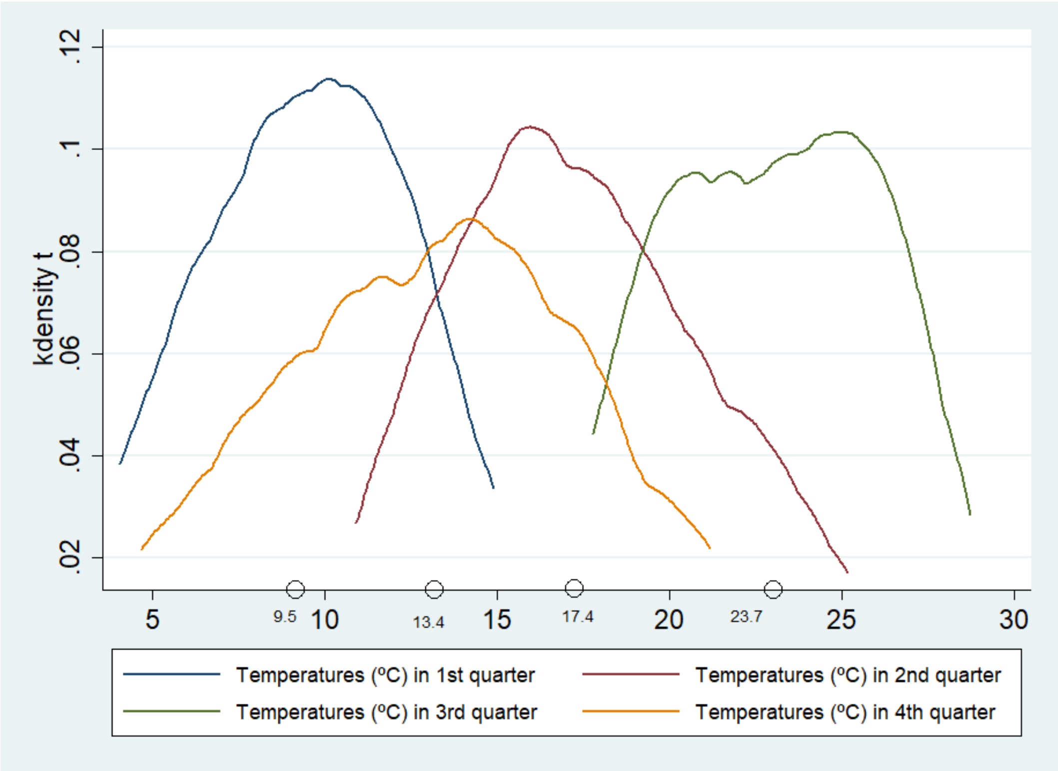

Figure A7.- Smooth kernel density plot for temperature per quarter

APPENDIX B

	RPLc-ECM	
Variable	Coef	St.Error
$REG1$	4.066***	0.4193
$REG2$	2.617***	0.4379
$REG3$	0.413	0.5373
$REG4$	2.266***	0.4347
$REG5$	2.128***	0.3985
$REG6$	1.417***	0.4275
$DIST$	–1.002***	0.0229
r_TEMP	0.804***	0.1778
$RAIN$	–0.310***	0.0684
$TCPI$	–0.006	5.9e-03
TOU_SPOTS	0.124***	9.3e03
NAT_PARKS	0.064***	8.6e-03
$SIZENAT$	0.001	9.0e-04
SKI_KM	0.001***	2.3e-04
$SKI\_KM * winter\_sport$	0.009***	5.0e-04
$COAST$	–1.544***	0.1817
$COAST * aquatic$	3.067***	0.1229
SD $DIST$	0.572***	0.019
SD r_TEMP	1.758***	0.3822
Cov($DIST$, r_TEMPT)	0.329**	0.1549
ϑ_1	0.338	0.3411
ϑ_2	1.045***	0.1267
ϑ_3	0.073	0.6617
ϑ_4	2.225***	0.2732
Log L	-12,134.37	
AIC	24,316.7	
Pseudo-R2	0.378	
N	6,661	

Table A2.- Parameter estimates for RPLc-ECM without taste shifters
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef	St.Error
$REG1$	3.429***	0.4165
$REG2$	2.116***	0.4339
$REG3$	–0.240	0.5237
$REG4$	1.769***	0.4292
$REG5$	1.476***	0.3911
$REG6$	0.592	0.4193
$DIST$	–0.519***	0.0598
r_TEMP	2.194***	0.7567
$RAIN$	–0.329***	0.0725
$TCPI$	–1.291**	0.6471
TOU_SPOTS	0.122***	0.0095
NAT_PARKS	0.054***	0.0089
$SIZE\_NAT$	0.002**	9.4e-04
SKI_KM	0.001***	2.3e-04
$SKI\_KM * winter\_sports$	0.009***	5.5e-04
$COAST$	–1.338***	0.1868
$COAST * aquatic$	2.799***	0.1274
SD $DIST$	0.439***	0.0181
SD r_TEMP	1.918***	0.3542
Cov($DIST, r\_TEMP$)	0.081	0.1353
ϑ_1	0.197	0.5507
ϑ_2	1.025***	0.1341
ϑ_3	0.011	2.3169
ϑ_4	2.322***	0.2527
$DIST$ Mean shifters		
age	0.002***	8.6e-04
inc2	0.074***	0.0261
inc3	0.169***	0.0384
parsize	–0.094***	0.0106
weekend	–0.578***	0.0289
q1	–0.014	0.0518
q2	–0.008	0.0426
q4	–0.208***	0.0546
d_warmorigin1	–0.024	0.0538
d_warmorigin2	–0.142***	0.0455
d_warmorigin3	–0.132***	0.0382
d_warmorigin4	0.114*	0.0646
winert_sports	–0.081	0.0737
mou_trek_nat	–0.188***	0.0285
rural	–0.041	0.0262
aquatic	0.295***	0.0282
advent	0.066**	0.0294
r_TEMP Mean shifters		
age	0.012	0.0093
inc2	0.205	0.2800
inc3	0.384	0.3579
parsize	–0.160	0.1039
weekend	–0.699***	0.2625
q1	–1.250**	0.5566

q2	–0.689	0.6200
q4	–1.397**	0.5647
d_warmorigin1	1.988***	0.4302
d_warmorigin2	1.022	0.6238
d_warmorigin3	–3.029***	0.6229
d_warmorigin4	–0.318	0.5990
winter_sports	–0.617	0.4200
mou_trek_nat	–0.723***	0.2784
rural	–0.814***	0.2899
aquatic	0.305	0.4067
advent	1.481***	0.3110
Log L	–11,170.0	
AIC	22,456.0	
Pseudo-R2	0.408	
N	6,661	

Table A3.- Parameter estimates for RPLc-ECM replacing $TCPI$ by r_TCPI
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef	St.Error
$REG1$	3.250***	0.3778
$REG2$	1.949***	0.3982
$REG3$	0.018	0.4694
$REG4$	1.610***	0.3959
$REG5$	1.313***	0.3542
$REG6$	0.489	0.3851
$DIST$	–0.520***	0.0516
r_TEMP	2.313***	0.4144
$RAIN$	–0.324***	0.0698
$TCPI$	–0.010*	0.0060
TOU_SPOTS	0.122***	0.0094
NAT_PARKS	0.056***	0.0087
$SIZE\_NAT$	0.002**	9.2e-04
SKI_KM	0.001***	2.3e-04
$SKI\_KM * winter\_sports$	0.009***	5.1e-04
$COAST$	–1.356***	0.1834
$COAST * aquatic$	2.787***	0.1255
SD $DIST$	0.434***	0.0168
SD r_TEMP	1.753***	0.3452
Cov($DIST$, r_TEMP)	0.106	0.1254
ϑ_1	0.061	0.6334
ϑ_2	0.964***	0.1369
ϑ_3	0.184	0.3782
ϑ_4	1.954***	0.2333
$DIST$ Mean shifters		
age	0.002***	8.2e-04
inc2	0.059**	0.0250
inc3	0.159***	0.0355
parsize	–0.093***	0.0101
weekend	–0.568***	0.0273
q4	–0.186***	0.0456
d_warmorigin2	–0.135***	0.0353
d_warmorigin3	–0.127***	0.0302
d_warmorigin4	0.108*	0.0616
mou_trek_nat	–0.188***	0.0269
aquatic	0.295***	0.0266
advent	0.073**	0.0285
r_TEMP Mean shifters		
weekend	–0.691***	0.2520
q1	–1.239***	0.3629
q4	–1.357***	0.3667
d_warmorigin1	2.067***	0.4116
d_warmorigin3	–2.984***	0.5116
mou_trek_nat	–0.592**	0.2590
rural	–0.720***	0.2790
advent	1.469***	0.2847
Log L	–11,189.1	
AIC	22,466.4	
Pseudo-R2	0.407	
N	6,661	

Table A4- Parameter estimates for RPLc-ECM imposing restrictions on some taste shifters
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef	St.Error
$REG1$	2.774***	0.1835
$REG2$	1.503***	0.2529
$REG3$	0.618**	0.2558
$REG4$	0.990***	0.2492
$REG5$	0.717***	0.1550
$DIST$	–0.544***	0.0596
r_TEMP	2.481***	0.7433
$RAIN$	–0.222***	0.0721
$TCPI$	–0.003	0.0065
TOU_SPOTS	0.104***	0.0097
NAT_PARKS	0.081***	0.0109
$SIZE\_NAT$	0.002***	9.3e-04
SKI_KM	0.001***	2.4e-04
$SKI\_KM * winter\_sports$	0.009***	5.7e-04
$COAST$	–1.380***	0.1837
$COAST * aquatic$	2.575***	0.1272
SD $DIST$	0.293***	0.0248
SD r_TEMP	1.779***	0.3501
Cov($DIST$, r_TEMP)	0.012	0.1118
ϑ_1	0.500**	0.2316
ϑ_2	0.370	0.2662
ϑ_3	0.031	1.3255
$DIST$ Mean shifters		
age	0.002***	8.5e-04
inc2	0.117***	0.0263
inc3	0.207***	0.0380
parsize	–0.077***	0.0098
weekend	–0.501***	0.0265
q1	–0.072	0.0511
q2	–0.029	0.0437
q4	–0.210***	0.0537
d_warmorigin1	–0.021	0.0546
d_warmorigin2	–0.106**	0.0463
d_warmorigin3	–0.126***	0.0389
d_warmorigin4	0.132**	0.0629
winter_sports	–0.071	0.0733
mou_trek_nat	–0.194***	0.0272
rural	–0.043*	0.0255
aquatic	0.257***	0.0291
advent	0.084***	0.0279
r_TEMP Mean shifters		
age	0.008	0.0090
inc2	0.270	0.2707
inc3	0.389	0.3440
parsize	–0.118	0.0994
weekend	–0.597**	0.2503
q1	–1.500***	0.5453
q2	–0.657	0.6063
q4	–1.671***	0.5517

d_warmorigin1	1.747***	0.4222
d_warmorigin2	1.178*	0.6045
d_warmorigin3	–2.956***	0.6194
d_warmorigin4	–0.111	0.5972
winter_sports	–0.602	0.4097
mou_trek_nat	–0.833***	0.2701
rural	–0.693**	0.2787
aquatic	–0.141	0.3903
advent	1.336***	0.2994
Log L	–10,273.2	
AIC	20,658.5	
Pseudo-R2	0.388	
N	6,207	

Table A5.- Parameter estimates for RPLc-ECM without the Balearic and the Canary Islands
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX C

The travel cost method has a long tradition in recreational demand studies as a way to compute the costs associated with travelling to a particular place. Typically, it consists of multiplying the geographic distance by some monetary indicator that considers both transportation costs and the value of travel time. When modelling same-day recreational demand, most people travel by car so that each kilometre between the place of residence and the destination is usually multiplied by €0.19 (Bujosa and Rosselló, 2013). Indeed, this corresponds to the official per diem allowance of transportation costs per kilometre in Spain. The opportunity cost of one hour of travelling time is generally defined as one third of the wage rate (Cesario, 1976; Hanauer and Reid, 2017; Swait et al., 2020), although there is some controversy about whether such fraction is appropriate (Amoako-Tuffour and Martínez-Espiñeira, 2012; Czajkowski et al., 2019).

However, when working with trips that involve long distances, we face the problem that we need to define distinct unitary cost per kilometre depending on the chosen mode of transport. This choice is likely to be endogenous with the choice of destination, since for instance the Balearic and the Canary Islands cannot be accessed by car, bus or train. Moreover, individuals might use several modes of transport to reach faraway destinations (see Voltaire et al., 2017 for a discussion). That is the reason why we opted for using geographical distance in the main analysis. Nevertheless, here we inspect whether our findings would change if we used the travel cost method.

Given the joint determination of the mode of transport and the destination, we follow Bujosa and Rosselló (2013) and Voltaire et al. (2017) restrict the sample only to tourists travelling to destinations that can be accessed by road-based transport modes. Therefore, the Balearic and the Canary Islands are excluded, both as destinations and potential origins. The resulting sample size is 6,207 individuals.

We do not have information on respondents' hourly wage. We only know households' income (in intervals). To have a continuous indicator, we first run an interval regression in which household income is regressed on standard sociodemographic characteristics including gender, age, education, household size, labour status, nationality, civil status, the size of the municipality of residence and regional fixed effects. The parameter estimates are available upon request. The fitted values are then divided by the number of household members to get a (continuous) estimation of *individual* monthly income. Then, for those who declare to be employed, this imputed individual income is divided by 150 under the assumption individuals work 37.5 hours per week (standard full-time workday in Spain). Therefore, the resulting value is an estimate of the hourly wage (in €). This way of deriving hourly wages from aggregated monthly or annual income is common in the travel cost literature in the absence of wage information (Hanauer and Reid, 2017; Swait et al., 2020). Next, assuming individuals travel on average at 90 kilometres per kilometre, estimated travel time to each destination would be given by $DIST_{ij}/90$. Altogether, the travel cost to reach region j for each individual i is defined as follows:

$$COST_{ij} = €0.19 * DIST_{ij} + \frac{1}{3} * \widehat{Income_i} * \frac{1}{housesize_i} * \frac{1}{150} * \frac{DIST_{ij}}{90} \qquad if\ employed = 1$$
$$COST_{ij} = €0.19 * DIST_{ij} \qquad if\ employed = 0 \tag{1}$$

where $\widehat{Income}$ is the continuous estimation of household income.

Figure A8 presents a kernel density plot of the imputed (net) hourly wage in the sample ($w_i = \widehat{Income_i} * \frac{1}{housesize_i} * \frac{1}{150}$). The mean estimate is €8.71, with is not far from the *gross* mean hourly wage in Spain that is equal to €11.9 (INE, 2021).

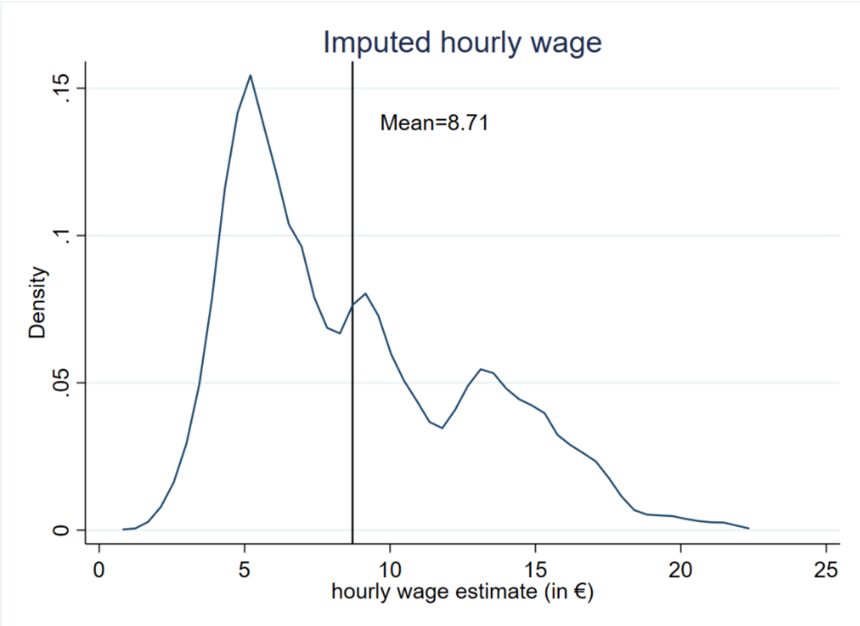

Figure A8.- Kernel density plot for imputed hourly wage (w_i)

We recognize this cost imputation is subject to the underlying assumptions, but studies using the travel cost method also make similar suppositions. Figure A9 depicts kernel density plots of DIST and COST. As shown there, given aside scale differences, the distribution of both variables in the sample is quite similar. The pairwise correlation between the two is 0.979. This is an expected result because of the following: the conversion of geographical distance to transportation costs is a simple scale adjustment. The second component (opportunity cost of time) adds little variation to the first one, especially as geographic distance to the origin grows relative to the hourly wage. As a result, the travel cost method allows for capturing individual-specific travel costs through the opportunity cost of time when the travel distances are reduced. In the context of long-distance trips, this method adds little to the use of geographic distance.

That might be the reason why the travel cost method is preferred in the recreational demand literature while origin-destination distances are used in studies concerned about tourism flows.

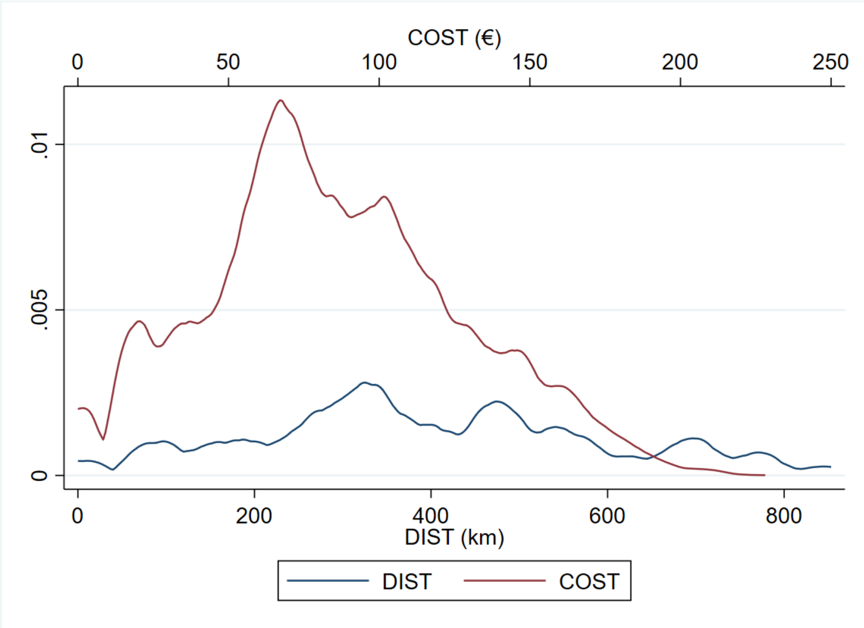

Figure A9.- Kernel density plots for DIST and COST

The variable $DIST_{ij}$ is replaced in the model by $COST_{ij}$, with everything else kept unchanged. Table A6 presents the corresponding parameter estimates. Note that since we are working with a subsample that excludes the Balearic and the Canary Islands, the estimates can be compared with those presented in Table A5 above. By comparing the estimation results from Tables A5 (with geographic distance, DIST) and A6 (with travel costs, COST), we document that the travel cost method produces consistent results.

	RPLc-ECM	
Variable	Coef	St.Error
$REG1$	2.749***	0.1828
$REG2$	1.475***	0.2523
$REG3$	0.563**	0.2558
$REG4$	0.968***	0.2489
$REG5$	0.701***	0.1553
$COST$	–0.022***	0.0028
r_TEMP	2.374***	0.7431
$RAIN$	–0.220***	0.0724
$TCPI$	–0.003	0.0065
TOU_SPOTS	0.104***	0.0098
NAT_PARKS	0.081***	0.0109
$SIZE\_NAT$	0.002***	9.4e-04
SKI_KM	0.001***	2.4e-04
$SKI\_KM * winter\_sports$	0.009***	5.7e-04
$COAST$	–1.385***	0.1841
$COAST * aquatic$	2.577***	0.1270

SD $COST$	0.014***	0.0011
SD r_TEMP	1.787***	0.3472
Cov($COST$, r_TEMP)	-5.3e-04	0.0052
ϑ_1	0.521**	0.2284
ϑ_2	0.357	0.2778
ϑ_3	0.085	0.7649
$COST$ Mean shifters		
age	1.0e-04***	4.1e-03
inc2	0.006***	0.0012
inc3	0.011***	0.0018
parsize	–0.005***	0.0004
weekend	–0.022***	0.0012
q1	–0.004*	0.0024
q2	–0.016	0.0020
q4	–0.009***	0.0024
d_warmorigin1	–0.001	0.0025
d_warmorigin2	–0.004**	0.0021
d_warmorigin3	–0.006***	0.0018
d_warmorigin4	0.005**	0.0028
winter_sports	–0.071	0.0733
mou_trek_nat	–0.008***	0.0012
rural	–0.001	0.0012
aquatic	0.011***	0.0013
advent	0.004***	0.0013
r_TEMP Mean shifters		
age	0.010	0.0090
inc2	0.244	0.2703
inc3	0.340	0.3437
parsize	–0.107	0.0990
weekend	–0.563**	0.2512
q1	–1.447***	0.5463
q2	–0.592	0.6075
q4	–1.597***	0.5527
d_warmorigin1	1.750***	0.4231
d_warmorigin2	1.219**	0.6054
d_warmorigin3	–2.824***	0.6197
d_warmorigin4	–0.068	0.5943
winter_sports	–0.618	0.4088
mou_trek_nat	–0.827***	0.2699
rural	–0.707**	0.2785
aquatic	–0.191	0.3899
advent	1.351***	0.2989
Log L	–10,304.6	
AIC	20,721.2	
Pseudo-R2	0.387	
N	6,207	

Table A6.- Parameter estimates for RPLc-ECM without the Balearic and the Canary Islands and replacing DIST by travel costs (COST)
*** p<0.01, ** p<0.05, * p<0.

REFERENCES

Amoako-Tuffour, J. and Martínez-Espiñeira, R. (2012). Leisure and the Net Opportunity Cost of Travel Time in recreation demand analysis: An application to Gros Morne National Park. *Journal of Applied Economics*, 15(1), 25-49.

Bujosa, A. and Rosselló, J. (2013). Climate change and summer mass tourism: the case of Spanish domestic tourism. *Climatic Change* 117: 363-375.

Cesario, F.J. (1976). The value of time in recreation benefit studies. *Land Economics*, 52, 32–41.

Czajkowski, M., Giergiczny, M., Kronenberg, J. and Englin, J. (2019). The individual Travel Cost Method with consumer-specific values of travel time savings. *Environmental and Resource Economics*, 74, 961-984.

Hanauer, M.M. and Reid, J. (2017). Valuing urban open space using the travel-cost method and the implications of measurement error. *Journal of Environmental Management*, 198, 50–65.

Swait, J., Franceschinis, C. and Thiene, M. (2020). Antecedent Volition and Spatial Effects: can multiple goal pursuit mitigate distance decay? *Environmental and Resource Economics* 75(2): 243-270.

Voltaire, L., Lévi, L., Alban, F. and Boncoeur, J. (2017). Valuing cultural word heritage sites: an application of the travel cost method to Mont-Saint-Michel. *Applied Economics*, 49(16), 1593-1605.

APPENDIX D

The estimates of the structural parameters in our Random Utility Model might provide an incomplete picture of individual's marginal utilities. The 'unconditional' mean of β_{ki} (only conditioning on Z) is simply:

$$E[\beta_{ki}|Z_i] = b_k + \delta_k{}'Z_i \qquad (2)$$

According to Greene (2004), the 'unconditional' mean estimator in (2) is an ambiguous estimator of the marginal sensitivities. A proper characterization of the MUs needs to also take into account the actual choices made by the individual, the existing correlation between the attributes assumed to be random (if any) and the correlation between similar destinations in the form of the error-components included in the model. In other words, we look for an estimator of the MUs that considers all available information about individual i.

Let $f(\beta|\Omega)$ be the distribution of the individual-specific parameters in the population, $P_{ij}(y_j|\beta)$ the probability that respondent i chooses destination j conditional on β, and $h(\beta|y_j,\Omega)$ the distribution of the individual-specific parameters for those who make the choices y_j. By applying Bayes Theorem, the conditional on choices in-sample marginal distribution $h(\beta|y_j,\Omega)$ for the RPL model can be derived as follows:

$$h(\beta|y_j,\Omega) = \frac{P_{ij}\big(y_j|\beta\big)f(\beta|\Omega)}{P_{ij}\big(y_j|\Omega\big)} = \frac{\frac{\exp(V_{ij})}{\sum_{j=1}^{J}\exp(V_{ij})} f(\beta_i|\Omega)}{\int \frac{\exp(V_{ij})}{\sum_{j=1}^{J}\exp(V_{ij})} f(\beta_i|\Omega)\, d\beta_i} \qquad (3)$$

Since the denominator in (B.2) is the integral of the numerator and is a constant, the conditional marginal distribution $h(\beta|y_j,\Omega)$ is proportional to the numerator. The expression in (3) becomes more complex when the model incorporates error-components. Since in the RPLc-ECM model $V_{ij} = ASC_j + X_{kj}{}'\beta_k + \vartheta_n En_{ij}$, the conditional distribution of the choices ($P_{ij}(y_j|\beta)$) needs to firstly eliminate the error components from the expression by integrating over its standard normal distribution.

The conditional expectation of the individual-specific marginal utilities is given by:

$$E\big(\beta_i|y_{ij},X_{kj},Z_i\big) = \frac{\int_{\beta_i}\int_{E_i}\beta_i\, P_{ij}\big(y_j|\beta,X_{kj},E_i\big)\, f(\beta_{i,}E_i\,|Z_i)dE_i\, d\,\beta_i}{\int_{\beta_i}\int_{E_i} P_{ij}\big(y_j|\beta,X_{kj},E_i\big)\, f(\beta_{i,}E_i\,|Z_i)dE_i\, d\,\beta_i} \qquad (4)$$

where $f\big(\beta_{i,}E_i\,|Z_i\big)$ is the joint marginal density of β_i and E_i . Since β_i and E_i are independent, the joint density equals the product of their separate marginal distributions. Therefore, for each individual i, the MUs are estimated as the mean of their conditional distribution[13]. Note that $E(\beta_i|y_{ij}, X_{kj}, Z_i)$ conditions on all available information about individual i whereas $E[\beta_i|Z_i]$ only conditions on the vector of taste shifters. This is why we have referred to the latter as the 'unconditional' distribution of the MUs. A detailed derivation of these conditional mean estimates can be found in Greene (2004; 2012, pp. 144-147), Greene et al. (2006), Train (2009, p. 262-264), and Hess (2010). Since the integral in (4) does not have a closed form solution, the conditional means of the parameters are approximated by Simulated Maximum Likelihood[14].

Figures A10 and A11 show a kernel plot of the estimated individual-specific parameter estimates for $DIST$ and r_TEMP[15].

As is evident from these Figures, the conditional mean estimates of the marginal utilities of $DIST$ and r_TEMP are heterogeneous. Regarding distance, although a small share attaches positive utility, in general it can be regarded as a dissuasive factor rather than a desirable feature. Interestingly, although the structural parameter of r_TEMP (b_2) is positive, a non-negligible share of the sample has a negative conditional marginal utility. It is important to highlight here that although the two distributions appear to be bimodal, these conditional mean estimates are not necessarily normally distributed (Greene, 2004).

---

[13] Greene (2012) warns that these estimates are conditioned on the observable information for individual i. Put another way, the $\widehat{\beta_i}$ would be the same for two individuals with exactly the same observable characteristics and observed choices, since the estimates are mean values for the subpopulation that have the same observables and made the same choice. In any case, these individual-specific parameter estimates are efficient estimates of β_i.
[14] Alternatively, the estimation could be performed under a hierarchical Bayes framework. Huber and Train (2001) provide a discussion on the Bayesian and the classical approaches to derive the individual estimates. They conclude that both procedures are virtually identical.
[15] These estimates have been computed for the RPLc-ECM model. We use a kernel density estimator instead of a histogram because the underlying distributions are continuous rather than discrete.

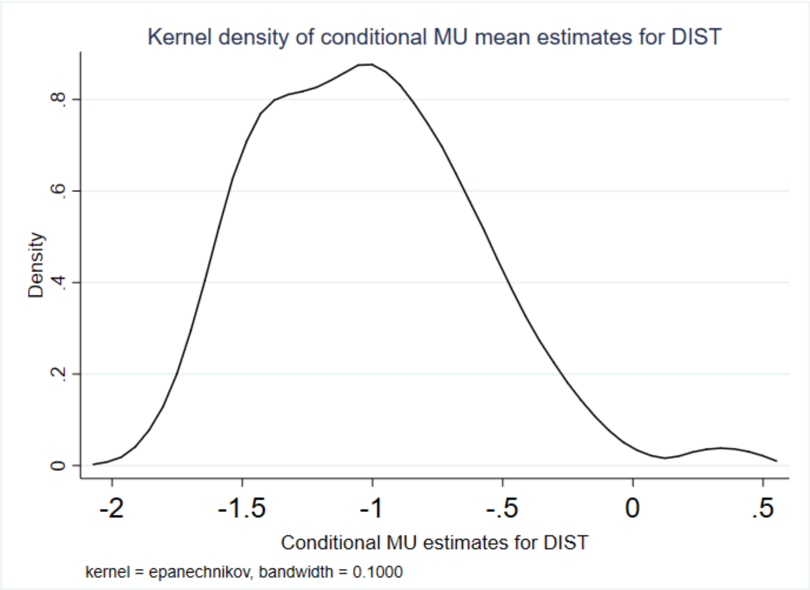

Figure A10.- Kernel density of conditional MU mean estimates for *DIST*

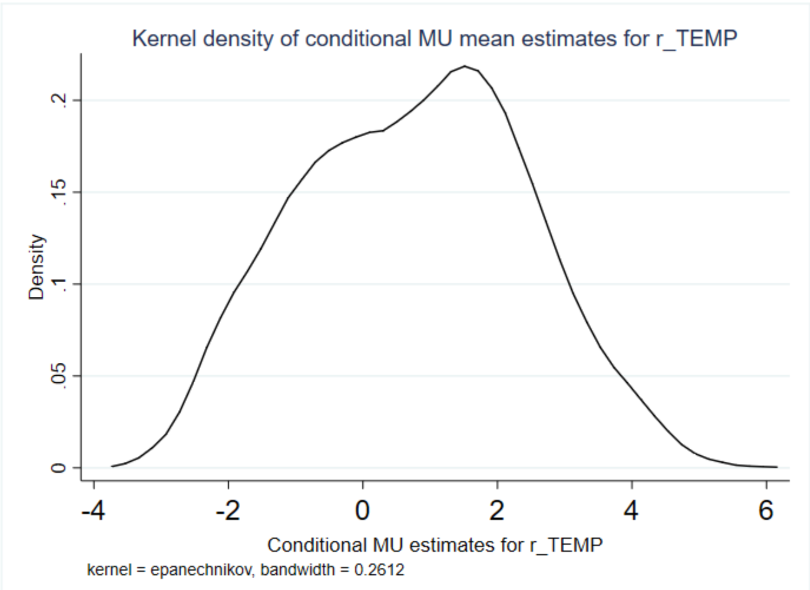

Figure A11.- Kernel density of conditional MU mean estimates for *r_TEMP*

REFERENCES

Greene, W.H. (2004). Interpreting estimated parameters and measuring individual heterogeneity in random coefficient models. Department of Economics, Stern School of Business, New York University.

Greene, W.H. (2012). *NLOGIT5 Reference Guide*. Chapter 29, 556-566.

Greene, W.H., Hensher, D.A. and Rose, J. (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B*, 40, 75-92.

Hess, S. (2010). Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool. *Journal of Choice Modelling*, 3(2), 134-152.

Huber, J. and Train, K. (2001). On the similarity of Classical and Bayesian estimates of individual mean partworths. *Marketing Letters*, 12(2), 259-269.

Train, K. (2009). *Discrete choice models with simulation*. Second Edition. Cambridge University Press.